

Un corpus structuré et hétérogène de textes latins médiévaux (Bourgogne, V^e-XV^e siècle)

Eliana MAGNANI
 CNRS - Lamop UMR 8589
 (février 2017)

Le projet de constitution d'un corpus structuré et hétérogène de textes latins médiévaux vise à faire profondément évoluer le contenu de plate-forme documentaire d'actes diplomatiques bourguignons, les CBMA (*Chartae Burgundiae Medii Aevi* - Chartes de la Bourgogne du Moyen Âge <http://www.cbma-project.eu>)¹. Il s'agit de mettre à la disposition des chercheurs, en libre accès, un corpus hétérogène de textes latins médiévaux, et accompagner ainsi les évolutions propres à la pratique de la recherche que l'analyse inédite de la masse de documents numérisés tend actuellement à bouleverser. Orienté, dans une première phase d'élargissement, vers les écrits hagiographiques, le corpus qui sera réuni est à même d'aider à faire la part des techniques discursives et des évolutions sémantiques à l'intérieur d'un système social en grande partie polarisé par les sanctuaires et les reliques des « morts très spéciaux » que sont les saints. L'objectif du programme est de se donner les moyens de comprendre de manière raisonnée, dans une société hétéronome aspirée par l'au-delà, l'articulation de ses productions idéelles et matérielles.

Situation

Depuis une dizaine d'années, le projet de recherche des CBMA numérise, enregistre et exploite la prolifique documentation diplomatique bourguignonne, parmi laquelle figurent, entre autres, les chartes des emblématiques abbayes de Cluny et de Cîteaux. Les réalisations et les résultats du projet sont mis à la disposition via une plate-forme documentaire et une base de données d'actes diplomatiques librement accessibles sur Internet, sous différents formats électroniques². Régulièrement enrichi, y compris par des transcriptions de documents inédits, ce corpus bourguignon qui a privilégié à ses débuts les documents antérieurs à 1300, entend se déployer jusqu'au XV^e siècle et couvrir ainsi l'ensemble du Moyen Âge. Comptant actuellement plus de 23 000 chartes médiévales, il est interrogeable avec le logiciel Philologic³ et constitue un terrain expérimental pour de nouvelles méthodes de recherche. Le projet s'intéresse en même temps à la production matérielle des écrits et propose ainsi plus de quarante cartulaires manuscrits originaux à feuilleter.

Le programme, qui s'appuie sur un large réseau de collaborateurs conduit également une réflexion épistémologique sur les transformations dans la pratique de la recherche, induites par la numérisation massive de textes médiévaux édités et par l'usage du support électronique dans la reproduction des documents et manuscrits. Des expériences nouvelles sont menées pour mieux connaître la société médiévale, pour la connaître autrement, en faisant toute la place aux études de sémantique structurelle, aux approches statistiques et spatiales, ainsi qu'aux implications cognitives de la littérature (*i.e.* les usages sociaux et la culture de l'écrit), en tant que technologie graphique et visuelle.

¹ La plateforme des CBMA est hébergée par le TGIR Huma-Num du CNRS.

² Ces actes sont également déposés dans une base de données diplomatique nationale, *Chartae Galliae* (<http://www.cn-telma.fr/chartae-galliae/index/>).

³ <http://philologic.cbma-project.eu/>

Vers un corpus hétérogène

Centré jusqu'à maintenant sur la documentation diplomatique bourguignonne, le programme des CBMA souhaite développer et consolider une évolution nécessaire, déjà amorcée, vers la constitution d'un corpus hétérogène de documents latins médiévaux bourguignons. En effet, si les 23 000 actes déjà numérisés permettent, et ont déjà suscité, plusieurs études sémantiques, statistiques et spatiales, générant des résultats novateurs, on ne peut se contenter des seules sources diplomatiques, en particulier dans une région également prolifique en matière de production de textes narratifs, théologiques et normatifs, propres à resserrer, voire à transformer, les interprétations.

L'enjeu est de taille car l'ambition à moyen terme du projet est de proposer un corpus raisonné, géographiquement circonscrit (la Bourgogne est comprise ici d'après les juridictions ecclésiastiques médiévales : les diocèses d'Auxerre, Langres, Nevers, Autun, Chalon, Mâcon, Besançon), facilement accessible et adaptable, qui pourra être utilisé pour lui-même mais aussi servir de point de comparaison avec d'autres corpus, ouvrant à des approches à différentes échelles : locales, régionales, européennes.

La constitution d'un corpus hétérogène contenant des Vies de saints, des chroniques, des traités exégétiques (ceux, par exemple, de l'école carolingienne d'Auxerre), des productions juridiques et réglementaires (conciles, collections canoniques, statuts, coutumiers...), entre autres, nécessite un travail préalable d'inventaire, de recherche des éditions disponibles, de préparation des numérisations-océrisations (saisie manuelle ou conversion automatique de l'image d'un document en fichier texte), d'intégration de métadonnées, avant leur mise en ligne sous forme de base de données, et dans différents formats électroniques, dont des formats propres à être employés dans des logiciels de fouille de texte.

Un sous-corpus hagiographique

Après les documents diplomatiques, les textes hagiographiques constituent sans doute le sous-corpus le plus aisément identifiable, passible d'être réuni relativement rapidement et numériquement représentatif. En effet, les éditions (et rééditions) de textes hagiographiques ont une longue histoire. Depuis le XVII^e siècle, des collections importantes ont été publiées et sont facilement accessibles et récupérables aujourd'hui via Internet (*Acta sanctorum*, *Patrologia latina*, *Monumenta Germaniae Historica*...). Même si la priorité sera donnée aux éditions scientifiques plus récentes et plus fiables, la masse de textes facilement disponibles est considérable. Le travail qui reste à faire concerne principalement la structuration de ces écrits en « corpus ».

L'inventaire des textes concernés est en cours à partir des notices des Bollandistes (BHL - *Bibliotheca Hagiographica Latina*, Bruxelles, 1898-1901, Suppl. 1911, Novum Suppl. 1986) et il devra être complété et mis à jour d'après l'historiographie récente⁴ (Fig. 1). Toutefois, une évaluation préliminaire des Vies de saints ayant été (probablement) rédigées dans les diocèses bourguignons pointe vers environ 130 saints et au moins autant de textes concernés. Ce chiffre est sans doute en deçà des écrits qu'on aura à traiter, probablement plus que le double, car on prendra en compte l'ensemble du dossier hagiographique de chaque saint – *vitae*, *miracula*, *translationes*... –, y compris, dans la mesure du possible, dans leurs versions liturgiques. Même si certaines pièces n'ont pas été forcément composées en Bourgogne, qui se caractérise pour avoir été une région d'« accueil » de reliques en translation (pensons, parmi d'autres, à l'exemple de l'itinéraire des reliques de saint Philibert,

⁴ Nous pensons, en particulier, aux travaux de la collection *Hagiographies* publiés ou en préparation, maintenant sous la direction de Monique Goulet, Brepols.

de Noirmoutier à Tournus), c'est seulement dans l'ensemble des compositions, et sur la longue durée, qu'elles peuvent être correctement situées. Ce travail préalable de recensement permettra également de proposer un bilan des textes hagiographiques encore inédits.

Dossier	Titres / Analyse CBMA	BHL	Id	Édition	Annotations / autres éditions	Auteur Vita Nom	Auteur Vita Lieu	Date de rédaction / retenue	Dates du saint	Lieu (du saint)	X	Y	Diocèse médiéval
Robertus ab. Molisemensis	Robert de Molesme. Vita a	7265	BHL7265_R Robertus	Patrologie latine, t. 157, col. 1269-1288	SPAHR K., Das Leben des hl. Robert von Molesme, Freiburg, 1944, p. 2-31 AASS, April, III, 669-676 AASS, 3e ed. 676-683 Auteur : moine anonyme de Molesme	moine anonyme	Molesme	1222	1028/1029-1110	Molesme	3.466667	47.616667	Langres
	Robert de Molesme. Vita b	7266	BHL7266_R Robertus	Surius, II (1571), 898-905 ; (1578), 1012-1018 ; IV (1618), 357-360 ; IV (1876), 659-678					1028/1029-1110	Molesme	3.466667	47.616667	Langres
	Robert de Molesme. Vita c	7266 b	BHL7266b Robertus	SPAHR K., Das Leben des hl. Robert von Molesme, Freiburg, 1944, p. 37-43					1028/1029-1110	Molesme	3.466667	47.616667	Langres
	Robert de Molesme. Inquisitio de miraculis	7267	BHL7267_R Robertus	Patrologie latine, t.c. 1289-1291, n. 25-29	Gallia christiana (Paris, 1620), 622-624 Labbe, Bibl. mss., I, 648-649, III Mansuete A. Gisteroniens...Annales, IV (1659), 213-215 AASS, t.c. 677 (3e ed. 684-685), n. 25-29 Martène et Durand, Thes. Nov. Anecd., I, 904-907				1028/1029-1110	Molesme	3.466667	47.616667	Langres
	Robert de Molesme. Epitome		BHL0000_R Robertus	Petrus de Nat., IV, 102	Hist. SS. (Colon 1483), f. 272a-272d Hist. SS. (Lovan 1485), f. 58v-59v				1028/1029-1110	Molesme	3.466667	47.616667	Langres
Odilo ab. Cluniacensis	Jotsaldus, Vita s. Odilonis	6281	BHL6281 Odilo	Jotsald von Saint-Claude, Vita des Abtes Odilo von Cluny, hg. von JOHANNES STAUB (MGH Scriptores rerum Germanicarum in usum scholarum separatim editi 68) Hannover 1999, p. 141-254. Rez.: TMR. Rez.: BE:Ca https://www.dmgp.de/de/61/0/bjct/display/bsb0000716_meta_titlePage.html?sortIndex=01007000680100000	Patrologie latine, t. 142, col. 915-936 — JOHANNES STAUB, Studien zu Jotsalds Vita des Abtes Odilo von Cluny (MGH Studien und Texte 24) Hannover 1999. Rez.: TMR — Jotsald, De vita et virtutibus sancti Odilonis abbatis, hg. von GEORG WAITZ, Teilabdruck in: MGH SS 15, Hannover 1888, n. 812-820; Text — Vita Odilonis abb. Cluniac. auct. Jotsald	Jotsald	Saint-Claude	1052	c. 962-c. 1048	Cluny	4.65	46.433333	Mâcon

Figure 1: Extrait de l'inventaire des textes hagiographiques.

La réalisation de cet inventaire s'accompagne du renseignement d'un certain nombre de métadonnées pour chaque texte, dont il s'agira d'établir la liste – datation(s), édition(s), etc. –, y compris les coordonnées de géolocalisation. L'un des objectifs de ce programme est, en effet, de rendre ce corpus exploitable aussi par des systèmes d'information géographique (SIG). Il deviendra ainsi possible de dresser une cartographie des productions écrites, une géographie des usages linguistes, entre autres applications⁵.

Les choix techniques et de diffusion : simplicité, multiplicité, ouverture, libre accès

Le sous-corpus hagiographique intégrera la plate-forme déjà existante des CBMA, hébergée par le TGIR Huma-Num du CNRS. L'utilisation de cette méta-plate-forme institutionnelle de diffusion collective permet à l'équipe du projet de bénéficier de compétences spécialisées sur la durée, d'être déchargée du souci de la pérennisation des contenus à moyen terme ainsi que de l'optimisation de l'indexation pour les moteurs de recherche.

En suivant les orientations qui ont marqué le programme dès ses débuts, les acquis du projet et les données seront mis à disposition en accès libre et ouvert, presque immédiatement, au fur et à mesure de leur acquisition grâce aux mises à jour fréquentes du site et des pages documentaires. De la base de données de travail aux unités textuelles, l'ensemble sera disponible dans des formats simples et multiples en vue de faciliter leur (ré)utilisation et leur appropriation par les chercheurs.

Les textes seront enregistrés aux formats .txt et .odt, identifiés par le titre du fichier (BHLn^o+nom.latin) et mis en ligne, chacun d'entre eux, sous cette forme élémentaire.

Dans la base de données de travail, sous FileMakerPro (logiciel propriétaire, mais de coût très modique, éprouvé et commode pour les exports/imports des données), chaque texte

⁵ Pour cette approche spatiale, le projet pourra s'appuyer notamment sur le travail de cartographie historique mené par le programme « Topoma » (Topographie de l'Antiquité et du Moyen Âge - dir. Th. Lienhard - <http://www.menestrel.fr/spip.php?rubrique1621&lang=fr>), partenaire de notre projet.

fera l'objet d'une fiche contenant les métadonnées (Fig. 2). Cette base et son export en format .csv, aisément récupérable, seront également mis en ligne pour téléchargement, y compris dans ses versions intermédiaires.

The screenshot shows a FileMaker Pro window with a form for 'CBMA- Chartes bourguignonnes du Moyen Age'. The form includes the following fields and content:

- type de document:** Hagiographie
- BHL:** 6281
- Date du saint:** c. 962-c. 1048
- Auteur:** Jotsald, moine
- Id:** BHL6281_Odilo
- référence édition:** Jotsald von Saint-Claude, Vita des Abtes Odilo von Cluny, hg. von JOHANNES STAUB (MGH Scriptores rerum Germanicarum in usum scholarum separatim editi 68) Hannover 1999, p. 141- 254 Rez.: TMR Rez.: BECH
- diocèse:** Mâcon
- lieu (latin):** Cluny (identification)
- coordonnées GPS:** 4.65, 46.433333
- date édition:** (empty)
- date retenue:** 1052
- annotations:** Patrologie latine, t. 142, col. 915-936 — JOHANNES STAUB, Studien zu Jotsalds Vita des Abtes Odilo von Cluny (MGH Studien und Texte 24) Hannover 1999 Rez.: TMR — Jotsald, De vita et virtutibus sancti Odilonis abbatis, hg. von GEORG WAITZ, Teilabdruck in: MGH SS 15, Hannover 1888, p. 812-820; Text — Vita Odilonis abb. Cluniac. auct. Jotsald mon. Cluniac. in: Acta Sanctorum quotquot orbe colluntur..., Antwerpen 1843, Januar I, p. 65-71 — Vita Odilonis abb. Cluniac. auct. Jotsald mon. Cluniac. in: LUC D'ACHERY — JEAN MABILLON, Acta sanctorum Ordinis S. Benedicti, 6, Paris 1680, p. 679-710 — Vita Odilonis abb. Cluniac. auct. Jotsald mon. Cluniac. in: LUC D'ACHERY — JEAN MABILLON, Acta sanctorum Ordinis S. Benedicti, 6 (saec. VI), 2 Venedig 1738, p. 598-623 — Vita Odilonis abb. Cluniac. auct. Jotsald mon. Cluniac. in: MARTIN BOUQUET, Recueil des historiens des Gaules et de la France 10, Paris 1760, p. 374 — Vita Odilonis abb. Cluniac. auct. Jotsald mon. Cluniac. in: Luc D' Achery — Jean Mabillon, Acta sanctorum

The text area on the right contains the Latin text of the prologue, starting with 'PROLOGUS IOTSALDI MONACHI AD STEPHANUM PONTIFICEM. Reverendo patri Stephano secundum sui nominis interpretatio nem pontificali gloria coronato...' and ending with 'EXPLICIT PROLOGUS. A. LUGUBRIS LAMENTATIO EIUSDEM MONACHI DE TRANSITU BEATI VIRI. Quis dabit capiti meo aquam et oculis meis fontem lacrimarum, et plorabo desolationem non unius hominis neque duorum vel trium sed multorum populorum promiscui sexus, diverse conditionis, diverse ætatis, diversi ordinis ? Plorabo, inquam, vastitatem nobilissime congregationis, ample familie destitute dormitione sui patris Odilonis, de cuius visceribus, de cuius magisterio et virtutibus quasi quibusdam vinculis et funibus religata dependebat et,'

Figure 2: Extrait d'une fiche d'un texte hagiographique, avec métadonnées (BHL 6281, Jotsaldus, Vita s. Odilonis).

Cette structuration élémentaire, facile et relativement rapide à mettre en œuvre, pourra servir à des réutilisations ultérieures et reste ouverte et maniable pour les balisages et les traitements plus raffinés. Le choix de ces formats multiples et d'une technologie simple est central car il permet de ne pas figer la base dans un état définitif, poussant ainsi à explorer les possibilités nouvelles offertes par des logiciels en cours de développement.

Sous ce format, le corpus de textes hagiographiques pourra être importé et analysé par des applications lexicométriques⁶. Dans une étape ultérieure, l'un des objectifs du projet est de procéder à la tokenisation (segmentation en mots) et à la lemmatisation (regroupement des différentes formes d'un mot sous leur lemme) de ce corpus de textes latins médiévaux et de le diffuser sous cette forme propre aux analyses sémantiques et statistiques⁷.

Parallèlement, et d'après l'expérience acquise avec le corpus bourguignon, il sera également envisagé la constitution d'autres corpus hagiographiques régionaux (Provence,

⁶ Nous pensons tout particulièrement à TXM (Textométrie - <http://textometrie.ens-lyon.fr/>), logiciel libre développé à l'École normale supérieure de Lyon, employé désormais par une large communauté de chercheurs et dont l'équipe des concepteurs assure la mise à jour régulière.

⁷ Les paramètres adaptés au latin médiéval existent déjà et sont librement accessibles, développés par l'ANR Omnia (Outils et Méthodes Numériques pour l'Interrogation et l'Analyse des textes médiolatins - <http://www.glossaria.eu/>), pour être utilisés avec le logiciel de lemmatisation TreeTagger (<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>).

Normandie, Péninsule ibérique, Italie...) propres à développer davantage les approches comparatives.

S'appuyant sur un large réseau de chercheurs et ouvert à tous les intéressés⁸, cette opération est portée par le Lamop (Laboratoire de Médiévistique Occidentale de Paris - UMR 8589), soutenue par le Labex HaStec en partenariat avec l'IRHT (Institut de recherche et d'histoire des textes – CNRS UPR 841), le CESOR (Centre d'Etudes de Sciences sociales du Religieux - UMR 8216), le Centre Jean Mabillon (EnC - EA 3624), LEM/CERCOR (Laboratoire d'Etudes sur les Monothéismes – UMR 8584 / Centre européen de recherche sur les congrégations et les ordres religieux).

⁸ Renaud Alexandre, Bruno Bon, François Bougard, Pierre-Henri Billy, Olivier Canteaut, Noëlle Deflou-Leca, Patrick Henriot, Thomas Lienhard, Michèle Gaillard, Monique Goulet, Anita Guerreau-Jalabert, Dominique Iogna-Prat, Nicolas Perreaux, Alain Rauwel, Coraline Rey. Tous les collègues intéressés à participer de ce projet sont invités à nous contacter (cbma.project@gmail.com).