



HAL
open science

L'eau, l'écrit et la société (IXe-XIIe siècle). Étude statistique sur les champs sémantiques dans les bases de données [C.B.M.A. et autres]

Nicolas Perreaux

► **To cite this version:**

Nicolas Perreaux. L'eau, l'écrit et la société (IXe-XIIe siècle). Étude statistique sur les champs sémantiques dans les bases de données [C.B.M.A. et autres]. Bulletin du Centre d'études médiévales d'Auxerre, 2011, 5, 10.4000/cem.12062 . halshs-01148894

HAL Id: halshs-01148894

<https://shs.hal.science/halshs-01148894v1>

Submitted on 5 Feb 2020

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.



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

15 (2011)

Varia

Nicolas Perreaux

L'eau, l'écrit et la société. Étude statistique sur les champs sémantiques dans les bases de données [CBMA]

Mémoire de Master 2 de l'université de Bourgogne,
sous la direction d'Eliana Magnani, Dijon, 2010.

Avertissement

Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur.

Les œuvres figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage soit personnel, soit scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document.

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

revues.org

Revues.org est un portail de revues en sciences humaines et sociales développé par le Cléo, Centre pour l'édition électronique ouverte (CNRS, EHESS, UP, UAPV).

Référence électronique

Nicolas Perreaux, « L'eau, l'écrit et la société. Étude statistique sur les champs sémantiques dans les bases de données [CBMA] », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA* [En ligne], 15 | 2011, mis en ligne le 25 août 2011, consulté le 17 juin 2015. URL : <http://cem.revues.org/12062> ; DOI : 10.4000/cem.12062

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

<http://cem.revues.org>

<http://www.revues.org>

Document accessible en ligne sur :

<http://cem.revues.org/12062>

Document généré automatiquement le 17 juin 2015. La pagination ne correspond pas à la pagination de l'édition papier.

© Tous droits réservés

Nicolas Perreux

L'eau, l'écrit et la société. Étude statistique sur les champs sémantiques dans les bases de données [CBMA]

Mémoire de Master 2 de l'université de Bourgogne, sous la direction d'Eliana Magnani, Dijon, 2010.

Pagination de l'édition papier : p. 439-449

- 1 Intitulé « L'eau, l'écrit et la société. Étude statistique sur les champs sémantiques dans les bases de données [CBMA] »¹, notre mémoire de master partait d'un constat relativement simple : les chartes, ces documents conservés plutôt en grande abondance, contiennent un nombre extrêmement important de termes relatifs à la description et donc, par extension, à la perception du monde par les médiévaux. Or, il est clair que ces formules, plus ou moins stéréotypées, n'ont presque jamais été considérées comme des informatrices² potentielles en ce qui concerne ce lien entre l'homme médiéval et le *mundus*, *i.e.* la Création qui l'entoure. Néanmoins, dès l'orée de l'étude, il est rapidement apparu qu'une telle recherche ne pouvait faire abstraction de plusieurs problématiques plus « classiques » relatives à ce type documentaire : questions de distribution, de géographie et de chronologie. Nous devons donc partir d'un large *corpus* – la chose étant grandement facilitée par la présence d'un outil exceptionnel au sein du laboratoire ARTeHIS (UMR 5594, CNRS – Université de Bourgogne – Ministère de la Culture) : les CBMA sous Philologic³ –, étudier initialement sa constitution et son degré d'homogénéité, avant de finalement rejoindre ces considérations d'ordre anthropologique. Fondamentalement, notre recherche s'est donc articulée autour de trois axes :
- 2 1. La question du développement inégal au sein du système féodal, tout d'abord à travers le problème de la production documentaire différenciée dans l'espace et dans le temps, mais aussi à partir d'une réflexion sur les zones de circulation des formulaires d'actes, zonations⁴ que nous avons proposé d'appeler « aires de scripturalités ». Partant systématiquement d'observations concernant le champ sémantique de l'eau, cette première partie, intitulée « L'écriture de l'eau » visait avant tout à déterminer une nouvelle chronologie de l'écriture diplomatique, tout en faisant apparaître des différences régionales importantes, que nous avons aussi essayé de catégoriser et d'explicitier en les rapprochant des structures sociales. Le problème fut envisagé à deux échelles bien distinctes, tout d'abord celle de la Bourgogne, mais aussi, avec des sondages plus ponctuels, au niveau d'une partie de l'Europe occidentale – royaume des Francs, Catalogne, Lombardie, Angleterre et Empire.
- 3 2. Une seconde question largement évoquée fut bien entendu celle du champ sémantique de l'élément aquatique à proprement parler⁵. Quand est-il employé ? Quelle est sa richesse ? *Aqua*, *fluvius*, *rivus*, *fluviolus*, *gutta* sont-ils utilisés de manière indistincte, comme on le suggère parfois ? L'ensemble de ces développements s'inspirèrent abondamment des travaux de Michel Zimmermann sur la Catalogne, en particulier ceux portant sur la question de la dénomination et de la localisation (chapitres « Situer et décrire / dire et inscrire : la mémoire et les cours d'eau » et « L'attention des hommes à l'élément aquatique »)⁶. Cette question a ainsi permis de nous interroger sur les catégories et la typologie attachées au champ sémantique de l'élément aquatique, en particulier avec une mise en parallèle des chartes, d'une part, et des textes d'Isidore de Séville puis de Raban Maur (*De Universo*) d'autre part : « *Aquarum naturae multa est diversitas*⁷ », écrit justement le premier de ces auteurs.
- 4 3. Enfin, un axe d'ordre plus anthropologique, inspiré, entre autres et pour ce champ de recherche, par les travaux de Claude Lévi-Strauss⁸, de Maurice Godelier⁹ et de Philippe Descola¹⁰, où l'on s'est interrogé sur l'articulation du champ sémantique de l'eau, non

seulement en ce qui concerne sa structuration interne, mais aussi à propos des liens qu'il entretenait avec les autres groupes du lexique. Cette dernière partie, intitulée « Le monde de l'élément aquatique », a permis de déterminer les principales valences que pouvait prendre l'élément aquatique dans le système de représentation de l'Occident médiéval (lien entre *aqua* et *scriptura* par exemple) et de comprendre en quoi elles pouvaient nous éclairer sur le rapport que ces collectifs humains – par ailleurs fortement différenciés – entretenaient avec le monde qui les entourait. Elle fut aussi l'occasion de s'interroger sur le rôle de l'analogie au sein de l'Occident médiéval, non seulement en ce qui concerne les pratiques de l'écrit (omniprésence de formules stéréotypées construites, déconstruites, copiées), mais aussi à propos de son influence sur la structuration et la création du sens au Moyen Âge carolingien et central.

Production documentaire et développement inégal

- 5 Dans un premier temps, l'emploi de la statistique a donc permis de mettre en lumière des rythmes de production documentaire intensément scindés, non seulement au sein de la Bourgogne, mais aussi à l'échelle de l'Europe. Ici, nous avons eu recours à différentes bases de données, de tailles variables : les CBMA bien entendu, celle des originaux de l'Artem¹¹, mais aussi d'autres fonds numérisés, disponibles grâce à la générosité d'un grand nombre de chercheurs et professeurs, faisant ainsi intervenir volontairement un jeu d'échelles complexe.
- 6 L'analyse, recourant abondamment à des techniques de fouille de données – *Data Mining* et *Text Mining* –, en particulier l'analyse factorielle des correspondances et l'analyse en composantes principales, a permis de montrer que la Bourgogne méridionale possédait un « excédent documentaire » très important pour le X^e et la première moitié du XI^e siècle – au-delà même du cas clunisien –, la partie septentrionale de la région voyant quant à elle plutôt arriver des chartes aux XII^e et XIII^e siècles (moment où elles deviennent par ailleurs relativement rares dans le sud de cet espace). Ces constatations furent également l'occasion d'explorer les rapports diplomatiques qu'entretenait Cluny avec certains *scriptoria* proches : Savigny, Sauxillanges, Saint-André-le-Bas de Vienne, mais surtout le chapitre Saint-Vincent de Mâcon, dont on a cherché à montrer le rôle clé dans la mise en place de certaines formules omniprésentes dans le Cluny des premiers temps ; en particulier à travers l'emploi d'un formulaire de description de parcelles endémique de cette zone : « *a mane [...] a medio die [...] a sero [...]* ». L'ensemble de ces observations nous a alors permis de revenir sur la question du lien entre Cluny et Tours et, pour notre part, de défendre l'idée d'une influence diplomatique avant tout locale, au niveau de la zone méridionale de cet espace.
- 7 Poussant plus loin nos investigations, on a pu montrer que cette double dynamique, qui scindait la Bourgogne et les CBMA en deux ensembles nettement distincts – l'effet de gradient nord/nord-est jouant à quelques dizaines de kilomètres seulement –, s'étendait en fait à l'ensemble du royaume des Francs, avec un grand nombre d'actes anciens dans les zones centrales, mais aussi méridionales : Auvergne, Limousin, Poitou, PACA, Midi-Pyrénées, Catalogne, etc. À l'inverse, les zones du nord de l'espace considéré – Picardie, Champagne, Normandie, etc. –, si elles possèdent un certain nombre d'actes mérovingiens et carolingiens (*grosso modo* antérieurs à 850), voient avant tout leurs fonds documentaires se remplir pour une période couvrant le XII^e et le XIII^e siècle. En définitive, on propose de rapprocher la production documentaire de plusieurs zones : Bourgogne, Auvergne et Rhône-Alpes, qui forment ainsi un premier groupe très proche, se développant très rapidement aux X^e et XI^e siècles, suivis de peu par le Poitou, le Limousin, la Provence mais aussi la Catalogne. L'avance de l'espace méridional se réduit quelque peu dans les zones du sud et sud-ouest, qui accusent parfois un certain retard, de plus en plus intense à mesure que l'on va vers l'ouest : Languedoc, Midi-Pyrénées et surtout Aquitaine. Le nord semble lui aussi connaître plusieurs rythmes, avec des espaces à développement plus rapide (Pays-de-la-Loire, Centre par exemple) et d'autres clairement plus en retard (Champagne, Picardie, etc.). Un phénomène de production documentaire inégale donc, que nous avons jugé bon, en première approximation, de rapprocher d'autres éléments caractéristiques de la société médiévale, en particulier architecturaux (roman/gothique). D'une manière générale, les chartes furent envisagées comme des *traces* de la *dynamique sociale*, traces qui montrent qu'au-delà

des pertes, il existe une véritable cohérence, une logique globale, dans la structure des fonds actuellement conservés¹².

- 8 Cette première analyse des zonations a ensuite été confirmée par l'étude de la circulation de formes/lemmes et autres formules spécifiques du champ sémantique de l'eau. De fait, nombre de termes issus de ce *Wortfeld* sont spécifiques à certaines régions, plus ou moins vastes, et permettent de (re)construire une « carte théorique » et chronologique de la diffusion de ces mots ; une carte qui ne doit sans doute rien à la fantaisie ou au prétendu arbitraire des scribes. C'est le cas des lemmes *molinum*, *fluviolus* ou *imber*, mais aussi celui de la formule « *Sicut aqua extinguit ignem, ita elemosinam extinguit peccatum* » – omniprésente pour la Bourgogne à Cluny, Saint-Vincent de Mâcon, Paray-le-Monial, mais strictement absente des *corpus* du nord de cet espace. Nous avons par la suite étudié sa répartition dans près de 80 000 chartes indexées, confirmant les remarques ci-dessus à propos des zonations et de l'opposition nord/sud. Ces questionnements ont été menés sur la totalité des chartes à notre disposition, en tenant compte de la difficulté qu'il y a à prendre en considération à la fois la dimension spatiale mais aussi chronologique de ce type de problèmes (il s'agit de biais générés par l'hétérogénéité des *corpus*, mais une hétérogénéité structurelle – et dont on cherchait précisément à montrer la logique structurale –, et pas seulement fonctionnelle ou méthodologique). Différents outils statistiques ont alors mis en œuvre, afin de s'assurer que les phénomènes observés n'étaient pas générés par de « simples » effets de *corpus* : regroupement de chartes en paquets égaux¹³, mais aussi *Data Mining* et analyses factorielles. Il ressort de cette double analyse – production documentaire, diffusion des syntagmes spécifiques – qu'il existait dans l'ensemble de l'espace étudié différentes zonations, très nettes, s'emboîtant les unes avec les autres, dans un jeu d'échelles complexe. Nous voyons ainsi la Bourgogne du sud se raccrocher nettement à la région rhodanienne, à l'Auvergne et au Poitou... mais aussi, d'une manière plus large, à la Provence et à la Catalogne. À l'inverse, la Bourgogne du nord, à travers des éditions telles que le *Cartulaire général de l'Yonne*¹⁴ ou encore les *Chartes et documents de Saint-Bénigne de Dijon*¹⁵, se rapproche nettement des productions d'une large part de l'Île-de-France¹⁶, du nord-est, de la Normandie, etc. Ce faisceau d'indices permet de conclure non seulement à l'existence de pratiques d'écritures résolument distinctes, aux différences stables sur le temps long (plusieurs siècles), mais aussi à la persistance de structures sociales clairement différenciées et auparavant rarement mises en valeur. Lors de cette réflexion initiale, deux concepts ont abondamment guidé notre réflexion : celui d'*échelle* et celui de *seuil*.

L'écriture de l'eau : chronologie, espace, mouvement

- 9 La seconde et la troisième parties furent alors consacrées à l'étude du champ sémantique de l'eau proprement dit, dans les chartes en particulier. Ici, les travaux d'Alain Guerreau, mais aussi ceux de Michel Zimmermann servirent – nous l'avons déjà mentionné – de point de départ. Dans un premier temps, il a fallu reconstruire la liste des lemmes composant le champ sémantique de l'eau – opération fastidieuse mais nécessaire –, à partir des sources (chartes et *Patrologie Latine*), mais aussi à partir des dictionnaires de latin médiéval. Au cours de cette étape, nous avons pu réaliser que le champ sémantique était non seulement passablement présent dans les textes hagiographiques et exégétiques, mais également dans les documents diplomatiques, sous des formes extrêmement variées même chez ces derniers, depuis la présence dans des formules de préambules (« *Ecclesia insuper Christi undique adversitatum procellis irruentibus divini auxilii anchora firmata ad æternæ securitatis littus tendens vix periculum evadat naufragii, hanc quam prædixi ecclesiam indignam alieni cognovi amminiculi.* » dans une charte du fonds d'Autun¹⁷), jusqu'à la désignation des cours d'eau.
- 10 Cette seconde partie nous a aussi permis de déterminer puis de nous intéresser aux liens entre ce champ sémantique et d'autres proches, en particulier ceux de l'espace et du mouvement. En effet, une étude plus approfondie des associations entre le mot *aqua* et les termes désignant l'espace (associations par cooccurrences telles *usque aquam*, *sub aquam*, *ultra aquam*) montre qu'elles sont trois fois plus présentes après 1030 qu'avant cette période. Plusieurs méthodes furent d'ailleurs mises en place afin de quantifier les évolutions de ces vocables. De fait, travailler sur de si larges *corpus* pose d'innombrables problèmes d'ordre statistique mais aussi

théorique, la masse n'arrangeant rien ou peu à l'affaire. Dans le cas présent, nous avons eu recours à deux techniques – souvent combinées – : d'une part, l'utilisation assez massive de logarithmes, permettant de limiter fortement divers effets liés à l'hétérogénéité des *corpus*, d'autre part, le comptage des formes à partir de paquets de chartes égaux. Cette dernière méthode, qui consiste à « découper » la chronologie en plusieurs moments contenant autant de chartes les uns que les autres – à quelques pour cents près –, permet de s'affranchir de la plupart des effets de *corpus* et de comparer des nombres d'occurrences dans des ensembles de mots relativement égaux.

11 Ainsi donc, on a pu vérifier une explosion du vocabulaire spatial passé 1030 dans le sud de la Bourgogne, un peu plus tard dans le nord. Cette augmentation de l'association entre eau et espace s'accompagne, en parallèle, de la diminution de l'association – toujours par cooccurrences – entre eau et mouvement (*aqua vergit, aqua discurrente*, etc.). Tout se passe donc comme si, entre 1030 et 1080, l'eau était de plus en plus associée à l'espace – avec l'emploi beaucoup plus fréquent des formes à l'accusatif, type *aquam*, qui permettent de désigner, de montrer – tandis que le mouvement prenait une place de plus en plus faible dans les documents diplomatiques. Cette dynamique est accompagnée par l'emploi de plus en plus concret des termes du champ sémantique. Une analyse du sens donné à *aqua*, en fonction de son degré d'abstraction (très abstrait : préambules et formules stéréotypées – très peu abstrait/concret : hydronymes et toponymes) montre que pour cette même période – 1030-1080 donc –, on note un très net infléchissement de la valeur attribuée aux termes employés et un passage d'un moment où l'eau est essentiellement utilisée de manière abstraite, inquantifiable, à un moment où on l'emploie avant tout comme une chose concrète et définie. D'un point de vue social, il nous paraît intéressant de rapprocher toutes ces observations concordantes du concept d'encellulement proposé par Robert Fossier¹⁸. Plus intéressant encore, il semble que cette émergence de l'espace connaît d'amples variations – ce qui nous ramène bien entendu à notre réflexion sur le développement inégal – d'une région à une autre, ne serait-ce qu'entre Bourgogne du sud et Bourgogne du nord. Le mémoire fut ainsi l'occasion de s'interroger sur l'opérabilité des modèles historiographiques du système féodal et de montrer – en substance – que l'opposition, souvent avancée comme canonique, entre la chronologie proposée par Georges Duby et celle avancée par Marc Bloch est illusoire, les deux étant en fait complémentaires, l'une et l'autre fonctionnant en fait à des échelles différentes, mais qui correspondent à des phénomènes de mise en place de structures similaires : en Bourgogne du sud pour le premier ; à l'échelle de l'Europe et en moyenne pour le second. À notre sens, les évolutions bien réelles d'effectifs relatifs à des champs sémantiques tels que l'eau, l'espace, le mouvement, sont des indices précieux de la mise en place de nouvelles structures matérielles, mais aussi d'un nouveau rapport au monde – en disant cela, nous pensons bien évidemment aux travaux de Maurice Godelier sur la part « idéelle du réel »¹⁹.

12 Poursuivant sur cette voie, nous avons aussi « testé » l'évolution que connaissaient les vingt lemmes les plus courants du champ sémantique de l'eau, en bloc, en recourant cette fois aux coefficients de corrélations. Le *corpus* a alors été découpé en quatre parties contenant un nombre égal de chartes, pour ensuite dénombrer les occurrences de chacun des vingt lemmes dans ces quatre moments. Il suffisait ensuite de calculer les coefficients de corrélations sur ces quatre colonnes/périodes pour observer celles qui étaient les plus liées : à nouveau, sans guère de surprise désormais, on note une rupture très nette entre les années 1030 et 1080. Cette flexion très forte est d'ailleurs d'autant plus remarquable qu'elle se produit en seulement quelques décennies et s'oppose, par exemple, à la grande stabilité des effectifs repérée pour les périodes précédentes. L'ensemble de cette réflexion a ainsi permis de revenir sur la question de la « mutation féodale », le champ sémantique de l'eau étant, en effet, un indicateur relativement neutre – *a contrario* des lemmes tels que *miles* ou *feodum* sur lesquels a porté le débat jusque-là – de l'évolution remarquable, certes dépendante des régions, de l'ensemble du vocabulaire entre le début du XI^e siècle et la fin du XII^e : une évolution d'autant plus frappante qu'à nouveau son intensité n'a d'égalé que sa rapidité, une fois le tournant lexical amorcé.

Le monde de l'élément aquatique : entre espace et non-espace²⁰

- 13 La dernière partie de notre travail – volontairement inspirée par les travaux de Philippe Descola sur la question des rapports à l'altérité et au monde – fut consacrée à l'étude de la structuration du champ sémantique de l'eau dans la Patrologie Latine. Partant d'une citation de Pierre le Vénérable relevée par Dominique Iogna-Prat – « *Aspice aquam in glaciem commutatam, glaciem in crystallum alteratam, et cernes in his aperte substantias permutari, formas substantiarum seu species reservari* »²¹ –, on a voulu savoir quelle valence était attribuée à chaque élément du champ et surtout quel était son dénominateur commun (chapitre « Les valences du champ sémantique de l'eau : comparatisme et formalisation »). On a pour cela réalisé une série d'analyses factorielles sur un tableau de cooccurrences généralisées pour vingt-et-un mots-clés relatifs au champ – en fait vingt-et-une espèces de l'élément aquatique : *aqua, fons, lacus, rivus, unda, oceanus, fluvius, mare, diluvium, lacrima, abyssus, nubes, pluvia, imber, ros, gutta, nix, etc.* Différents traitements ont alors été appliqués à ces vastes tableaux de cooccurrences, en particulier en employant un « codage logique » ou « codage disjonctif complet »²².
- 14 Très nettement, ces expériences attestent que le champ sémantique est marqué par une forte opposition entre ce que nous avons nommé « espèces terrestres » et « espèces célestes », chacune de ces grandes catégories s'organisant de part et d'autre des analyses factorielles. Nous avons ainsi pu montrer de manière automatique et non *a priori* que c'était bien l'opposition du charnel et du spirituel, si bien relevée par Anita Guerreau-Jalabert, qui gouvernait la structuration du champ sémantique de l'élément aquatique et ceci sur le temps long. Cette première clé sémantique déterminée, elle fut appliquée à plusieurs cas précis, afin de vérifier son opérabilité.
- 15 Dès lors, une analyse plus « qualitative » a été engagée, recourant abondamment au *De miraculis* de Pierre le Vénérable²³ (qui contient plusieurs passages renvoyant à la question de l'articulation entre eau et espace), mais aussi à l'introduction du *Cartulaire de l'abbaye d'Uzerche*²⁴, ou encore au préambule clunisien – que l'on retrouve par ailleurs au même moment chez Odon de Cluny – relatif à la parabole de Lazare et du pauvre²⁵ – et dont l'articulation des principaux éléments a par la suite été analysée : eau / feu / pauvreté / richesse / intériorité / extériorité / immatériel / matériel / sein d'Abraham / enfer / mie de pain / goutte d'eau/haut (céleste) / bas (infernale). L'ensemble de ces investigations, recourant de manière secondaire à des sources iconographiques²⁶, permet de montrer que l'eau était presque systématiquement associée : 1. à la pauvreté (et par extension au don). 2. au rejet de l'extériorité et du matériel. 3. à l'intériorité (spirituelle). En parallèle, d'autres investigations ont été poursuivies sur la nature (*i.e.* les qualités) de l'eau baptismale en tant qu'élément de *transitus* (chapitre : « La double nature des eaux : correspondances et *visio* dans le sacrement du baptême »).
- 16 D'une manière plus large, les multiples valences - essentiellement positives - attribuées à l'eau (et à son champ sémantique), qui agit donc un peu comme un « joker » au sein du système de représentation, nous ont permis de revenir sur l'opérabilité du modèle des ontologies et du régime analogique tels qu'ils sont définis par Philippe Descola. Cette réflexion fut ainsi l'occasion de revenir sur la préface des *Historiae* de Raoul Glaber à propos de la *Divine quaternité*²⁷ et de tenter de montrer le rôle fondamental qu'y jouent non seulement l'élément aquatique, mais aussi l'*analogisme généralisé*. En définitive, l'eau, élément clé du *passage* au sein du système de l'Occident féodal – aussi bien au plan idéal que social –, tissait sans cesse des liens avec la plupart des autres éléments du système de représentation, est, à notre sens, le paradigme même de ce fonctionnement par hiérarchisation et analogie. L'hypothèse posée est donc que les inégalités d'écriture et de production documentaire à l'échelle européenne traduisent non seulement le développement inégal du système féodal – phénomène qui se note par exemple dans l'emploi de termes, de *discours*²⁸, très nettement différenciés – mais aussi des rapports (sémantiques) au monde et à l'autre relativement dissemblables dans le temps et dans l'espace, le tout cohabitant au sein d'un système pourtant cohérent. C'est cette

hypothèse globale que nous souhaitons explorer lors d'une thèse en cours²⁹, en recourant abondamment aux techniques de fouilles de données (*Data/Text Mining*) et en employant cette fois un *corpus* documentaire déjà réuni par nos soins, comprenant, outre les textes théologiques plus de 150 000 chartes numérisées.

Notes

1 Nous tenons à remercier chaleureusement Eliana Magnani et Alain Guerreau pour leur relecture attentive et pour leurs conseils lors de la rédaction de ce résumé de notre mémoire. Chantal Palluet, Pierre Bonnerue et Daniel Russo ont participé à la relecture finale de cet article : qu'ils en soient remerciés eux aussi. Pour des raisons évidentes de concision, les références bibliographiques de l'article ont été limitées au strict minimum.

2 Le terme « informatrice » est employé à dessein. Il s'agit d'insister sur la dimension interactive de la recherche et sur le fait que la « réaction » des documents n'est jamais passive, mais dépend étroitement du questionnaire qu'on leur applique.

3 À propos des *CBMA*, voir l'ensemble des articles d'Eliana Magnani et de Marie-José Gasse-Grandjean dans le *Bucema* [<http://cem.revues.org/>] : n° 9 (2005), p. 179 [index751.html] ; 11 (2007), p. 163-167 [index1064.html] ; 12 (2008), p. 237-244 [index6962.html] ; 13 (2009), p. 245-251 [index11077.html] ; 14 (2010), p. 197-207 [index11433.html] et p. 273 [index11582.html].

4 Il s'agit d'un concept emprunté à la géographie et à l'écologie scientifique, servant habituellement à désigner l'étendue d'un écosystème, d'un climat ou d'un phénomène naturel en général.

5 J. TRIER, *Der Deutsche Wortschatz im Sinnbezirk des Verstandes: Die Geschichte eines sprachlichen Feldes, Band 1: Von den Anfängen bis zum Beginn des 13 Jahrhunderts*, Heidelberg, 1931 ; *Id.*, « Über Wort- und Begriffsfelder », in L. SCHMIDT (éd.), *Wortfeldforschung : zur Geschichte und Theorie des sprachlichen Feldes*, Darmstadt, 1973, p. 1-38 ; A. GUERREAU, « Le champ sémantique de l'espace dans la *vita* de Saint Maieul (Cluny, début du XI^e siècle) », *Journal des savants*, 1997, p. 363-419 ; *Id.*, *L'avenir d'un passé incertain. Quelle histoire du Moyen Âge au XXI^e siècle ?*, Paris, 2001 ; *Id.*, « “Textus” chez les auteurs latins du XII^e siècle », in L. KUCHENBUCH et U. KLEINE (éd.), “Textus” im Mittelalter, Göttingen, 2005, p. 149-178 ; A. GUERREAU-JALABERT, « Parole/Parabole : analyse d'un champ lexical et sémantique », in R.-M. DESSI et M. LAUWERS (éd.), *La parole du prédicateur*, Nice, 1997, p. 311-339 ; A. GUERREAU-JALABERT et B. BON, « *Pietas* : réflexions sur l'analyse sémantique et le traitement lexicographique d'un vocable médiéval », *Médiévales*, 42 (2002), p. 73-88 ; *Eid.*, « Le trésor au Moyen Âge : étude lexicale », in L. BURKART, P. CORDEZ, P.-A. MARIAUX et Y. POTIN (dir.), *Le trésor au Moyen Âge. Discours, pratiques et objets*, Florence, p. 11-32.

6 M. ZIMMERMANN, « Glose, tautologie ou inventaire ? L'énumération descriptive dans la documentation catalane du X^e au XII^e siècle », *Cahiers de linguistique et de civilisation hispaniques médiévales*, 14-15 (1989-1990), p. 309-338 ; *Id.*, « Les noms de la route et du chemin dans la Catalogne médiévale », in C. DESPLAT (dir.), *Terres et hommes du sud. Hommage à Pierre Tucoo-Chala*, Biarritz, 1992, p. 387-413 ; *Id.*, *Écrire et lire en Catalogne (IX^e-XII^e siècle)*, Madrid, 2003.

7 ISIDORUS HISPALENSIS, *Etymologiae. Libro XIII, De mundo et partibus*, éd. PL 82, col. 73-0728C, ici col. 482A-482B ; édition plus récente dans G. GASPAROTTO (éd. et trad.), *Etymologiae, Libro XIII, De mundo et partibus*, Paris, 2004.

8 Ici en particulier : C. LÉVI-STRAUSS, « La geste d'Asdiwal », in *Anthropologie structurale II*, Paris, 1973, p. 175-233 [1^{re} édition de l'article dans *École pratique des hautes études. Section des sciences religieuses. Annuaire 1958-1959*, Paris, 1958, p. 3-43 ; aussi in *Les temps modernes*, n° 179, mars 1962] ; *Id.*, *La Pensée sauvage*, Paris, 1962.

9 Cf. *infra*.

10 P. DESCOLA, *Par-delà nature et culture*, Paris, 2005.

11 B.-M. TOCK (dir.), M. COURTOIS, P. DEMONTY et M.-J. GASSE-GRANDJEAN (réalisé par), *La diplomatie française du haut Moyen Âge : inventaire des chartes originales antérieures à 1121 conservées en France*, 2 vol., Turnhout, 2001.

12 Certaines de nos expériences plus récentes tendent à montrer – après un dépouillement assez large de cartulaires – que la structure des originaux conservés est extrêmement proche de celle de l'ensemble des documents édités. Un point qui devra être exploré et vérifié lors d'une thèse en cours.

13 Cette pratique, clairement inspirée par les travaux et remarques d'Alain Guerreau (en particulier son cours intitulé *Statistique pour historiens*, professé à l'École des chartes en 2003-2004, disponible en ligne : <http://elec.enc.sorbonne.fr/statistiques/stat2004.pdf>), part d'une série de constatations sur la complexité des distributions lexicales. À la lecture de quelques ouvrages de référence, on comprend aisément qu'il est strictement impensable de se baser sur les effectifs bruts ou même – c'est sans

aucun doute pire – sur des pourcentages, afin d'effectuer des comparaisons entre *corpus* documentaires ou même entre différentes périodes. La plupart des outils restent donc à inventer et ces deux années de recherches furent aussi l'occasion d'en appliquer et de tester l'efficacité d'un certain nombre. Outre le cours mentionné ci-dessus, quelques jalons essentiels dans : H. R. BAAYEN, *Word Frequency Distributions*, Dordrecht, 2001 (Text, Speech and Language Technology n°18) ; S. EVERT, *The statistics of word cooccurrences: Word pairs and collocations*, Stuttgart, 2005 ; A. GUERREAU, « Pourquoi (et comment) l'historien doit-il compter les mots ? », *Histoire & mesure*, 4 (1989), p. 81-105 ; B. MANDELROT, *Les objets fractals. Forme, hasard et dimension*, Paris, 1975 ; *Id.*, *Fractales, hasard et finance*, Paris, 1997 (en particulier le chapitre 4, « Aléas du discours »).

14 M. QUANTIN (éd.), *Cartulaire général de l'Yonne*, 2 vol. et supplément, Auxerre, 1854-1860 et 1873.

15 G. CHEVRIER, M. CHAUME, R. FOLZ et J. MARILIER (éd.), *Chartes et documents de Saint-Bénigne de Dijon : prieurés et dépendances, des origines à 1300*, 2 vol., Dijon, 1943-1986.

16 Nous avons utilisé la base de données créée par l'École des chartes, qui contient plusieurs milliers d'actes franciliens.

17 A. DE CHARMASSE (éd.), *Cartulaire de l'Église d'Autun*, Paris, 1865-1900, n° 14 – 1033.

18 R. FOSSIER, *Enfance de l'Europe. Aspects économiques et sociaux*, 2 vol., Paris, 1982.

19 M. GODELIER, « La part idéale du réel », *L'Homme*, 18 (1978), p. 155-187 ; *Id.*, *L'idéal et le matériel – Pensée, économies, sociétés*, Paris, 1984.

20 A. GUERREAU, « Le champ sémantique... », *op. cit.*

21 PETRUS CLUNIACENSIS, *Adversus Petrobrusianos haereticos*, éd. PL 189, col. 719-0850D, ici col. 805A-805B ; voir aussi l'édition plus récente de J. FEARNES (éd.), *Contra Petrobrusianos hereticos*, Turnhout, 1968 (CCCM, 10) ; passage mentionné dans D. IOGNA-PRAT, *Ordonner et exclure. Cluny et la société chrétienne face à l'hérésie, au judaïsme et à l'islam (1000-1150)*, Paris, 1998, p. 148.

22 L'ensemble du tableau a été converti en binaire selon une règle du type : présence de la cooccurrence pour le mot-clé = 1 ; absence = 0.

23 D. BOUTHILLIER (éd.), *De Miraculis libri duo*, Turnhout, 1988 (CCCM, 83) ; D. BOUTHILLIER et J.-P. TORRELL (trad.), *Les Merveilles de Dieu*, Paris/Freiburg, 1992.

24 J.-B. CHAMPEVAL (éd.), *Cartulaire de l'abbaye d'Uzerche (Corrèze) du X^e au XIV^e siècle*, Paris, 1901.

25 « [...] *pro micis panis quas ulceroso Lazaro negavit, stillam aquae in inferno ardens desideravit [...]* » dans A. BERNARD et A. BRUEL (éd.), *Recueil des chartes de l'abbaye de Cluny*, 6 vol., Paris, 1876-1903, t. 1, n° 269, en 926. Nous avons aussi consacré plusieurs pages à cette charte, rédigée par un clerc présent à Saint-Vincent de Mâcon et sans doute proche de l'évêque de cette même ville, qui eut un impact remarquable sur les pratiques de l'écrit clunisien. Elle montre que les clunisiens ont avant tout puisé dans un répertoire de formules locales et ceci dès leurs origines (mais comment aurait-il pu en être autrement ?).

26 J. BASCHET, *Le sein du père. Abraham et la paternité dans l'Occident médiéval*, Paris, 2000.

27 RODULFUS GLABER, *Historiarum libri quinque*, éd. et trad. J. FRANCE, *Rodulfi Glabri. Historiarum libri quinque. Vita domni Willelmi abbatis (The life of St William, by the same author)*, Oxford, 1989 ; aussi éd. PL 142, col. 611-698 ; traduction française in M. ARNOUX (éd.), *Histoires*, Turnhout, 1996.

28 M. FOUCAULT, *Les mots et les choses*, Paris, 1966 ; *Id.*, *L'archéologie du savoir*, Paris, 1969.

29 Sous la direction d'Eliana Magnani et de Daniel Russo, ainsi qu'avec l'aide d'Alain Guerreau.

Pour citer cet article

Référence électronique

Nicolas Perreux, « L'eau, l'écrit et la société. Étude statistique sur les champs sémantiques dans les bases de données [CBMA] », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA* [En ligne], 15 | 2011, mis en ligne le 25 août 2011, consulté le 17 juin 2015. URL : <http://cem.revues.org/12062> ; DOI : 10.4000/cem.12062

Référence papier

Nicolas Perreux, « L'eau, l'écrit et la société. Étude statistique sur les champs sémantiques dans les bases de données [CBMA] », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 15 | -1, 439-449.

Droits d'auteur

© Tous droits réservés

Entrées d'index

Index de mots-clés : base de données, champ sémantique, eau, écrit, étude statistique, société