



HAL
open science

Qu'est-ce qu'un corpus? Compte-rendu de la journée d'études

Eliana Magnani

► **To cite this version:**

Eliana Magnani. Qu'est-ce qu'un corpus? Compte-rendu de la journée d'études. 2017,
<https://irht.hypotheses.org/3187>. halshs-01610087

HAL Id: halshs-01610087

<https://shs.hal.science/halshs-01610087>

Submitted on 4 Oct 2017

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.

Qu'est-ce qu'un corpus ? Compte-rendu de la Journée d'études.

par Eliana Magnani (CNRS - Lamop UMR 8589)

octobre 2017

paru dans *Les carnets de l'IRHT* : <https://irht.hypotheses.org/3187>

La journée d'études « Qu'est-ce qu'un corpus ? » a été organisée par l'équipe des CBMA (*Chartae Burgundiae Medii Aevi*) le 7 novembre 2016 à Paris¹, accueillie par l'IRHT, avec le soutien du Lamop et du Consortium Cosme. Elle avait pour point de départ le fait que, depuis les années 1980, mais surtout les années 2000, l'usage du mot « corpus » par les historiens médiévistes s'est largement diffusé sans que cet emploi généralisé ait été accompagné d'une quelconque réflexion sur ce qui se trouve derrière le terme, et encore moins d'un effort de conceptualisation. S'inscrivant en parallèle avec l'informatisation et la recrudescence des numérisations et des bases de données de documents médiévaux, cette situation contraste avec d'autres disciplines des sciences humaines et sociales, à commencer par la linguistique, qui depuis longtemps réfléchissent et débattent sur cette notion.

Entre mises au point, synthèses et études de cas, les communications et les vives discussions qui ont marqué cette fructueuse rencontre, confirment le constat préalable d'un usage impensé et d'une conception lâche du corpus qu'il conviendrait, à mon avis, de resserrer un tant soit peu. En effet, l'un des aspects récurrents soulevés par les participants lors des discussions sont les mots alternatifs, et parfois plus explicites, qu'on pourrait utiliser dans certaines occasions à la place de « corpus » : dossier, collection, bibliothèque... Il était évident que le mot « corpus » ne revêtait pas le même sens pour tous les présents, et ce compte-rendu reflète cette diversité.

¹ Ont participé à cette journée : Renaud Alexandre, Michael I. Allen, Sébastien Barret, Melissa Barry, Philippe Bernardi, Paul Bertrand, François Bougard, Caroline Bourlet, Olivier Bruand, Frédérique Cahu, Olivier Canteaut, Charles de Miramon, Vincent Debiais, Sébastien Fray, Patrice Frigault-Hamel, Hélène Gautier, Jean-Philippe Genet, Magali Guaresi, Maria Gurrado, Marlène Helias-Baron, Jeanne Imhauser, Estelle Ingrand-Varenne, Eliana Magnani, Joseph Morsel, Donatella Nebbiai, Claire Pernuit, Nicolas Perreaux, Birgitte Prod'homme, Coraline Rey, Daniel Russo, Sergio Torres, Bastien Tourenc, Cécile Treffort, Ismérie Triquet, Anaïs Wion.

Je remercie Marlène Helias-Baron de m'avoir communiqué ses notes détaillées prises au cours de cette journée ainsi que les intervenants qui ont partagé leurs supports de présentation et revu ce compte-rendu.

En guise d'introduction, je suis revenue sur les initiatives institutionnelles qui font du « corpus » un sujet d'actualité pour les chercheurs, à commencer par les premiers appels à projet ANR « Corpus et outils de la recherche en SHS » en 2006 et 2007, et la multiplication des initiatives qui s'ensuivirent, allant de la constitution d'axes de recherche et l'organisation de séminaires dans des laboratoires à la tenue de rencontres chez les médiévistes et ailleurs. C'est justement à l'occasion d'une journée d'études, dont les actes sont parus en 2014, qu'émerge une réflexion d'ensemble sur le problème, due à Cécile Treffort, qui définit alors le corpus comme un objet documentaire à portée heuristique, caractérisé « par sa recherche d'exhaustivité ou d'une représentativité raisonnée », « par une approche à la fois sérielle et comparative », « obligeant à définir, à choisir et à éliminer », mais aussi à délimiter, à transformer et finalement organiser « intellectuellement et matériellement, pour faire de la collection un corpus ». S'adressant à de jeunes chercheurs préparant leur thèse, mais responsable aussi de l'important *Corpus des inscriptions de la France Médiévale*, Cécile Treffort distingue deux types de corpus, d'un côté le corpus à visée collective, réalisé par un individu ou par une équipe mais destiné à un usage public, et de l'autre le corpus à visée personnelle, destiné à porter un propos scientifique déterminé (par exemple, pour être l'objet d'une thèse)². On ne saurait donc pas minimiser la part que la pratique de la recherche elle-même et les programmes collectifs représentent dans la nécessité de réfléchir sur la notion de « corpus » : la présente journée d'études a été pensée justement pour commencer à répondre aux interrogations de l'équipe des CBMA alors que ce programme, après s'être consacré essentiellement à réunir des documents diplomatiques, s'apprête à évoluer profondément en intégrant différents types de documents médiévaux en vue de constituer un corpus hétérogène, mais structuré, de textes latins médiévaux relatifs à la Bourgogne des V^e-XV^e siècles³.

Pour fournir un cadre d'ensemble aux débats j'ai présenté les résultats d'une enquête préliminaire sur le mot « corpus » dans les dictionnaires courants et dans les travaux

² C. Treffort, « Le corpus du chercheur, une quête de l'impossible ? Quelques considérations introductives », *Le corpus. Son contour, ses limites et sa cohérence*, *Annales de Janua*, Actes des Journées d'études, n°2, 2014 [En ligne] <http://annaesdejanua.edel.univ-poitiers.fr/index.php?id=725> (Consulté le 28/09/2017).

³ Cf. E. Magnani, « Un corpus structuré et hétérogène de textes latins médiévaux (Bourgogne, V^e-XV^e siècle) », *Bulletin du CERCOR - Centre Européen de recherches sur les congrégations et ordres religieux*, 41 (2017), p. 59-65 [En ligne] <https://halshs.archives-ouvertes.fr/halshs-01529451> (Consulté le 28/09/2017).

des médiévistes. Ne considérant que des exemples français, il faudra compléter cet aperçu par des sondages similaires dans d'autres langues. D'une manière générale il faut tenir compte du fait que *corpus* en latin est un terme polysémique et que cette polysémie se retrouve également dans les dérivés français « cors, corps » et « corpus ». *Corpus* passe en français depuis les syntagmes *Corpus Dei*, *Corpus Domini*, *Corpus Christi* et désigne l'hostie jusqu'au XVIII^e siècle. Au XIX^e siècle, le syntagme « corpus juris » est utilisé pour désigner la « collection du droit romain » (attesté en 1826 et dans la lexicographie en 1845) ou simplement le corpus (1863). L'usage de corpus dans le sens d'un « ensemble de données d'un certain type (notamment en lettres) réunies en vue de leur étude scientifique », attesté en 1809 ne proviendrait pas de la généralisation du sens et de l'analogie avec « corpus juris ». Il s'agirait plutôt d'un transfert linguistique depuis l'allemand *Corpus/Korpus*, attesté en 1787, car en français on utilisait alors le mot « corps ». « Corpus » en français apparaît donc comme un synonyme de l'une des acceptions de « corps », terme qu'il « va concurrencer puis évincer ».

L'acception de « corpus » pour un usage propre au domaine de la linguistique, en tant qu'« ensemble de textes établis selon un critère thématique en vue de leur étude linguistique » est attestée depuis 1855, mais il n'entre dans les dictionnaires français que plus d'un siècle plus tard (1968, pour la première attestation lexicographique). Depuis, les dictionnaires proposent généralement deux acceptions pour « corpus » : la première proviendrait de la langue du droit — « Recueil de pièces, de documents concernant une même discipline » —, la seconde de la linguistique — « Ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique » (*Le Petit Robert*, 1989). L'écart entre les deux acceptions est considérable, car au caractère construit et délimité du « corpus » des linguistes s'oppose le sens beaucoup plus lâche de « corpus » en tant qu'une forme de collection, d'assemblage.

C'est plutôt cette acception distendue qu'on retrouve sous la plume des historiens médiévistes. L'enquête réalisée dans les articles parus dans la *Bibliothèque de l'École des Chartes* (BEC, 1840-2007) et dans les *Actes des congrès de la SHMESP - Société des Historiens Médiévistes de l'Enseignement Supérieur Public* (1970-2007), confirme l'importante augmentation des occurrences à partir des années 1980, et surtout depuis les années 2000 (Fig. 1). Mais ce sont surtout les co-occurents de « corpus » qui portent des enseignements, avec l'expression de sa possession (« mon corpus », « notre corpus », « corpus d'Untel », « son propre corpus », « corpus personnels », « corpus de

l'historien»), de sa taille (de « petit », « limité », à « très gros », « pratiquement exhaustif », « sans cesse accru »), ainsi que d'une série de ses qualités et des opérations et actions qu'il incite à engager. Toutefois jamais une définition de « corpus » n'est proposée dans les 127 articles concernés.

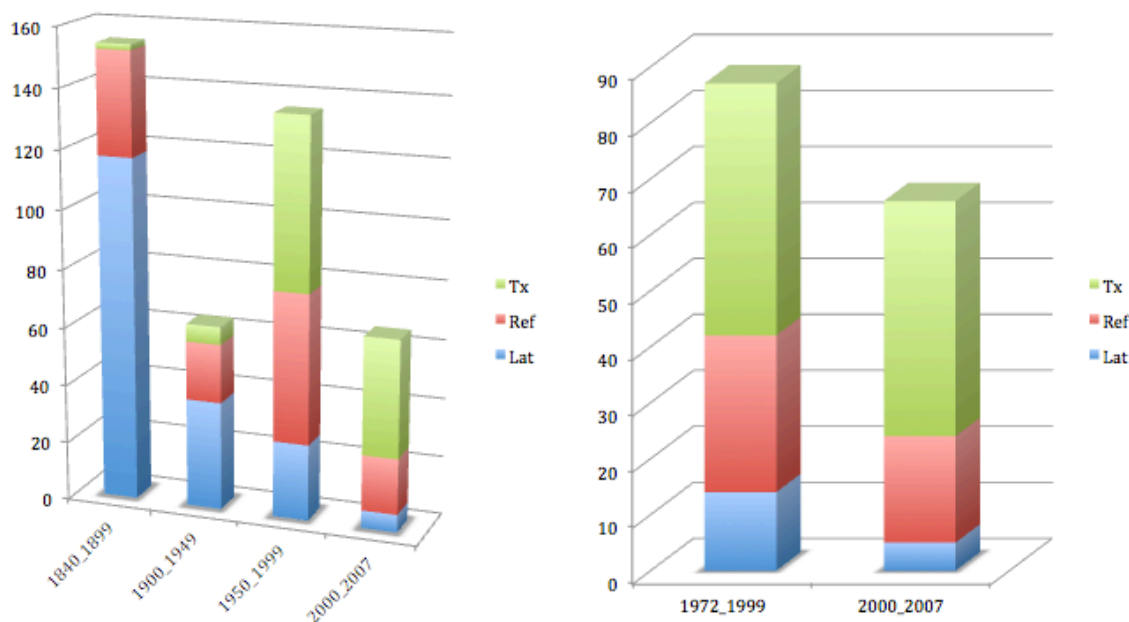


Figure 1: Chronologie des occurrences du mot « corpus » dans la BEC et dans les Actes des congrès de la SHMESp. (Légende - Tx: en français dans le corps du texte; Ref: en latin ou en français dans le titre de références bibliographiques; Lat: en latin dans la citation de syntagmes et d'extraits de documents médiévaux) (© E. Magnani).

La situation est assez différente chez les linguistes, et chez les historiens s'inspirant de leurs méthodes, dont les principaux jalons ont été présentés par Magali Guaresi. L'idée de « corpus » comme un objet fabriqué en suivant une méthode et une théorie préalables afin d'orienter une recherche déterminée date des années 1950 et est rattachée aux travaux de John R. Firth et de la *London School*. Elle est au départ de la « linguistique de corpus » (*corpus linguistic*), qui considère, en rupture avec la linguistique de la Langue, que le sens d'un mot n'est pas un donné a priori mais se situe dans ses usages avérés et en contexte, d'où la nécessité de réunir en corpus ces attestations. Deux approches principales se distinguent actuellement, d'après la formulation d'Elena Tognini-Bonelli⁴. La première, *corpus-based*, utilise le corpus pour vérifier une théorie ou une hypothèse afin de la valider, la réfuter ou l'améliorer. La deuxième, *corpus-driven*, plus inductive, revendique une exploration sans a priori du

⁴ E. Tognini-Bonelli, *Corpus Linguistics at Work*, Amsterdam-Philadelphia, 2001 (Studies in corpus linguistics, 6).

corpus faisant émerger depuis lui-même des hypothèses sur le langage. Cela revient, par conséquent, à considérer que le sens se situe à l'échelle du corpus, pas dans celle du mot, de la phrase ou du texte isolés. Le corpus est ainsi, selon l'expression de Damon Mayaffre, « la forme maximale du contexte »⁵. C'est alors à cette échelle, du corpus comme « matrice du sens », que les intérêts des linguistes et ceux des historiens se trouvent réunis, les opérations de description et d'interprétation d'une unité textuelle en vue de sa saisie matérielle dans le corpus étant indissociables, se façonnant ainsi en véritable objet interdisciplinaire. Construction de l'analyste, le corpus est la configuration matérielle propre à une observation contrôlée, une réification formalisable de l'intertextualité qui serait insaisissable autrement. Cette objectivation est d'autant plus structurante dans le cadre actuel des traitements numériques auxquels les corpus sont susceptibles d'être soumis, conduisant aux nouvelles formes de lectures, dites « hyperlectures », quantitatives, paradigmatiques, réticulaires.

Par l'exemple de sa thèse sur les formes d'accès des femmes à la représentation politique sous la Cinquième République, Magali Guaresi a explicité la méthode et la démarche historique adoptée en faisant valoir le cadre épistémologique de la linguistique de corpus/textuelle et de la textométrie. Le corpus-matrice-contexte constitué par les professions de foi des députées élues entre 1958 et 2007, a ainsi été comparé avec un corpus raisonné de professions de foi d'hommes dans la perspective d'une analyse du genre en politique (*Parler au féminin : les professions de foi des députées sous la Cinquième République (1958-2007)*, Univ. Nice-Sophia Antipolis, 2015).

Mais à l'inverse d'un corpus textuel contemporain auquel on peut toujours agréger des nouvelles attestations, un corpus textuel ancien (médiéval, par exemple) est en principe fini. Dans quelle mesure, dans ces conditions, les corpus réunis par les historiens médiévistes peuvent-ils être considérés comme représentatifs de la production sociale médiévale envisagée (et dont le corpus n'est en théorie qu'un moyen d'appréhension), et quels sont les méthodes heuristiques et les outils contemporains qui peuvent être employés pour le vérifier ? Telle était l'interrogation de départ de Nicolas Perreaux. En articulant la dichotomie *corpus-based/corpus-driven* à « structure historique/structure

⁵ D. Mayaffre, *Vers une herméneutique matérielle numérique. Corpus textuels, Logométrie et Langage politique*, mémoire pour l'Habilitation à diriger des recherches, Université Nice-Sophia Antipolis, 2010, vol. 1, p. 13, 18, 97 [En ligne] <http://tel.archives-ouvertes.fr/tel-00655380> (Consulté le 28/09/2017).

historique », il montre l'existence de deux conceptions opposées. La première considère le corpus comme une masse documentaire non-représentative et incohérente, mais dans laquelle il est possible de puiser des exemples. La deuxième prône la cohérence et la représentativité du corpus, du moment qu'il est analysé dans sa globalité, en vue de faire émerger sa structure.

L'hypothèse ici adoptée que le corpus soit « un ensemble articulé de relations » (*i.e.* une structure) est mise à l'épreuve à partir des recherches réalisées dans les chartes de Bourgogne, la mise en parallèle des originaux et des copies, les variations d'échelles géographiques et la confrontation du corpus textuel diplomatique au corpus archéologique des édifices « romans ». Le tout étant soumis aux différentes méthodes d'exploration numériques : *text mining*, analyses statistiques (AFC) et spatiales (SIG). En dehors de toute problématique pré-formulée, l'observation de la répartition quantitative, chronologique et spatiale des chartes réunies par le programme des CBMA a montré la forte cohérence de l'ensemble. En fait, ce qui pourrait être compris comme des biais du corpus réuni (une région non couverte à une telle époque, la surreprésentation de certaines institutions) s'est avérée au contraire être la mise en évidence de tendances sociohistoriques qu'il convenait d'interpréter en tant que telles.

Ainsi, en faisant varier l'échelle du corpus, de la Bourgogne à l'Europe, donc de 10 000 actes à l'époque de l'enquête à 140 000, mais aussi en le confrontant avec un autre corpus européen, celui des 8 400 édifices « romans », on peut constater la même cohérence sur une échelle plus vaste, postuler la dynamique différenciée du système et modéliser le phénomène de densification sociale. Des enquêtes plus fines utilisant des algorithmes d'analyse lexicale montrent par ailleurs que le corpus est constitué de sous-ensembles nettement inscrits géographiquement, indice supplémentaire de la cohérence générale (cf. la thèse *L'écriture du monde. Dynamique, perception, catégorisation du mundus au moyen âge (VII^e-XIII^e siècles). Recherches à partir de bases de données numérisées*, Univ. Bourgogne, 2014).

Enfin, les mêmes méthodes de fouille de texte s'avèrent utiles pour vérifier l'attribution d'auteur (*authorship attribution*) des textes médiévaux. L'expérience réalisée avec des textes hagiographiques clunisiens (X^e-XI^e siècles) et les vies rédigées par Jonas de Bobbio (*ca.* 600-659) dont ferait partie la vie de Jean de Réôme a ainsi révélé que cette dernière se détache des deux groupes, son attribution devant ainsi être reconsidérée (Fig. 2). Ces résultats conduisent par conséquent à prendre au sérieux le potentiel

heuristique des documents réunis en corpus et les méthodes numériques pour les exploiter, alors que leur définition et leur analyse sont indissociablement articulées.

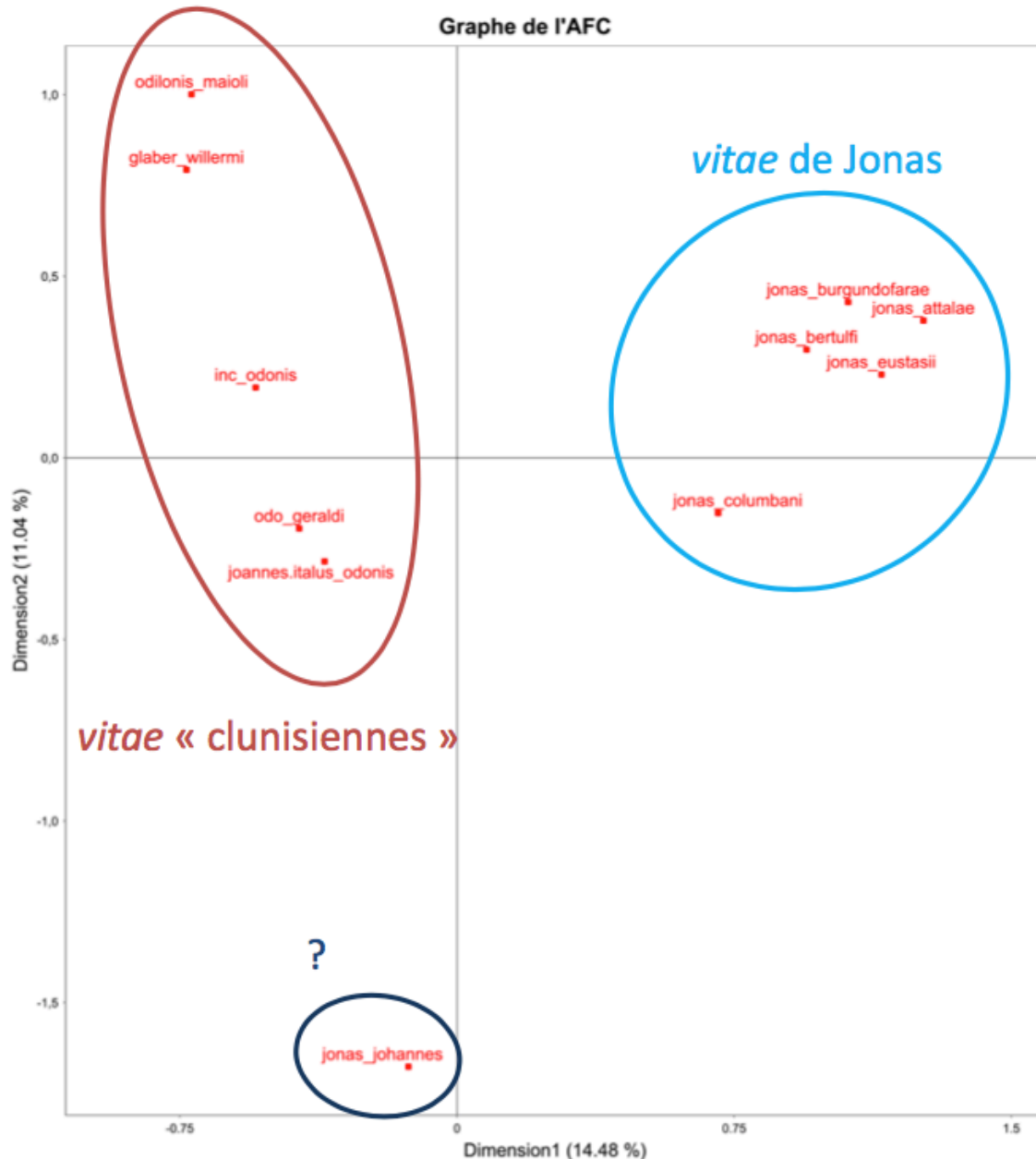


Figure 2: AFC. Espace lexicographique des vies attribuées à Jonas de Bobbio et des *vitae* « clunisiennes » (© N. Perreaux).

Avant que la linguistique ne donne une signification conceptuelle au mot « corpus », le syntagme « corpus juris » pouvait sembler le plus emblématique de l'acception ancienne du vocable dans le sens d'« un ensemble d'écrits ». La présentation de Charles de Miramon visait justement à dégager les usages du vocable et de l'expression depuis leurs premières attestations dans le domaine du droit canonique et du droit romain au Moyen Âge. Or on observe que les premiers cas, isolés, d'utilisation de *corpus iuris* dans le titre

des éditions des Corpus juridiques n'apparaît qu'à la fin du XVI^e siècle, dans le contexte propre à l'humanisme juridique et au marché des éditions qu'il suscite : le *Corpus iuris civilis* de Denys Godefroy paraît à Lyon en 1583 tandis que le *Corpus iuris canonici* est édité en un seul volume à Paris en 1597 (Fig. 3). Mais à l'inverse des éditions du début du même siècle, comme celles de Jean Chappuis (*Decretum aureum domini Gratiani* en 1512, *Institutiones Imperiales* en 1513), on publie sous ce titre de *corpus* les textes de droit romain ou de droit canonique sans la glose médiévale⁶.

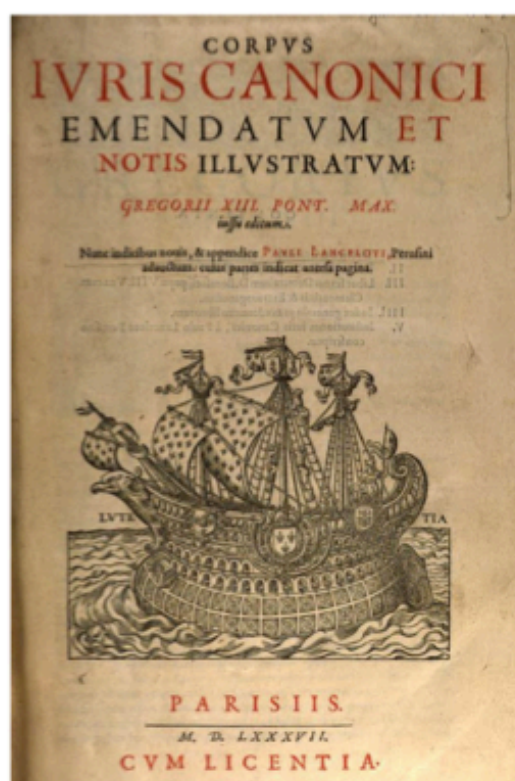


Figure 3: *Corpus iuris civilis*, éd. D. Godefroy (Lyon, 1583). *Corpus iuris canonici* (Paris, 1597).

Il n'en reste pas moins que dans les textes (et pas seulement comme titre), l'usage de *corpus* associé à *iuris* (« corps du droit »), même s'il est rare, remonte à Tite-Live (III.34 - *corpus omnis Romani iuris*). Au VI^e siècle, le Code Justinien le mentionne une fois (5.13.1), alors qu'aux VII^e et VIII^e siècles l'expression vient à désigner certaines collections canoniques (Hispana, Dionysio-Hadriana). À ces époques les syntagmes comme « collection des canons », « livre des canons » ou « codex des canons » sont beaucoup

⁶ Sur ce point, la question du *corpus* vient enrichir celle du *textus* sur laquelle s'était penché le colloque de Göttingen en 2001 (cf. L. Kuchenbuch, U. Kleine (dir.), *Textus im Mittelalter. Komponenten und Situationen des Wortgebrauchs im schriftsemantischen Feld*, Göttingen, 2006).

plus usités. L'expression « corps de canons » (*corpus canonum*) survient peu après l'an mil dans la préface du *Decretum* (1008–1012) de Burchard de Worms, mais dans toutes ces occurrences *corpus* est employé comme synonyme de livre, collection, codex, volume. L'idée d'un ensemble textuel de référence, *ex corpore canonum*, apparaît dans les collections canoniques au début du XII^e siècle. Elle est en correspondance avec le « tournant interprétatif » de l'histoire du droit (Paul Fournier)⁷ où, de la « tradition canonique » on passe à considérer « les canons et ce que l'on peut argumenter à partir de ceux-ci », le canon tendant ainsi à devenir une *sententia* (une règle de droit).

Toutefois, aucun code juridique ne prend alors, on l'a vu, le titre de *corpus*, le sens juridique le plus récurrent du vocable est celui d'*universitas*, tandis qu'au pluriel il prend une acception collective et désigne plutôt les lois, les décrets, les décrétales. Finalement, ce n'est que vers le milieu du XIII^e siècle qu'on vient à concevoir le « corps du droit » comme un ensemble en cours de clôture. Tout ajout d'une nouvelle norme générale par le prince ou par le pape devait désormais considérer le droit existant et être insérée de façon explicite et publique.

Ce premier ensemble de communications qui entendait esquisser un cadre historique et conceptuel à la notion de « corpus », a été complété par des appréhensions diverses du corpus à partir d'expériences de recherche vécues ou en cours.

Sébastien Fray est revenu sur l'enjeu sensible qui constitue la définition d'un corpus en vue de la réalisation d'une thèse, en examinant rétrospectivement sa propre démarche pour sa thèse sur *L'aristocratie laïque au miroir des récits hagiographiques des pays d'Olt et de Dordogne (X^e-XI^e siècles)* (Univ. Paris IV, 2011) et en prenant pour contrepoint d'autres thèses récentes basées sur le même type documentaire. Alors qu'à l'époque de son travail doctoral, S. Fray concevait implicitement le corpus comme « un rassemblement de textes ou une

⁷ P. Fournier, « Un tournant de l'histoire du droit : 1060-1140 », *Nouvelle revue historique de droit français et étranger*, 41 (1917), p. 129-180 [En ligne] <http://gallica.bnf.fr/ark:/12148/bpt6k158070/f131.image> (Consulté le 28/09/2017).

collection de textes regroupés sur la base d'hypothèses de travail en vue de les interroger », il se pose à présent la question de la requalification de ce corpus en « corpus réflexif », où chaque élément participe à « un *réseau sémantique* performant dans un tout (le corpus) cohérent et auto-suffisant » (D. Mayaffre)⁸. Dans le contexte propre à l'état de la documentation du haut Moyen Âge et son incidence sur la taille du corpus, l'opération centrale s'avère être l'étude critique de chaque unité, en prenant en compte l'épaisseur chronologique de la transmission documentaire. L'opération critique régit la décision d'intégration ou non de chaque unité documentaire dans le corpus, mais une large place est également laissée à une approche pragmatique qui conduit à « enrichir » le corpus avec des éléments extérieurs, entre les choix raisonnés du chercheur, les contraintes et les hasards contingents. En l'occurrence, l'analyse interne des textes hagiographiques retenus d'après une aire géographique de production préalablement fixée a conduit à dépasser le cadre initial, à prendre en compte les documents diplomatiques, à réévaluer les liens qu'ils entretenaient avec d'autres textes hagiographiques exclus d'un premier abord, à les réinjecter dans le corpus, tout en considérant les acquis des fouilles archéologiques des sites nommés dans ces mêmes documents (Fig. 4). Cette circularité qui traverse l'ensemble peut être comprise comme une forme de réflexivité inhérente au corpus ainsi constitué.

Dossier hagiographique	Cœur du corpus	Documents ajoutés au corpus
Géraud	<i>Vita Geraldi (Vita Prolixior Prima, BHL 3411)</i> : Aurillac	<i>Vita Prolixior Secunda</i> (BHL vacat) : Cluny
	<i>Vita Brevior</i> (BHL 3412-3414) : Aurillac	Diplôme de Charles le Simple (899)
	<i>Miracula beati Geraldi</i> (BHL vacat) : Aurillac	Testament de 909
		Fausse « charte » de Landeyrat
		Fouilles archéologiques

Figure 4 : Ancien corpus de thèse traité comme corpus réflexif (© S. Fray).

⁸ D. Mayaffre, « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus*, 1 (2002), § 35 [En ligne] <http://corpus.revues.org/11> (Consulté le 28/09/2017).

Ismérie Triquet s'est également penchée sur le corpus réuni à l'occasion de sa thèse sur *Les manuscrits enluminés de la 'Grande Chronique de Normandie' aux XIV^e et XV^e siècles* (Univ. Rennes II, 2014) et aux corpus parallèles qu'il a fallu constituer au fur et à mesure du développement de la recherche – « corpus » étant ici entendu comme l'ensemble des exemplaires d'un même objet. I. Triquet distingue ainsi un corpus de base : les 18 manuscrits enluminés (11 illustrés et 7 présentant des réserves pour l'illustration) sur les 43 manuscrits connus de la *Grande Chronique de Normandie*, comportant un total d'environ une centaine d'enluminures. Pour comprendre cette production il a fallu réunir des corpus complémentaires contextuels (ou « textes annexes ») : celui des textes historiographiques normands et anglo-saxons, ainsi que les généalogies en rouleau. La *Grande Chronique de Normandie* étant le premier et le seul texte historiographique normand ayant été mis en images en utilisant au départ des « scènes typologiques » (mariages, hommages, couronnements, etc.) (Fig. 5), il a fallu la comparer aux chroniques régionales enluminées de la même époque ou encore aux chroniques ayant incorporé la tradition iconographique mise en place par les chroniques normandes elles-mêmes. Du fait de ces extensions successives, la définition de corpus de départ était une notion appelée à être dépassée et à être reconstruite dans une perspective transdisciplinaire.



Figure 5: Grande chronique de Normandie. Rouen, Bibliothèque municipale, ms. 1233 (Y 26), f° 24v (1410) - Mariage de Guillaume Longue Épée. (Source : <http://www.archexpo.fr/fr/contenu/chroniques-de-normandie>).

La communication de Michael I. Allen a mobilisé l'acception de « corpus » en tant qu'un ensemble d'écrits d'un auteur et sa compilation dans un manuscrit. La préparation d'une

nouvelle édition des lettres de Loup de Ferrières a conduit à considérer le manuscrit de Paris (Bibliothèque nationale de France, lat. 2858), réalisé après 862, juste après la mort de Loup, comme le témoin à publier, dans la mesure où il révèle un corpus factice, authentique et organique, sélectionné et réuni par Loup lui-même (Fig. 6). Le *libellum* de 64 feuillets contient 128 lettres (les lettres 1 à 127, plus la lettre 115 bis), regroupées en différents ensembles, ou paquets (*distinctiones*) contenant un choix de lettres à visée mémorielle. Certains cahiers, entre les feuillets 56 et 57, ont toutefois été ôtés dès le IX^e siècle (*deperdita*), et le manuscrit ne contient pas l'ensemble de l'œuvre de Loup, qui a composé également des traités théologiques et des hagiographies (*extravagantes*).

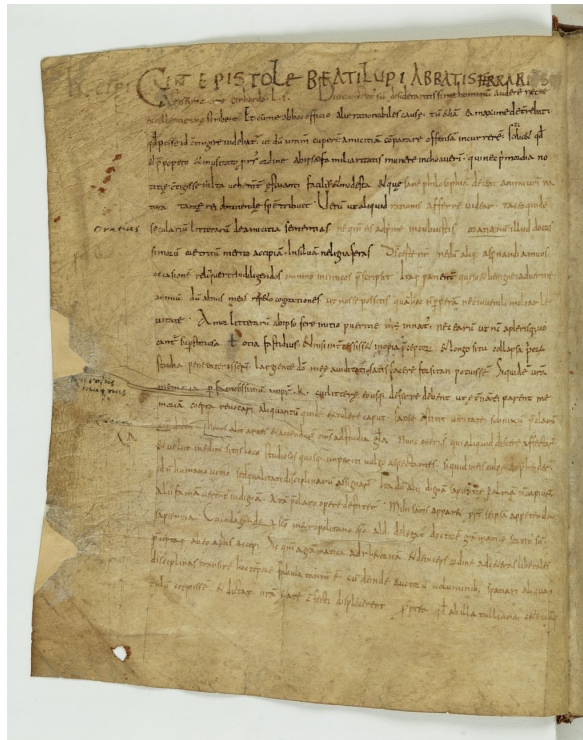


Figure 6: Lettres de Loup de Ferrières. Paris, Bibliothèque nationale de France, lat. 2858, f^o 1v (IX^e siècle). (Source : <http://gallica.bnf.fr/ark:/12148/btv1b10318625w>).

Le tour de table qui a suivi les communications a donné la parole à tous les présents pour réagir sur les échanges tenus au cours de la journée. Il a été rappelé, entre autres, que l'importance acquise par le mot « corpus » serait une spécificité française, absente ailleurs. Le mot corpus serait devenu un « mot de prestige » entré dans le jargon du médiéviste, entre autres, et dont on attend désormais qu'il soit utilisé. D'une manière générale, les propos tenus permettent de relever une tension principale entre les participants : pour les uns il faudrait maintenir le flou autour du mot, considérer le corpus comme une notion ouverte à la fonction que l'on veut bien lui donner ; pour les autres, il faudrait le définir et ne pas ignorer que dans le champ des sciences humaines

et sociales, et de la linguistique en particulier, il existe déjà un effort de conceptualisation important. Ces positions pourraient correspondre à des tendances méthodologiques diverses et renvoyer au moment historiographique actuel, avec le déplacement de la « nouvelle érudition » des années 1990 vers le traitement des grands corpus, chantier ouvert par la révolution technique, les numérisations et les outils informatiques disponibles d'aujourd'hui. Cependant plutôt qu'opposer les deux approches il faudrait œuvrer à leur compatibilité, voire à leur complémentarité.

Lors de la conclusion finale, Joseph Morsel a problématisé les travaux de la journée autour de quatre questions. 1) Faut-il définir, qualifier le corpus ? Si l'on peut reconnaître le caractère heuristique du flou, il faudrait rester attentif à l'ensemble des notions mobilisées car des quasi-synonymes ne sont pas des équivalents. Les tentatives pour préciser et objectiver les concepts utilisés par les historiens visent à saisir les implications que ces termes ont sur notre travail. Des notions comme « sources », « texte », « trace », « information », « donnée »... sont souvent de simples métaphores, qui véhiculent cependant des connotations particulières et demandent alors une réflexivité permanente de la part de l'historien s'il veut rester maître du sens de ses résultats.

2) Qu'est-ce alors qu'un corpus ? Quel concept est-il passé dans le vocabulaire courant des historiens et comment ? S'il y a eu fossilisation en français par la voie du droit et de la théologie par le rôle éventuel de l'hostie, le binôme *corpus-anima*, le rapport entre la matière et l'esprit n'est-il pas encore présent ? Le matériau converti en « corpus » ne fait-il pas disparaître une partie du travail de l'historien, en l'occurrence le choix documentaire ? Sur quoi travaille-t-on, alors : sur le corpus ou sur les éléments qui composent le corpus ? De quel type d'objet s'agit-il par rapport au travail de l'historien ? Le corpus est-il un objet ou un moyen, comme on pourrait l'envisager dans la distinction entre *corpus-based* et *corpus-driven* – le premier correspondant à une manière traditionnelle de constituer un corpus puis de le piller, le deuxième engageant à penser la manière d'exploiter un corpus. En reprenant la métaphore du corps-esprit, on pourrait considérer que le corpus est à la fois matériau et concept – et en tant que tel, par définition, une production historique puisque c'est par sa démarche scientifique de l'historien se saisit du document, le fait exister comme le matériau de son travail.

D'autres distinctions sont toutefois possibles, entre un « corpus indigène », comme un cartulaire ou un *libellus* constitué au Moyen Âge, un « corpus historique », comme la *Patrologie latine*, prise dans sa totalité et érigée en tant que corpus par l'historien, et un « corpus historien » taillé en fonction d'une problématique préalable. La distinction est aussi possible entre « corpus mou » et « corpus dur » (selon qu'il peut ou non admettre des enrichissements ou recompositions en fonction des besoins). Et l'on peut aussi envisager d'intégrer la question des conditions de transmission.

3) À propos de l'étendue du corpus, doit-il être exhaustif ou représentatif ? À quelle échelle se situer ? Un macro-corpus est-il plus significatif ou représentatif qu'un micro-corpus, voire qu'un document unique ? En fait, toutes les échelles sont probablement nécessaires (Bernard Lepetit)⁹, à condition de les faire varier.

4) Où se trouve enfin la cohérence du corpus ? S'il y a cohérence évidente dans le *corpus juris*, celle qui régit les milliers de chartes européennes est un postulat de départ qu'il faut essayer de faire apparaître. La cohérence étant différente de l'homogénéité : un corpus hétérogène dans la typologie des documents réunis n'est pas pour autant incohérent.

Rappel du programme :

- ✚ Introduction. Les médiévistes et le corpus - ELIANA MAGNANI (CNRS, Lamop UMR 8589)
- ✚ Histoire du corpus et corpus en histoire : éléments de présentation théoriques et pratiques - MAGALI GUARESI (Univ. Nice Sophia Antipolis, CMMC EA 1193, BCL UMR 7320)
- ✚ Le corpus : une structure historique ou historique ? NICOLAS PERREAUX (SFB 1095 Ressourcenregime und Schwächediskurse, Frankfurt am Main)
- ✚ Le Corps du droit. La lente sédimentation du corpus en droit canonique et en droit - CHARLES DE MIRAMON (CNRS, CRH UMR 8558)
- ✚ Essai de définition pragmatiste et ouverte du corpus en histoire du haut Moyen Âge - SEBASTIEN FRAY (Univ. Saint-Etienne)

⁹ Entre autres, voir B. Lepetit, « De l'échelle en histoire », in J. Revel (dir.), *Jeux d'échelle. La micro-analyse à l'expérience*, Paris, 1995, p. 71-95.

- ✚ La problématique du corpus dans l'étude de l'historiographie Anglo-normande - ISMERIE TRIQUET (Univ. Rennes II)
- ✚ Corpus, distinctions, extravagantes, deperdita : les lettres de Loup de Ferrières, les défis et les nouveaux acquis - MICHAEL I. ALLEN (University of Chicago)
- ✚ Tour de table et conclusions - JOSEPH MORSEL (Univ. Paris 1, Lamop UMR 8589)