

Hee-Soo CHOI

Doctorante en Sciences du Langage au LORIA / ATILF - ATER à l'IDMC
Nancy, France — hee-soo.choi@loria.fr — (+33) 618137231 — <https://hschoi4.github.io/>

FORMATION

Doctorat mention Sciences du Langage Laboratoires ATILF - LORIA, Université de Lorraine, Nancy, France Sous la direction de Pr. Mathieu Constant et Pr. Karèn Fort Lier des ressources lexicales du français en vue d'une interopérabilité entre niveaux linguistiques	2021 — Aujourd'hui
Master Langue et Informatique , mention bien Sorbonne Université, Paris, France Mémoire M2 : Vérification d'universaux linguistiques sur des corpus multilingues annotés en syntaxe de dépendances	2019 — 2021
Licence Langue Française et Techniques Informatiques (LFTI) , mention très bien Sorbonne Université, Paris, France	2016 — 2019
Première Année Commune aux Etudes de Santé (PACES) Université Pierre et Marie Curie, Paris, France	2014 — 2016

EXPÉRIENCES PROFESSIONNELLES

Attachée Temporaire d'Enseignement et de Recherche (ATER) à temps plein IDMC, Université de Lorraine, Nancy, France Enseignements en Licence MIASHS et Master TAL (Erasmus Mundus)	septembre 2024 — août 2025
Doctorante invitée SapienzaNLP, Sapienza University of Rome, Rome, Italie Sous la direction du Pr. Roberto Navigli Projet de recherche sur la levée d'ambiguïté lexicale en français par intégration de ressources lexico-sémantiques	février 2024 — mai 2024
Stagiaire de recherche Laboratoire LORIA, Nancy, France	avril 2021 — septembre 2021
Stagiaire de recherche Laboratoire LORIA, Nancy, France	juillet 2020 — août 2020

ENSEIGNEMENTS

Tous mes supports de cours personnels sont disponibles sur ce lien : <https://hschoi4.github.io/teaching/>

Institut des sciences du Digital, Management Cognition (IDMC)	2024 — 2025
Licence MIASHS	
• L1 - Bases de données	TD 24h
• L1 - Langages de script	TD 20h
• L2 - Bases de données	TD 48h
• L3 Parcours TAL - Programmation orientée objet Python avancé Création de tous les supports de cours.	CM 22.5h / TD 15h
Master NLP (Enseignements en anglais)	
• M1 - Corpus écrits	TD 30h
• M2 - Projets logiciels	CM 12h / TD 12h
Télécom Nancy	2023
1A - Bases de données	TD 4h / TP 8h
• Modélisation conceptuelle et logique des données	
• Algèbre relationnel et calcul relationnel de tuples	
• Langages SQL et PL/PgSQL	
Université de Lorraine	2021 — 2023
L2 Sciences du Langage - Introduction à Python Création de la totalité des supports du cours, des TD et d'évaluation.	TD 24h/an

- Introduction à l'algorithmique
- Structures de données et de contrôle
- Expressions régulières
- Manipulation de fichiers et fonctions
- Traitements et analyses statistiques de textes

PUBLICATIONS

Conférences internationales avec comité de lecture

- **Mitigating Data Scarcity in Semantic Parsing across Languages: the Multilingual Semantic Layer and its Dataset**
Abelardo Carlos Martínez Lorenzo, Pere-Luís Huguet Cabot, Karim Ghonim, Lu Xu, [Hee-Soo Choi](#), Alberte Fernández Castro, Roberto Navigli
In Findings of ACL 2024, août 2024, Bangkok, Thaïlande
- **Beyond Model Performance: Can Link Prediction Enrich French Lexical Graphs?**
[Hee-Soo Choi](#), Priyansh Trivedi, Mathieu Constant, Karën Fort, Bruno Guillaume
LREC-COLING 2024, mai 2024, Turin, Italie (44 % taux d'acceptation)
- **Corpus-based Language Universals Analysis using Universal Dependencies**
[Hee-Soo Choi](#), Bruno Guillaume, Karën Fort
Quantitative Syntax (QUASY), SyntaxFest 2021, mars 2022, Sofia/distanciel, Bulgarie
- **Investigating Dominant Word Order on Universal Dependencies with Graph Rewriting**
[Hee-Soo Choi](#), Bruno Guillaume, Karën Fort, Guy Perrier
Recent Advances in Natural Language Processing (RANLP) 2021, septembre 2021, Varna/distanciel, Bulgarie

Conférences nationales avec comité de lecture

- **Au-delà de la performance des modèles : la prédiction de liens peut-elle enrichir des graphes lexico-sémantiques du français ?**
[Hee-Soo Choi](#), Priyansh Trivedi, Mathieu Constant, Karën Fort, Bruno Guillaume
Actes de la 31e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), juillet 2024, Toulouse, France
- **Des ressources lexicales du français et de leur utilisation en TAL : étude des actes de TALN**
[Hee-Soo Choi](#), Karën Fort, Bruno Guillaume, Mathieu Constant
Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), juin 2023, Paris, France
- **État de l'art : Liage de ressources lexicales du français**
[Hee-Soo Choi](#)
Actes de la 24e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL), juin 2022, Avignon, France

Séminaires invités

- **French Lexical Semantic Graphs: Enrichment by Link Prediction and Integration in a WSD model**
[Hee-Soo Choi](#)
Séminaire annuel ANR SELEXINI, juin 2024, Paris, France
- **Corpus-based Language Universals Analysis using Universal Dependencies**
[Hee-Soo Choi](#)
Séminaire Traitement Automatique du Langage Ecrit et Parlé (TALEP)
Laboratoire d'Informatique et Systèmes (LIS), mars 2023, Marseille, France
- **Investigating Dominant Word Order on Universal Dependencies with Graph Rewriting**
[Hee-Soo Choi](#)
Séminaire Research Group in Mathematical Linguistics (GRLMC), mai 2022, Tarragona/distanciel, Espagne

Conférences nationales sans comité de lecture

- **Analyse orientée corpus d'universaux de Greenberg sur Universal Dependencies**
[Hee-Soo Choi](#), Bruno Guillaume, Karën Fort
GDR LIFT Linguistique Informatique, Formelle et de Terrain 2021, décembre 2021, Grenoble, France

Ouvrages de vulgarisation

- **PENSE-M'EN**
Hee-Soo Choi, Marie Cousin, Amandine Decker, Valentin D. Richard
THINK BEFORE LOADING, 2024

CO-ENCADREMENT DE MÉMOIRE DE MASTER

- Anisa Amourat et Assia Bouznad (M1 Langue et Informatique, Sorbonne Université) février 2024 - juin 2024
Annotation semi-automatique en sens d'un corpus du français par un modèle de levée d'ambiguïté lexicale multilingue (co-encadrement avec Karën Fort)

AUTRES ACTIVITÉS

- Relectrice** 2024
- LREC-COLING 2024
 - GDR LIFT 2024
- Membre du comité d'organisation de la conférence IWCS 2023** 2023
IDMC, Nancy, France
- Membre du Bureau des Doctorants du LORIA** 2023
Responsable activités au week-end d'intégration (WEI) 2023
- Membre du comité d'organisation du tournoi de ping-pong interne du LORIA** 2022 - 2024
4 éditions - 30 participants / édition
- Évaluatrice de traduction pour la revue Programming Historian en français** 2022
Leçon Python sur la réutilisation de code et modularité
(<https://programminghistorian.org/fr/lecons/reutilisation-de-code-et-modularite>)

MÉDIATION

- La Nouvelle de la classe** juin 2024
Atelier autour du TAL pour des classes de CE2-CM1-CM2, ATILF, Nancy, France
Introduction à la syntaxe avec des puzzles créés en découpe laser et présentation d'outils de TAL
- Intervention CHICHE - 1 scientifique, 1 classe**
Présentation du métier de chercheur en informatique et du quotidien d'une doctorante.
Discussion sur les parcours académiques après le lycée
- 4 classes - Lycée Saint-Exupéry, Fameck, France janvier 2024
 - 3 classes - Lycée Varoquaux, Tomblaine, France décembre 2023
 - Classe Scientifilles - Lycée Claude Gellée, Épinal, France janvier 2023

BOURSES

- Bourse **ATALA** - TALN 2024
- Bourse **LUE DrEAM 13** - Mobilité dans un laboratoire étranger, SapienzaNLP, Sapienza University of Rome

COMPÉTENCES INFORMATIQUES

- **Langages de programmation** : Python, bash, L^AT_EX, SQL, PL/PgSQL, Java
- **Autres** : Git, Environnement Linux

LANGUES

- **Français** : langue maternelle
- **Anglais, Coréen** : niveau C2
- **Espagnol** : niveau B1
- **Chinois, Grec moderne, Italien** : notions