

 Montréal
 +33 (0)6 35 35 83 09
 alexia.schneider@umontreal.ca
 [Git](#)

Alexia Schneider

Linguiste computationnelle avec un intérêt pour :

- les Humanités Numériques
- les études stylistiques computationnelles.

Expériences professionnelles

- Avril - Juillet 2024
stage **Reconnaissance des entités nommées pour l'encodage automatique en XML-TEI de notices bibliographiques d'ouvrages des XVIe et XVIIe siècles.**
Bibliothèque Nationale et Universitaire | Strasbourg
 - annotation manuelle en XML-TEI d'un sous-corpus pour l'entraînement de modèles
 - création et entraînement de modèles de reconnaissance d'EN
 - alignement automatique (désambiguïsation) avec le référentiel IdRef
- Janvier - Mars 2024
stage **Développement d'un "assistant IA" avec des LLMs génératifs**
noota | Paris
 - développement d'un système RAG
 - *prompt engineering* (rédaction et test de prompts) et évaluation des réponses générées
 - évaluation des LLM et SLM pour la mise en production
- Août - Septembre 2023
stage **Classification automatique du type de didascalie**
Supervision : Pablo Ruiz Fabo
Laboratoire "Linguistique, Langue, Parole" LiLPa UR 1339 | Unistra | Strasbourg
Projet : Thealtres ([website](#)) ([GitHub](#))
 - extraction depuis le French Drama Corpus (TEI) des didascalies via le langage de requête XQuery.
 - composition d'une typologie à 13 classes généralisable
 - entraînement d'algorithme d'apprentissage automatique (RandomForest, LogisticRegression, Naive Bayes) avec la bibliothèque Python scikit-learn
 - affinage de grands modèles de langue (LLMs) de type BERT
 - évaluation de l'impact de la réduction du nombre d'exemples sur la qualité de l'apprentissage des modèles
- Mai - Juillet 2023
stage **Annotation manuelle d'un corpus de vaudeville et identification automatique des "airs".**
Supervision : Pablo Ruiz Fabo
Laboratoire "Linguistique, Langue, Parole" LiLPa UR 1339 | Unistra | Strasbourg
Projet : Thealtres ([website](#)) ([GitHub](#))
 - annotation manuelle et mise en accord d'une prosopographie avec une autre annotatrice
 - développement d'un programme pour l'identification de titres d'"airs" à base de règles ("regex" ou motifs dans la chaîne de caractères).
 - alignement des titres avec un référentiel à partir d'une recherche de similarité soit sur une mesure de distance d'édition (Levenshtein) soit une mesure de distance (cosinus) à partir de représentation du contenu avec des grands modèles de langue de type Transformers
 - intégration du projet sur la plateforme de gestion de code et de projet informatique GitLab

Productions scientifiques

- Juin 2024
Schneider, A., & Wingert, R. (2024). Le projet Ritter augmenté : Reconnaissance d'Entités Nommées pour les données bibliographiques. [Rencontre régionale IA4LAM « Intelligence Artificielle, patrimoine et Humanités Numériques »](#), BNU, Strasbourg.
- Mars 2024
Schneider, A., & Ruiz, P. (2024). Stage Direction Classification in French Theater : Transfer Learning Experiments. Proceedings of [LaTeCH-CLfL 2024](#),
[DOI & publication](#)
- Novembre 2023
Schneider, A. (2023). Classification automatique du type de didascalie. [Journée d'étude : « Données de théâtre, petit et grand : Le bon, le brut et le futé »](#), MISHA, Strasbourg.
[DOI](#)
- Schneider, A., & Nugues, L. (2023). Détecter automatiquement des airs de vaudeville. [Journée d'étude : 'Données de théâtre, petit et grand : Le bon, le brut et le futé', MISHA](#), Strasbourg
[DOI](#)

Formation

2025/en cours	Doctorat en Littératures option Humanités Numériques sous la direction de Marcello Vitali-Rosati <i>Titre : Enrichissement sémantique de recherches documentaires en SHS.</i> <i>Université de Montréal Montréal</i>
2022/2024	Master Technologie des Langues : mention "Traitement Automatique des Langues" <i>Unistra Strasbourg</i>
2016/2017	Licence Professionnelle dans les Métiers du Document en Environnement Numérique <i>IUT Robert Schuman, Unistra Strasbourg</i>
2015/2016	Master 1 de Littérature <i>Université Paris-Sorbonne Paris</i> <i>En Erasmus à St. Andrews Ecosse</i>
2014/2015	Licence de Lettres Modernes <i>Université Paris-Sorbonne Paris</i>
2012/2014	Classes préparatoires aux grandes écoles : section A/L <i>Lycée Fustel de Coulanges Strasbourg</i>

Langages de programmation

Python	<i>bibliothèques fréquemment employées :</i> pandas, scikit-learn, keras, bs4, lxml, streamlit, spaCy, NLTK
R	<u>Mesures statistiques linguistiques: analyse de corpus</u>
Java	<u>Programme d'extraction depuis une base de données</u> extraite d'ImDB
XML (TEI)/XPath /XSLT/XQuery	Extraction depuis les corpus existants de la plateforme Drama Corpora et annotation manuelle en XML-TEI.
HTML/CSS/JS	
SQL	

Langues

Français	langue maternelle
Anglais	deuxième langue (940/1000 TOEIC)
Espagnol	vécu et travaillé au Mexique et en Colombie pendant un an.