

Stefen Thelier

Master's degree in Language Sciences, with highest honours · corpus linguistics & political discourse analysis



CONTACT

stefenthelierpro@gmail.com
Reims, France
LinkedIn
ResearchGate

TECHNOLOGIES

Python
TXM
R
HTML / CSS
Git

METHODS

Corpus linguistics
Discourse analysis
Textual statistics
NLP

LANGUAGES

French · native
English · C1

EDUCATION

2026

Master's degree in Language Sciences

University of Reims Champagne-Ardenne (URCA) · CIRLEP research centre

Highest honours (*mention très bien*), ranked first in the year group. Specialisation in political discourse analysis through corpus linguistics.

MASTER'S THESIS, 2ND YEAR 20/20

“The ‘Extreme Centre’ in UK Political Discourses, 2017–2025: Boundaries, Shifts, and Intersections with the Far Right”, supervised by Kimberly Oger. Corpus study (2.5 million tokens) using Python, R and TXM.

MASTER'S THESIS, 1ST YEAR 20/20

“A contrastive study of personal pronouns in US State of the Union addresses and French prime ministers’ general policy speeches” (in French), supervised by Kimberly Oger.

2021–2024

Bachelor's degree in English Studies (LLCER)

University of Reims Champagne-Ardenne (URCA)

Foreign and regional languages, literatures and civilisations, English track.

RESEARCH

2025

Research internship at CIRLEP

University of Reims Champagne-Ardenne

Analysis of an archive of interviews about the wine-growing and winemaking world of the Villa Bissinger. Built a summary table, took part in methodological discussions, transcribed interviews by hand and with AI-assisted tools.

Sept. 2025

Henry Sweet Society Colloquium

University of Reims Champagne-Ardenne

Attended the colloquium as a member of the audience.

TEACHING & OUTREACH

Oct. 2026

Fête de la science 2026 (French national science festival)

Faculty of Arts and Humanities, URCA · CIRLEP research centre

Ran the language awareness workshop “Where fruit names come from in European languages”, one of CIRLEP’s three workshops, during a science outreach day that welcomed 500 pupils aged 3 to 15. Pupils compared fruit names across languages to discover their origins and the main branches of the Indo-European family.

2025–2026

Peer tutor, BA in English Studies

University of Reims Champagne-Ardenne

Supporting undergraduate students: understanding the lectures, organisation, study skills, corpus-building methodology and statistical interpretation of data. Subjects: introduction to linguistics, morphosyntax, corpus linguistics.

2022–2024

Pre-service teaching assistant (AED)

Collège Le Lac (lower secondary school), Sedan

Supervising pupils, one-to-one support and homework help. Designed English lessons and taught classes of pupils aged 11 to 15.

PUBLICATIONS

2026

Mémoire orale champenoise : transformation d'un fonds d'archives en corpus [Oral memory of Champagne: turning an archive collection into a corpus]

Kimberly Oger, Stefen Thelier, Malika Pedley · submitted to *Territoires du vin*, under review (in French)

How to turn oral archives into a research corpus: a look back at the CADO'C collection, interviews with professionals of the Champagne wine industry. The article describes the inventory of the collection, the choice of a transcription convention that non-linguists can read too, what automatic transcription with Whisper brings and where it falls short, and the ethical questions raised by anonymising life stories.

TALKS

7 April 2026

From archive to corpus: turning the Villa Bissinger oral archives into a corpus

CIRLEP "Language and Corpus" seminar · with Kimberly Oger and Malika Pedley · URCA (in French)

Making the most of the CADO'C oral archive (Villa Bissinger, Aÿ-Champagne): 154 interviews about the wine world, 120 speakers, 167 hours of recordings. Processing pipeline (inventory, anonymisation, automatic transcription with Whisper, then checking), transcription conventions adapted from LesVocaux, and technical, ethical and funding questions.

31 March 2026

Analysing Keir Starmer's speeches: presenting my research thesis

"Multilingualism, Plurilingualism & Intercomprehension" seminar · URCA (in French)

Presentation of the thesis and the work behind it: the genealogy of the concept of the "extreme centre" (Serna, Deneault), the definition used in the study, the critical discourse analysis framework (Wodak) and the main methodological difficulty: combining a qualitative discourse analysis with a corpus approach covering more than 500 texts.

24 March 2026

Automating corpus collection and linguistic analysis

CIRLEP "Language and Corpus" seminar · URCA (in French)

Why automate corpus collection, how a web page works, and how to extract its text with Python (libraries, APIs, Selenium), illustrated with UK parliamentary debates: 244 files collected in under 10 minutes. Then a presentation of LinguistiX alongside existing software (TXM, Sketch Engine, WordSmith, LancsBox, AntConc): features and limits.

Dec. 2024

The evolution of personal pronouns in State of the Union addresses

Talk

Diachronic analysis of the use of "I", "we" and "you" in the US presidential corpus (1790–2023).

TOOLS I HAVE BUILT

2026

Hansard Scraper

Browser extension · [learn more](#)

Extracts UK parliamentary debates with their metadata (CSV export) and direct links to the time-stamped Parliament TV videos, to make corpus building easier.