



**HAL**  
open science

# Antiquité tardive et littératures latines : corpus et perspectives numériques

Thibault Clérice

► **To cite this version:**

Thibault Clérice. Antiquité tardive et littératures latines : corpus et perspectives numériques. *Κοινωνία = Koinōnia*, 2022, 46, pp.207-215. hal-03943557

**HAL Id: hal-03943557**

**<https://hal.science/hal-03943557>**

Submitted on 17 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Antiquité tardive et littératures latines : corpus et perspectives numériques

Thibault Clérice

Juin 2022

## Introduction

Une rétrospective de trente ans sur le champ des études latines tardo-antiques peut paraître très courte pour un domaine aussi étudié, mais elle englobe presque la totalité du travail fait dans le domaine numérique, du moins du côté latin. Il existe très peu de travaux avant les années 1990 qui ne concernent pas les « classiques », si ce n'est du côté de la *Vulgate* ou de l'épigraphie<sup>1</sup>.

Pour autant, la situation du domaine tardo-antique doit intriguer dans l'histoire de l'informatisation du champ des études latines. Il est comme « l'oublié » d'un champ qui s'est informatisé très tôt. Dès 1949, Roberto Busa, sorte d'Hérodote des *Digital Humanities* pour bon nombre de chercheurs, lance un projet d'annotation du corpus de Thomas d'Aquin sur fiches perforées. Le 13 septembre 1961, le Laboratoire d'Analyse Statistique des Langues Anciennes (LASLA) est fondé avec comme premier objectif de traiter les oeuvres complètes de Sénèque, sous la direction de Louis Delatte et Étienne Évrard<sup>2</sup>. Après cette percée des équipes italiennes et belges, ce sont les corpus grecs, dont le *Thesaurus Linguae Graecae* (1972), qui sont au centre de l'attention. Le latin ne reprend le devant de la scène qu'avec PHI (1987) et une mise à jour de *Perseus* (1996)<sup>3</sup>.

Nous commencerons donc notre réponse par un « bilan », qui se résumerait presque en une histoire rapide de la naissance des corpus latins tardo-antiques depuis 1991. Nous esquisserons ensuite quelques perspectives et soulignerons les enjeux qui existent pour l'étude de l'antiquité tardive dans le contexte des *digital humanities* et des études classiques.

## 1 Une brève histoire de corpus

Latin médiéval et latin classique ont donc une histoire et forment presque avec Busa et le LASLA des points d'origine pour les humanités numériques. Des corpus numériques tardo-antiques latins apparaissent cependant dès 1991 et la période connaît, par à-coups, quelques importantes phases de croissance, de l'adoption de la période par *Perseus* à l'avènement du DigilibLT.

---

1. Ce court rappel historique nous permet d'ailleurs d'avertir nos lecteurs que nous ne traiterons ni de l'épigraphie ni de la papyrologie numérique, tout simplement par manque de connaissance d'un champ qui connaît une histoire extrêmement riche du point de vue numérique et qui s'est développé assez indépendamment du domaine littéraire.

2. Louis Delatte et Étienne Évrard, « Un laboratoire d'analyse statistique des langues anciennes à l'Université de Liège », *L'Antiquité Classique*, 30 (1961), p. 429-444, doi : 10.3406/antiq.1961.3417.

3. Il existe un corpus antérieur à ces deux derniers, l'*American Philological Association's Repository of Latin Texts in Machine-Readable Form*, mais son maintien n'était, semble-t-il, plus assuré en 1986 d'après J. J. Hughes. Cf. John Jay Hughes, *Bits, bytes & biblical studies : a resource guide for the use of computers in biblical and classical studies*, 1987, URL : <http://archive.org/details/1s/bitsbytesbiblica0000hugh> (visité le 12/10/2021), p. 593 ; Theodore F. Brunner, « Classics and the Computer. The History of a Relationship », dans *Accessing Antiquity : the Computerization of Classical Studies*, dir. John Solomon, 1993.

## 1.1 Premiers corpus : l'influence de la patristique

En 1968, Paul Tombeur, après sept années d'expérience et de collaboration avec le LASLA, fonde à Louvain le CEntre de Traitement Electronique des DOcuments (CETEDOC), dont le projet annoncé est de se charger du traitement des oeuvres médiévales<sup>4</sup>. En 1981, au « Congrès international informatique et sciences humaines » organisé par le LASLA, P. Tombeur annonce vouloir publier un *Thesaurus Patrum Latinorum*, reprenant ainsi les textes tardo-antiques et médiévaux des trois séries imprimées<sup>5</sup>. Il faut en fait attendre 1991 pour que soit publié le corpus sous le nom de *CETEDOC Library of Christian Latin Texts on CD-ROM*, plus souvent connu sous son acronyme de CLCLT, chez Brepols. Contrairement au PHI, au TLG ou à *Perseus*, le CLCLT n'est pas un corpus dont l'interface est pensée pour la lecture ou l'impression : on lit des comptes-rendus faisant état d'une limite de trente lignes pour l'impression ou de l'impossibilité de sauter à un chapitre précis sans avoir à faire défiler le reste du texte<sup>6</sup>. Le CLCLT sert d'abord à produire des concordanciers et à chercher des occurrences et ne peut agir qu'en support à une bibliothèque papier, permettant tout de même de rendre certaines tâches beaucoup plus rapides<sup>7</sup>.

Dans les mêmes années, un éditeur privé, l'entreprise Chadwyck-Healey, met à disposition du public une version numérisée de la *Patrologia Latina* de Migne. Basée sur une interface DynaText et SGML TEI<sup>8</sup>, la *Patrologia Latina Database* (PLD) permet alors de lire et chercher à l'intérieur d'un corpus immense pour l'époque. On peut parler à ce moment précis de l'histoire des corpus d'une collision, tant les éditeurs des deux corpus se feront face. À demi-mot, mais non sans violence, P. Tombeur reproche à la PLD de n'être qu'une version informatisée de la PL, scientifiquement douteuse pour de nombreuses attributions, et d'être extrêmement chère si l'on prend en compte l'objectif premier de Migne qui « la présentait comme une *bibliotheca oeconomica*<sup>9</sup> ». Si les comptes-rendus de l'époque abondent dans le sens de P. Tombeur quant à la question du sérieux scientifique de l'oeuvre de Migne au regard des attendus contemporains, certains comme R. W. Crown y voient tout de même une alternative sérieuse à une bibliothèque papier, contrairement au CLCLT<sup>10</sup>.

## 1.2 Corpus ouverts et mise en ligne

Il faut attendre de nombreuses années avant l'arrivée sur le « marché » des corpus des avancées sur le terrain tardo-antique latin. Nous identifions cinq acteurs majeurs : *Perseus* et sa composante *Open Greek and Latin* (OGL) pendant le passage de Gregory Crane à Leipzig, *Musisque Deoque* (MQDQ), la *Digital Library of late antique Latin texts* (DigilibLT), *Hyperdonat* et les corpus thématiques.

*Perseus* est rentré tardivement dans la course à la mise à disposition de corpus latins, mais s'est très peu intéressé aux textes tardifs en dehors de quelques classiques (Augustin, l'*Histoire Auguste*, Ausone). En 2013, G. R. Crane obtient la « chaire Humboldt pour les Humanités Numériques » à Leipzig et profite de grandes améliorations de la reconnaissance de texte automatique et d'un financement conséquent pour lancer une nouvelle campagne de mise à disposition de textes à travers la captation des textes du CSEL<sup>11</sup> et de la PL. Si le projet est géré par la chaire, la production des données est externalisée

4. Michel Gueret, « Analyse informatique et compréhension textuelle. Le CETEDOC : Le laboratoire de recherches textuelles par ordinateur à l'Université Catholique de Louvain. » *Revue LASLA*, 3-1977 (1977), p. 45-72, URL : <http://web.philo.ulg.ac.be/rissh/wp-content/uploads/sites/10/pdf/Annee1977/03/MGueret.pdf>.

5. *Corpus Christianorum, Series latina et Continuatio Mediaevalis*, cf. Paul Tombeur, « La constitution d'un thésaurus pour les pères de l'église latine », dans *Actes du Congrès international informatique et sciences humaines*, Liège, 1981, p. 870-880.

6. Tim Bucknall et Christopher M. McDonough, « Review of The CETEDOC Library of Christian Latin Texts on CD-ROM », *The Classical Journal*, 90-1 (1994), p. 90-96.

7. On note par exemple l'expérience de Zahn identifiant un fragment grâce au CLCLT, Peter Zahn, « Kirchenväter-Texte im CD-ROM-Speicher : » 16-3 (janv. 1992), p. 419-427, DOI : 10.1515/bfup.1992.16.3.419.

8. MacKenzie Smith, « DynaText : An Electronic Publishing System », *Computers and the Humanities*, 27-5/6 (1993), p. 415-420.

9. Elle est annoncée à 70 000\$ en 1990 contre les 3 800\$ du CLCLT de 1994 ou les 1 035\$ du TLG de 1993.

10. Ron W. Crown, « Comparing the Patrologia Latina and the CETEDOC Library of Christian Latin Texts Databases from a User's Perspective », *Journal of Religious & Theological Information*, 3-1 (juin 2000), p. 85-109, DOI : 10.1300/J112v03n01\_06.

11. Gregory Ralph Crane, Alison Babeu, Lisa Cerrato, Matthew Munson, Stella Dee, Annette Gessner, Bruce Robertson, Greta

(OCR, correction et mises en XML), dans la grande lignée des premiers projets gréco-latins. Les textes présentent encore des erreurs de reconnaissance, en particulier des césures malheureuses pour la recherche de termes en plein texte comme *sae-culum* ou *taberna-culum (sic)*<sup>12</sup>. Une lecture « humaine » n'est pas véritablement freinée par ce genre d'erreurs, et, avec précaution, ils peuvent servir à faire des études quantitatives. Malheureusement, le travail sur la *Patrologia Latina* de mise en forme suivant les mêmes standards que le reste de *Perseus* n'a pas été achevé, ne lui permettant pas d'être exploité avec les mêmes outils<sup>13</sup>. La PL connaît cependant une seconde vie sur le site du *Corpus Corporum*<sup>14</sup> où il semble avoir bénéficié d'une forme de correction.

Il faut attendre les CDs de *Poesis* (1995, 1999<sup>15</sup>) puis *Musisque Deoque* (MQDQ, 2005<sup>16</sup>) pour sortir de la patrologie sur support physique puis sur internet. Les CDs incluent dans leur sélection de poésie latine une assez bonne représentation de la période tardo-antique, tant est qu'elle est même citée comme un outil remarquable pour contextualiser des auteurs tardifs<sup>17</sup>. Le projet MQDQ s'étend dans sa première mise en place de 2005 à 2010 via deux financements successifs. Il a l'originalité de proposer dès 2010 des notes pour 30% des vers qui le composent. Cet appareil est constitué à la fois de notes informatives, mais aussi des variantes dont l'importance est jugée suffisante par l'équipe.

Dans la même veine que MQDQ, il existe plusieurs corpus thématiques qui couvrent uniquement ou en partie la période tardoantique. Le plus connu, mais aussi le plus inaccessible jusqu'en 2022 était le *Corpus Grammaticorum Latinorum*<sup>18</sup>. Ce corpus a connu une période d'inaccessibilité de plusieurs années<sup>19</sup>, et c'est seulement à travers un partenariat avec le *Corpus Corporum* de Roelli que les textes nous étaient restés accessibles. Éditions fac-similaires des textes latins de grammairiens du III<sup>e</sup> au VIII<sup>e</sup> siècle réunis par Heinrich Keil pendant la seconde moitié du XIX<sup>e</sup> siècle, ils sont corrigés et augmentés à partir d'éditions plus récentes dont les sources sont citées par les responsables scientifiques Alessandro Garcea et Valeria Lomanto. Aux côtés du CGL, on trouvera aussi le corpus *Droit Romain*, qui, à défaut de présenter une interface de recherche, nous permet d'accéder à une quantité non négligeable de textes juridiques, dont le corpus théodosien ou justinien.

Les deux derniers corpus représentent à eux deux de grandes révolutions. Tout d'abord, la *Digital Library of late antique Latin texts* (DigilibLT, 2010) est le premier corpus et bibliothèque numérique dont l'objectif est de mettre à disposition des textes tardoantiques, en particulier ceux inaccessibles et donc en dehors de la patrologie. On y trouve tour à tour grammairiens, *vitae*, textes de médecine, d'astronomie ou de science, poésie, etc. Véritable mine d'or, elle offre par-dessus tout la possibilité de télécharger tout ou partie de ses textes, mis à disposition librement<sup>20</sup> en suivant les principes d'« *Accesso aperto, licenze*

Franzini, Simona Stoyanova, Tabea Selle, et al., *CSEL - Open Greek And Latin*, Corpus, 2021, URL : <https://github.com/OpenGreekAndLatin/csel-dev> (visité le 27/12/2021).

12. L'avantage de *Perseus* étant qu'il est *open-source*, il est tout à fait possible d'aider à corriger ce genre d'erreurs quand on les croise : rien n'est imprimé *ad vitam aeternam*.

13. Les textes latins et grecs de *Perseus* sont édités suivant les *guidelines* Capitains, afin d'être lisibles et interprétables par la machine. Pour une introduction sur Capitains, lire Thibault Clérice, « Les outils CapiTainS, l'édition numérique et l'exploitation des textes », *Médiévales. Langues, Textes, Histoire*, 73-73 (déc. 2017), p. 115-131, DOI : 10.4000/medievales.8211.

14. Philipp Roelli, « The Corpus Corporum, a new open Latin text repository and tool », *Archivum latinitatis medii aevi : Bulletin du Cange-72* (2014), p. 289-304, URL : <https://doi.org/10.5167/uzh-171105>.

15. Paolo Mastandrea et Luigi Tessarolo, *Poesis 2 CD-ROM dei testi della poesia latina*, 1999.

16. Federico Boschetti, Angelo Mario Del Grosso et Linda Spinazzè, « La galassia Musisque Deoque : storia e prospettive » (, déc. 2021), DOI : 10.30687/978-88-6969-557-5/026.

17. Jeanne-Marie Demarolle, « Anne-Marie Turcan-Verkerk, Un poète latin chrétien redécouvert : Latinus Pacatus Drepanius, panégyriste de Théodose. » *L'Antiquité Classique*, 75-1 (2006), Publisher : Persée - Portail des revues scientifiques en SHS, p. 369-369, URL : [https://www.persee.fr/doc/antiq\\_0770-2817\\_2006\\_num\\_75\\_1\\_2609\\_t1\\_0369\\_0000\\_2](https://www.persee.fr/doc/antiq_0770-2817_2006_num_75_1_2609_t1_0369_0000_2) (visité le 07/06/2022).

18. Alessandro Garcea, Franck Cinato et Clément Plancq, « Corpus Grammaticorum Latinorum. Un projet de traitement informatique autour des grammairiens latins », *Rivista di Cultura Classica e Medievale*, 52-2 (2010), p. 377-400, URL : <https://halshs.archives-ouvertes.fr/halshs-00735167>.

19. La nouvelle adresse est <https://htldb.huma-num.fr/exist/apps/cgl/index.html>.

20. C'est grâce à cette politique d'ouverture que leur inclusion dans un travail de traitement automatique des textes dans mes recherches a été possible.

*Creative Commons, software libero* »<sup>21</sup>. Comme *Perseus*, la *DigilibLT* continue de vivre et de s'agrandir, permettant chaque année de retrouver de nouveaux textes dans ses collections.

Enfin, il reste à présenter le projet *Hyperdonnat* de Bruno Bureau, Christian Nicolas et Maud Ingarao. Il est à notre connaissance la seule édition nativement numérique de textes classiques ou tardo-antiques en latin. Le projet est développé en TEI pour mettre à profit la possibilité de constituer des témoins « virtuels » pour l'oeuvre de Donat dont la tradition est complexe et qui ne peut s'éditer comme un texte linéaire<sup>22</sup>. Ainsi, le projet représente une première mise à profit des possibilités du numérique pour dépasser les limites imposées par le papier à l'ecdotique latine<sup>23</sup>. À l'image des CGL, *Hyperdonnat* connaît une légère pause dans son développement informatique, et si les données déjà publiées sont encore en ligne, d'autres textes sont encore bloqués.

Trente ans auront suffi à mettre en place plusieurs projets de corpus, passant d'un vide important de la fin des années quatre-vingt à une période avec des corpus imposants (OGL, *DigilibLT*) et des oeuvres innovantes (MQDQ et *Hyperdonnat*) gratuitement accessibles, facilitant ainsi l'accès à cette période. Il est probable que d'autres projets existent, et il faudra excuser l'auteur de leur oubli, qui s'est limité aux projets assez généralistes et de corpus. Il est important pour la communauté d'exploiter les outils à sa disposition pour faire connaître ces corpus, en particulier le Wiki des *Digital Classicists*<sup>24</sup>.

## 2 Perspectives

Trente années ont certes suffi à mettre à disposition des corpus, mais de nombreux défis restent à disposition de la communauté. Dans un contexte de renouvellement des méthodes et pratiques de recherche, et en particulier avec l'avènement des méthodes computationnelles et d'apprentissage machine, la question des corpus ouverts, lisibles pour la machine et de qualité doit devenir centrale et permettre de dépasser la question de l'accumulation des textes de mauvaise qualité ou celles des textes uniquement prévus pour un usage « humain ».

### 2.1 Le problème de la *Patrologia Latina*

Dans son article de 1993, P. Tombeur reproche surtout à la *PLD* de Chadwick d'être une adaptation du travail de Migne, qui serait truffé de fautes, où les annotations des mauristes sont incorporées au texte principal et où l'attribution des textes est datée, ce qui poserait de véritables problèmes pour le chercheur n'étant pas au fait des avancées sur ce point<sup>25</sup>. Ce problème, toujours présent dans les versions de la PL librement disponible sur *Perseus* ou le *Corpus Corporum*, vient s'ajouter à celui de textes en XML TEI où les références font partie du texte sans être exclues par un quelconque artifice techniquement interprétable par la machine. Les citations directes dans les commentaires ou les références données par les éditeurs telles que « (Luc. XVI, 16) » partagent le même statut numérique que le corps de texte des auteurs, posant un problème pour le levier statistique. Ainsi, si la patristique a été tout d'abord l'enfant chéri de l'antiquité tardive du point de vue numérique, elle est celle qui, assurée d'une avance, est désormais en retard en termes de qualité des données. Si l'on veut aborder cette période dans son

21. Maurizio Lana, « Metodologie e problematiche per una biblioteca digitale. Il caso di digilibLT », *Digitalia*, 1–0 (nov. 2012), p. 40–64, URL : <http://digitalia.sbn.it/article/view/516> (visité le 14/12/2021).

22. Bruno Bureau, Christian Nicolas et Maud Ingarao, *Hyperdonnat, commentaire attribué à Aelius Donat aux comédies de Térence*, 2008, URL : <https://hal-univ-lyon3.archives-ouvertes.fr/hal-00879555>.

23. B. Bureau, C. Nicolas et Ariane Pinche, « Encoding (inter)textual insertions in Latin "grammatical commentary" », *Journal of Data Mining and Digital Humanities*, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages (déc. 2017), DOI : 10.46298/jdmdh.1392.

24. En particulier, [https://wiki.digitalclassicist.org/Category:Late\\_Antiquity](https://wiki.digitalclassicist.org/Category:Late_Antiquity).

25. P. Tombeur, « La « Patrologia Latina Database (PLD) » ou la reproduction informatisée d'un corpus textuel établi selon une heuristique et une méthodologie éditoriale dépassées », *Bulletin de Philosophie Médiévale*, 35 (janv. 1993), p. 223–228, DOI : 10.1484/j.bpm.3.478.

intégralité, il faudra dépasser le simple fac-similé de la patrologie et retravailler ces données dans un contexte de corpus ouverts et mis à jour, une véritable nouvelle PL, accessible et complète.

## 2.2 Le temps des éditions nativement numériques ?

Nous avons beaucoup parlé de corpus et finalement assez peu d'édition. Aucun des corpus cités n'est un corpus d'éditions critiques, et si MQDQ apporte quelques annotations, nous restons encore bien loin de ce que peuvent offrir, en termes d'approche des traditions textuelles, les collections imprimées des différents grands éditeurs.

Or, la production d'éditions critiques, nativement numériques ou adaptées et traduites de documents originaux, semble aujourd'hui une condition *sine qua non* de l'exploitation pleine et entière des outils computationnels pour l'étude de la philologie classique, tardoantique ou médiévale<sup>26</sup>. Par exemple, du côté de la philologie vernaculaire, la stematologie computationnelle a permis d'appuyer dans deux cas au moins des études de traditions<sup>27</sup>. Seule l'édition numérique peut aider à produire ce type d'études et à confronter les approches et ainsi faire avancer nos connaissances. Sur le même thème, le projet *Open Stemmata*, à travers une collection des *stemmata codicum*, promet de renouveler ou appuyer notre perception des phénomènes de transmission des textes<sup>28</sup>. Une approche nativement numérique permet aussi d'envisager de nombreuses révolutions : étude des variations dialectales, du passage aux langues vernaculaires, des systèmes graphiques et abrégatifs, de paléographie numérique, etc. Ces dernières touchent plus particulièrement la réception et la transmission du tardo-antique à la période médiévale, mais restent des enjeux importants qui traversent les études latines.

## 2.3 Données linguistiques

Le dernier enjeu des prochaines années, du point de vue numérique, concerne l'angle mort que représentent les données linguistiques pour la période tardo-antique. En dehors de la *Vulgate* du projet PROIEL<sup>29</sup>, d'une excursion du LASLA vers l'*Histoire Auguste* et d'un travail mené avec Anthony Glaise sur des échantillons du I<sup>er</sup> millénaire<sup>30</sup>, il n'en existe pas d'autres à notre connaissance. Or, avec l'apparition de nombreux termes spécifiques à l'Église romaine via des emprunts grecs ou simplement de noms propres à la construction inédite pour la période classique, le latin tardif représente une rupture du point de vue quantitatif face au latin classique. Dans une étude sur l'extensibilité d'un modèle de lemmatisation du latin classique à des données tardives, une baisse de quinze points de pourcentage pouvait être remarquée, en particulier si les phénomènes sus-cités étaient fort importants dans les textes vérifiés<sup>31</sup>. Pire encore, l'apparition de textes dont la forme n'est pas normalisée, comme l'édition du *Gynaeciorum Sorani* de Caelius Aurelianus disponible sur DigilibLT, pose des problèmes de confusions lorsque

26. Sur ce sujet, je ne peux que renvoyer vers l'excellent travail de mon collègue : Jean-Baptiste Camps, « Où va la philologie numérique ? », *Fabula LHT (Littérature, histoire, théorie)*, Le Moyen Âge pour laboratoire, dir. Florent Coste et Amandine Mussou 20 (janv. 2018), URL : <https://halshs.archives-ouvertes.fr/halshs-01674953>.

27. A. Pinche, *Edition nativement numérique du recueil hagiographique "Li Seint Confessor" de Wauchier de Denain d'après le manuscrit 412 de la Bibliothèque nationale de France*, Thèse de doctorat, Lyon, Lyon 3, 2021, URL : <http://www.theses.fr/s150996> (visité le 29/05/2019); J.B. Camps et Florian Cafiero, « Genealogical variant locations and simplified stemma : a test case », dans *Analysis of Ancient and Medieval Texts and Manuscripts : Digital Approaches*, dir. Tara Andrews et Caroline Macé, Turnhout, 2014 (Lectio, 1), p. 69-93.

28. J.B. Camps, Simon Gabay et Gustavo Fernández Riva, « Open Stemmata : A Digital Collection of Textual Genealogies », dans *EADH2021 : Interdisciplinary Perspectives on Data, 2nd International Conference of the European Association for Digital Humanities*, Krasnoyarsk, 2021, URL : <https://halshs.archives-ouvertes.fr/halshs-03260086> (visité le 15/10/2021).

29. Dag Trygve Truslew Haug et Marius L. Jøhndal, « Creating a Parallel Treebank of the Old Indo-European Bible Translations », dans *Proceedings of the second workshop on language technology for cultural heritage data (LaTeCH 2008)*, 2008, p. 27-34.

30. Anthony Glaise et T. Clérice, *Du II<sup>ème</sup> siècle à Thomas More, un corpus gold de latin lemmatisé et annoté en morpho-syntaxe*, déc. 2021, DOI : 10.5281/zenodo.6383162.

31. T. Clérice, *Détection d'isotopies par apprentissage profond : l'exemple de la sexualité en latin classique et tardif*, Thèse de doctorat, Lyon, Lyon 3, 2022.

le -ae devient -e. Ce travail, nécessaire pour l'exploitation statistique, mais tout aussi important pour des modes de recherche avancés, doit prendre sa place pour assurer une meilleure reconnaissance des formes à l'avenir.

## Conclusion

Il y a trente ans, tout était encore à bâtir pour l'étude de l'antiquité tardive avec des outils numériques, à part peut-être pour l'étude de la *Vulgate*. Des travaux impressionnants ont permis de mettre à niveau le domaine en termes d'accessibilité des textes (CLCLT, PLD, OGL et DigilibLT) voire ont pris de l'avance sur ce qui se faisait pour les périodes classiques ou médiévales (MQDQ et Hyperdonat). Cependant, l'accessibilité des textes en lecture ne peut plus être une fin en soi : il faut prendre en compte le changement de paradigme que représentent les humanités computationnelles et l'application de l'intelligence artificielle aux textes latins. De ce point de vue, la mise à niveau des textes ouverts comme la construction de nouveaux modèles d'édition ou de création de données restent à être mis en chantier pour pouvoir espérer répondre à ce nouvel enjeu.

## Références

- BOSCHETTI (Federico), GROSSO (Angelo Mario Del) et SPINAZZÈ (Linda), « La galassia Musicque Deoque : storia e prospettive » (, déc. 2021), DOI : 10.30687/978-88-6969-557-5/026.
- BRUNNER (Theodore F.), « Classics and the Computer. The History of a Relationship », dans *Accessing Antiquity : the Computerization of Classical Studies*, dir. John Solomon, 1993.
- BUCKNALL (Tim) et McDONOUGH (Christopher M.), « Review of The CETEDOC Library of Christian Latin Texts on CD-ROM », *The Classical Journal*, 90-1 (1994), p. 90-96.
- BUREAU (Bruno), NICOLAS (Christian) et INGARAO (Maud), *Hyperdonat, commentaire attribué à Aelius Donat aux comédies de Térence*, 2008, URL : <https://hal-univ-lyon3.archives-ouvertes.fr/hal-00879555>.
- BUREAU (Bruno), NICOLAS (Christian) et PINCHE (Ariane), « Encoding (inter)textual insertions in Latin "grammatical commentary" », *Journal of Data Mining and Digital Humanities*, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages (déc. 2017), DOI : 10.46298/jdmdh.1392.
- CAMPS (Jean-Baptiste), « Où va la philologie numérique ? », *Fabula LHT (Littérature, histoire, théorie)*, Le Moyen Âge pour laboratoire, dir. Florent Coste et Amandine Mussou 20 (janv. 2018), URL : <https://halshs.archives-ouvertes.fr/halshs-01674953>.
- CAMPS (Jean-Baptiste) et CAFIERO (Florian), « Genealogical variant locations and simplified stemma : a test case », dans *Analysis of Ancient and Medieval Texts and Manuscripts : Digital Approaches*, dir. Tara Andrews et Caroline Macé, Turnhout, 2014 (Lectio, 1), p. 69-93.
- CAMPS (Jean-Baptiste), GABAY (Simon) et RIVA (Gustavo Fernández), « Open Stemmata : A Digital Collection of Textual Genealogies », dans *EADH2021 : Interdisciplinary Perspectives on Data, 2nd International Conference of the European Association for Digital Humanities*, Krasnoyarsk, 2021, URL : <https://halshs.archives-ouvertes.fr/halshs-03260086> (visité le 15/10/2021).
- CLÉRICÉ (Thibault), « Les outils CapiTainS, l'édition numérique et l'exploitation des textes », *Médiévales. Langues, Textes, Histoire*, 73-73 (déc. 2017), p. 115-131, DOI : 10.4000/medievales.8211.
- *Détection d'isotopies par apprentissage profond : l'exemple de la sexualité en latin classique et tardif*, Thèse de doctorat, Lyon, Lyon 3, 2022.

- CRANE (Gregory Ralph), BABEU (Alison), CERRATO (Lisa), MUNSON (Matthew), DEE (Stella), GESSNER (Annette), ROBERTSON (Bruce), FRANZINI (Greta), STOYANOVA (Simona), SELLE (Tabea), *et al.*, *CSEL - Open Greek And Latin*, Corpus, 2021, URL : <https://github.com/OpenGreekAndLatin/csel-dev> (visité le 27/12/2021).
- CROWN (Ron W.), « Comparing the Patrologia Latina and the CETEDOC Library of Christian Latin Texts Databases from a User's Perspective », *Journal of Religious & Theological Information*, 3-1 (juin 2000), p. 85-109, DOI : 10.1300/J112v03n01\_06.
- DELATTE (Louis) et ÉVRARD (Étienne), « Un laboratoire d'analyse statistique des langues anciennes à l'Université de Liège », *L'Antiquité Classique*, 30 (1961), p. 429-444, DOI : 10.3406/antiq.1961.3417.
- DEMAROLLE (Jeanne-Marie), « Anne-Marie Turcan-Verkerk, Un poète latin chrétien redécouvert : Latinius Pacatus Drepanius, panégyriste de Théodose. » *L'Antiquité Classique*, 75-1 (2006), Publisher : Persée - Portail des revues scientifiques en SHS, p. 369-369, URL : [https://www.persee.fr/doc/antiq\\_0770-2817\\_2006\\_num\\_75\\_1\\_2609\\_t1\\_0369\\_0000\\_2](https://www.persee.fr/doc/antiq_0770-2817_2006_num_75_1_2609_t1_0369_0000_2) (visité le 07/06/2022).
- GARCEA (Alessandro), CINATO (Franck) et PLANCQ (Clément), « Corpus Grammaticorum Latinorum. Un projet de traitement informatique autour des grammairiens latins », *Rivista di Cultura Classica e Medievale*, 52-2 (2010), p. 377-400, URL : <https://halshs.archives-ouvertes.fr/halshs-00735167>.
- GLAISE (Anthony) et CLÉRICE (Thibault), *Du IIème siècle à Thomas More, un corpus gold de latin lemmatisé et annoté en morpho-syntaxe*, déc. 2021, DOI : 10.5281/zenodo.6383162.
- GUERET (Michel), « Analyse informatique et compréhension textuelle. Le CETEDOC : Le laboratoire de recherches textuelles par ordinateur à l'Université Catholique de Louvain. » *Revue LASLA*, 3-1977 (1977), p. 45-72, URL : <http://web.philo.ulg.ac.be/rissh/wp-content/uploads/sites/10/pdf/Annee1977/03/MGueret.pdf>.
- HAUG (Dag Trygve Truslew) et JØHNDAL (Marius L.), « Creating a Parallel Treebank of the Old Indo-European Bible Translations », dans *Proceedings of the second workshop on language technology for cultural heritage data (LaTeCH 2008)*, 2008, p. 27-34.
- HUGHES (John Jay), *Bits, bytes & biblical studies : a resource guide for the use of computers in biblical and classical studies*, 1987, URL : <http://archive.org/details/bitsbytesbiblica0000hugh> (visité le 12/10/2021).
- LANA (Maurizio), « Metodologie e problematiche per una biblioteca digitale. Il caso di digilibLT », *DigItalia*, 1-0 (nov. 2012), p. 40-64, URL : <http://digitalia.sbn.it/article/view/516> (visité le 14/12/2021).
- MASTANDREA (Paolo) et TESSAROLO (Luigi), *Poesis 2 CD-ROM dei testi della poesia latina*, 1999.
- PINCHE (Ariane), *Edition nativement numérique du recueil hagiographique "Li Seint Confessor" de Wauchier de Denain d'après le manuscrit 412 de la Bibliothèque nationale de France*, Thèse de doctorat, Lyon, Lyon 3, 2021, URL : <http://www.theses.fr/s150996> (visité le 29/05/2019).
- ROELLI (Philipp), « The Corpus Corporum, a new open Latin text repository and tool », *Archivum latinitatis medii aevi : Bulletin du Cange-72* (2014), p. 289-304, URL : <https://doi.org/10.5167/uzh-171105>.
- SMITH (MacKenzie), « DynaText : An Electronic Publishing System », *Computers and the Humanities*, 27-5/6 (1993), p. 415-420.
- TOMBEUR (Paul), « La constitution d'un thésaurus pour les pères de l'église latine », dans *Actes du Congrès international informatique et sciences humaines*, Liège, 1981, p. 870-880.
- « La « Patrologia Latina Database (PLD) » ou la reproduction informatisée d'un corpus textuel établi selon une heuristique et une méthodologie éditoriale dépassées », *Bulletin de Philosophie Médiévale*, 35 (janv. 1993), p. 223-228, DOI : 10.1484/j.bpm.3.478.

ZAHN (Peter), « Kirchenväter-Texte im CD-ROM-Speicher : » 16-3 (janv. 1992), p. 419-427, DOI :  
10.1515/bfup.1992.16.3.419.