

Traduire la chute de Jérusalem : une approche par *topic modeling* comparé

Complément d'analyse

Alice Leflaëc

Ces quelques lignes sont le fruit d'une réflexion menée au cours de l'été 2024 en vue de la soutenance de mon mémoire pour le certificat de spécialisation en humanités numériques à l'Université de Genève et formaient la conclusion de ma présentation. Alors que tout le travail a été réalisé conjointement avec Mathilde Schwoerer, cette réflexion est personnelle.

Cette étude a été menée alors que nous avions déjà une connaissance plutôt bonne des deux œuvres comparées, en particulier du livre V des *Historiae*. Notre objectif n'était donc pas de découvrir les documents et les sujets qu'ils contiennent, mais de comparer certaines de nos intuitions et de nos impressions à la lecture des textes avec les résultats produits par la machine, et de croiser ainsi le *Close* et le *Distant Reading*.

En fin de compte, ce sont les limites de cette étude et de la méthode employée qui nous sont rapidement apparues et nous avons tâché de les exposer dans notre article. En effet, en dépit des solides algorithmes mathématiques utilisés, il est nécessaire de lire les résultats avec précaution. Cela tient avant tout à la part très forte de l'intervention humaine dans ce type d'analyse. Dans l'interprétation même des résultats, il peut être difficile pour l'œil humain d'estimer la distance réelle entre les *topics* déterminés par la machine en raison de leurs ressemblances et il faut reconnaître que, dans notre cas, la connaissance préalable des œuvres nous a bien aidées à distinguer les thèmes. Mais surtout, comme l'a montré notre expérience sur les pronoms personnels et les déterminants possessifs, le paramétrage des mots-vides à retirer de l'analyse, que nous effectuons manuellement, a de véritables répercussions sur la modélisation de sujets. Si ces limites concernent toute analyse de *topic modeling*, leur incidence est bien plus forte sur des corpus de taille aussi réduite que les nôtres.

En reprenant cette étude au cours de l'été, il m'a semblé, toutefois, que nous pouvions dresser un bilan plus positif de cette analyse. Certes, le *topic modeling* appliqué aux textes littéraires pose un certain nombre de problèmes et il convient de lire les résultats produits et leur interprétation avec précaution, mais l'expérience réalisée n'est pas dénuée d'utilité, à mon avis, pour l'analyse des *Historiae* du Pseudo-Hégésippe et pour une réflexion plus générale sur la manière d'étudier les réécritures littéraires.

- Le *topic modeling*, tout d'abord, permet de s'intéresser uniquement aux thématiques des œuvres là où le *Close Reading*, assez rapidement, regarde aussi les techniques de réécriture (par exemple l'amplification d'un champ lexical, le passage du discours direct au discours indirect...) et se concentre inévitablement sur les différences plutôt que sur les ressemblances. Pour une analyse à la fois plus nuancée et

plus riche d'une réécriture, il y a donc un intérêt à grouper les deux approches.

La modélisation de sujets réalisée sur *La Guerre des Juifs* et les *Historiae* nous a permis de nous éloigner de la construction des textes et de mieux mesurer la fidélité globale du Pseudo-Hégésippe par rapport à sa source grecque sur le plan thématique.

- Les tests effectués sur le maintien ou le retrait des pronoms personnels et des déterminants possessifs dans le livre V des *Historiae* ont également permis d'apprécier davantage l'importance et surtout le rôle des discours comme support de l'idéologie anti-juive et pro-romaine dans ce livre.
- Enfin, cette étude d'une traduction latine d'un texte grec nous pousse à réfléchir davantage aux éléments qui agissent sur l'esprit du lecteur dans un texte et, dans notre cas, à notre propre pratique de la traduction sur la perception d'un texte. En effet, notre esprit est marqué par l'esthétique de l'horreur qui imprègne le livre V des *Historiae* et pourtant, l'observation de la fréquence des *tokens* révèlent que les termes relevant du champ lexical du sang et de la mort ne font pas partie des mots les plus fréquents. Notre entreprise de traduction du livre V constitue sans doute un biais puisque l'exercice requiert d'accorder une attention particulière à chaque mot du texte et nous amène à nous arrêter sur certains termes sur lesquels nous aurions glissé plus rapidement lors d'une simple lecture. Nous constatons, cependant, que les scènes de famine s'ancrent davantage dans notre esprit que la description, par exemple, des différentes factions qui s'affrontent dans Jérusalem. C'est donc que certains termes semblent marquer davantage l'esprit du lecteur que d'autres et, alors que dans notre groupe de travail nous abordons désormais le commentaire du texte, l'expérience réalisée nous incite à étudier avec plus d'attention les mécanismes à l'œuvre dans le livre V des *Historiae* pour transmettre une idée et créer une atmosphère.

En fin de compte, toute cette expérience relevant du *Distant Reading* a été pensée en dialogue étroit avec le *Close Reading* que nous pratiquons par la traduction et le commentaire de texte. L'analyse de *topic modeling* comparé nous a permis de prendre du recul sur les œuvres étudiées pour y revenir de plus près avec un regard quelque peu renouvelé.