



OMNIA – Outils et Méthodes Numériques pour l'Interrogation et l'Analyse des textes médiolatins

Bruno Bon



Édition électronique

URL : <http://journals.openedition.org/cem/11086>
DOI : 10.4000/cem.11086
ISSN : 1954-3093

Éditeur

Centre d'études médiévales Saint-Germain d'Auxerre

Édition imprimée

Date de publication : 15 août 2009
Pagination : 291-292
ISSN : 1623-5770

Référence électronique

Bruno Bon, « OMNIA – Outils et Méthodes Numériques pour l'Interrogation et l'Analyse des textes médiolatins », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA* [En ligne], 13 | 2009, mis en ligne le 01 octobre 2009, consulté le 01 mai 2019. URL : <http://journals.openedition.org/cem/11086> ; DOI : 10.4000/cem.11086

Ce document a été généré automatiquement le 1 mai 2019.



Les contenus du *Bulletin du centre d'études médiévales d'Auxerre (BUCEMA)* sont mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

OMNIA – Outils et Méthodes Numériques pour l'Interrogation et l'Analyse des textes médiolatins

Bruno Bon

- 1 Depuis le 1^{er} janvier 2009, et pour une durée de quatre ans, l'UMR 5594 – Archéologie, terre, histoire, sociétés (ARTeHIS, Dijon) – est associée à l'École nationale des chartes (ENC, Paris) et à l'équipe de Lexicographie latine de l'UPR 841 – Institut de recherche et d'histoire des textes (IRHT, Paris) – pour mettre en œuvre le projet ANR « Omnia » – Outils et Méthodes Numériques pour l'Interrogation et l'Analyse des textes médiolatins [<http://www.glossaria.eu/>].
- 2 Le projet « Omnia » se fonde sur les quatre constats suivants :
 - la majeure partie des documents médiévaux est rédigée en latin ;
 - les outils de travail des médiévistes sont notoirement insuffisants ;
 - la réflexion sur l'organisation sémantique de ces langues est encore embryonnaire ;
 - une quantité importante de documents a été numérisée au cours de la dernière décennie.
- 3 Trois perspectives y sont étroitement liées :
 - la récupération des outils existants, pour en faciliter la consultation et l'utilisation, grâce à une interface conviviale en libre accès complet : dans un premier temps, le *Glossarium mediæ et infimæ latinitatis* de C. du Cange (1678, dans son édition de Niort ; L. Favre, 1883-1887), et le *Novum Glossarium Mediae Latinitatis* (en cours de rédaction sous les auspices de l'Union académique internationale, Bruxelles) ;
 - l'organisation d'une réflexion collective des médiévistes sur les rapports entre l'évolution des sociétés médiévales et les évolutions des pratiques discursives et des lexiques ;
 - l'utilisation des documents numérisés pour développer des procédures de recherche formalisées et largement assistées, élaborées en fonction des réflexions sur les rapports entre structures linguistiques et structures sociales.
- 4 La récupération des outils existants (perspective 1) est actuellement en cours de mise en œuvre, les lettres O-Z du *Glossarium* de Du Cange étant déjà librement accessibles, et

interrogeables en plein texte, sur le site de l'École des chartes [<http://ducange.enc.sorbonne.fr/>], dans une interface dynamique et conviviale, réalisée rapidement pour permettre l'interrogation de cet instrument de travail avant même la réalisation de la plate-forme « Omnia ». Les lettres A-N du *Glossarium* et les lettres couvertes par le *Novum Glossarium* complèteront d'ici la fin de l'année 2010 le corpus de base à partir duquel une encyclopédie interactive pourra se construire.

- 5 L'organisation d'une réflexion collective sur le lien entre structures de sens et structures sociales (perspective 2) a conduit le groupe de travail ¹ à privilégier le modèle interactif du « wiki » pour intégrer les fonctionnalités propres à l'édition d'un travail en cours (work in progress). Les données récupérées ci-dessus seront intégrées dès que possible dans cette interface.
- 6 Le développement de procédures de recherche formalisées et assistées (perspective 3) n'est pas dissociable de la réflexion qu'il permet et qui le rend possible. Le groupe de travail se réunit donc régulièrement pour tenter d'améliorer progressivement l'exploitation des documents médiévaux numérisés. Parmi les quelques logiciels d'interrogation (et de comparaison) des textes disponibles, le choix (au moins provisoire) du groupe s'est finalement porté sur « PhiloLogic » [<http://philologic.uchicago.edu/>], développé par l'université de Chicago (ARTFL – American and French Research on the Treasury of the French Language). Ce logiciel libre permet d'envisager l'intégration de nouvelles fonctionnalités adaptées aux spécificités du latin médiéval, en particulier la lemmatisation, rapidement apparue comme l'obstacle essentiel à lever au plus vite, pour pouvoir obtenir des résultats pertinents et réaliser des statistiques à grande échelle. Pour cette lemmatisation, le logiciel libre TreeTagger [<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>], développé par l'université de Stuttgart, et déjà associé pour le français et l'anglais à « PhiloMine » (extension de « PhiloLogic »), s'est imposé. Cet outil nécessitant des fichiers de paramètres spécifiques de la langue considérée, le groupe s'est donc orienté vers la réalisation, pour le latin médiéval, d'un fichier de formes associées à leur lemme – et éventuellement à un marquage morphosyntaxique pour enrichir les procédures de recherche –, et d'un fichier de textes lemmatisés (et éventuellement marqués). Cette étape est actuellement en cours de réalisation, avec le soutien d'Achim Stein (université de Stuttgart), professeur invité à l'École des chartes.
- 7 Le projet « Omnia » vise ainsi à construire une encyclopédie interactive du latin médiéval, disponible librement sur internet, proposant non seulement toute la richesse cumulée des instruments de travail traditionnels, mais aussi les travaux et les réflexions en cours, ainsi que les outils d'une sémantique historique fondée sur le traitement des corpus numérisés.

NOTES

1. Bruno Bon, Olivier Canteaut, Marie-José Gasse-Grandjean, Frédéric Glorieux, Anita Guerreau-Jalabert et Alain Guerreau.

INDEX

Mots-clés : OMNIA