

# ADT vs Deep\_Learning dans la recherche de l'intertexte : le cas de Jérôme

Arianna Rosa

University of Liège – [Arianna.Rosa@uliege.be](mailto:Arianna.Rosa@uliege.be)

## Abstract

This contribution aims to assess the importance of literary genre both in identifying similarities between authors and texts and in identifying areas of intertextuality. The main corpus consists of the letters of Jerome (4th-5th century), one of the principal Latin Fathers of the Church, in comparison with a corpus composed of classical prose works belonging to different literary genres. The aim is to investigate which authors and, in particular, which works belonging to which literary genre influenced the style of Jerome's letters. For this purpose, we apply two different methodological approaches to these corpora, one based on traditional ADT methods (Hyperbase, Text Data Analysis software), the other on neural networks (Hyperdeep, Deep Learning model), to highlight their advantages and disadvantages.

**Keywords:** ADT, Deep Learning, intertextuality, Jerome, style.

## Resumé

La présente contribution vise à évaluer le poids du genre littéraire tant dans l'identification de proximités entre auteurs et textes que dans le repérage de zones d'intertextualité. Nous appuyons la réflexion sur les lettres de Jérôme (4ème-5ème siècle), l'un des principaux Pères latins de l'Église, en comparaison avec un corpus composé d'œuvres classiques en prose appartenant à différents genres littéraires. L'objectif est de rechercher quels auteurs et, en particulier, quelles œuvres appartenant à quel genre littéraire ont influencé le style des lettres de Jérôme. À cette fin, nous appliquons à ces corpus deux approches méthodologiques différentes, l'une fondée sur les méthodes traditionnelles de l'ADT (Hyperbase, logiciel d'Analyses des Données Textuelles), l'autre sur les réseaux neuronaux (Hyperdeep, modèle de Deep Learning), pour souligner leurs atouts et inconvénients.

**Mots clés :** ADT, Deep Learning, intertextualité, Jérôme, style.

## 1. Introduction

L'efficacité de la statistique textuelle pour rechercher des proximités langagières et stylistiques entre œuvres latines n'est plus à démontrer depuis longtemps (cf. Fialon et al. 2015). Par ailleurs, diverses études (cf. Vanni et al. 2024) ont récemment montré que la déconvolution des réseaux neuronaux pouvait servir à détecter des zones d'intertextualités. Notre contribution vise à évaluer le poids du genre littéraire tant dans l'identification de proximités entre auteurs et textes que dans le repérage de zones d'intertextualité. Dès lors, se pose la question de savoir dans quelle mesure les résultats varient en fonction de l'homogénéité ou de l'hétérogénéité du corpus du point de vue du genre : lors de la constitution d'un corpus de comparaison, vaut-il mieux se limiter à choisir œuvres d'un même genre littéraire ou ce choix pourrait-il conduire à l'exclusion de points de contact avec des œuvres de genres différents ?

Nous appuyerons la réflexion sur les lettres de Jérôme de Stridon, l'un des principaux Pères latins de l'Église : son corpus épistolaire (124 lettres) est comparable, pour la clarté de sa langue et l'ampleur des sujets traités, à celui des grands épistoliers classiques latins, Cicéron, Sénèque et Plinius (Hritz 1939; Canellis 2002). Dans la recherche de proximités stylistiques avec les auteurs classiques, nous pourrions donc penser à une influence de ces grands épistoliers, en

partant du principe qu'ils partagent une sorte de « manière d'écriture épistolaire » qui a pu influencer la manière dont Jérôme a composé ses lettres. Il faut en effet préciser que Jérôme était le Père de l'Église le plus érudit (Stoico 1972: 148), ayant été initié dès son plus jeune âge à la lecture des prosateurs et des poètes latins classiques, puis s'étant intéressé à l'exégèse biblique en langues grecque et hébraïque. Dans la composition de ses lettres, il est donc possible qu'il ait eu à l'esprit celles des grands épistoliers qui l'ont précédé.

Pour essayer de répondre à cette hypothèse, nous avons constitué un corpus de comparaison d'œuvres classiques en prose appartenant à différents genres littéraires (historiographie, épistolographie, traité, roman, etc.) avec l'objectif de rechercher quels auteurs et quelles œuvres appartenant à quel genre littéraire ont influencé le style épistolaire de Jérôme. À cette fin, nous appliquons à ces corpus deux approches méthodologiques différentes, l'une fondée sur les méthodes traditionnelles de l'ADT (Hyperbase, logiciel d'Analyses des Données Textuelles), l'autre sur les réseaux neuronaux (Hyperdeep, modèle de Deep Learning), pour souligner leurs atouts et inconvénients.

## **2. Les deux corpus**

### **2.1 Corpus de base**

Le corpus épistolaire de Jérôme est composé de 124 lettres écrites entre 370 et 420 : elles ont été lemmatisées avec une annotation morphosyntaxique de manière automatique grâce au programme Collatinus-LASLA. Ce dernier lemmatise automatiquement chaque mot des textes en latin en se basant sur les analyses déjà effectuées au sein de Collatinus et dans les fichiers du LASLA. Collatinus est un programme qui a été développé au début des années 2000 par Yves Ouvard pour l'enseignement avec l'objectif d'aider les étudiants en fournissant pour chaque forme d'un texte la traduction et des analyses morphologiques. Le LASLA (Laboratoire d'Analyse Statistique des Langues Anciennes) de l'Université de Liège a commencé un projet de lemmatisation et annotation morphosyntaxique des textes latins en 1961, avec l'objectif de créer des index et étudier les textes avec des méthodes statistiques et quantitatives. Ces deux projets ont uni leurs ressources en fournissant un vaste répertoire d'analyses et de lemmatisations sur lequel repose le lemmatiseur créé par Philippe Verkerk (cf. Verkerk et al. 2020), atteignant un taux d'efficacité global supérieur à 90 % (97 % sur les textes de Cicéron, cf. Verkerk 2024).

Pour ce qui concerne les analyses du texte de Jérôme, une partie des données produites par Collatinus-LASLA a été vérifiée par nos soins et nous avons constaté que le programme atteignait un haut niveau de précision, à l'exception des noms propres hébraïques. Ils font en fait partie du lexique de l'époque chrétienne, qui a été influencé par l'étude et l'exégèse des textes sacrés. Puisque la quasi-totalité des analyses sur lesquelles se base le lemmatiseur provient de textes latins de la période classique (78 av. JC-14), il n'est pas surprenant qu'ils n'aient pas été détectés, de même que certaines formes lexicales du latin tardif chrétien. Quoi qu'il en soit, notre objectif ne portant pas sur les noms propres, cette imprécision n'entrave pas nos recherches.

### **2.2 Corpus de comparaison**

Le corpus de comparaison est composé des œuvres classiques latines en prose de différents genres, déjà présentes dans les fichiers du LASLA et qui ont donc déjà été lemmatisées et annotées morphosyntaxiquement selon la procédure LASLA, LEI (LASLA Encoding Initiative

interface). Il s'agit d'une lemmatisation semi-automatique : un lemmatiseur compare chaque mot du texte à toutes les formes présentes dans le dictionnaire de formes LASLA. Après ce prétraitement, le texte est présenté sous forme d'une liste de tokens avec toutes les lemmatisations et analyses sous forme alphanumérique. Ensuite, le philologue doit sélectionner l'analyse correcte et, dans le cas où l'analyse n'était pas fournie, il doit l'ajouter (cf. Verkerk et al. 2020).

Les différents genres et œuvres littéraires inclus dans le corpus sont : le **genre historique** (César : *Bellum Civile*, *Bellum Gallicum* ; Salluste : *Catilina*, *Jugurtha*, *Fragmenta Historiarum* ; Tite-Live : *Ab Urbe Condita* ; Tacite : *Annales*, *Historiae*, *Agricola* ; Quinte-Curce : *Historiae Alexandri Magni*), les **traités** (Caton : *De Agri Cultura*, *Origines*, *Fragmenta reliqua* ; Cicéron<sup>1</sup> ; Sénèque : *Dialogi* ; Tacite : *Dialogus de Oratoribus*, *Germania* ; Cicéron), les **discours** (Cicéron : *Orationes*<sup>2</sup> ; Caton : *Orationum Fragmenta*), l'**épistolographie** (Sénèque : *Ad Lucilium* ; Pline : les lettres ; Cicéron : *Ad Brutum*, *Ad Atticum* I-III, *Ad Quintum*, *Ad Familiares* I-VI), la **biographie** (Cornélius Népos : *De Viris Illustribus* ; Svétone : *Vitae Caesarum* ; Tacite : *Agricola*), le **roman** (Pétron : *Satyricon*, Apuleius : *Metamorphoses*). Cette sélection d'œuvres constitue un panorama vaste et représentatif des œuvres principales des différents genres littéraires de la littérature latine classique, nous offrant un vaste champ de recherche.

### 3. Les méthodes traditionnelles de l'ADT : Hyperbase

Nous commençons notre exploration à partir de la fonction « distance intertextuelle » d'Hyperbase. Ce dernier est un logiciel qui s'inscrit dans la tradition de l'ADT depuis l'essor de l'informatique en France dans les années 80s. La fonction « distance intertextuelle » est associée à une analyse arborée et calculée selon la méthode Jaccard, qui se base sur le rapport entre les mots communs aux différents textes et ceux qui n'appartiennent qu'à un seul texte (cf. Vanni 2024).

Comme nous pouvons le voir sur la figure 1, Jérôme est placé sur la même branche que Sénèque (divisée entre « senecat », c'est-à-dire ses traités, et « senecalt », c'est-à-dire ses lettres), Caton, Pétrone et Apulée. Même s'il se trouve sur la même branche, il reste toutefois isolé par rapport aux autres auteurs mentionnés qui forment deux noyaux, celui de l'ensemble des œuvres de Sénèque et celui de Caton, Pétrone et Apulée. Cela signifie qu'au niveau lexical, aucune convergence significative n'a été trouvée entre Jérôme et les textes du corpus de comparaison. Le vocabulaire d'un auteur chrétien a en effet intégré des hellénismes et des hébraïsmes, dus à l'étude et à l'exégèse des textes sacrés, et c'est probablement pour cette raison qu'il a été isolé.

<sup>1</sup> *De Amicitia*, *De Imperio Cn Pompei*, *De Haruspicum Responso*, *De Officii*, *De Provinciis Consularibus*, *De Senectute*.

<sup>2</sup> *In Catilinam*, *De Lege Agraria*, *In Verrem Actio I-II*, *De Domo suan*, *Divinatio in Q. Caecilium*, *In L. Calpurnium Pisonem*, *In Vatinius*, *Post Reditum ad Quirites et in Senatum*, *Pro Ligario*, *Pro Caecina*, *Pro Cluentio*, *Pro Balbo*, *Pro Caelio*, *Pro C. Rabirio Perduellionis Reo*, *Pro Flacco*, *Pro Marcello*, *Pro M. Fonteio*, *Pro Milone*, *Pro M. Tullio*, *Pro Murena*, *Pro Plancio*, *Pro P. Quinctio*, *Pro Roscio Comoedo*, *Pro Rabirio Postumo*, *Pro Rege Deiotario*, *Pro Sestio*, *Pro Roscio Amerino*, *Pro Sulla*, *Philippicae Orationes*, *Pro Archia*.

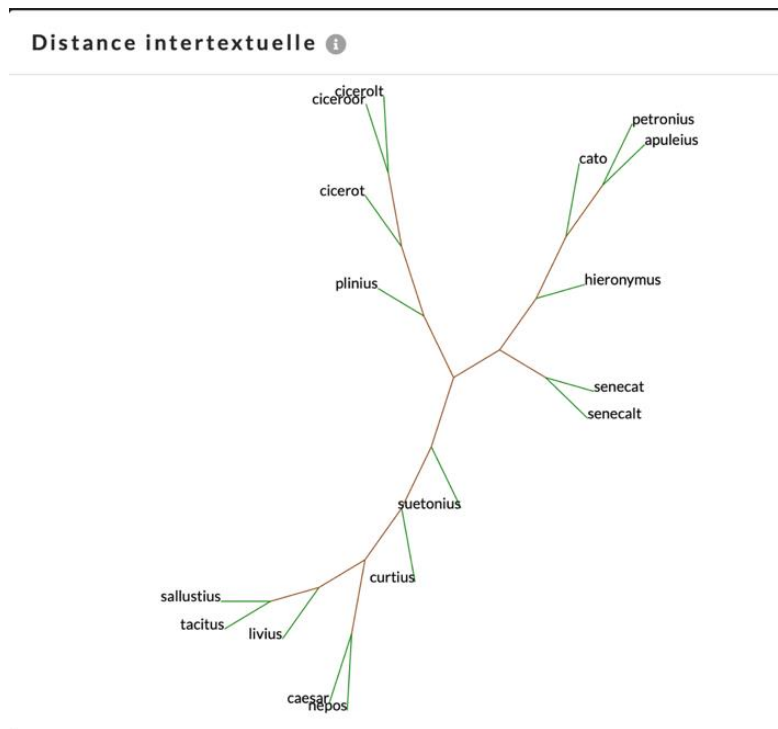


Fig. 1 Distance intertextuelle

Nous poursuivons l'analyse lexicale par l'Analyse Factorielle des Correspondances (AFC, fig. 2) sur la catégorie grammaticale des substantifs : les auteurs (en rouge) sont organisés sur deux axes, en fonction des mots les plus fréquemment utilisés dans leurs textes (en bleu). Nous pouvons constater que Jérôme continue d'être isolé, même s'il est un peu proche des discours de Cicéron (« ciceroor » sur le graphique). Ce qui pourrait sembler être un résultat négatif, constitue en réalité un bon point de départ pour caractériser l'auteur. En effet, sur le graphique AFC, plus un auteur ou une œuvre est éloigné du centre, plus les résultats qui ressortent le caractérisent. Nous avons en effet effectué une exploration du lemme « IVDICIVM », qui apparaît à côté du Jérôme (fig. 2) : à partir de la fonction « Thème », nous avons recherché des possibles cooccurrences entre formes, lemmes et catégories grammaticales qui pourraient définir le style de Jérôme. Cette recherche a abouti à mettre en évidence ce qui semble être un « motif textuel » qui lui est typique. Par « motif », nous désignons un patron linguistique multidimensionnel consistant en l'association récurrente de "n" éléments du texte pourvu de sa structure linéaire (Legallois 2006). Dans notre cas, il s'agit du motif « interpellatricem duri iudicis », attesté dans la littérature latine uniquement dans deux de ses lettres (lettres 60, 79) et dans un de ses commentaires bibliques (*Commentarii in Isaiam* 17,21)<sup>3</sup>. De plus, il est utilisé de la même manière dans les deux lettres : en faisant référence à la figure de la veuve insistante (dans la parabole racontée par Jésus dans Luc 18, 1-8), Jérôme compare l'insistance d'une personne (Avitus dans la lettre 79, Nepotianus dans la lettre 60) à demander une lettre de la part de Jérôme et à l'obtenir ainsi.

<sup>3</sup> Ces résultats proviennent de la base de données en ligne « Library of Latin Texts » (<https://clt.brepolis.net/llta/Search>) (accès : 07 mars 2026).

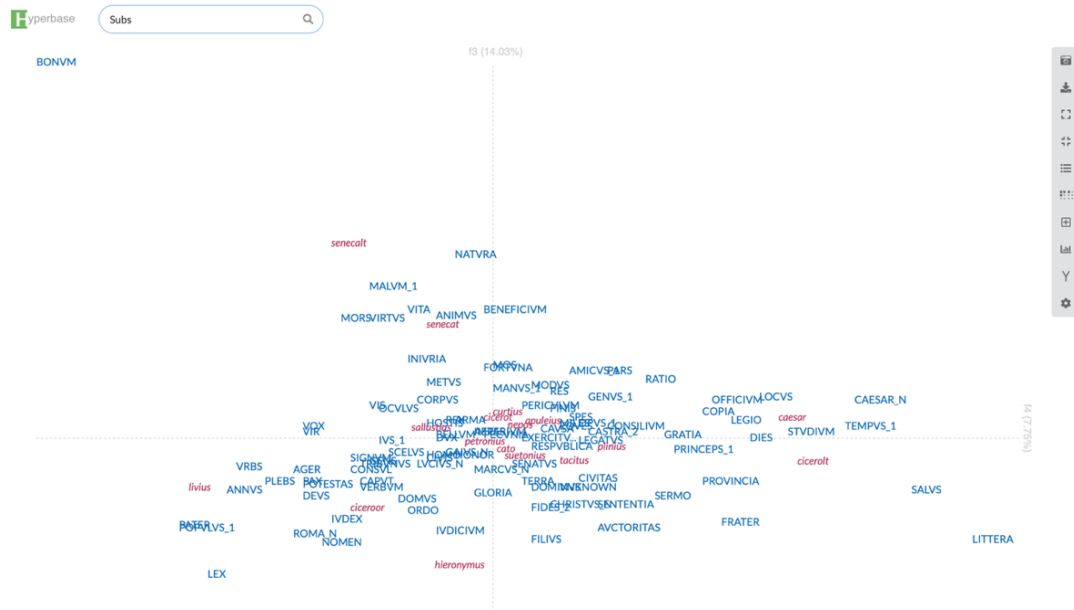


Fig. 2 Analyse Factorielle des Correspondances : substantif

Afin de neutraliser le vocabulaire chrétien de Jérôme et, par conséquent, de trouver des points communs avec le corpus de comparaison, nous avons procédé ensuite à une analyse morphosyntaxique, au moyen d'une Analyse Factorielle des Correspondances des Parts of Speech (POS). Comme dans le précédent AFC (fig. 2), Jérôme semble encore isolé par rapport aux autres auteurs, un peu rapproché de Cornélius Népos, mais proche des POS « subordonnée » et « pronom relatif » (fig. 3). En analysant ces parties du discours, on constate qu'effectivement, dans le corpus créé sur Hyperbase, la POS « pronom relatif » est spécifique à Jérôme, présentant le taux d'utilisation le plus élevé par rapport aux autres auteurs (en particulier, le lemme QVI\_1). Cette analyse nous conduit donc, une autre fois, à une caractérisation du style de Jérôme plutôt qu'à des points de convergence avec le corpus de comparaison.

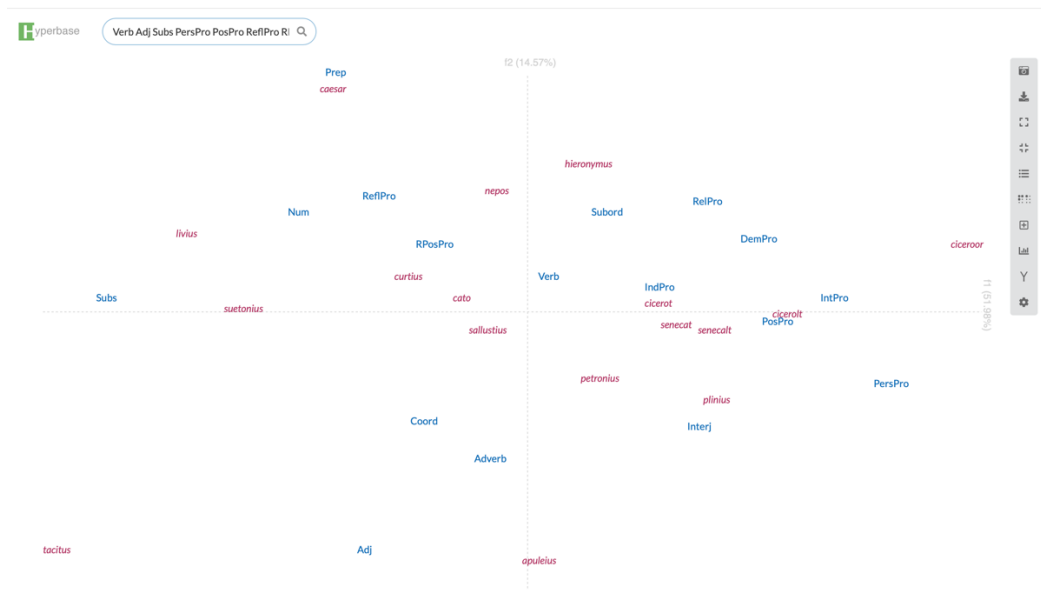


Fig. 3 Analyse Factorielle des Correspondances : Parts of Speech

Nous pouvons donc conclure provisoirement que les résultats obtenus avec l'approche d'ordre multidimensionnel de la lexicométrie (analyse arborée et AFC) nous ont plutôt conduit à une caractérisation de l'auteur étudié, grâce à la découverte d'un motif textuel, qu'à des proximités linguistiques entre le corpus principal et celui de comparaison. En outre, la méthode ne permet pas de prendre en compte les aspects du texte (lexique et morphosyntaxe) ensemble, mais seulement de manière isolée. S'agit-il d'une limite des méthodes traditionnelles de l'ADT ou plutôt d'une conséquence de la diachronie de notre corpus, c'est-à-dire du grand écart temporel entre Jérôme et les autres auteurs ? Les résultats obtenus grâce au Deep Learning permettront peut-être de répondre à cette question.

#### 4. Le Deep Learning : HyperDeep

Hyperdeep est un modèle de Deep Learning qui repose sur une architecture de neurones profonds (cf. Vanni et al. 2020) : il a été entraîné sur le corpus de comparaison et nous lui avons ensuite demandé de prédire l'attribution de la paternité des lettres de Jérôme pour retrouver des traces d'intertextualité d'autres auteurs. Les résultats obtenus, qui prend en compte trois facteurs principaux (forme, lemme et Part of Speech), indiquent, parmi les premières positions, Sénèque (23% ses traités, 22 % les lettres), Pline (19%) et Pétrone (7.6%) (colonne de gauche de la fig. 4). Si l'on pouvait s'attendre à ce que Sénèque et Pline, en tant qu'épistoliers, apparaissent hauts dans la liste, on ne s'attendait pas à ce que Pétrone, auteur de romans, soit également mentionné. En effet, dans l'analyse arborée (Fig.1), nous avons constaté une proximité avec Sénèque et Pétrone, mais pas une réelle convergence. De plus, Caton et Apulée ont maintenant chuté vers les dernières positions.

En explorant les passages clé mentionnés, c'est-à-dire des parties du texte où les réseaux neuronaux ont détecté un degré de certitude plus élevé sur les classes repérées (forme/code/lemme ; cf. Vanni et al. 2020), nous devons toujours vérifier les informations sur la base créée sur Hyperbase afin de s'assurer qu'ils correspondent à une utilisation effective par l'auteur mentionné et, surtout, qu'ils lui sont spécifiques dans le corpus choisi. Dans le cas des deux premiers auteurs mentionnés, les proximités repérées par le Deep Learning semblent surtout reposer sur certaines tournures linguistiques et cooccurrences spécifiques à ces auteurs et présentes dans la correspondance de Jérôme.

Dans le cas de Sénèque (lettres), nous trouvons, par exemple, la cooccurrence entre les formes « bonum » et « malum » (en bleu dans la figure 4) et le lemme du verbe être (en vert) : « bonum \*\*\* malum \*\*\* LEM:SVM\_1 ». Cette cooccurrence, spécifique des lettres de Sénèque dans le corpus sélectionné (fig. 5), apparaît à plusieurs reprises dans le corpus épistolaire de Jérôme. Une reprise qui semble directe est la cooccurrence « *nec bonum nec malum est* » présente dans trois passages de Sénèque (*De beneficiis* 1, 6 ; *Ad Marciam* 19, 5 ; *Ad Lucilium* 99, 12) et deux de Jérôme (*Comm. In Ezechielem* 6, 20 ; *Epist.* 112, 16)<sup>4</sup>.

<sup>4</sup> Ces résultats proviennent de la base de données en ligne « Library of Latin Texts » (<https://clt.brepolis.net/llta/Search>) (accès : 07 mars 2026).

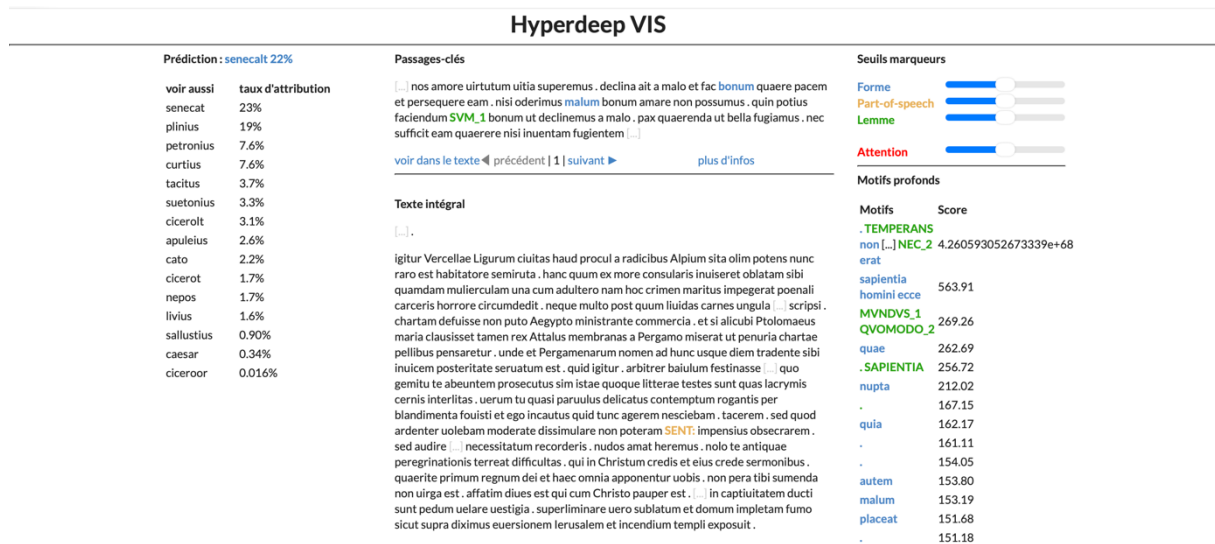


Fig. 4 Hyperdeep et l'influence de Sénèque sur Jérôme : un passage clé

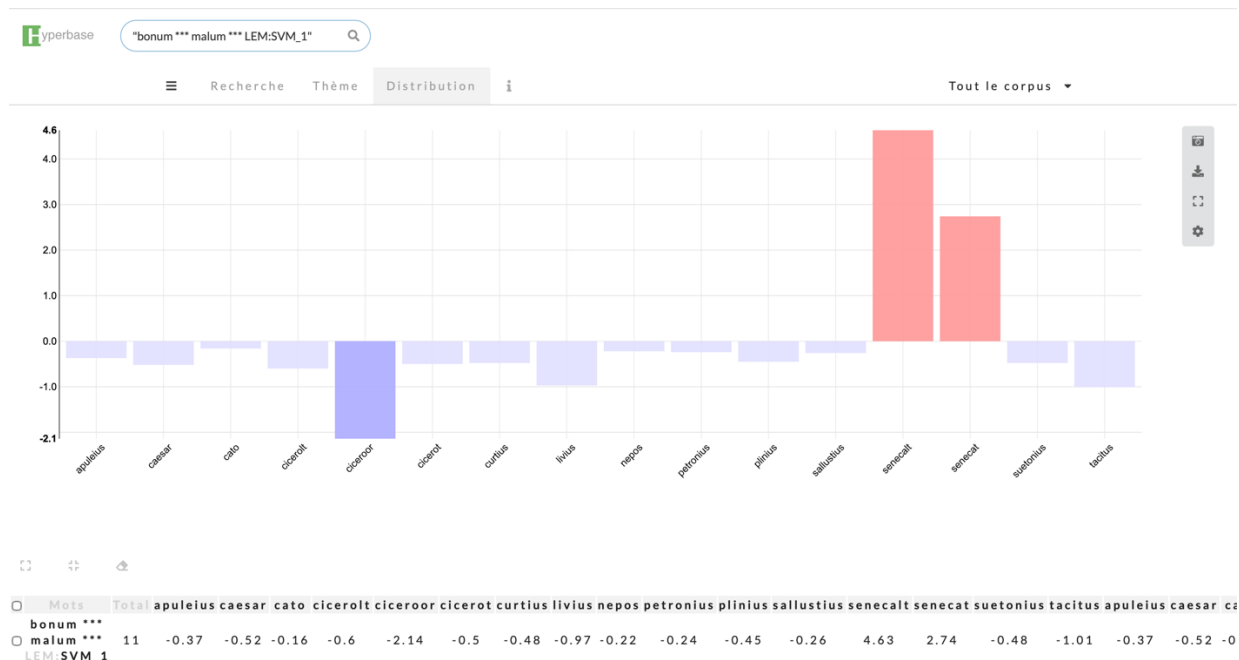


Fig. 5 Fréquence absolue de la cooccurrence « bonum \*\*\* malum \*\*\* LEM:SVM\_1 » chez Sénèque dans le corpus de comparaison sur Hyperbase

En ce qui concerne Pline, nous trouvons également des tournures linguistiques qui, dans ce cas, semblent caractériser le genre épistolaire, comme la cooccurrence « LEM:EXPRIMO \*\*\* LEM:VERBVM ». Il est spécifique à Pline dans le corpus sur Hyperbase et nous le retrouvons plusieurs fois dans le corpus épistolaire de Jérôme et dans celui de Sénèque. Il s'agit en effet d'une tournure que nous continuons nous aussi à utiliser souvent dans les cartes, les messages ou les lettres (par exemple, « je ne sais pas exprimer avec des mots... »).

Dans le cas de Pétrone, en revanche, la situation semble assez différente : ce qui le rapproche de Jérôme ne semble pas être des traits stylistiques, mais plutôt un type précis de lexique. Il s'agit d'un lexique concret, qui dénote la réalité quotidienne. Parlons de mots tels que « manduco » (manger), « tonsor » (tondeur), « discumbo » (se coucher). Ces lemmes concernent

une réalité « ordinaire », qui est généralement exclue des œuvres littéraires dans le but d'« élever » le ton et le contenu. Ce lexique est donc plus facilement observable dans les genres littéraires moins élevés, comme les comédies et le roman. Dans le cas particulier du *Satyricon* de Pétrone, le lexique non seulement bas, mais aussi vulgaire est volontairement bien exploité pour des raisons thématiques. Pour ce qui concerne Jérôme, cependant, la situation est différente : dans ce type de lexique, nous trouvons les traces de ce que l'on peut appeler le « latin chrétien ancien » (cf. Schrijnen 1977). Il s'agit d'un type de latin qui se forme progressivement sous différentes influences : il est caractérisé par la continuité avec le latin des grands auteurs classiques (Cicéron, Sénèque, Pline) par d'érudits, tels que Jérôme, qui sont, également, les premiers exégètes latins des textes sacrés, qui ont lu et connu la *Vetus Latina*, l'ensemble des traductions latines de la *Bible* réalisées entre le II<sup>e</sup> et le IV<sup>e</sup> siècle, antérieures à la *Vulgate*, la traduction effectuée par Jérôme lui-même. La langue de la *Vetus Latina* est une langue populaire, la langue des couches les plus basses de la population, celles qui sont les premières à adhérer au christianisme et auxquelles la religion devait être accessible et compréhensible. Les résultats que nous avons obtenus nous ont montré ces deux facettes de la langue de Jérôme, grand connaisseur des textes latins classiques comme des textes sacrés, en nous permettant aussi de les explorer en détail.

En conclusion, les résultats obtenus grâce au Deep Learning nous ont permis d'explorer davantage les points de contact entre Jérôme et les auteurs du corpus de comparaison, non seulement au niveau lexical, mais aussi au niveau de l'intersection entre différentes catégories linguistiques (lemmes, formes, POS). Si, dans le cas de Sénèque et de Pline, nous avons pu retracer des tournures linguistiques qui pourraient être identifiées comme des traces d'intertextualité, dans le cas de Pétrone, une influence semble moins plausible, mais les résultats nous ont permis d'approfondir une partie de son vocabulaire et celui du latin chrétien ancien.

## 5. Conclusions

En conclusion, les résultats obtenus avec les deux méthodes s'avèrent différents, mais complémentaires. Grâce à l'ADT classique, nous n'avons pas trouvé de points de contact forts avec les textes du corpus de comparaison, mais plutôt une caractérisation de l'auteur examiné. Par contre, grâce aux neurones profonds du Deep Learning, nous avons révélé des points de convergence avec les œuvres du corpus de comparaison, même avec des œuvres appartenant à un genre littéraire différent. Le Deep Learning semble donc capable de neutraliser le vocabulaire chrétien, en nous permettant de comparer des auteurs chrétiens et classiques, sans perturbations liées aux différences lexicales et diachroniques entre les corpus. La deuxième méthodologie s'avère donc plus efficace que la première dans la recherche d'intertextualité : le Deep Learning, en fait, permet de prendre en compte les différents niveaux linguistiques (lexique et morphosyntaxe) dans leur ensemble, ce qui n'est pas possible dans les méthodes de l'ADT traditionnelle.

Il est important toutefois de souligner que les deux méthodologies s'entrecroisent : il faut toujours s'assurer sur la base créée dans Hyperbase que les résultats obtenus grâce au Deep Learning sont spécifiques à l'auteur auquel ils sont liés. De plus, une autre étape indispensable dans l'utilisation des deux méthodologies est l'interprétation des résultats. Tant sur Hyperbase que sur Hyperdeep, il est important de comprendre pourquoi certaines formes ou traces d'intertextualité apparaissent : dans notre cas, par exemple, une intertextualité avec Pétrone est peu probable, mais comprendre les sorties de la machine nous a permis d'approfondir le lexique de Jérôme. De la même manière, la constitution du corpus de comparaison semble donc dépendre de la méthodologie choisie. Si, dans Hyperbase, il semble préférable d'utiliser un

corpus présentant des points de contact formels et diachroniques avec le texte examiné, dans Hyperdeep, il semble préférable d'utiliser un corpus plus large afin de retrouver des traits saillants communs à des œuvres hétérogènes en termes de genre littéraire et de chronologie.

## Références

Brepols. (2026). *Library of Latin Texts – Search*. Brepols, <https://clt.brepols.net/llta/Search>. Accessed 7 Mar. 2026.

Canellis, A. (2002). La lettre selon saint Jérôme: l'épistolarité de la correspondance hiéronymienne. In Nadjó, L. and Gavoille, É., editors, *Epistulae antiquae. 2, Actes du IIe colloque international « Le genre épistolaire antique et ses prolongements européens » : (Université François-Rabelais, Tours, 28-30 septembre 2000)*, pp. 311-332.

Fialon, S., Longrée, D. and Pietquin, P., editors. (2015). *Langues anciennes et analyse statistique : cinquante ans après – Distances textuelles et intertextualités*, Les Études classiques, 82.

Hritz, J. N. (1939). *The style of the Letters of St. Jerome*. Catholic University of America Press.

Schrijnen, J. (1977) *I caratteri del latino cristiano antico*: Pàtron.

Stoico, G. (1972). *L'epistolario di San Girolamo: studio critico-letterario di stilistica latina*. Giannini.

Vanni, L. (2024). Hyperbase Web. (Hyper)Bases, Corpus, Langage. *Corpus*, 25.

Vanni, L., Corneli, M., Longrée, D., Mayaffre, D. and Precioso, F. (2020). Deep learning descriptif pour l'analyse de données textuelles. In *Proc. of JADT 2020 15èmes Journées Internationales d'Analyse statistique des Données Textuelles, Jun 2020, Toulouse, France*.

Vanni, L., Mahmoudi, H., Longrée, D. and Mayaffre, D. (2024). Multichannel Convolutional Transformer and Intertextuality: A Latin Case Study. In Giordano, G., Misuraca, M., editors, *New Frontiers in Textual Data Analysis. JADT 2022. Studies in Classification, Data Analysis, and Knowledge Organization*, pp. 197-207.

Verkerk, P. (2024). Elaboration of a practical lemmatiser for Latin using Artificial Intelligence. *Archivum Latinitatis Medii Aevi*, 80.

Verkerk, P., Ouvrard, Y., Fantoli, M. and Longrée, D. (2020). L.A.S.L.A. and Collatinus: a convergence in lexis. In Tesconi, L., editor, *Studi e saggi linguistici 2020* (1), pp. 1-26.