HAL open science

Le corpus épigraphique bourguignon (VIIIe-XVe siècle). Des catalogues aux applications numériques.

Eliana Magnani, Estelle Ingrand-Varenne

To cite this version:

Eliana Magnani, Estelle Ingrand-Varenne. Le corpus épigraphique bourguignon (VIIIe-XVe siècle). Des catalogues aux applications numériques.. Bulletin du Centre d'études médiévales d'Auxerre, 2018. halshs-01946701

HAL Id: halshs-01946701

https://shs.hal.science/halshs-01946701v1

Submitted on 6 Dec 2018

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.

Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Bulletin du centre d'études médiévales d'Auxerre | BUCEMA
Collection CBMA

Le corpus épigraphique bourguignon (VIIIe-XVe siècle). Des catalogues aux applications numériques

Compte-rendu de la journée d'études. Paris, 23 octobre 2018

Eliana Magnani et Estelle Ingrand-Varenne

Édition électronique
URL : http://journals.openedition.org/cem/15591
ISSN : 1954-3093

Éditeur
Centre d'études médiévales Saint-Germain d'Auxerre

Référence électronique
Eliana Magnani et Estelle Ingrand-Varenne, « Le corpus épigraphique bourguignon (VIIIe-XVe siècle). Des catalogues aux applications numériques », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA* [En ligne], Collection CBMA, Les journées d'études, mis en ligne le 15 novembre 2018, consulté le 15 novembre 2018. URL : http://journals.openedition.org/cem/15591

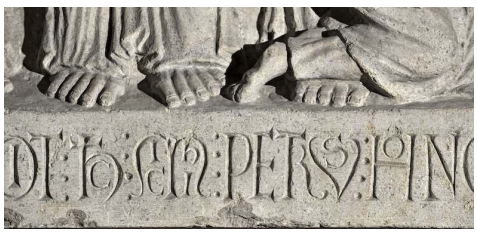
Ce document a été généré automatiquement le 15 novembre 2018.

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.

Le corpus épigraphique bourguignon (VIIIe-XVe siècle). Des catalogues aux applications numériques

Compte-rendu de la journée d'études. Paris, 23 octobre 2018

Eliana Magnani et Estelle Ingrand-Varenne



1 Les équipes des programmes consolidés CIFM (*Corpus des Inscriptions de la France Médiévale* - CESCM-Poitiers)[1] et CBMA (*Corpus Burgundiae Medii Aevi* - LaMOP - Paris)[2] ont mis en place, en 2018, un projet commun de constitution d'un corpus numérisé d'inscriptions épigraphiques bourguignonnes médiévales.

2 D'un côté, le CBMA s'est engagé dans la réalisation d'un corpus structuré et hétérogène de textes latins médiévaux, dont l'objectif est, à moyen terme, de pouvoir interroger de manière simultanée différents types documentaires[3]. D'un autre côté, le CIFM, depuis sa création dans les années 1960, a réuni un nombre considérable de notices avec le texte et les références des inscriptions épigraphiques connues dont celles antérieures au XIIIe siècle sont publiées, pour la Bourgogne, dans trois volumes de la collection (n° 19, 20, 21, relatifs aux départements du Jura, Nièvre, Saône-et-Loire, de la Côte-d'Or, et de l'Yonne)[4]. Le recensement initial a couvert également la fin du Moyen Âge et se trouve consigné dans un fichier papier.

3 Les deux équipes étant intéressées par les procédés de lemmatisation, participant aux travaux du groupe « Lemmes » du Consortium Sources Médiévales 2 (COSME[2])[5], plusieurs conditions étaient réunies pour réaliser rapidement le corpus : les volumes étaient déjà océrisés et disponibles sur la plateforme Persée, ce qui a permis l'extraction des données, de même que, avec le soutien du consortium, il a été possible de réviser ces informations

et d'adjoindre les inscriptions encore inédites recensées. En quelques mois, un corpus de 1 418 inscriptions bourguignonnes du VIIIe au XVe siècle, écrites en différentes langues, latin et ancien français en particulier, a été constitué (Fig. 1). La journée d'études[6] entendait revenir sur le processus d'acquisition et de structuration des données du corpus et d'évaluer son potentiel d'exploitation pour la recherche historique au regard des différents traitements informatiques possibles, linguistiques et spatiaux. Il s'agissait aussi, sur la base de différentes expériences réalisées, d'affiner le jeu de métadonnées retenues[7] et de discuter, aussi bien des caractéristiques propres du matériel linguistique des inscriptions que de leurs usages dans l'espace et dans le temps, comme l'ont rappelé Eliana Magnani et Estelle Ingrand-Varenne en introduction.

Fig. . Répartition des inscriptions par langue (E. Magnani).

Langue	nombre d'inscriptions	*nombre de mots*
latin	774	*~21 033*
latin-français	63	
latin-grec	3	
français	499	*~22 257*
français-latin	71	
latin ou français	1	/
grec	1	/
incertain	6	/
TOTAL	1418	/

4 En ce qui concerne les caractéristiques générales du corpus, les trois volumes édités comportent un total de 471 inscriptions avec des états divers de conservation et de restitution possible, réparties entre le VIIIe et le XIIIe siècle[8], tandis que le nombre d'inscriptions inédites, notamment des XIIIe, XIVe et XVe siècles s'élève à 947 inscriptions. La répartition chronologique de l'ensemble indique que plus de 80% des inscriptions du corpus ainsi constitué datent des trois derniers siècles du Moyen Âge et ont été produites principalement dans des contextes urbains (Dijon, Auxerre, Sens...) (Fig. 2 et 3). Le classement typologique de ces inscriptions, qu'il faut encore affiner et structurer, montre aussi la prépondérance des inscriptions en contexte funéraire qui correspondent à plus de 60% de l'ensemble.

Fig. . Répartition chronologique des inscriptions (E. Magnani).

siècle	**VIIIe**	**IXe**	**Xe**	**XIe**	**XIIe**	**XIIIe**	**XIVe**	**XVe**	**XVIe**	**s.d.**
nombre	2	11	13	47	144	326	376	454	11	34

%	0,1%	0,7%	0,9%	3,3%	10%	23%	27%	32%	0,7%	2,3%

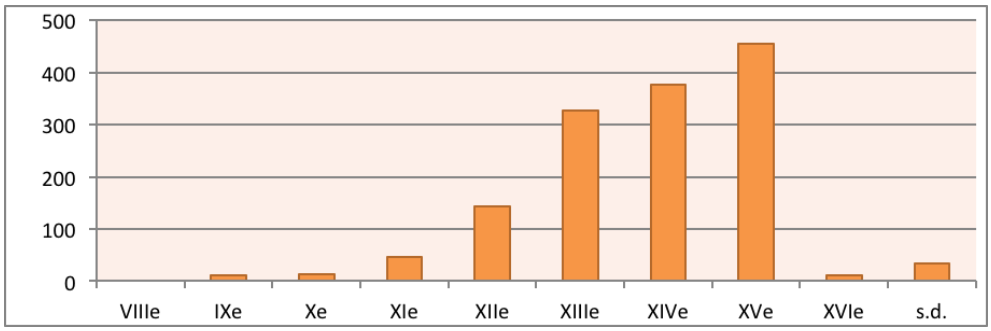
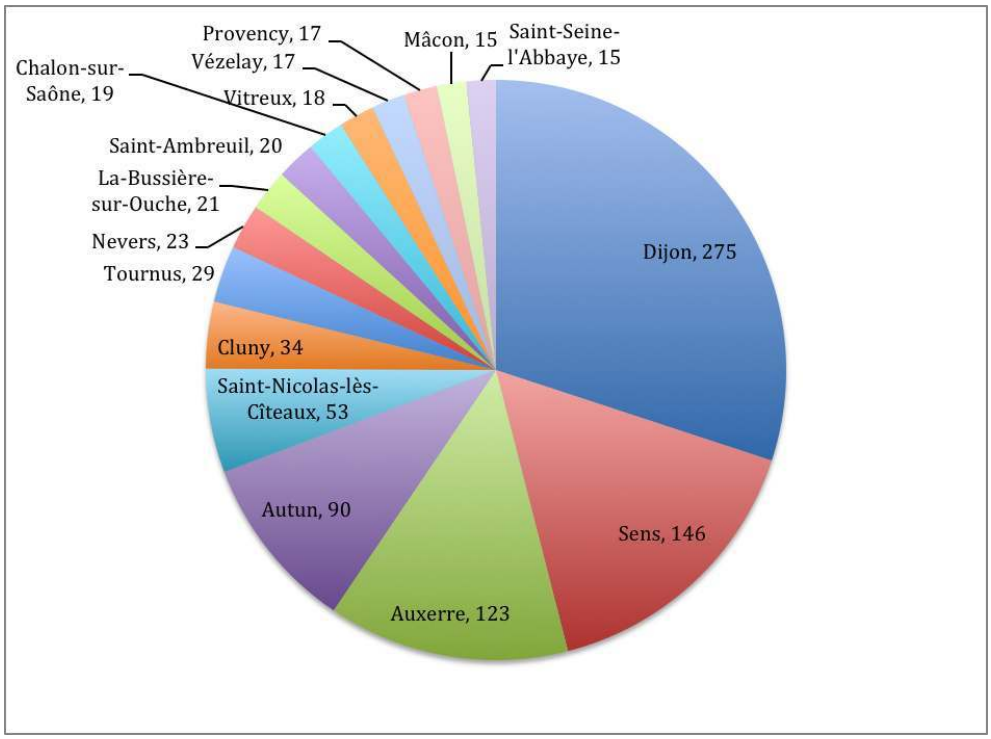


Fig. . Répartition du corpus par « ville » (E. Magnani).



5 Par ailleurs, au regard de la production épigraphique française, un rapide coup d'œil aux 1 418 textes réunis, couvrant 5 départements « bourguignons », permet de dégager trois pistes de recherche – quantitative, typologique et linguistique – qui mériteraient d'être creusées.

6 La première est un double accroissement : l'augmentation continue du nombre d'inscriptions au fil des siècles (triplant du XIe au XIIe s., puis du XIIe au XIIIe s., puis multiplié par 1,15 du XIIIe au XIVe s., et par 1,2 du XIVe au XVe s.), alliée à un allongement des textes. Les inscriptions de la fin du Moyen Âge pèsent donc fortement par leur masse textuelle. Si cette croissance peut sembler évidente et corroborer l'explosion documentaire constatée dans la culture écrite de ces mêmes siècles, elle n'a pas été relevée partout, notamment en Poitou. La deuxième piste de recherche concerne la typologie. Les inscriptions bourguignonnes sont en majorité lapidaires et funéraires (65% au XVe s.) et l'allongement des textes touche particulièrement les épitaphes (au sens large

du terme, texte relatant la mort d'un individu), qui passe de l'éloge funèbre à de véritables biographies/*curriculum vitae*. Le vocabulaire religieux, lié à la prière, recule, tandis que les noms propres, les titres, les charges se multiplient.

7 La troisième piste de recherche qu'ouvre ce projet est liée à l'entrée en lice de la langue vernaculaire, à partir du XIIIe s., employée dans les mêmes proportions que le latin au XIVe s., et dépassant celui-ci au XVe s. Cette pénétration du français, qui vient cohabiter avec et non remplacer la langue latine, se fait à plusieurs échelles : d'un seul mot (le plus souvent un nom propre) à un texte entier, créant différentes strates de « bilinguisme », laissant des formules de prière dans leur langue d'origine, sans parler de la syntaxe, ou de la dissimulation de lexèmes romans sous une déclinaison latine. L'augmentation et l'allongement des textes alliés à l'emploi croissant du français font que le corpus réuni contient plus de mots français que de mots latins. Dans le corpus bourguignon comme ailleurs, c'est par l'entremise des laïcs que le vernaculaire s'insinue dans les textes funéraires, mais la Bourgogne se démarque par un emploi « genré » très clair : au XIIIe s., les épitaphes concernant les femmes sont majoritairement écrites en français (32 sur 56), et la proportion va en s'accroissant au bas Moyen Âge. Ce constat statistique pose une question : Les femmes favorisent-elles la langue maternelle ? Ou encore leur attribue-t-on une langue d'expression particulière parce qu'elles sont femmes ?

8 Ces quelques éléments montrent déjà les perspectives fructueuses que le travail de lemmatisation permettra de mener afin de mieux cerner l'emploi des langues des inscriptions, mais aussi le vocabulaire, les croisements avec les typologies textuelles, matérielles, contextuelles, fonctionnelles, et surtout les autres écrits de Bourgogne.

Épigraphie médiévale et lemmatisation. Des index au projet TITULUS

9 Afin de mieux comprendre et mettre en contexte le matériau assemblé, un retour à l'histoire de la collection du CIFM s'avère aussi indispensable. Estelle Ingrand-Varenne a insisté sur le fait que même si elles sont restées secondaires et limitées jusqu'ici, les réflexions sur la lemmatisation des inscriptions médiévales en France ne sont pas si récentes et peuvent se résumer en trois étapes. Comme tout éditeur de sources, le Corpus des inscriptions de la France médiévale a proposé dès la publication des premiers volumes un index des mots des inscriptions ; ce qui constitue un premier travail de lemmatisation. Dans les volumes du CIFM parus de 1974 à 2002, il s'agissait d'« un index des matières », mêlant vocabulaire et entrées thématiques. Le vocabulaire, lui-même, ne recevait pas un traitement exhaustif, seuls certains mots jugés clés apparaissaient, tandis que ceux à moindre valeur sémantique étaient délaissés (mots-outils, pronoms relatifs, prépositions, verbe être). Parallèlement au travail éditorial, une concordance informatique fut réalisée (fin des années 1970), fournissant le contexte de chaque attestation (Fig. 4). Plusieurs concordances furent ainsi réalisées : pour les *Carmina Latina Epigraphica*, les inscriptions en prose latine et française, les inscriptions latines en vers, les dédicaces et consécrations, ainsi que les inscriptions espagnoles. Ces concordances (consultables à l'inscripthèque du CESCM) ont permis l'étude des formules épigraphiques et de leur fonctionnement.

Fig. . CIFM - Concordance informatique de la fin des années 1970 (E. Ingrand-Varenne).

10 Le vrai changement dans le traitement de la lemmatisation des inscriptions vient du projet TITULUS[9], suivant les traces de l'ANR ORIFLAMMS[10]. TITULUS, édition électronique en XML-TEI des textes épigraphiques, propose en effet une lemmatisation complète, univoque et systématique de toutes les nouvelles inscriptions éditées. Grâce à l'attribut @lemmaRef inséré dans la balise <w>, l'ensemble des lemmes (français et latin) de toutes les inscriptions encodées est regroupé dans un lexique commun à toutes les notices (entité). Dans ce lexique (actuellement composé de 837 lemmes latins, et 22 français correspondants à 149 inscriptions), chaque mot lemmatisé est accompagné d'une information sur sa langue et sa catégorie grammaticale ; les noms propres sont également distingués grâce à une balise @subtype.

Le matériel épigraphique révisé et complété

11 Cet atelier a été l'occasion pour Aurore Menudier de présenter la documentation du CIFM qui a servi de base pour la création du corpus, qui tend à l'exhaustivité mais reste incomplet, d'inscriptions bourguignonnes du VIIIe s. au XVe s. Il a également permis de revenir sur la constitution de ce fonds documentaire, sur les sources et sur les méthodes utilisées par les épigraphistes pour relever et éditer les inscriptions. Enfin, des exemples ont permis d'aborder les difficultés rencontrées et les limites des interventions possibles pour compléter et corriger les textes.

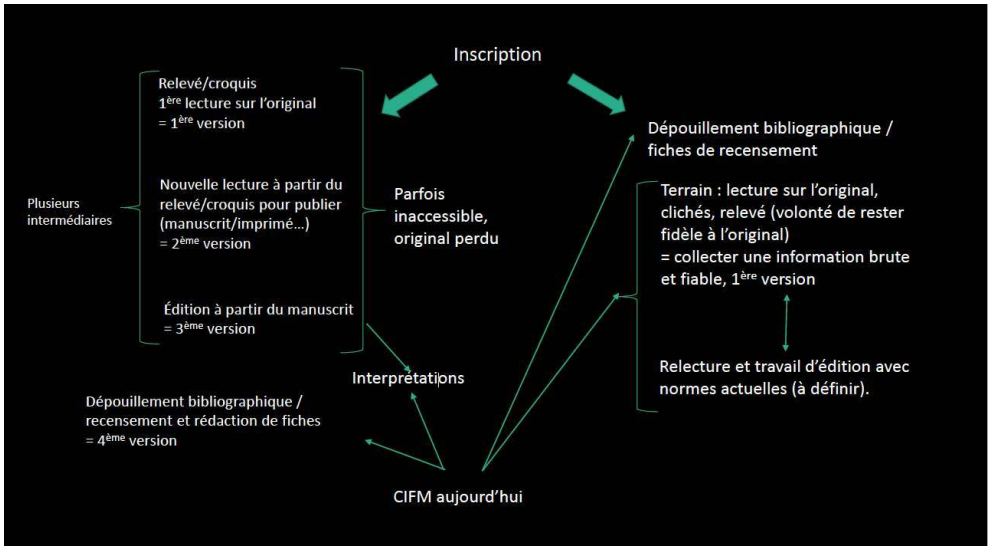
12 La documentation utilisée se compose de versions océrisées de trois volumes parus après 1999 et qui ont bénéficié d'une expertise scientifique par l'équipe du CIFM : visite sur le terrain, photographies, édition critique avec restitution des abréviations et des lacunes (lorsque cela était possible), traduction et commentaires paléographiques et historiques. Ces textes représentent numériquement moins du tiers de l'ensemble proposé. Pour compléter ce corpus (l'édition s'arrêtant en 1300), étaient disponibles principalement des notices rédigées après un recensement bibliographique de publications parues à partir du XIXe s. et qui reprenaient des travaux plus anciens : divers manuscrits, dessins de la collection Gaignières, *Gallia christiana*, etc. Ces textes ont été présentés en l'état dans lequel ils ont été trouvés (Fig. 5). Il a été possible de proposer quelques inscriptions

inédites, qui ont été communiquées au CIFM par des musées et diverses personnes, ainsi que des inscriptions inédites de Dijon qui ont bénéficié d'une mission de terrain en 2015.

Fig. . CIFM - Exemple de fiche de recensement (A. Menudier).

13 Ce matériau hétérogène a dû être retravaillé pour parvenir au résultat attendu (Fig. 6) : des textes complétés (lacunes et abréviations) et corrigés si nécessaire pour pouvoir être lemmatisés automatiquement. Lorsque les témoins de la source originale sont à notre disposition et en tenant compte des connaissances et des critères d'édition actuels, qui se veulent moins interventionnistes que ceux retenus par les érudits des périodes précédentes, il est possible de proposer des versions relativement fiables du texte, aussi bien dans le fond que dans la forme. Lorsqu'il ne reste que des témoins secondaires (dessins, éditions sans relevés), il est difficile, sinon impossible, de déterminer quelles interventions (corrections, restitutions) ont été effectuées par les auteurs. Par conséquent, le degré de fiabilité est moins élevé, et ont été proposés des textes dont l'ensemble des mots est très probablement correct mais dont la variante orthographique retenue est sujette à caution.

Fig. . Circuits des opérations de traitement des inscriptions (A. Menudier).



14 La présentation de cas concrets a permis de constater l'hétérogénéité du fonds documentaire, les difficultés qui en résultent pour le traitement de ces textes, mais aussi de mesurer les implications pour le travail des autres membres de l'équipe (pour la lemmatisation, par exemple) et pour les résultats produits.

Les opérations informatiques. De l'OCR au web sémantique

15 La constitution rapide du corpus et la possibilité de l'interroger d'ores et déjà avec un logiciel de fouille de texte relève aussi des traitements informatiques qui ont été opérés. Pierre Brochard explicite la série d'opérations et les élargissements réalisés, dont certains n'étaient pas prévus au départ. Les volumes du CIFM donnent pour chaque inscription ou fragment une fiche d'édition documentée bien structurée. À partir des fichiers texte océrisés, les séparateurs graphiques des champs récurrents ont été identifiés et les données aspirées et organisées dans un fichier final .csv. Après la relecture et l'addition des inscriptions inédites, ce fichier contenait, outre des métadonnées temporelles, linguistiques et documentaires, des informations géographiques auxquelles ont été associées les coordonnées de géolocalisation (WGS 84 - EPSG 4326), donnant lieu à la réalisation d'un fichier shapefile (.shp) utilisable dans les logiciels SIG. Des identifiants wikidata ont également été joints en vue de stabiliser l'identification des entités dans le temps face à la pluralité de leurs désignations. Toutefois, d'autres métadonnées géographiques importantes, comme le diocèse médiéval d'où proviennent les inscriptions, restent encore à acquérir. L'interopérabilité avec les grandes bases de données nationales a également été pensée, notamment via l'inclusion des liens vers les fiches de la base Mérimée sur le patrimoine architectural français[11].

16 Déposés sur GitLab, tous les produits et phases du projet sont librement accessibles[12]. Le corpus et les métadonnées ont ainsi été diffusés à l'équipe dans une première version, sous la forme d'un tableau .csv et d'un fichier texte (.txt) pour chaque inscription, dont le nom est composé de la date simplifiée, de la provenance et de l'identifiant de l'inscription (1239_tilchatel_co106.txt), répartis ensuite en dossiers linguistiques et géographiques.

Cette structuration a permis l'importation du corpus, en vue de son interrogation avec le logiciel Philologic[4], dont la TEI a été détournée pour pouvoir l'adapter aux caractéristiques propres de l'ensemble épigraphique[13].

17 Le grand nombre d'inscriptions funéraires contenant des noms de personne ainsi que des dates de décès ont ouvert une autre possibilité d'utilisation et d'exploitation du corpus sous la forme d'une base prosopographique, grâce à la mise en place de procédés propres au web sémantique (fichiers RDF) (Fig. 7).

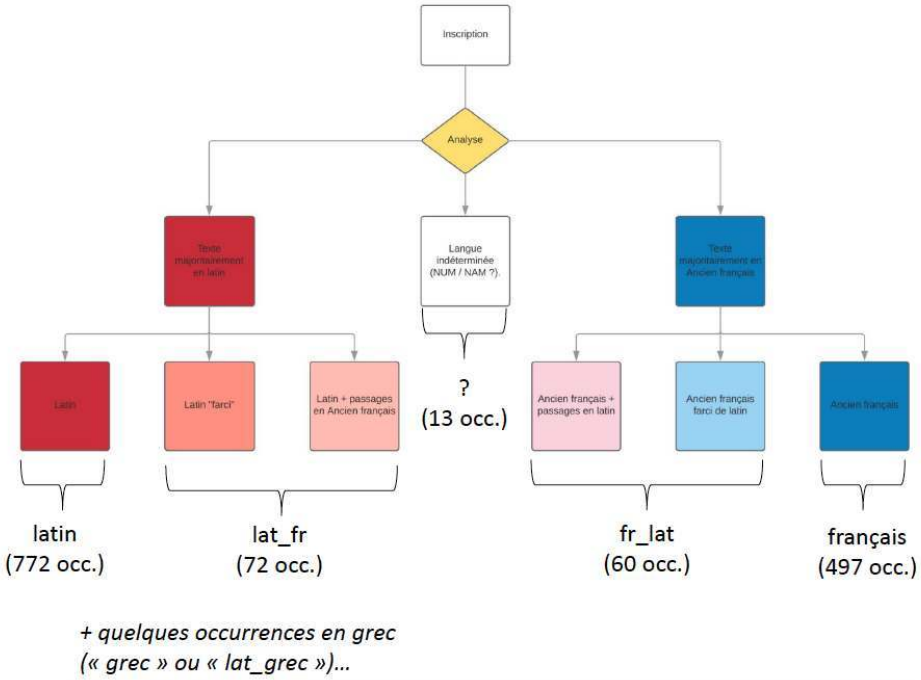
Fig. . Web sémantique. Base de données prosopographique. Exemple de fichier RDF (P. Brochard).

```
- <rdf:RDF>
  - <foaf:Document rdf:about="http://www.cbma-project.eu/CO300">
      <dcterms:creator rdf:resource="http://www.cbma-project.eu/CBMA"/>
      <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2018-10-22</dcterms:created>
      <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2018-10-22T04:15:33</dcterms:modified>
    </foaf:Document>
  - <bibo:Manuscript rdf:about="http://www.cbma-project.eu/CO300">
    - <dc:title>
        Inscription funéraire pour Philippe le Bon et deux de ses épouses, Isabelle de Portuel et Bonne d'Artois
      </dc:title>
      <dc:date>1472</dc:date>
      <dcterms:Location>Dijon</dcterms:Location>
      <dc:subject rdf:resource="http://www.cbma-project.eu/CO3001"/>
      <dc:subject rdf:resource="http://www.cbma-project.eu/CO3002"/>
      <dc:subject rdf:resource="http://www.cbma-project.eu/CO3003"/>
    </bibo:Manuscript>
  - <foaf:Person rdf:about="http://www.cbma-project.eu/CO3001">
      <owl:sameAs rdf:resource="https://www.wikidata.org/wiki/Q239337"/>
      <foaf:name>Philippe Le Bon</foaf:name>
      <foaf:givenname>Philippe</foaf:givenname>
      <foaf:family_name>Le Bon</foaf:family_name>
    - <bio:event>
      - <bio:Death>
          <bio:date>1467</bio:date>
        </bio:Death>
      </bio:event>
```

Lemmatisation automatisée des inscriptions

18 L'un des objectifs du projet étant de proposer un corpus lemmatisé, face à un ensemble plurilingue, composé de textes courts, stylisés et remplis de noms propres et de dates, les premières expériences de lemmatisation menées par Nicolas Perreaux visaient à la fois à affiner la formalisation du corpus, à tester les paramètres existants et à proposer des voies possibles de traitement automatisé. La formalisation du corpus a consisté principalement à séparer des inscriptions multiples d'un lieu réunies en une seule fiche au moment de leur publication, à réaliser la réduction de la datation à l'expression d'une année et à traiter un certain nombre de chaines de caractères ajoutées par les éditeurs. Sur la base du nouveau tableau ainsi constitué, l'enjeu était d'observer dans quelle mesure le jeu d'étiquettes linguistiques attribuées par les épigraphistes/philologues aux inscriptions (latin, français, lat_français = prédominance du latin, fr_latin = prédominance du français) correspondrait ou pas à la granularité issue d'une analyse des textes par l'ordinateur dont les paramètres sont établis par langue (Fig. 8).

Fig. . Composition linguistique du corpus (N. Perreaux).



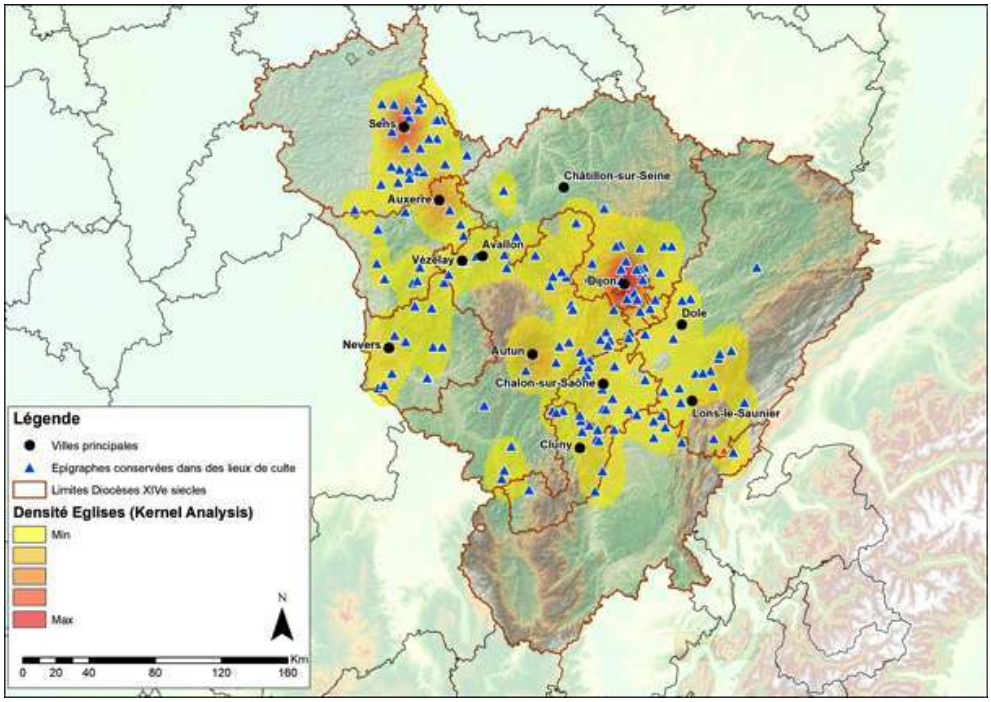
19 Des résultats globalement bons ont été obtenus avec les inscriptions latines étiquetées avec le logiciel TreeTagger[14] et les paramètres pour le latin médiéval OMNIA (dir. Alain Guerreau)[15], en utilisant le tokenizer développé pour le corpus CEMA (N. Perreaux)[16] et qui contourne le problème des noms propres en attribuant systématiquement la valeur NAM aux mots possédant une majuscule et ne se situant pas au début d'une phrase. Pour les inscriptions en ancien français ont été utilisés deux outils, les paramètres pour TreeTagger issus du travail sur le *Nouveau Corpus d'Amsterdam* (Achim Stein)[17] et le lemmatiseur Pandora (Mike Kestemont, Jean-Baptiste Camps, Thibault Clerice)[18], qui demande quelques adaptations du tokenizer, avec une reconnaissance un peu meilleure par le premier.

20 En fin de parcours, l'une des solutions envisageables pour lemmatiser les inscriptions et surmonter les difficultés posées par les différents degrés de mélange linguistique serait de laisser les lemmatiseurs départager les langues en comparant le nombre de termes non identifiés (*unknows*) pour ne retenir, *in fine*, que la version la plus « efficace ».

Géolocalisation des inscriptions

21 Profitant du fait que les limites des diocèses bourguignons au Moyen Âge ont été revus à nouveaux frais par Davide Gherdevich dans le cadre de son travail au sein de l'ANR Col&Mon[19], l'exploitation par lui des données de géolocalisation des inscriptions a été l'occasion d'observer leur distribution à l'aide des SIG. La répartition chronologique, outre l'accroissement du nombre de témoins au cours des siècles, montre aussi le passage d'une production dispersée au haut Moyen Âge vers une concentration de plus en plus marquée dans et autour des centres urbains du nord de la Bourgogne (Fig. 9).

Fig. . Répartition des inscriptions. Carte de densité (D. Gherdevich).



22 Comme pour la lemmatisation, l'utilisation pratique des métadonnées qui accompagnent le corpus conduit à réfléchir à leur amélioration progressive. La possibilité de rétroinjecter depuis les cartes le diocèse d'origine de chaque inscription semble une priorité pour le projet afin de proposer une répartition du corpus sur des critères médiévaux. Les inscriptions étant souvent conservées dans des dépôts lapidaires, comme les musées, il faudra aussi préciser, dans la mesure du possible, leur provenance. D'autres caractéristiques, comme l'institution productrice, le support, le type d'écriture et la taille des caractères, la typologie simplifiée, la conservation ou disparition, pourront être renseignés progressivement. Il n'en reste pas moins que dans son état actuel, avec la mise au point réalisée à la suite de cette journée d'études, le corpus d'inscriptions médiévales de la Bourgogne peut être diffusé et nourrir les recherches à venir.

Rappel du programme[20] :

Acquisition et traitements

- Introduction. Autour du corpus épigraphique bourguignon : chiffres, chronologie, langues - Eliana Magnani (CNRS, Lamop) et Estelle Ingrand-Varenne (CNRS, CESCM)
- De l'OCR à Philologic. Un parcours informatique - Pierre Brochard (CNRS, Lamop)
- Le matériel épigraphique révisé et complété. Le traitement des inscriptions de la fin des années 1990 à aujourd'hui - Aurore Menudier (CESCM)

Exploitations

- La lemmatisation des inscriptions est-elle possible ? - Nicolas Perreaux (Université de Francfort)

- Les inscriptions d'après des expériences en SIG - Davide Gherdevich (Univ. Versailles-Saint-Quentin, DYPAC)

NOTES

1. http://cescm.labo.univ-poitiers.fr/publi/corpus-des-inscriptions-de-la-france-medievale/.
2. http://www.cbma-project.eu/.
3. E. MAGNANI, « Un corpus structuré et hétérogène de textes latins médiévaux (Bourgogne, Ve-XVe siècle) », *Bulletin du CERCOR - Centre Européen de recherches sur les congrégations et ordres religieux,* 41, 2017, p. 59-65 - https://halshs.archives-ouvertes.fr/halshs-01529451.
4. R. FAVREAU, J. MICHAUD, B. MORA, *Jura, Nièvre, Saône-et-Loire*, Paris, 1997 (Corpus des inscriptions de la France médiévale, 19) - http://www.persee.fr/doc/cifm_0000-0000_1997_cat_19_1 ; ID., *Côte-d'Or*, Paris, 1999 (Corpus des inscriptions de la France médiévale, 20) http://www.persee.fr/doc/cifm_0000-0000_1999_cat1; R. FAVREAU, J. MICHAUD, *Yonne*, Paris, 2000 (Corpus des inscriptions de la France médiévale, 21) - http://www.persee.fr/doc/cifm_0000-0000_2000_cat_21_1.
5. https://cosme.hypotheses.org/.
6. Ont participé à cette journée d'études : Mathieu Beaud (Univ. Paris1 - LaMOP), Sébastien Brisbois (Univ. Strasbourg), Pierre Brochard (CNRS - LaMOP), Hélène Caillaud (Univ. Limoges ; CBMA - LaMOP), Olivier Canteaut (EnC), Davide Gherdevich (UVSQ - DYPAC), Alain Guerreau (CNRS), Estelle Ingrand-Varenne (CNRS - CESCM), Laurent Jégou (Univ. Paris 1 - LaMOP), Eliana Magnani (CNRS - LaMOP), Aurore Menudier (CESCM), Nicolas Perreaux (Goethe Universität ; LaMOP), Clara Renedo (EnC ; Titulus), Leandro Alves Teodoro (Fapesp ; Unicamp).
7. Les métadonnées de départ sont calquées sur la structure de l'édition du CIFM – Département, Ville, Localisation, Titre, Date CIFM – auxquelles ont été ajoutées : ID projet, id wikidata, Date simplifiée, nomfichiertext, Latitude, Longitude, Texte développé, Langue, Corrections/commentaires, Auteur de la notice, URL Persée, URL Mérimée.
8. 18 pour le Jura, 20 pour la Nièvre, 90 pour la Saône-et-Loire, 112 pour la Côte d'Or, 231 pour l'Yonne.
9. http://titulus.huma-num.fr/recherche/recherche.php.
10. https://oriflamms.hypotheses.org/.
11. http://www2.culture.gouv.fr/public/mistral/dapamer_fr?ACTION=NOUVEAU&.
12. https://gitlab.huma-num.fr/lamop/cbma-epigraphie/commits/master.
13. http://philologic.lamop.fr/epigraphie/.
14. http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/.
15. http://glossaria.eu/.
16. N. PERREAUX, « L'écriture du monde (I). Les chartes et les édifices comme vecteurs de la dynamique sociale dans l'Europe médiévale (VIIe-milieu du XIVe siècle) », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA* [En ligne], 19.2 | 2015, mis en ligne le 11 juillet 2016, consulté le 31 octobre 2018. URL : http://journals.openedition.org/cem/14264 ; DOI : 10.4000/cem.14264.
17. https://sites.google.com/site/achimstein/research/resources.
18. https://github.com/hipster-philology/pandora.
19. D. GHERDEVICH, « Les limites des diocèses au Moyen Âge : sources historiques et outils d'interprétation SIG », *Carnet de recherche COL&MON*, 2017, https://colemon.hypotheses.org/21.

20. https://cosme.hypotheses.org/765.

INDEX

Index géographique : Bourgogne
Index chronologique : Moyen Âge

AUTEURS

ELIANA MAGNANI

CNRS - LaMOP

ESTELLE INGRAND-VARENNE

CNRS - CESCM