TAL et Humanités Numériques

M1 Humanités Numériques

2025-2026

Projet

Gaël Lejeune, Sorbonne Université

Diagnostic de faisabilité d'une étude de corpus avec des outils numériques

Les humanités numériques offrent un large éventail d'outils pour l'analyse de corpus textuels, iconographiques ou audiovisuels. Cependant, la mise en œuvre d'une étude numérique suppose des choix méthodologiques, techniques et épistémologiques précis. Avant toute analyse automatisée, il est essentiel d'évaluer la faisabilité d'un projet de recherche fondé sur un corpus donné.

Ce mini-mémoire a pour objectif de réaliser un **diagnostic de faisabilité** d'une étude de corpus mobilisant des outils numériques. Il ne s'agit pas nécessairement de mener une analyse exhaustive, mais d'évaluer rigoureusement les conditions de possibilité d'un tel projet.

1 Objectifs pédagogiques

Votre rendu devra montrer que vous êtes capable de :

— formuler une question de recherche relevant des humanités
— identifier et délimiter un corpus pertinent pour cette question
— évaluer la faisabilité technique et scientifique d'une analyse numérique
— justifier le choix des outils numériques adaptés à votre question
— identifier les limites, biais et enjeux épistémologiques du projet.

2 Corpus

Le corpus est *a priori* textuel (œuvres littéraires, presse, archives, discours, réseaux sociaux) mais il pourrait être aussi :

— iconographique (images patrimoniales, affiches, photographies)
— audiovisuel (entretiens, vidéos, podcasts)
— multimodal.

Le corpus peut être issu de bibliothèques numériques, d'archives ouvertes ou être en cours de constitution. Dans ce dernier cas, la faisabilité de la collecte devra être explicitement analysée.

3 Travail attendu

Le mini-mémoire (environ *5 pages*) devra comporter les parties suivantes :

3.1 Présentation du projet de recherche

— contexte scientifique
— question(s) de recherche
— intérêt du recours aux humanités numériques.

3.2 Description du corpus

— nature, taille et formats des données
— sources et conditions d'accès
— problèmes éventuels (droits, qualité des données, hétérogénéité).

3.3 Diagnostic de faisabilité numérique

— prétraitements nécessaires (OCR, nettoyage, normalisation, annotation, etc.)
— outils numériques envisagés (par exemple : Python, AntConc, Tesseract)
— compétences techniques requises
— estimation du temps et des ressources nécessaires.

Une expérimentation exploratoire sur un sous-corpus sera proposée afin d'illustrer le diagnostic, sans obligation d'analyse exhaustive.

3.4 Analyse critique et limites

— limites méthodologiques : qu'est-ce qui est difficile à "traduire" en méthodologie numérique ?
— biais liés aux outils ou aux données : quelle déformation ?
— alternatives possibles (méthodes qualitatives ou hybrides).

3.5 Conclusion

La conclusion proposera une évaluation globale de la faisabilité du projet et ouvrira sur des perspectives d'évolution possibles.

4 Critères d'évaluation

Le travail sera évalué selon les critères suivants :
— clarté et pertinence de la problématique
— qualité de la description du corpus
— compréhension des outils et méthodes numériques
— esprit critique face aux données et aux technologies
— qualité de la rédaction et de l'argumentation.