

Stefen Thelier

Diplômé d'un Master Sciences du Langage, mention très bien · linguistique de corpus & analyse du discours politique



CONTACT

stefenthelierpro@gmail.com
Reims, France
LinkedIn
ResearchGate

TECHNOLOGIES

Python
TXM
R
HTML / CSS
Git

MÉTHODES

Linguistique de corpus
Analyse de discours
Statistiques textuelles
TAL / NLP

LANGUES

Français · langue maternelle
Anglais · C1

FORMATION

2026

Master Sciences du Langage

Université de Reims Champagne-Ardenne (URCA) · laboratoire CIRLEP

Mention très bien, major de promotion. Spécialisation en analyse du discours politique par la linguistique de corpus.

MÉMOIRE DE M2 20/20

« *The 'Extreme Centre' in UK Political Discourses, 2017–2025 : Boundaries, Shifts, and Intersections with the Far Right* », sous la direction de Mme Kimberly Oger. Étude de corpus (2,5 M de tokens) avec Python, R et TXM.

MÉMOIRE DE M1 20/20

« *Étude contrastive de l'utilisation des pronoms personnels dans les discours sur l'État de l'Union aux États-Unis et dans les discours de politique générale en France* », sous la direction de Mme Kimberly Oger.

2021–2024

Licence LLCER Anglais

Université de Reims Champagne-Ardenne (URCA)

Langues, littératures et civilisations étrangères et régionales, parcours anglais.

RECHERCHE

2025

Stage au CIRLEP

Université de Reims Champagne-Ardenne

Analyse d'un fonds d'archives d'entretiens sur le monde viticole et vinicole de la Villa Bissinger. Constitution d'un tableau de synthèse, participation aux discussions méthodologiques, transcription d'entretiens manuelle et assistée par outils d'IA.

Sept. 2025

Henry Sweet Society Colloquium

Université de Reims Champagne-Ardenne

Participation au colloque en tant qu'auditeur : suivi des communications.

ENSEIGNEMENT & MÉDIATION

Oct. 2026

Fête de la science 2026

UFR Lettres et Sciences Humaines, URCA · laboratoire CIRLEP

Animation de l'atelier d'éveil aux langues « Origine des noms des fruits dans les langues européennes », l'un des trois ateliers du CIRLEP, lors d'une journée de médiation scientifique qui a accueilli 500 élèves de la maternelle à la 3^e. Les élèves comparaient les noms des fruits d'une langue à l'autre pour découvrir leurs origines et les grandes familles des langues indo-européennes.

2025–2026

Tutorat étudiant LLCER Anglais

Université de Reims Champagne-Ardenne

Accompagnement d'étudiants de Licence : compréhension des cours, organisation, méthodes de travail, méthodologie de constitution de corpus et interprétation statistique des données. Matières : introduction à la linguistique, morphosyntaxe, linguistique de corpus.

2022–2024

AED en préprofessionnalisation

Collège Le Lac, Sedan

Encadrement d'élèves, soutien individuel et aide aux devoirs. Création de cours d'anglais et prise en charge de classes de 6^e à 3^e.

PUBLICATIONS

2026

Mémoire orale champenoise : transformation d'un fonds d'archives en corpus

Kimberly Oger, Stefen Thelier, Malika Pedley · soumis à *Territoires du vin*, en cours d'évaluation

Comment transformer des archives orales en corpus de recherche : retour sur le fonds du CAdo'C, des entretiens avec des professionnels de la filière champenoise. L'article décrit l'inventaire du fonds, le choix d'une convention de transcription lisible aussi par des non-linguistes, les apports et les limites de la transcription automatique avec Whisper, et les questions éthiques que pose l'anonymisation de récits de vie.

COMMUNICATIONS

7 avril 2026

De l'archive au corpus : projet de transformation des archives orales de la Villa Bissinger

Séminaire Langue et Corpus du CIRLEP · avec Kimberly Oger et Malika Pedley · URCA

Valorisation du fonds d'archives orales du CAdo'C (Villa Bissinger, Aÿ-Champagne) : 154 entretiens sur le monde vitivinicole, 120 locuteurs, 167 h d'enregistrement. Chaîne de traitement (inventaire, anonymisation, transcription automatique avec Whisper puis vérification), conventions de transcription adaptées de LesVocaux, et questions techniques, éthiques et de financement.

31 mars 2026

Analyse des discours de Keir Starmer : présentation de mon mémoire de recherche

Séminaire Multilinguisme, Plurilinguisme & Intercompréhension · URCA

Présentation du mémoire et du travail mené : généalogie du concept d'« extrême centre » (Serna, Deneault), définition retenue pour l'étude, cadre de l'analyse critique du discours (Wodak) et principale difficulté méthodologique : articuler une analyse de discours qualitative avec une approche de corpus portant sur plus de 500 textes.

24 mars 2026

Automatisation de la collecte de corpus et de l'analyse linguistique

Séminaire Langue et Corpus du CIRLEP · URCA

Pourquoi automatiser la collecte de corpus, comment fonctionne une page web, et comment en extraire le texte avec Python (bibliothèques, API, Selenium), illustré sur les débats du Parlement britannique : 244 fichiers collectés en moins de 10 minutes. Puis présentation de Linguistix face aux logiciels existants (TXM, Sketch Engine, WordSmith, LancsBox, AntConc) : fonctionnalités et limites.

Déc. 2024

Évolution des pronoms personnels dans les discours sur l'État de l'Union

Communication

Analyse diachronique de l'usage de « I », « we » et « you » dans le corpus présidentiel américain (1790–2023).

OUTILS DÉVELOPPÉS

2026

Hansard Scraper

Extension de navigateur · en savoir plus

Extraction des débats du Parlement britannique avec leurs métadonnées (export CSV) et liens directs vers les vidéos horodatées de Parliament TV, pour faciliter la constitution de corpus.