



HAL
open science

Les documents pontificaux dans la masse : Machine learning, stylométrie et sémantique historique

Nicolas Perreaux

► **To cite this version:**

Nicolas Perreaux. Les documents pontificaux dans la masse : Machine learning, stylométrie et sémantique historique. Laurent Morelle; Rolf Große; Olivier Guyotjeannin. Les actes pontificaux : un trésor à exploiter, Göttingen University Press, pp.39-90, 2024, Abhandlungen der Niedersächsischen Akademie der Wissenschaften zu Göttingen. Neue Folge (55), 978-3-86395-611-0. 10.17875/gup2024-2515 . hal-04839515

HAL Id: hal-04839515

<https://cnrs.hal.science/hal-04839515v1>

Submitted on 16 Dec 2024

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.

Ceci est une **version d'auteur** de l'article sous le titre : **Nicolas Perreaux**, « **Les documents pontificaux dans la masse : *Machine learning*, stylométrie et sémantique historique** », dans GROSSE R., GUYOTJEANNIN O. et MORELLE L. (dir.), *Les actes pontificaux : un trésor à exploiter*, Göttingen, Niedersächsischen Akademie der Wissenschaften zu Göttingen, 2024, p. 39-90.

La pagination de l'article imprimé apparaît ci-dessous en bleu : {xx}

Les actes pontificaux

Un trésor à exploiter

Études réunies par Rolf Große,
Olivier Guyotjeannin et Laurent Morelle



Abhandlungen der
Niedersächsischen Akademie der Wissenschaften
zu Göttingen
Neue Folge

Les documents pontificaux dans la masse (VII^e-XIII^e siècle)

Machine learning, stylométrie et sémantique historique

« De toutes les sources diplomatiques, celles qui proviennent de la cour romaine sont certainement de beaucoup les plus nombreuses. Le gouvernement spirituel de la chrétienté, si étroitement uni, au Moyen Âge surtout, à toutes les affaires politiques et à tous les intérêts privés, a donné lieu depuis les premiers siècles de l'Eglise à une quantité immense de documents qui, du siège de la papauté, se sont dispersés dans tout l'univers et dont un très grand nombre se sont conservés. »

Arthur Giry, *Manuel de diplomatique*, 1894.

{39} C'est une évidence pour tous les historiens : il n'existe que peu d'ensembles textuels médiévaux dont la densité soit comparable à celle de la production pontificale¹. L'estimation généralement admise, concernant le nombre de documents expédiés par cette chancellerie avant 1200, varie le plus souvent entre 25 000 et 30 000². En dépit de recherches assidues depuis plus de deux siècles³, ce nombre fait qu'il est bien difficile {40} de dégager une vision cohérente du corpus des papes, de sa richesse lexico-sémantique, de sa distribution chrono-géographique, de sa qualité de conservation et même de son contenu. Les grandes entreprises d'éditions de documents pontificaux achoppent d'ailleurs depuis leurs origines à ces difficultés : comment identifier et réunir cet ensemble monumental, afin d'en permettre l'analyse scientifique ? Ici comme ailleurs, la numérisation progressive des textes légués par l'Europe médiévale n'a pas fondamentalement modifié la

¹ À différents stades de la recherche et/ou de l'écriture, j'ai reçu les riches conseils de Paul Bertrand, François Bougard, Geneviève Bühner-Thierry, Fabrice Delivré, Laurent Morelle, Joseph Morsel, Alain Rauwel et Julien Théry, que je remercie tous chaleureusement. Les analyses qui suivent ne seraient en outre pas les mêmes sans l'aide amicale de Sébastien de Valeriola, qui a partagé avec moi sa grande connaissance de R, tant pour les méthodes de classification que pour les représentations graphiques.

² Frenz, Thomas: *Papsturkunden des Mittelalters und der Neuzeit*, 2^e édition actualisée. Stuttgart 2000 (*Historische Grundwissenschaften in Einzeldarstellungen* 2), p. 15-43 ; Guyotjeannin, Olivier, Jacques Picke et Benoît-Michel Tock (Dir.): *Diplomatique médiévale*, 3^e édition actualisée. Turnhout 2006, p. 113.

³ Quelques jalons fondamentaux : Poole, Reginald L.: *Lectures on the History of the Papal Chancery*. Cambridge 1915, qui constitue une bonne porte d'entrée à cette documentation, aux côtés des manuels de diplomatique ; Rabikauskas, Paul: *Diplomatica pontificia. Praelectionum lineamenta*, 5^e édition augmentée. Rome 1994 ; Frenz: *Papsturkunden* (voir n. 2) ; *Papsturkunden des frühen und hohen Mittelalters. Äußere Merkmale, Konservierung, Restaurierung*. Éd par Irmgard Fees, Andreas Hedwig et Francesco Roberg. Leipzig 2011.

problématique. Même si elle pourrait potentiellement le faire. À l'heure du développement des méthodes numériques et des grands corpus historiques digitalisés, ce *trésor documentaire* continue en effet de poser les mêmes défis, souvent bien difficiles à relever⁴. Les grandes caractéristiques des documents émanant des papes et/ou de la chancellerie pontificale sont particulièrement nettes : une immense richesse qualitative et quantitative, une complexité dans la tradition, enfin, un traitement historiographique foisonnant, mais aussi inégal. Corpus textuel, pape(s) et papauté forment ainsi un nœud historique et historiographique d'une complexité redoutable.

Parallèlement, la définition typologique de ce vaste ensemble se présente sous la forme d'une aporie : les *actes pontificaux* sont constitués de *lettres* (*cum serico* ou *cum filo canapis*)⁵, de *privilèges*, de *breufs*, ou encore de *bulles* – ces dernières constituant l'une des typologies les plus englobante et donc incertaine de la diplomatie pontificale⁶. Faut-il y voir des ensembles distincts, ou bien une *masse*, voire un *trésor* – pour {41} reprendre la terminologie du présent volume –, dont le sens serait précisément donné (en partie) par sa cohérence ? Les hésitations des éditeurs sont ici assez suggestives : tel document est tantôt classé comme « privilège », puis comme « lettre », enfin comme « bulle », en fonction de ses incarnations éditoriales successives. Si bien qu'une enquête systématique sur la documentation pontificale doit choisir entre différents présupposés, qui ne sont pas sans impact sur l'analyse, quitte à les rectifier ensuite. Deux options principales s'offrent en effet à nous : analyser le corpus comme un tout, et risquer des hypothétiques effets typo-chronologiques ;

⁴ Un point déjà signalé par Rolf Große, dans son article *Gallia Pontificia online* : « Eine digitale Plattform für die Edition der Papsturkunden? ». Dans : *Francia* 40 (2013), p. 265-273, ici p. 266. Voir de même Herbers, Klaus et Viktoria Trenkle (dir.) : *Papstgeschichte im digitalen Zeitalter. Neue Zugangsweisen zu einer Kulturgeschichte Europas*. Köln 2018 (Beihefte zum Archiv für Kulturgeschichte 85).

⁵ Sur les typologies de lettres, voir Birnstiel, Andrea et Diana Schweitzer : *Nicht nur Seide oder Hanf! Die Entwicklung der äußeren Merkmale der Gattung Litterae im 12. Jahrhundert*. Dans Irmgard : *Papsturkunden des frühen und hohen Mittelalters* (voir n. 3), p. 305-334 ; Falkenstein, Ludwig : *Beispiele für Mischformen päpstlicher litterae in der Kanzlei Alexanders III. Mit einer Liste bislang datierter Briefe und Mandate*. Dans : *Francia* 41 (2014), p. 335-380. Sur les lettres pontificales du haut Moyen Âge, en général, voir Jasper, Detlev et Horst Fuhrmann : *Papal Letters in the Early Middle Ages*. Washington 2001. Toujours sur les frontières complexes entre lettres et privilèges, au-delà même des documents pontificaux, voir Guyotjeannin, Olivier : *Lettre ou titre ? Le modèle épistolaire dans les chancelleries médiévales*. Dans : Lefèvre, Sylvie (Éd.) : *La lettre dans la littérature romane du Moyen Âge*. Orléans 2008, p. 19-26 ; Sirantoine, Hélène (Dir.) : *La lettre diplomatique. Écriture épistolaire et actes de la pratique dans l'Occident latin médiéval*. Madrid 2018. Une des manifestations récentes de cette perméabilité typologique est d'ailleurs l'inclusion de nombreuses lettres au sein de grands corpus d'actes diplomatiques numérisés, comme *Chartae Galliae* ou *Diplomata Belgica*. Ainsi, dans son article cité dans cette note, Olivier Guyotjeannin indique : « La lettre est partout et nulle part dans les chartiers médiévaux » (p. 21).

⁶ Étymologiquement, elle devrait bien entendu ne désigner que les documents « bullés ». En pratique, elle concerne souvent l'ensemble des documents émanant des papes et de la chancellerie pontificale jusqu'à une date avancée. Concernant les hésitations typologiques, voir les remarques de Jamme, Armand : *Écrire pour le pape du XI^e au XIV^e siècle. Formes et problèmes*. Dans : *Mélanges de l'École française de Rome. Moyen Âge 128-1* (2016), qui évoque la « [d]iversité morphologique et diplomatique, [la] démultiplication des procédures de production [qui] indiquent des changements de politique documentaire plus fréquents qu'on ne l'imagine » et une « absence de linéarité » du corpus pontifical. Thomas Frenz distingue quant à lui 12 catégories textuelles différentes pour les documents pontificaux de la période allant de Léon IX à Pie IX, dans Frenz : *Papsturkunden* (voir n. 2), p. 19-41. Voir aussi Rabikauskas, Paul : *De significatione verborum « bulla », « breve »*. Dans : *Periodica de re morali, canonica, liturgica*, 55 (1966), p. 85-92, qui écrit : « *Documenta autem ita bullata privilegia, litterae apostolicae, litterae consistoriales, litterae executoriae, constitutiones, decreta, etc. – secundum eorum momentum et qualitatem – appellantur.* ». Une définition on ne peut plus large. Sur les *bullaires*, voir nos remarques note 16.

ou bien découper a priori dans la documentation, en fonction de typologies préexistantes, mais floues, et risquer des biais historiographiques – ces biais pouvant d’ailleurs varier d’une période à l’autre. Puisque les deux voies constituent un risque, nous avons choisi de réunir le plus largement possible la documentation, tout en gardant à l’esprit que, s’il nous faut distinguer dans la masse, deux ensembles documentaires dominant le corpus : pour le dire bien trop rapidement, il s’agit des « lettres » et des « privilèges »⁷. Ce choix, conscient, nous permettra par la suite de discuter des hypothèses établies à partir de la totalité du corpus, tout en contrôlant, lorsque la chose est possible et nécessaire, ce qui se passe lorsque l’on compare l’une et l’autre des hypothétiques typologies. Cette décision a en outre l’avantage de ne pas imposer a priori une frontière imperméable entre les types d’écrits produits à la chancellerie pontificale – frontière d’autant plus illusoire que les deux mondes, lettres et privilèges, se sont souvent influencés l’un l’autre⁸.

{42} Cette contribution propose ainsi d’appliquer une série de méthodes numériques aux documents des pontifes, avant de donner différentes pistes pour l’exploration systématique du corpus. Elle s’interroge sur la signification à donner à cette *masse* documentaire au début du XXI^e siècle – tout en sachant que le terme de *masse* peut ici faire référence aux documents pontificaux eux-mêmes, à la place de ceux-ci dans l’ensemble des textes diplomatiques, ou encore à leur situation au sein de l’écrit médiéval au sens large. Cette approche rejoint une préoccupation plus profonde : celle de la construction de la *papauté* en tant qu’institution, mais aussi comme concept historien. Du pape à la papauté, les étapes sont en effet nombreuses, complexes, et passent précisément par la création et l’analyse de ce corpus pontifical – dont l’existence justifie d’une certaine façon l’institution elle-même⁹. Faire comme si ce corpus était le produit d’une papauté intangible et immuable constitue un

⁷ Les lettres pontificales se distinguent en effet non pas nécessairement par leur formalisme (voir les références données note 5), mais potentiellement par le contenu. Très variées, elles peuvent être juridiques, dogmatiques, pastorales, canoniques, organisationnelles, etc. Concernant les privilèges, le *Vocabulaire international de la diplomatie* (p. 97) donne la définition suivante : « Un privilège (lat.: *privilegium*) est un acte solennel, émané généralement d’une autorité ecclésiastique (privilège pontifical, épiscopal, synodal) et constituant ou confirmant une situation juridique spéciale (une *privata lex*) qui le fait échapper au droit commun. » Dans le cas des pontifes, il est vrai que la notion de privilège répond en diplomatie à des critères précis (présence du *Bene valete*, de la rota à partir de Léon IX, liste de souscriptions, etc.) et que l’on distingue parfois entre privilèges solennels et privilèges mineurs (à partir du XII^e siècle). Ici, le terme est entendu dans le sens le plus large : un privilège diplomatique au sens large. Le séminaire 2020-2021 de Laurent Morelle, auquel nous avons pu assister régulièrement cette année, était précisément consacrée à la question des documents pontificaux et à leurs typologies.

⁸ Dans son important article, Olivier Guyotjeannin parle d’une chancellerie pontificale « qui, au plan formel, n’a jamais produit *que* [nous soulignons] des lettres, et par millions, jusqu’au privilège le plus solennel [...] », dans Guyotjeannin: La lettre diplomatique (voir note 5), p. 23. Voir aussi plus loin la remarque de Laurent Morelle, en lien avec la note 89.

⁹ Voir en la matière l’excellent article de D’Alberto, Claudia, Chrystal Bernat, Frédéric Gabriel, Dominique Iogna-Prat, Dominic Moreau, Dan Muresan, Bénédicte Sère et Jean-Marc Ticchi: Papauté. Dans: Iogna-Prat, Dominique, Frédéric Gabriel et Alain Rauwel (Dir.). Dictionnaire critique de l’Église. Paris (à paraître) (nous remercions vivement les auteurs d’avoir accepté de nous confier une copie de ce travail). Plus largement, notre proposition s’inspire donc de l’historiographie de la scripturalité médiévale, en particulier des travaux portant sur une typologie ou un mode de conservations particuliers : Morsel, Joseph: En guise d’introduction: les chartriers entre « retour aux sources » et déconstruction des objets historiens. Dans: Les chartriers seigneuriaux: défendre ses droits, construire sa mémoire (XIII^e-XXI^e siècle). Paris 2010, p. 9-34 ; Id: La noblesse contre la ville ? Comment faire l’histoire des rapports entre nobles et

anachronisme qui masque l'élément central et la dynamique institutionnelle. Il est possible de faire l'hypothèse inverse, en suggérant que cet ensemble textuel est précisément la porte d'entrée permettant d'étudier le phénomène d'institutionnalisation des papes¹⁰.

{43} Une enquête dans différentes bases de données indique d'ailleurs que le terme « papauté » est assez peu fréquent avant 1800, mais qu'il se développe radicalement entre cette date et 1870 – autrement dit au moment même où se mêlent les discussions sur les états pontificaux et celles sur les nationalismes européens. D'ailleurs, c'est aussi le moment de la cristallisation du dogme de l'infailibilité pontificale. Pour s'en convaincre, il suffit de regarder la très complète base *Regesta Imperii*, qui recense les titres d'ouvrages et d'articles en médiévisique, publiés depuis le XVIII^e siècle au moins. Dans celle-ci, le terme « papauté » n'apparaît dans aucun titre avant 1845¹¹. Parallèlement (nous y reviendrons), les textes médiévaux n'évoquent que très rarement cette « papauté », au moins avant le XIII^e siècle : *papatus* est excessivement rare, de même que les expressions *sedes sancti Petri* et *sedes apostolica*, qui sont peu ou moyennement fréquentes¹². {44} Faire de la papauté une catégorie

citadins (en Franconie, vers 1500) ? . Mémoire d'habilitation à diriger des recherches inédit, Paris 2009. En particulier, ce qui nous intéresse en abordant le corpus pontifical de cette fois, c'est bien le passage des « archives en sources », pour reprendre l'expression de Joseph Morsel. Ici, on fait l'hypothèse que ce passage est largement invisibilité par le rapport à aux objets historiques « pape(s) » / « papauté », dont on présuppose très souvent l'immutabilité.

¹⁰ Voir les remarques lexicographiques qui suivent. L'historiographie rappelle sans cesse l'immutabilité (relative) de la fonction pontificale, sans toujours voir que cette immutabilité est elle-même un discours, une construction idéologique. Dire qu'elle se fonde sur Pierre, ce n'est pas nécessairement attester de son acception immédiate et totale dans la pensée médiévale. Sur l'historicité et la réception du discours pétrinien dans l'idéologie des pontifes, pour les premiers siècles chrétiens, voir Demacopoulos, George E.: *The Invention of Peter. Apostolic Discourse and Papal Authority in Late Antiquity*. Philadelphia 2013. Il exprime parfaitement les enjeux du problème ainsi posé : « *The problem, of course, is that there was no unbroken ascent of Roman episcopal authority – the expansion of Rome's hegemony in the Western Church was intermittent and often contested. What is more, the bishops of the late ancient period, whatever their rhetorical claims, simply could not have imagined and therefore did not call for the economic and political force that the papacy would enjoy in the later Middle Ages. Even employing the terms "pope" and "papacy" to describe the earliest bishops of Rome and their administrations runs the risk of suggesting a level of institutional development and international recognition that developed only much later [...]* », p. 8. Pour des périodes plus tardives, on retrouvera différents éléments de cette dynamique historique dans Delivré, Fabrice: *Succession apostolique: autorité des évêques et pouvoir des clés dans l'Occident médiéval (fin XI^e-milieu XV^e siècle)*. Dans: Genet, Jean-Philippe (Dir.): *La légitimité implicite*. Paris 2015, p. 121-144. Dans son étude sur les *rotae*, Edouard Bouyé s'interroge à juste titre sur cet élément comme rhétorique sur la prétendue stabilité pontificale : Bouyé, Edouard: *Les papes, leurs roues et leurs devises (XI^e-XXI^e siècles)*. Dans: Meissonnier, Jacques (Dir.): *De la Bourgogne à l'Orient. Mélanges offerts à Monsieur de Doyen Jean Richard*. Dijon 2020, p. 669-694.

¹¹ André, Jean-François: *Histoire politique de la monarchie pontificale au XIV^e siècle. Ou la papauté à Avignon*. Paris 1845. Toujours dans la même base, le terme anglais *papacy* apparaît dans un titre pour la première fois en 1866 : Guettée, Wladimir: *The Papacy. Its Historic Origin and Primitive Relations with the Eastern Churches*. New York 1866. De la même façon, en Allemagne, le terme *Papstum* fait son apparition dans les titres recensés en 1854 : Jäger, Albert: *Über Kaiser Maximilians I. Verhältnis zum Papsttum*. Dans: *Sitzungsberichte. Akademie der Wissenschaften in Wien, Philosophisch-Historische Klasse 12 (1854)*, p. 195-323, 409-441. De la même façon en italien, en 1847 : Albéri, Eugenio: *Del papato e dell'Italia*. Firenze 1847. Ce synchronisme parfait ne saurait être le fruit du hasard, et montre les liens étroits entre les débats sur le territoire pontifical, l'émergence des états-nations européens et la perception historiographique. Au sein du SUDOC, on note cependant que des titres d'ouvrages emploient le terme « papauté » dès le XVI^e siècle. Ils ne relèvent cependant pas de l'histoire, et les ouvrages de ce type se développent principalement au XIX^e siècle, dès les années 1830. En Allemagne, on pense bien entendu aux travaux de von Ranke.

¹² Sur les titres attribués aux pontifes, voir Congar, Yves: *Le pape, patriarche d'Occident. Approches d'une réalité trop négligée*. Dans: *Istina 28 (1983)*, p. 374-390 (repris dans *Église et papauté*. Paris 1994) ; Id.: *L'ecclésiologie du haut moyen âge. De saint Grégoire le Grand à la désunion entre Byzance et Rome*. Paris 1968. Congar indique que *papatus* (qui désigne d'ailleurs beaucoup plus le « pontificat » que la « papauté ») est un terme rare au IX^e siècle, mais qu'il se développe fortement au XI^e siècle. Or, *papatus* ne se rencontre que 43 fois en tout et pour tout dans les CEMA (sur ce

immuable gomme ainsi en quelque sorte *les papes* (au pluriel), et leurs actions extrêmement variables, y compris au plan documentaire.

Dans ce cadre, trois axes seront privilégiés : dans un premier temps, on s'attachera à définir une méthode pour identifier le corpus disponible des documents pontificaux au format numérique, grâce en particulier aux méthodes d'intelligence artificielle. Cette étape paraît essentielle : si des progrès conséquents ont été réalisés ces dernières années autour de cette documentation, plusieurs entreprises éditoriales, qui sont des marqueurs historiographiques, doivent aujourd'hui faire face à un avenir incertain, entre sous-financement des institutions et manque de personnel¹³. Il s'agit donc de partir de ce qui est utilisable dès aujourd'hui, si possible en trouvant des bases communes à l'ensemble – ce qui soulève différentes questions méthodologiques. L'article sera ensuite l'occasion de se pencher sur la distribution chronologique du corpus, mais aussi la richesse de son lexique, ainsi que sur l'évolution de cette richesse entre le VII^e et le XIII^e siècle. Enfin, on tentera une analyse globale du vocabulaire pontifical, avec pour objectif la détermination des ruptures et des tendances au sein du corpus, non pas à partir de caractères isolés ou « diplomatiques » au sens strict¹⁴, mais sur la totalité des mots que les documents contiennent. Cette partie donnera lieu à l'évocation de différentes hypothèses concernant l'évolution sémantique des documents pontificaux.

I. Les documents pontificaux : un « trésor » à identifier

I.1. Réunir les documents : les bullaires et la Patrologie Latine

La première étape avant d'exploiter le « trésor » des documents pontificaux consiste bien entendu à l'identifier. Il ne s'agit pas d'une mince affaire, aussi bien avant qu'après le numérique¹⁵.

corpus, voir les développements qui suivent). L'expression *sedes sancti Petri* est elle aussi très rare, avec quelques occurrences seulement dans les corpus consultés. Quant au syntagme *sedes apostolica*, on rencontre certes 388 mentions dans les CEMA, mais ce score est à comparer avec les 70 millions de mots du corpus – et par ailleurs les mentions proviennent massivement des XII^e et surtout XIII^e-XIV^e siècles. On retrouve certes l'expression plus fréquemment dans les lettres pontificales des XIV^e-XV^e siècles, mais la proportion reste relative (un millier d'occurrences selon Julien Théry, que je remercie vivement). Il serait cependant intéressant de voir si cette proportion n'augmente pas avec le temps, ce qui est l'hypothèse la plus probable. Pour comparaison, le lemme *papa* possède plus de 16 400 occurrences dans le même corpus européen. Concernant les termes relatifs à la « papauté », l'entrée du *Dictionnaire critique de l'Église* (voir n. 9) mentionne « [les] conciles de la fin du Moyen Âge ; [les] œuvres de Jean Huss, de Luther, de Calvin ou des polémistes catholiques qui leur répondent comme Laurent de Brindes ». Autrement dit, le concept paraît se développer tardivement, dans un contexte essentiellement polémique – ce que l'on retrouve ensuite à l'époque contemporaine.

¹³ Le colloque de 2019 a été l'occasion de dresser un bilan assez sombre – mais néanmoins énergique – des perspectives liées aux entreprises éditoriales des documents pontificaux. Au-delà des difficultés liées au moyens, on pourrait aussi s'interroger sur les causes de la désaffection pour ce corpus, qui est probablement indissociable du rapport que nos sociétés entretiennent avec l'Église en tant qu'institution, gouvernée par un pontife.

¹⁴ Voir par exemple Kortüm, Hans-Henning: *Zur päpstlichen Urkundensprache im frühen Mittelalter : die päpstliche Privilegien*, 896-1046. Sigmaringen 1995.

¹⁵ L'historiographie des grandes entreprises d'éditions et de registes de documents pontificaux est relativement dense. Outre les ouvrages classiques déjà mentionnés, qui reviennent généralement sur la question, et les articles dans le présent volume, voir en particulier Santifaller, Leo: *Neuere Editionen mittelalterlicher Königs- und Papsturkunden*. Eine

Ces dernières années, les propositions pour constituer un corpus di-[{45}](#)-gital de documents des papes se sont concentrées sur les éditions spécialisées, que l'on qualifiera pour simplifier de *bullaires*¹⁶. C'est le cas par exemple de bases textuelles telles que les *Regesta Pontificum Romanorum online*¹⁷, du corpus *Ut per litteras Apostolicas*¹⁸, mais aussi plus récemment d'*Aposcripta*¹⁹. On pourrait dans ces cas parler de *bullaires numériques*, puisque leur logique est de proposer un corpus constitué exclusivement de documents pontificaux. Cette approche puise dans la tradition, riche et ancienne, d'éditions de textes et de regestes spécifiquement consacrés aux papes, qui remonte essentiellement au XVII^e siècle²⁰.

Parallèlement, d'autres corpus diplomatiques 2.0 ont aussi intégré certaines éditions de documents pontificaux. C'est le cas par exemple du *Bullarium sacri ordinis Cluniacensis*, que l'on retrouve dans le corpus des chartes bourguignonnes (*Corpus {46} Burgundiae Medii Aevi*, désormais CBMA)²¹, des trois volumes du *Butllari de Catalunya*, disponible sur le site de la Fundació

Übersicht. Wien 1958 ; Hiestand, Rudolf: Das Göttinger Papsturkundenwerk. Dans *Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland* 1988 (1989), p. 13-17 ; Poncet, Olivier: Les entreprises éditoriales liées aux Archives du Saint-Siège : histoire et bibliographie (1880-2000). Rome 2003 ; Lohrmann, Dietrich: Vingt-cinq ans de Gallia Pontificia. Note sur l'avancement des travaux (mai 2007). Dans: *Revue historique de l'Église de France* 94 (2008), p. 117-125 ; Guyotjeannin: Diplomatique médiévale (voir n. 2), p. 333-338 ; Große, Rolf: De l'utilité des regestes. Dans: Canteaut, Olivier et Rolf Große (Dir.): Pourquoi éditer des textes médiévaux au XXI^e siècle ?. Paris 2014 (en ligne sur Perspectivia.net).

¹⁶ Quelques éléments de définition dans Thurston, Herbert: Bullarium. Dans: *The Catholic encyclopedia*, volume 3. New York 1908, p. 48-50 ; Ortolan, Théophile: Bullaire. Dans: *Dictionnaire de théologie catholique*, tome 2/1. Paris 1923, col. 1243-1255 ; Bertrand, Paul: Les écritures ordinaires. *Sociologie d'un temps de révolution documentaire (1250-1350)*. Paris 2015, p. 30-32. Si des collections antérieures de bulles existent (par exemple *Bullae diversorum pontificum incipientes a Ioanne XXII. usque ad Sanctiss. D.N.D. Paulum Papam III [...]. Rome 1542 ; Bullae diversorum Romanorum Pontificum incipientes a Bonifacio VIII usque ad Paulum IV. Rome 1559*), c'est Laerzio Cherubini qui emploie le terme de bullarium dans un titre d'ouvrage pour la première fois, avec son édition du *Bullarium, sive collectio diversarum constitutionum multorum pontificum a Gregorio VII usque ad Sixtum V. Rome 1586*. Il semble en outre que les manuscrits-bullaires se développent essentiellement à l'époque moderne : dans la base CartulR, que nous avons analysée, 95% des *codices* classés comme tels datent en fait des XIV^e-XVIII^e siècles, dont 75% pour les XV^e-XVII^e siècles.

¹⁷ Dirigée par Klaus Herbers, l'équipe propose à ce jour le téléchargement des *Papsturkunden in Portugal*, en mode texte, ainsi que divers ensembles issus des *Papsturkunden in Frankreich*, cette fois en mode image. Elle prépare la publication de nombreux regestes, qui viendront compléter ceux déjà disponibles sur le très riche site *Regesta Imperii*. Voir une présentation dans Herbers, Klaus et Thorsten Schlauwitz: Annotationen ohne Ende? Auszeichnungsprozesse am Beispiel der *Regesta Pontificum Romanorum online*. Dans: *Magazin für digitale Editionswissenschaften* 1 (2015), p. 35-42. Les *Regesta pontificum Romanorum* de Potthast, bien qu'extrêmement importants pour l'étude (historique et historiographique) des documents pontificaux, ne sont évidemment pas utilisables dans notre perspective – qui retient uniquement les documents en plein texte.

¹⁸ Disponible sur le site de l'éditeur Brepols, le site contient les 32 volumes des *Registres et lettres des Papes du XIII^e siècle* (Rome 1883-), ainsi que les 48 tomes des *Registres et lettres des Papes du XIV^e siècle* (Rome 1899-).

¹⁹ Dirigée par Julien Théry et hébergé sur la plateforme Telma (IRHT), la base contenait 3 048 documents propres (hors des corpus déjà diffusés par ailleurs) au départ de notre enquête, c'est-à-dire en juillet 2019. En février 2021, alors que nous terminons cet article, elle en contient plus de 12 000. Le corpus nous semble extrêmement important, d'une part parce qu'il est ouvert, d'autre part car il rassemble le plus largement possible les différentes typologies d'actes.

²⁰ Nous renvoyons aux éditions du XVI^e siècle, mentionnées à la note 16. Cependant, tous comme les cartulaires-bullaires tardo-médiévaux, liés à un ordre ou à une institution particulière, ils nous paraissent correspondre à une logique apologetique, spirituelle et/ou mémorielle, propre à cette période. Dans tous les cas, une logique différente de celle qui prévaut pour les collections historiques qui émergent au XVII^e-XVIII^e siècles.

²¹ Le corpus est dirigé par Eliana Magnani. On le trouvera à l'adresse suivante : <http://www.cbma-project.eu/>.

Noguera²², ou encore d'une partie de la série des *Papsturkunden in Frankreich*, dispersée sur différents sites. Depuis 2009, nous nous sommes attaché à réunir l'ensemble des textes diplomatiques numérisés dans un corpus unique intitulé *Cartae Europae Medii Aevi* (désormais CEMA)²³. Cet ensemble contient aujourd'hui environ 240 000 documents, dont l'ensemble des bullaires numérisés en accès ouvert : ils ont tous été retenus pour la présente analyse.

Un second gisement à considérer pour les privilèges et lettres pontificales est bien entendu la *Patrologie latine* (désormais *PL*), qui couvre de façon inégale la période allant du III^e au début du XIII^e siècle. En divisant pièce à pièce les 85 volumes (sur 217 au total) contenant les documents pontificaux²⁴, nous constatons que celle-ci contient plus de 12 000 privilèges ou lettres des pontifes – ce qui est évidemment considérable. La plupart de ces documents imprimés par Migne proviennent d'éditions reconnues, bien qu'elles ne correspondent plus aux normes éditoriales actuelles, réalisées par Mabillon, Baluze, Pertz, Jaffé, Löwenfeld, Dronke, ou encore Muratori²⁵. En définitive, seulement 17 papes possèdent plus de cent documents édités dans la *PL*, toutes typologies confondues : il s'agit par ordre décroissant, d'Innocent III²⁶, Alexandre III, Grégoire I, Eugène III, Innocent II, Pascal II, Jean VIII, Lucius III, Urbain II, Calixte II, Adrien IV, Nicolas I, Clément III, Alexandre II, Urbain III, Honorius II et enfin Léon IX. Ils représentent à eux tous presque 10 200 lettres ou privilèges, soit près de 90% du corpus pontifical dans la *PL*²⁷.

{47} Afin de faciliter les repérages de ces documents, nous nous sommes fondés sur les titres donnés par Migne. Ce choix implique que quelques rares textes, en particulier ceux des III^e-V^e siècles, n'ont pas été inclus dans le corpus d'analyse. C'est par exemple le cas de certaines lettres et pseudo-lettres de Zosime, contenues au tome 45 de la *PL*, qui se trouvent classées dans une sous-partie consacrée aux textes anti-pélagiens, intitulée « *Varia scripta ad historiam Pelagianorum*

²² Sabanés I Fernández, Roser, et Tilamnn Schmidt (Éd.), *Butllari de Catalunya: documents pontificis originals conservats als arxius de Catalunya (1198-1417)*. Barcelone 2016.

²³ Le corpus des CEMA étant en cours de diffusion en accès ouvert, il sera bientôt possible d'interroger le sous-corpus d'actes pontificaux réuni pour l'article – de façon séparée, ou en comparaison avec les autres documents de la base.

²⁴ Leo Santifaller propose une liste des tomes de la *PL* qui contiennent des éditions de bulles, mais sans information précise sur le contenu ou les éditeurs, dans Santifaller: *Neuere Editionen*, voir n. 14, p. 57-62. Il faut en outre distinguer entre les volumes incluant une édition substantielle, et ceux mentionnant seulement des passages ou pièces isolés. En définitive, une cinquantaine de tomes de la *PL* paraissent contenir des éléments importants pour notre enquête.

²⁵ Il resterait toutefois à mesurer l'impact de la sélection/réimpression des éditions par Migne, ce dernier modifiant régulièrement certaines graphies, normalisant des passages, en évacuant d'autres dits « formulaires », dans des proportions que nous ignorons à ce stade.

²⁶ Dont les fameux registres occupent quasiment l'intégralité des quatre derniers volumes de textes de la *PL* (214-217). Concernant la production pontificale sous Innocent III, voir la synthèse de Fossier, Arnaud: Innocent III. Un pape de papier. Dans *Médiévales* 65 (2013), p. 179-188.

²⁷ Suivent 35 papes possédant au moins 10 documents édités dans la *PL*, soit un peu plus de 1 000 textes au total. Les 123 autres pontifes édités dans la collection de Migne y sont donc faiblement représentés, le plus souvent parce que leur production conservée est elle-même limitée. Malgré ces écarts très importants, la *PL* propose tout de même des documents pour 156 pontifes, sur un total de 176 pour la période allant saint Corneille [† 253] à Innocent III [† 1216].

pertinentia »²⁸. Dans le même volume et toujours dans cette sous-partie, on trouve encore une lettre d’Innocent I^{er}, intégrée sous le titre « *De eadem, ut creditur, vexatione Catholicorum a Pelagianis* »²⁹. L’identification de ces textes, par exemple sur la base de l’index donné par Leo Santifaller, nécessiterait une indexation complète de la *PL*, soit un examen critique de chaque texte et de sa tradition. Cela est d’autant plus hors de notre propos ici, que le nombre de textes ajoutés au corpus serait mineur face à la masse déjà identifiée³⁰. Enfin, nous avons préféré ne pas intégrer le tome 130 de la *PL*, qui contient les *Décrétales* pseudo-Isidoriennes³¹. En réunissant ces volumes aux bullaires précités, on arrive à un corpus conséquent³², mais aussi partiellement dépassé et biaisé.

I.2. *Cartulaires et intelligence artificielle*

Afin de constituer un corpus plus large et varié, une possibilité peu explorée jusqu’ici consiste à inverser la logique de recherche, en partant non pas des bullaires ou des collections spécialisées, mais des corpus de chartes numérisés, afin d’en extraire les documents pontificaux. Malgré la richesse des grandes entreprises déjà évoquées, on peut en effet postuler que la plupart des documents pontificaux, au moins depuis l’époque mérovingienne jusqu’au XII^e siècle inclus, ont déjà été édités dans les cartulaires concernant une institution ou une région. Il suffit d’observer quelques recueils de bulles récents pour s’en convaincre : dans son édition de référence, Harald Zimmermann donne 630 numéros (dont 556 correspondent à des textes complets) pour la période allant de 896 à 1046³³. Or, la totalité de ces textes avait précédemment été édités dans des collections diverses et variées³⁴. Il ne s’agit certes pas de nier l’intérêt des travaux éditoriaux récents ou en cours, mais de souligner que dans notre perspective maximaliste, la tradition imprimée des documents pontificaux est souvent l’histoire {48} d’une multi-édition – ce qui appuie *de facto* l’idée d’un récolement général regroupant les travaux antérieurs. Comment dès lors repérer ces documents, au sein des milliers d’éditions et des 240 000 textes inclus dans les CEMA, qui contiennent évidemment d’autres typologies diplomatiques ? L’analyse manuelle du corpus est bien entendu inenvisageable, car extrêmement chronophage.

²⁸ *PL* 45, col. 1719-1723.

²⁹ *PL* 45, col. 1710-1711.

³⁰ Mais aussi problématique, car les plus anciennes bulles de la *PL* proviennent en partie du corpus pseudo-Isidorien et d’autres ensembles documentaires tout aussi incertains.

³¹ Jasper: *Papal Letters* (voir n. 5), p. 135-196.

³² À l’issue de ces récolements numériques, tous les documents en question ont été intégrés définitivement aux CEMA.

³³ Zimmermann, Harald: *Papsturkunden 896-1046*, 3 volumes. Wien 1984-1989.

³⁴ En particulier par des éditeurs de renom tels Mabillon, Muratori, Ughelli, Kehr, Mansi, Wattenbach, Cocquelines, Tomassetti, Migne, von Harttung, von Sickel, Wierderhold, Ramackers, Wampach, etc. Ce constat avait déjà été fait dans certains comptes rendus de l’édition, dont Barbiche, Bernard: Harald Zimmermann. *Papsturkunden 896-1046*. I. Band : 896-996. II. Band : 996-1046. Wien: Verlag der Oesterreichischen Akademie der Wissenschaften, 1984-1985. 2 vol. in-4°, 1182 pages [...]. Dans: *Bibliothèque de l’École des chartes* 145-1 (1987), p. 214-216, ici p. 216.

Grâce aux procédures de classification supervisée (*Machine learning*), il est possible d'apprendre à l'ordinateur à distinguer les bulles des autres documents diplomatiques³⁵. Cette approche repose sur une hypothèse élémentaire, selon laquelle les textes pontificaux se distinguent des autres documents diplomatiques par leur contenu. Le processus numérique consiste en différentes moments : a) la constitution d'un corpus indexé pour l'apprentissage et les tests du modèle ; b) l'apprentissage lui-même, soit la création du modèle selon différents paramètres statistiques ; c) l'application du modèle à un corpus de test, n'ayant pas servi à l'apprentissage, mais qui est déjà indexé par ailleurs ; d) le contrôle de l'efficacité du modèle, à partir des résultats obtenus sur le corpus de test ; e) l'application du modèle final à l'ensemble du corpus non-indexé. On peut ajouter à cela une dernière étape, soit le contrôle manuel ou semi-automatique des résultats finaux.

Il s'agit donc en premier lieu de créer un corpus d'entraînement à partir d'un jeu de données préexistant, dans lequel les documents pontificaux sont indexés par différentes métadonnées, et donc opposables aux autres textes diplomatiques³⁶. Si ces {49} métadonnées n'étaient pas disponibles pour l'intégralité des bases incluses dans les CEMA, c'était le cas pour une partie des CBMA³⁷, mais aussi

³⁵ En 2011, lors du colloque *Digital Diplomats* de Naples, nous avons proposé d'employer ces méthodes de classification automatique pour les bulles et les diplômes. Voir la publication résultante dans Perreaux, Nicolas: De l'accumulation à l'exploitation ? Expériences et propositions pour l'indexation et l'utilisation des bases de données diplomatiques. Dans *Digital Diplomats. The Computer as a tool for diplomatist?*. Éd. par Antonella Ambrosio, Sébastien Barret et Georg Vogeler. Köln, Weimar, Wien 2014, p. 187-210 et 321-322. Les expériences restent assez rares dans ce domaine. Voir cependant Dahllöf, Mats: Classification of Medieval Documents: Determining the Issuer, Place of Issue, and Decade for Old Swedish Charters. Dans: DHN 2020. *Digital Humanities in the Nordic Countries: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*. Éd. par Sanita Reinsone, Inguna Skadiņa, Anda Baklāne et Jānis Daugavietis. Riga 2020, p. 12-23, où l'on trouvera une bibliographie concernant la classification des chartes suédoises.

³⁶ Se pose alors une question fondamentale : faut-il retenir un modèle distinguant dès le départ toutes les catégories diplomatiques (documents pontificaux, diplômes, actes épiscopaux, actes abbatiaux, donations de particuliers, etc.), ou bien un modèle plus rudimentaire, ne proposant que deux (documents pontificaux vs. autres documents) voire trois (documents pontificaux vs. diplômes vs. autres documents) catégories ? Nos expériences antérieures, ainsi que celles réalisées à l'occasion de cet article, nous font pencher pour un modèle à deux catégories, du moins en ce qui concerne les algorithmes retenus (*i.e.* le SVM, cf. plus bas). La raison est simple : les bulles constituent la catégorie diplomatique qui se singularise le plus fortement, suivie de près par les diplômes royaux et impériaux. Dans la perspective de catégoriser par type d'auteur l'intégralité des CEMA (un travail que nous souhaitons réaliser prochainement), on pourrait ainsi procéder en plusieurs étapes, en enchaînant en quelque sorte en « cascade » une série d'algorithmes : d'abord pour distinguer les bulles des autres documents grâce un premier modèle ; puis en distinguant les diplômes dans le corpus duquel nous aurions déjà soustrait les bulles ; enfin, trier dans l'ensemble restant entre les actes épiscopaux et les autres catégories diplomatiques, tout en sachant que ce dernier modèle devra probablement intégrer des catégories chronogéographiques (en particulier parce que les actes de certaines zones se distinguent, comme les actes italiens ou anglo-saxons). Cette hypothèse d'une classification par étapes est liée au fait que les écarts entre les types documentaires ne sont pas égaux : par exemple la différence entre les bulles ou les diplômes et le reste du corpus est généralement plus grande qu'entre toutes les autres catégories. Il convient donc de « découper » à chaque étape ce qui se distingue le plus dans la masse documentaire. Pour prendre une analogie cartographique, si nous devons distinguer par la distance à Paris un corpus contenant les villes de Bordeaux, Évreux, Marseille, Meaux, Orléans, Tours et Versailles, il faudrait d'abord isoler Marseille, puis dans l'ordre Bordeaux, Tours, Orléans, Meaux, Évreux et enfin Versailles. On trouvera différentes pistes sur la catégorisation algorithmique des chartes dans Perreaux: De l'accumulation (voir n. 35). Sur les liens entre les papes et l'Italie, nous renvoyons à Théry, Julien et Gilli, Patrick: *Le Gouvernement pontifical et l'Italie des villes au temps de la théocratie (fin XII^e-mi-XIV^e siècle)*. Montpellier 2010.

³⁷ Le corpus indexe ainsi un peu plus de 1 000 bulles, sur un total d'environ 30 000 textes.

pour la totalité de l'Artem³⁸ ou encore du corpus *Diplomata Belgica*³⁹. Un examen supplémentaire des analyses données par les éditeurs a permis d'étoffer ce corpus d'apprentissage, en repérant d'autres documents pontificaux grâce à des descriptions éditoriales typiques : « *Acte du pape...* », « *Bulle de...* », « *Le pape confirme que...* », etc. Enfin, certains bullaires disponibles ont eux aussi été versés dans le corpus d'apprentissage. Nous n'avons toutefois décidé de ne retenir qu'un corpus d'environ 4 926 textes, dont 188 documents pontificaux (soit 3,8%), sélectionnés aléatoirement dans l'ensemble des CEMA indexés. On répond ainsi à un triple impératif : que le corpus d'apprentissage ne soit pas trop vaste, qu'il prenne en compte la totalité de la chronologie impliquée, et enfin que la proportion de documents pontificaux inclus soit représentative de ce que l'on observe dans l'ensemble des CEMA. Dans le cas contraire, le risque serait grand d'arriver à un modèle de classification moins efficace⁴⁰.

L'étape suivante consiste alors à entraîner l'ordinateur sur les 2/3 de ce jeu de données, puis à contrôler l'efficacité du modèle statistique ainsi établi sur le 1/3 restant. L'intérêt de ce découpage est de faciliter le contrôle de l'efficacité du modèle, car il suffit alors d'aligner les résultats obtenus sur le 1/3 non-entraîné avec ses métadonnées. On voit ainsi rapidement la proportion d'erreur et la validité globale du modèle. Le processus est itératif : de nombreuses combinaisons algorithmiques ont ainsi été testées, afin d'aboutir à un modèle reconnaissant environ 95% des bulles. En pratique, c'est l'application d'un algorithme SVM (*Support Vector Machine*)⁴¹, sur un tableau pré-`{50}`-traité de 20 000 bi-formes qui s'est révélé être le plus efficace⁴². Une fois ce modèle établi, contrôlé et validé, il suffit de l'appliquer aux documents non-étiquetés pour obtenir la classification et donc le repérage des documents pontificaux sur la totalité des CEMA. Pour plus de fiabilité, les résultats de cette classification ont été contrôlés manuellement⁴³.

³⁸ Ce second ensemble indexe 272 bulles, soit 188 provenant des originaux avant 1121, et 84 de l'édition en cours pour la période allant de 1121 à 1220 (concernant uniquement la Lorraine à l'heure actuelle). On remarquera au passage qu'entre les deux chronologies, la proportion passe d'environ 3% de bulles à environ 10%.

³⁹ Le corpus propose et indexe un peu plus de 1 800 bulles.

⁴⁰ Il faut en effet que le modèle couvre la totalité du corpus, puisque c'est sur cette totalité que nous allons ensuite l'appliquer. Par ailleurs, afin que les caractéristiques statistiques du modèle soient proches de ceux de l'ensemble des CEMA, afin de sous-représenté ou surreprésenté les caractères lexicaux propres à tel ou tel ensemble documentaire.

⁴¹ En français : « Machine à vecteurs de support ». Il s'agit d'algorithmes linéaires, mais qui se révèlent trop robustes en cas de paramètres très nombreux (ce qui est le cas ici : nos paramètres étant les bi-formes sélectionnés). En l'occurrence, c'est la variante « `svmLinear3` » du package R `Caret` qui a été retenu, après de nombreux essais sur d'autres algorithmes (dont des réseaux neuronaux). Il s'agit d'un algorithme que nous avons déjà retenu pour la classification des actes diplomatiques en 2011/2014.

⁴² Retenir les bi-formes (des groupes de deux formes) possède un double avantage : on entraîne le modèle à la fois sur des informations lexicales (termes, graphies), mais aussi diplomatiques (enchaînement de formes). Afin de gagner en efficacité, les bi-formes très peu fréquents ont été retirés du calcul.

⁴³ Ce contrôle consiste à vérifier que l'attribution typologique est bonne, en examinant en particulier les résultats pour les documents dont la suscription et l'adresse ne sont pas conformes au canon de la l'écriture pontificale.

In fine, cette méthode a permis de repérer la presque totalité des documents pontificaux contenus dans les CEMA⁴⁴. En combinant les bullaires numérisés, la *Patrologie latine*, les textes pré-étiquetés, et surtout ceux repérés grâce à l'Intelligence artificielle, nous arrivons à un corpus composé d'environ 26 000 documents pontificaux, pour la période allant du IV^e au XVI^e siècle – dont presque 18 000 avant le décès d'Innocent III, et environ 22 000 avant la mort d'Alexandre IV, en mai 1261. Une telle quantité valide l'hypothèse selon laquelle il faut partir des recueils édités pour réunir le plus grand nombre de documents pontificaux possibles, lorsque les bullaires ne sont pas disponibles/pas numérisés. Un simple test permet d'aller plus loin dans le contrôle des résultats : le *Bullaire du pape Calixte II*, édité par Ulysse Robert, contient en effet 530 numéros (dont quelques-uns sont vides)⁴⁵. Or, sans même intégrer cette édition, la cumulation des méthodes employées et en particulier la classification automatique a permis de réunir 423 bulles de ce pape (soit 80% du corpus)⁴⁶. Ces actes ne proviennent bien entendu pas d'un seul corpus, mais de la *PL*, des *CBMA*, de *Chartae Galliae*, du Corpus des originaux (avant et après 1121), d'*Aposcripta*, du *Württembergisches Urkundenbuch*, des *Early Scottish Charters*, du *Deeds Project*, de *Diplomata Belgica*, du *Registrum Petri Diaconi*, des *Papsturkunden in Portugal*, etc.⁴⁷. Un score sensiblement identique est obtenu pour l'édition des *Papsturkunden* de Zimmermann⁴⁸, puisque la période 896-1046, couverte par les deux volumes, correspond à 485 actes dans les CEMA – soit près de 77% du corpus complet. Bien entendu, de tels scores sont inenvisageables au-delà d'Innocent III, lorsque les éditions de documents pontificaux ne sont pas intégrés directement aux CEMA⁴⁹. Mais ils laissent tout de même entrevoir qu'avant 1200, le corpus réuni est relativement dense.

{51} I.3. *Datations, abréviations, doublons*

Subsistent différents problèmes, en partie propres à cette documentation, mais aussi aux méthodes mises en œuvres, pour la datation, les actes abrégés et les doublons. En premier lieu, les questions de datation pourraient paraître moins cruciales que pour d'autres ensembles documentaires, car la présence des pontifes, mais aussi souvent de multiples indications chronologiques⁵⁰, facilitent

⁴⁴ La question des actes des légats est ici intéressante, car les algorithmes ont tendance à les associer automatiquement aux documents pontificaux. Cela confirme ainsi ce que nous savons des circulations entre ces types documentaires.

⁴⁵ Robert, Ulysse (Éd.): *Bullaire du pape Calixte II : 1119-1124*, essai de restitution, 2 volumes. Paris 1891.

⁴⁶ Quarante doublons ont été éliminés de nos calculs.

⁴⁷ Soixante pourcents de ces documents proviennent de la *PL*.

⁴⁸ Zimmermann: *Papsturkunden 896-1046* (voir n. 33).

⁴⁹ Outre les travaux déjà mentionnés, sur l'explosion de l'écrit pontifical au-delà des bulles, voir Fossier, Arnaud: *Le bureau des âmes. Écritures et pratiques administratives de la Pénitencerie apostolique (XIII^e-XIV^e siècle)*. Rome 2018 ; Fossier, Arnaud ; Petitjean Johann et Revest Clément (Dir.): *Écritures grises. Les instruments de travail des administrations (XII^e-XVII^e siècle)*. Paris-Rome 2019.

⁵⁰ Même si la datation des documents avant 1187 ne va évidemment pas sans soulever des difficultés – c'est-à-dire avant l'introduction de l'indiction sous Grégoire VIII, puis de l'année du pontificat sous Clément III.

l'attribution d'une date aux actes. Toutefois, le problème est ici éditorial, car certains éditeurs ne fournissent que le pontificat, qui n'est pas directement exploitable par les outils numériques. Ce problème touche spécialement la *PL*, qui classe les documents par pontife, donne des analyses latines, mais pas de date. Afin de rendre le corpus exploitable, nous avons donc décidé de systématiquement remplir un champ numérique des CEMA indiquant quel pape était concerné pour chaque document, et ceci pour les 26 000 textes identifiés. Cette étape permet ensuite d'attribuer la date médiane du pontificat au document⁵¹, dans le cas où celle-ci est manquante⁵².

Il est tout aussi important de détecter les textes abrégés, car cette pratique éditoriale peut fortement impacter des études comme la nôtre, portant sur l'évolution des formulaires, du lexique et de la sémantique. Ces documents sont en effet relativement nombreux (21% du corpus réuni), en particulier pour des périodes tardives⁵³. Mais ces {52} abréviations sont inégales et peuvent toucher de façon imprévisible des éléments très ponctuels... ou la majeure partie de l'acte⁵⁴. Les textes abrégés se présentent en effet sous des formes très diverses, ce qui complique fortement leur repérage : on trouve ainsi, pour les registres, des insertions de citations de documents originaux au sein

⁵¹ Les pontificats ont généralement une durée assez courte, en raison de l'âge assez avancé auquel on accède au siège de saint Pierre. Sur la durée des pontificats, voir aussi nos remarques note 132. En médiane, la durée des pontificats du V^e au XIII^e siècle (inclus) est de 4,5 années – ce qui fait que, statistiquement, la date centrale du pontificat n'est de toute façon guère éloignée de la date réelle à laquelle les documents sont produits.

⁵² Il s'agit de l'approche que nous appliquons systématiquement dans les CEMA pour tous les actes datés à partir d'une fourchette. De nombreuses procédures algorithmiques nécessitent en effet que les documents soient datés, même s'ils ne le sont que grossièrement. Il s'agit en quelque sorte de déterminer un « indicateur de tendance centrale », au sens statistique. Cette méthode d'attribution d'une date médiane présente différents avantages, en particulier face au *terminus ad quem* ou *terminus a quo*. En effet, dans ce dernier cas, lorsque nous sommes face à un document pour lequel toutes les années de la fourchette sont équiprobables, le choix des *termini* induit une plus forte probabilité d'erreur cumulée. Par exemple si la fourchette est de 25 ans, le fait de retenir un des *termini* n'offre qu'une chance sur vingt-cinq de tomber juste, mais aussi une grande probabilité d'erreur forte : si la date réelle s'avère être le *terminus a quo*, alors que nous avons retenu le *terminus ad quem*, l'erreur est grande (de 24 ans donc). Tandis qu'en prenant la date médiane de la fourchette, même si la probabilité de tomber juste n'est pas plus grande (1/25), la probabilité de faire une erreur lourde est fortement limitée. Dans le pire des cas, l'erreur sera en effet de 12 ans. En procédant de la sorte, on peut donc estimer que les erreurs seront moins nombreuses et les modélisations fréquentielles basées sur la diachronie plus fiables.

⁵³ C'est évidemment la stabilité des formulaires pontificaux qui induit indirectement cette attitude chez certains éditeurs. Dans le corpus pontifical réuni, nous avons ainsi pu dénombrier 12 240 mentions de « etc. », répartis dans 5 392 textes. Cela signifie que presque 21% des textes pontificaux inclus dans les CEMA sont abrégés, d'une façon ou d'une autre, à différents degrés (le plus souvent mineurs : les abréviations sont particulièrement localisées dans certaines parties du discours diplomatique). Si l'on compare ce score avec celui obtenu sur les diplômes royaux et impériaux, on trouve que 3 190 documents sur 23 947 de cette typologie contiennent une mention « etc. », soit 13,3%. Enfin, dans les autres documents des CEMA (207 972 textes), seuls 9 820 intègrent une telle occurrence, soit 4,7%. Cela signifie que les textes pontificaux sont quatre fois plus souvent abrégés que les autres, hors diplômes – eux aussi assez régulièrement abrégés pour des raisons de stéréotypicalités.

⁵⁴ Des variations étant courantes au sein d'une même édition, en fonction de l'intérêt de l'éditeur. Bien entendu, ces pratiques ont été fort heureusement largement abandonnées au cours du XX^e siècle.

de résumés⁵⁵, mais aussi des mentions de dates seules, des occurrences de témoins isolés⁵⁶, des renvois à des éditions antérieures, et enfin des abréviations plus ou moins poussées d'intitulations ou de clauses comminatoires⁵⁷, etc. Là encore l'hétérogénéité domine. La seule solution envisageable, qui, dans le cadre du présent article, ne soit pas particulièrement chronophage, est aussi simple qu'insatisfaisante : elle consiste à retirer des analyses numériques tous les documents contenant moins de 100 mots, ainsi que ceux débutant par une formule de datation (qui sont de fait abrégés). Les CEMA intègrent par ailleurs assez peu de regestes, qui sont massivement abrégés. Le cas des actes où seulement certaines formules sont abrégées resterait à traiter en propre, car ils sont particulièrement courants en diplomatie pontificale. Il s'agit en effet d'un problème pour lequel des solutions potentielles existent, mais nous n'avons pas encore réalisé ce travail⁵⁸.

La question des documents présents sous la forme de doublons est tout aussi délicate, et d'une portée plus générale encore. Il est pourtant essentiel de s'y atteler, car les analyses pourraient être faussées par ces textes multi-édités, d'ailleurs plus fréquents pour le haut Moyen Âge, mais aussi a priori plus nombreux que dans d'autres typologies diplomatiques⁵⁹. La tâche est loin d'être évidente car les variantes induites par les {53} éditeurs successifs – dans les graphies, la ponctuation ou la transcription des éléments visuels par exemple, mais aussi à travers la présence ou non d'abréviations – peuvent être importantes. Ceci sans même prendre en compte la nature de la source éditoriale, qui peut être dans certains cas un original, et dans d'autres une de ses copies (cartulaires, vidimus, papiers d'érudits, éditions anciennes, etc.). Les variantes pour un même acte vont donc de différences mineures, comme dans le cas de l'acte d'Honorius III pour Jacques d'Aragon (concernant la ville de Montpellier et ses dépendances), en 1219, édité à la fois par Teulet dans ses « Layettes du Trésor des chartes », mais aussi par Rouquette et Villemagne dans le *Bullaire de Maguelone*⁶⁰ ; à des différences plus importantes, comme dans un acte de Grégoire VII pour Cluny, en 1076, édité par Bernard et

⁵⁵ Il s'agit d'un cas particulièrement problématique, car il devient alors difficile de distinguer (pour un ordinateur) ce qui relève du document ancien et de l'éditeur, puisqu'il s'agit de la même langue. Sur l'intérêt propre des regestes, voir cependant Große, Rolf: De l'utilité des regestes (voir note 15). Ils restent toutefois peu utiles dans notre perspective.

⁵⁶ « *Ego Aurasicensis episcopus Gilelmus suscripsi.* », dans un acte d'Urbain II (Albanes, Joseph-Hyacinthe et Chevalier, Ulysse (Éd.): Histoire des archevêchés, évêchés et abbayes de France, Gallia Christiana Novissima, tome VI : Orange. Valence, 1916, col. 182-183).

⁵⁷ « *Ad notitiam vestram, etc.* », dans un acte d'Anastase IV (PL vol. 88, n° 20).

⁵⁸ Il s'agirait de compléter les actes avec les formules typiques du pontificat. Mais cette opération ne va pas sans poser de problème, tant on ne peut exclure des variations ponctuelles, que ce soit dans la graphie, dans l'ordre des termes ou même dans les formules. C'est donc aussi par prudence que nous avons pour le moment différé cette opération, qui devra de toute façon être dûment renseignée dans les CEMA (par exemple par la présence de crochets, lorsque l'édition n'en use pas).

⁵⁹ Mais peut-être moins que dans le cas de la diplomatie royale ou impériale.

⁶⁰ Teuley, Alexandre (Éd.): Layettes du Trésor des chartes. Paris 1863, n° 1345 : « *Honorius episcopus [...] grato concurrentes assensu, villam Montis Pessulani et de Latis et de Omelar castra cum omnibus terris [...]* » ; Rouquette, Julien et Augustin Villemagne (Éd.): Bullaire de l'Église de Maguelone. Paris 1911-1914, volume II, n° 240 : « *Honorius episcopus [...] grato concurrentes assensu, villam Montispessulani, et de Latis et de Omelar castra, cum omnibus terris [...]* ».

Bruel, mais aussi présent dans l'Artem⁶¹. Dans ce deuxième cas, outre des différences sur la ponctuation et les graphies, la datation de l'acte apparaissant dans la copie de Bernard et Bruel n'est pas présente sur l'original de l'Artem. Ce genre de « troncature » rend difficile le repérage algorithmique des doublons.

Nous avons donc eu ici recours à une autre bibliothèque de fonctions R, intitulée TextReuse⁶². La méthode est assez similaire à celle employée pour l'apprentissage supervisé : elle consiste à comparer les bi-formes de toutes les paires possibles de documents pontificaux. On obtient ainsi un score, allant de 0 à 1 : 1 si la similitude des deux textes est parfaite (il s'agit alors d'un doublon au sens strict) ; 0 s'ils sont totalement dissociés. Afin d'améliorer ce repérage, différents traitements préparatoires ont été réalisés : transformation de « j » et « i », des « v » et « u », simplification des ligatures, suppression de la ponctuation, etc. On gomme ainsi les potentiels choix éditoriaux, pour se concentrer sur les textes⁶³. Le résultat du processus a été examiné manuellement : un travail fastidieux mais nécessaire. Lorsqu'un doublon a été détecté, l'édition la plus récente, en particulier lorsqu'elle était donnée à partir de l'original, a été préférée. La méthode révèle parallèlement des cas délicats à trancher. Que faire en effet des actes vidimés : doit-on les exclure même s'ils sont assortis de commentaires ou d'ajouts ?⁶⁴

{54} Au total, ce sont 1 376 doublons de documents pontificaux qui ont été repérés dans les CEMA. C'est assez peu, comparé au stock des 26 000 « bulles » identifiées initialement (environ 5,3% de l'ensemble). Une des causes majeures de ces doublons est la présence des chartes originales de l'Artem, souvent déjà éditées par ailleurs, mais aussi des bullaires qui redoublent les cartulaires. Nous passons ici rapidement, mais ces enquêtes pourraient ouvrir de nouvelles perspectives, tant historiques qu'historiographiques⁶⁵.

⁶¹ Bernard, Alexandre et Bruel, Alexandre (Éd.): Recueil des chartes de l'abbaye de Cluny, Tome 4 : 1027-1090. Paris 1888, n° 3498 : « *[Gregorius Episcopus, servus servorum Dei, dilecto in Christo filio Hugoni Abbati Monasterii [...] iudicium præmia æternæ pacis inveniant. Datum Lateranis quinto Idus Decembris [...]]* » ; Artem n° 1234 « *GREGORIUS, EPISCOPUS, SERVUS SERVORUM DEI, DILECTO in Christo filio HUGONI, abbati monasterii [...] iudicem premia æternæ pacis inveniant.* ».

⁶² Mullen, Lincoln: textreuse: Detect Text Reuse and Document Similarity. 2020, <https://github.com/ropensci/textreuse>.

⁶³ Sans bien entendu que le texte de l'acte inclus dans les CEMA ne soit modifié : il s'agit seulement de faciliter la détection des doublons.

⁶⁴ Il est vrai que l'identification manuelle des apports du *vidimus* n'est pas, la plupart du temps, une difficulté. Mais il n'en va évidemment pas de même pour une machine – que l'on pourrait cependant entraîner à la tâche. Comme souvent, l'idéal serait ici de réaliser des essais avec et sans ces actes. Ils jouent cependant un rôle mineur dans les calculs qui suivent, car la gangue ajoutée lors de la rédaction d'un *vidimus*, à l'échelle du corpus, reste marginale.

⁶⁵ Par exemple sur les pratiques éditoriales, en particulier les copies de copies.

II. Mesurer la richesse du corpus

II.1. Quelles chronologies pour les documents pontificaux ?

Une fois ce corpus réuni, identifié et grossièrement nettoyé, il est possible d'en donner une analyse⁶⁶. L'objectif est ici de montrer dans quelle mesure les méthodes numériques permettent de jeter une lumière originale sur l'écrit pontifical, confirmant ou infirmant des hypothèses antérieures, mais proposant aussi des pistes nouvelles. Une première possibilité offerte par la réunion du corpus est l'étude de la répartition chronologique des documents⁶⁷.

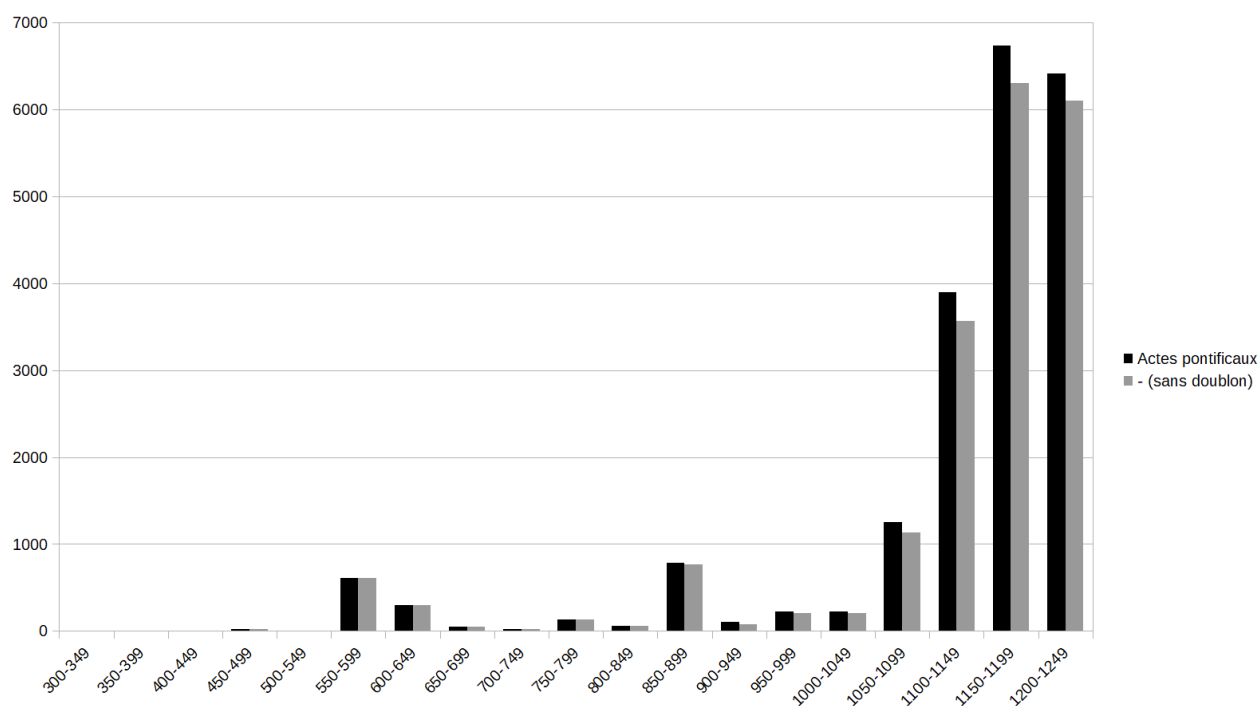


Fig. 1 : Répartition des documents pontificaux, privilèges et lettres, dans les CEMA (IV^e-milieu du XIII^e siècle).

⁶⁶ Les histoires généralistes des papes offrent ici des éléments contextuels utiles. Nous renvoyons en dernier lieu à Herbers, Klaus: *Geschichte des Papsttums im Mittelalter*. Darmstadt 2012 ; Frenz, Thomas: *Das Papsttum im Mittelalter*. Köln 2010 ; Tanase, Thomas: *Histoire de la papauté en Occident*. Paris 2019, où l'on trouvera de nombreuses références bibliographiques. On trouvera aussi différents éléments dans Weinfurter, Stefan (Dir.): *Päpstliche Herrschaft im Mittelalter: Funktionsweisen - Strategien - Darstellungsformen*. Ostfildern 2012. Une des difficultés déjà évoquées reste que la « papauté » est encore paradoxalement souvent abordée comme une réalité immuable, une institution qui serait coextensive de la société médiévale.

⁶⁷ Outre les ouvrages déjà mentionnés, voir d'autres estimations chiffrées dans l'article de Jamme: *Écrire pour le pape* (voir n. 6). Se fondant sur les calculs de Robert Fawtier (*Les registres de Boniface VIII*, éd. G. Digard, M. Faucon, A. Thomas, R. Fawtier, Paris, 1884-1939), Armand Jamme donne des chiffres impressionnants pour la production générale de l'écrit pontifical aux XIII^e-XIV^e siècles, avec 35 500 lettres par an en moyenne (malgré des variations interannuelles remarquables).

L’histogramme présenté ci-dessous comptabilise presque 21 000 documents. Il est donc plutôt représentatif des textes conservés, même s’il ne les intègre pas tous. La figure permet de relever différents phénomènes, plus ou moins attendus. Tout d’abord, malgré la présence de certains faux, la relative rareté de ces documents avant le milieu du XI^e siècle est confirmée⁶⁸. Il existe certes deux pics documentaires antérieurs, {55} dans la seconde moitié des VI^e puis IX^e siècle. Attendus, ils correspondent respectivement aux lettres de Grégoire le Grand [590-604]⁶⁹, puis aux actes de Nicolas I^{er} [858-867] et Jean VIII [872-882] (fig. 1 et 2a/b), pour lesquels nous conservons aussi un nombre élevé de privilèges mais surtout de lettres, à travers des registres ou pseudo-registres⁷⁰. Par ailleurs, on constate que la distribution des doublons est beaucoup plus {56} homogène que prévue⁷¹. L’enseignement principal du graphe reste l’explosion plus ou moins attendue du nombre des documents pontificaux conservés à partir de Léon IX [1048-1054]⁷². {57}

⁶⁸ La critique de tous les actes, à partir de leurs traditions éditoriales, étant évidemment irréaliste dans le cadre du présent article. Les plus anciens documents intégrés dans le décompte, avant Grégoire le Grand, proviennent principalement de Loewenfeld, Samuel (Éd.): *Epistolae pontificum romanorum ineditae*. Leipzig 1885 ; Albanès, Joseph-Hyacinthe (Éd.): *Gallia christiana novissima. Histoire des archevêchés, évêques et abbayes de France. Provinces d’Albi, Aix, Arles, Avignon et Auch. Montbéliard 1899-1920 ; Von Pflugk-Harttung, Julius (Éd.): Acta pontificum Romanorum inedita. Urkunden der Päpste vom Jahre 748 bis zum Jahre 1197, II. Stuttgart 1884 ; De Manteyer, Georges: Les chartes du pays d’Avignon (439-1040). Mâcon 1914 ; Quantin, Maximilien (Éd.): *Cartulaire général de l’Yonne. Auxerre 1854-1860*. Les actes antérieurs à Grégoire le Grand intégrés dans la *PL* n’ont pas été mentionnés, ainsi que nous l’avons déjà mentionné – car la masse des faux y est écrasante, en particulier dans les fausses décrétales. Sur la critique des documents pontificaux, nous renvoyons à Große, Rolf (Dir.): *L’acte pontifical et sa critique*. Bonn 2007.*

⁶⁹ *PL* 77 ; Martin, Jean-Marie, Pierre Chastang, Errico Cuozzo, Laurent Feller, Giulia Orofino, Aurélie Thomas et Matteo Viliani (Éd.): *Registrum Petri Diaconi*, 4 volumes. Roma 2015, n° 1, 65, 68, 69, 70, 71, 72 et 73.

⁷⁰ Pour Nicolas I^{er}, la majeure partie de nos textes provient de *PL* 119, qui les reprend de Jaffé. Voir cependant l’édition des lettres dans Perels, Ernest (Éd.): *Nicolai I. papae epistolae*. Berlin, 1925 (MGH *Epistolae Karolini Aevi* 4), qui donne 170 numéros. Sur ce pontife, nous renvoyons à Bougard, François: *Anastase le bibliothécaire ou Jean Diacre ? Qui a écrit la vie de Nicolas Ier et pourquoi ?*. Dans: Martin, Jean-Marie, Bernadette Martin-Hisard et Agostino Paravicini Bagliani (Éd.): *Vaticana et mediaevalia. Études en l’honneur de Louis Duval-Arnould*, Firenze 2008, p. 27-40 ; Id.: *Niccolò I, papa, santo*. Dans: *Dizionario biografico degli italiani* 78 (2013), https://www.treccani.it/enciclopedia/niccolo-i-papa-santo_%28Dizionario-Biografico%29/. Dominique Iogna-Prat insiste de la même façon sur l’importance de ce pape canonisé, dans Iogna-Prat, Dominique: *La maison Dieu. Une histoire monumentale de l’Église au Moyen Âge*. Paris 2006, p. 198-199, où il rappelle que le *Liber Pontificalis* mentionne que sous son pontificat « comme jamais le Siège apostolique est consulté par diverses régions ». Concernant Jean VIII, nos documents proviennent principalement de *PL* 126, mais aussi de Loewenfeld, du *Cartulaire de l’Église d’Autun*, des actes pour Saint-Philibert de Tournus, de Flavigny ou encore de Fleury (l’édition récente de Lohrmann, Dietrich (Éd.): *Das Register Papst Johannes VIII. 872-882*. Tübingen 1968, n’a malheureusement pas été incluse, car elle n’est pas disponible actuellement en version numérique). Les actes d’Adrien II [867-872] sont moins nombreux (l’édition d’Ernest Perels ne donne que 44 numéros). Du reste, comme l’indique François Bougard, l’influence scripturaire d’Anastase le Bibliothécaire est connue pour ces trois papes. Sur la scripturalité pontificale au IX^e siècle, nous renvoyons au remarquable travail de Unger, Veronika: *Päpstliche Schriftlichkeit im 9. Jahrhundert*. Archiv, Register, Kanzlei. Köln 2018 (Regesta Imperii, Beihefte: Forschungen zur Kaiser- und Papstgeschichte des Mittelalters, Bd. 45).

⁷¹ Du VII^e au XV^e siècle (inclus), la médiane des pourcentages de doublons par demi-siècle est de 3%, tandis que la moyenne se situe à 4,3%. C’est toutefois au X^e siècle (14% en moyenne) et plus globalement aux X^e-XIII^e siècles (8,5%) que ceux-ci sont plus présents : d’une part à cause des originaux de l’Artem, qui redoublent les cartulaires numérisés, d’autre part parce que la *PL* vient elle aussi proposer ces mêmes cartulaires.

⁷² On rejoint ici les remarques d’une partie de l’historiographie, qui aborde cette évolution par d’autres caractères (formes externes, formulaire diplomatique, etc.), par exemple dans Bischoff, Frank: *Urkundenformate im Mittelalter. Größe, Format und Proportionen von Papsturkunden in Zeiten expandierender Schriftlichkeit* (11.-13. Jahrhundert). Marburg 1996.

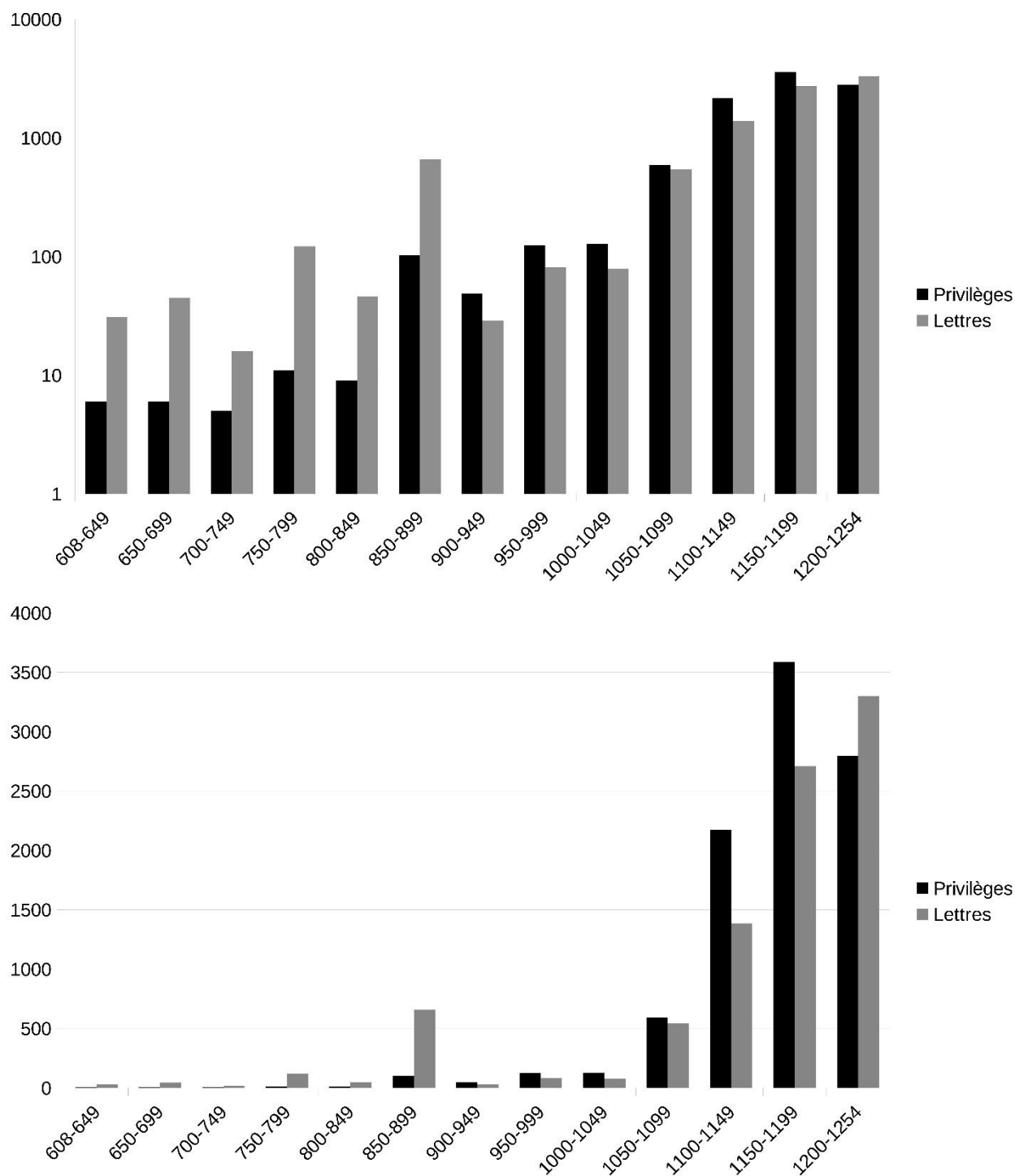


Fig. 2a et b : Répartition des privilèges et des lettres, dans les CEMA, 608-1254 (a : échelle logarithmique ; b : échelle arithmétique)⁷³.

⁷³ Sur ces figures 2a/2b, ainsi que les 4a/4b qui suivent, la division chronologique par demi-siècle n'est pas idéale, car elle ne fait pas apparaître les ruptures propres à certains pontificats. On consulera donc en parallèle la figure 3, qui donne une autre vision de l'évolution de la production textuelle, avec un décompte du nombre de documents par année et par pontife.

Une seconde figure, distinguant cette fois entre privilèges et lettres, pour la période allant du pontificat de Boniface IV [608-615] à celui d’Innocent IV [1243-1254], employant une échelle logarithmique, montre de la même façon une forte croissance des documents dans la seconde moitié du XI^e siècle. Une autre information complète cette tendance : le ratio entre les documents considérés comme « privilèges » et ceux renseignés comme « lettres » bascule totalement au cours de la période. Jusqu’au début du X^e siècle en effet, les lettres constituent en médiane 85% du corpus pontifical réuni, ce ratio étant d’ailleurs relativement stable du VI^e au IX^e siècle inclus. Or, à partir du X^e siècle, ce sont les privilèges qui occupent en médiane 60,5% du corpus. Loin d’invalider les observations précédentes, ces éléments les complètent largement : la croissance très rapide du corpus à partir du XI^e siècle est ainsi précédée par une évolution typologique déterminante, qui a lieu dans le courant du X^e siècle. Les lettres et les privilèges possèdent ainsi des caractères chronologiques en partie distinctifs, avec une surreprésentation des lettres jusqu’au X^e siècle, incluant des variations très importantes toutefois ; une très grande rareté des privilèges jusqu’à la seconde moitié du IX^e siècle⁷⁴ ; une évolution du ratio entre lettres et privilèges au X^e siècle (où les premières se font d’ailleurs très rares). Cela n’empêche pas une tendance globale, marquée d’une part par une relative rareté des documents jusqu’au XI^e siècle (avec une variabilité forte pour les lettres en fonction des pontifes et des dossiers conservés), d’autre part une explosion des documents après 1050. Même si l’on concède que ces typologies documentaires – lettres et privilèges – sont floues, cela signifie que le contenu et la formalisation des documents pontificaux évolue fortement aux X^e-XI^e siècles, ce que confirme par ailleurs l’historiographie. {58}

⁷⁴ Le privilège pontifical au sens strict n’existait diplomatiquement qu’à partir du VIII^e siècle, avec le pontificat d’Hadrien I^{er}.

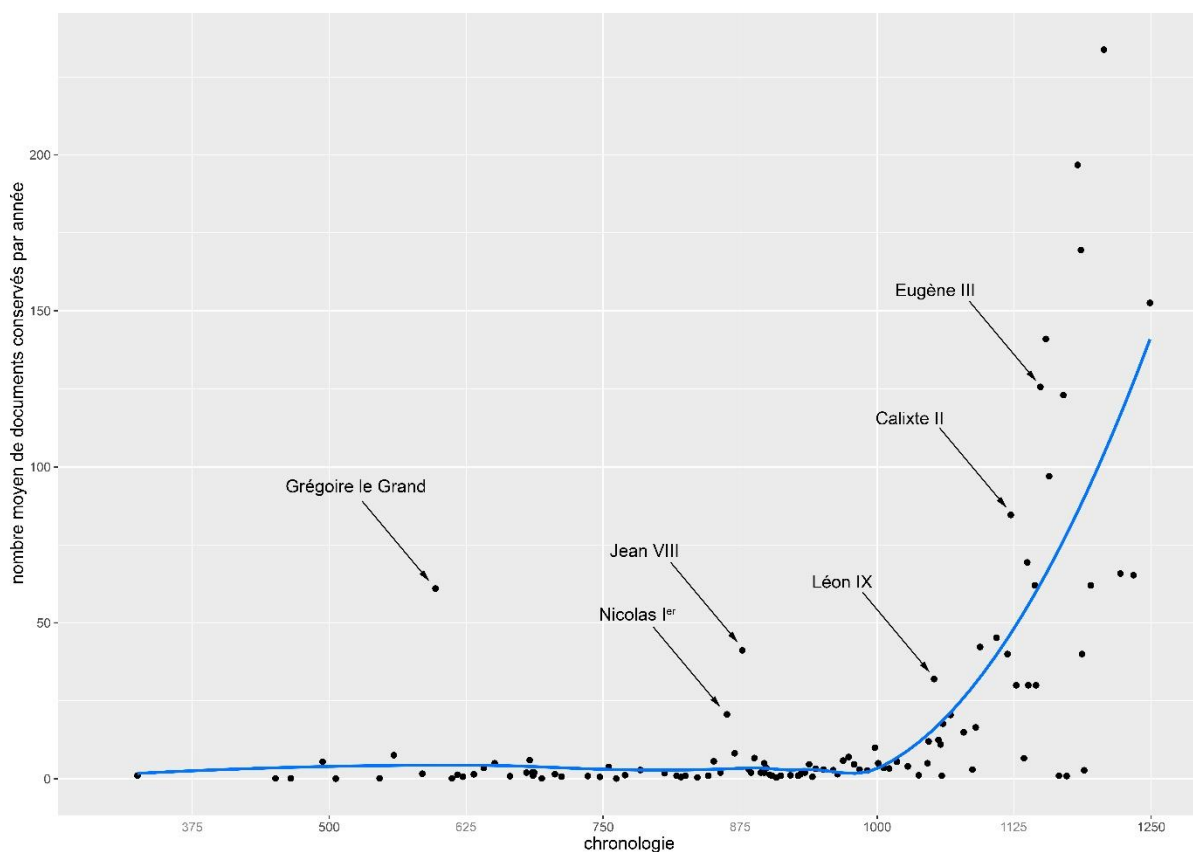


Fig. 3 : Nombre moyen de documents pontificaux (privilèges et lettres) par année, pour chaque pontife dans les CEMA (325-1250). Chaque point correspond à un pape : en abscisse, le milieu de son pontificat ; en ordonnées, le nombre d’actes moyen par année pour ce pontife. La ligne bleue correspond à la moyenne mobile. Ce lissage permet que des cas exceptionnels, comme les lettres de Grégoire le Grand, ne déforment pas la tendance globale.

Cette explosion du nombre de textes produits et conservés est confirmée par un troisième graphique (fig. 3). Sur cette figure, on relève peut-être plus facilement la profonde rupture quantitative que constitue le milieu du XI^e siècle – au-delà des variations déjà évoquées, liées à certains pontificats (Grégoire le Grand, Nicolas I^{er}, Jean VIII)⁷⁵ mais aussi aux évolutions typologiques. Une autre observation intéressante concerne la très grande variabilité du nombre des actes par année en fonction des pontifes, qui a tendance à s’accroître à partir du XI^e siècle. La différence est immense, entre, par exemple, Calixte II [1119-1124], qui possède en moyenne 85 actes par an dans les CEMA, et Victor III [1086-1087] qui compte seulement 3 documents dans la base. De telles distinctions ne peuvent a priori s’expliquer par le corpus lui-même, car {59} nous avons recueilli tous les documents

⁷⁵ Autrement dit, aux pontifes du haut Moyen Âge pour lesquels nous conservons des collections, des registres ou fragments de registres.

issus des cartulaires édités numérisés⁷⁶ – mais plutôt par l’action variable des pontifes, et la perception tout aussi changeante des bénéficiaires de documents. {60}

⁷⁶ L’extension des CEMA permettrait toutefois des enquêtes complémentaires, en particulier sur l’évolution chronogéographique de la répartition des documents pontificaux. Pour le moment, seule la Lombardie et une partie de l’Italie centro-méridionale sont bien couvertes par le corpus. L’ajout de cartulaires italiens, mais aussi hispaniques, permettrait sans doute de faire émerger des « zonations » plus précises pour l’activité pontificale, variable en fonction de la chronologie.

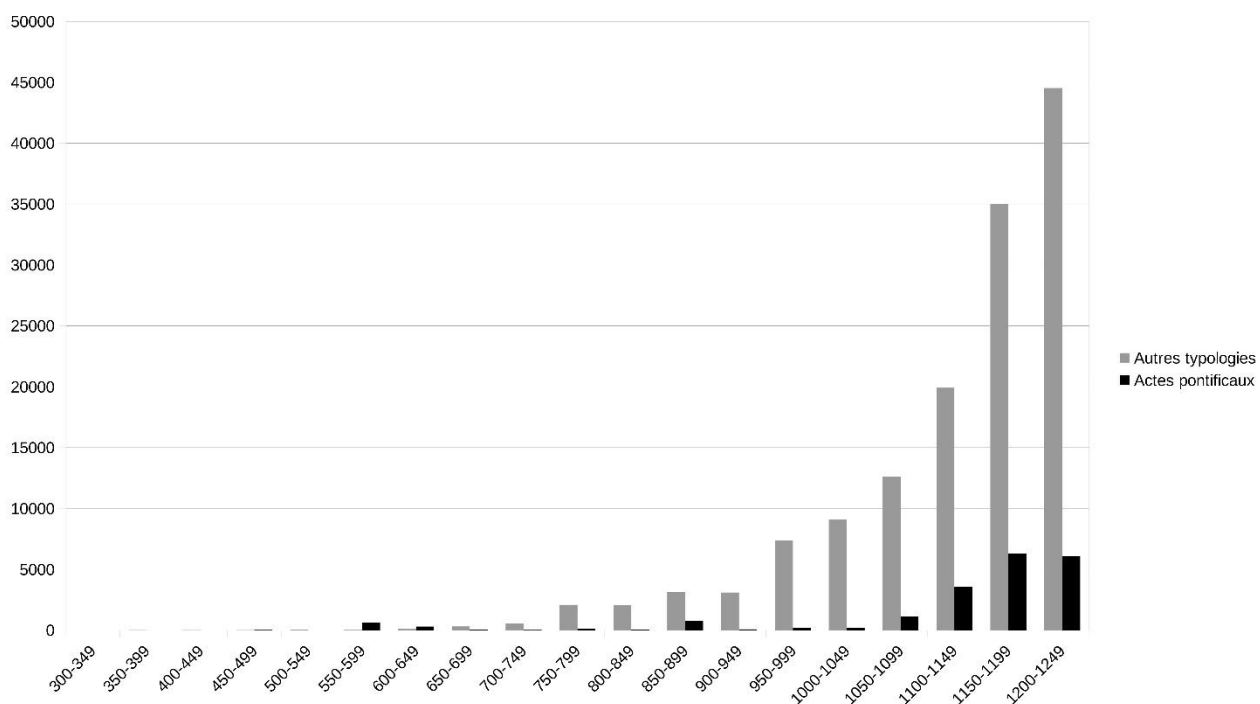
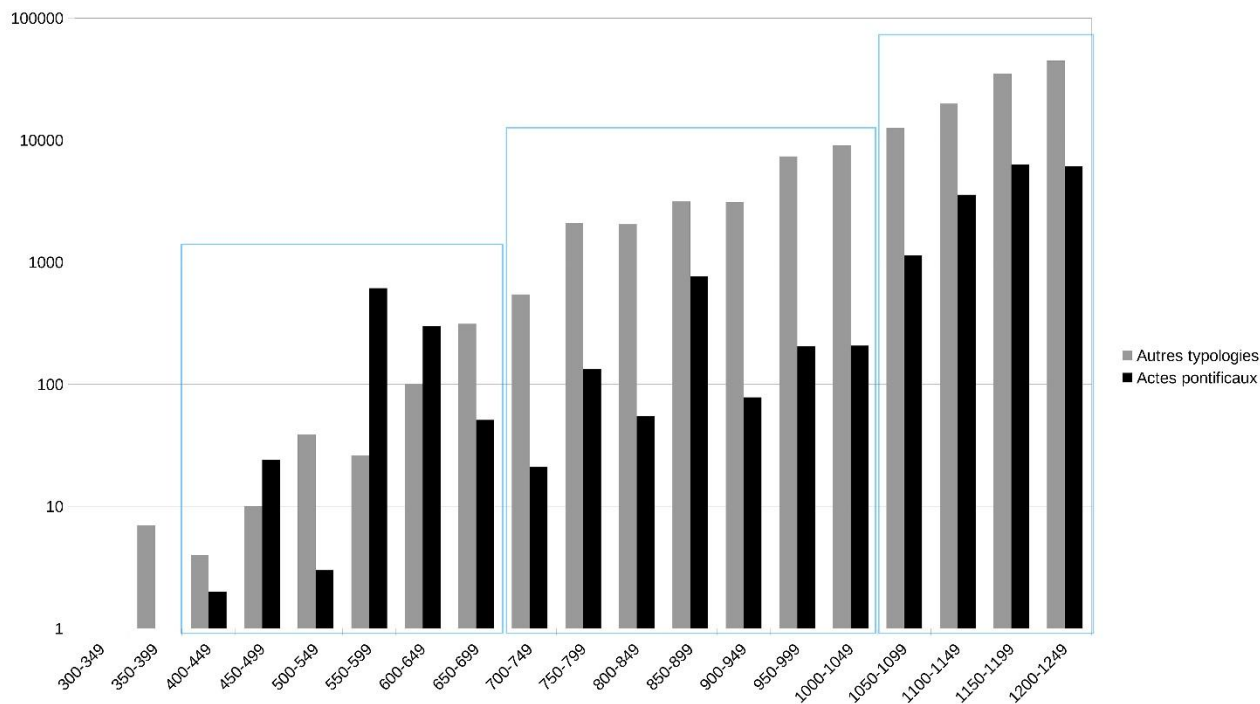


Fig. 4a et b : Documents pontificaux et autres typologies diplomatiques dans les CEMA, IV^e-XIII^e siècle (a : échelle logarithmique ; b : échelle arithmétique). Sur la figure 4a, les trois phases identifiées sont encadrées : elles sont partiellement liées à l'évolution typologique du corpus pontifical lui-même (ratio lettres vs. privilèges).

Il est possible d'observer l'évolution de la proportion de documents pontificaux dans la totalité des CEMA, par demi-siècle (fig. 4a/b). De tels changements traduisent certes la nature

variable de la documentation incluse dans la base, mais aussi, croyons-nous, des phénomènes proprement historiques. Sur ce plan, nous pouvons distinguer trois phases : 1) du V^e au VII^e siècle⁷⁷ ; 2) du VIII^e au milieu du XI^e siècle ; 3) du milieu du XI^e siècle à la fin du XIII^e siècle. Lors de la première phase, on observe que les documents pontificaux représentent environ la moitié du corpus, soit en moyenne 49,3% sur ces six demi-siècles (52% en médiane). Cette situation s'explique par la présence des lettres pontificales conservées, déjà évoquée à plusieurs reprises. La seconde phase constitue une profonde évolution, puisque les documents pontificaux n'y représentant plus que 5,6% du corpus en moyenne (2,7% en médiane). Cette tendance peut s'expliquer de différentes façons : d'une part, par le développement progressif mais néanmoins sensible de la production des actes diplomatiques, toutes typologies confondues ; d'autre part, par une moins bonne conservation des documents pontificaux ; enfin, peut-être, par un recul relatif du rôle de la fonction pontificale lors de certaines périodes au cours de ces siècles. Cette dernière possibilité n'est certes pas généralisable⁷⁸, mais elle pourrait néanmoins expliquer une part de la tendance observée. Enfin, la dernière phase voit le taux de documents pontificaux au sein des CEMA augmenter sensiblement, puisque ceux-ci y représentent en moyenne 10,3% (9,4% en médiane). Elle apparaît donc avant la conservation systématique des registres⁷⁹, et {61} traduit donc, au-delà d'une meilleure conservation, un rôle accru pour le pape et la curie pontificale, en tant qu'intercesseur, protecteur et arbitre.

⁷⁷ Pour la période antérieure, la documentation n'est probablement pas redevable d'une étude quantitative, tant chacune des rares pièces appelle une critique exhaustive et de longue haleine.

⁷⁸ Ainsi la seconde moitié du IX^e siècle voit la proportion de documents pontificaux augmenter en flèche, atteignant 19,57% du corpus (soit 763 documents pontificaux). Il s'agit encore une fois des lettres issues des importants pontificats de Nicolas I^{er}, Jean VIII, mais aussi Adrien II et Étienne V. Outre les références déjà mentionnées note 5, voir Herbers, Klaus: *Agir et écrire : les actes des papes du IX^e siècle et le Liber pontificalis*. Dans: Bougard, François et Sot, Michel (Éd.): *Liber, Gesta*, histoire : écrire l'histoire des évêques et des papes, de l'Antiquité au XXI^e siècle. Turnhout 2009, p. 109-126.

⁷⁹ On trouve certes des collections antérieures, conservées de façon fragmentaire, mais encore une fois elles ne déterminent pas la tendance observée lors de la phase n° 2 – d'autant plus que ces collections de furent probablement pas systématiques. Sur ces registres avant le XIII^e siècle, voir Delivré, Fabrice: *Les registres pontificaux du XII^e siècle : l'apport des Libri de primatu Toletane ecclesie*. Dans: *Mélanges de l'Ecole française de Rome. Moyen Âge 120-1* (2008), p. 105-138 ; Schieffer, Rudolf: *Die päpstlichen Registervor 1198*. Dans : Herbers, Klaus et Jochen Johrendt (Éd.): *Das Papsttum und das vielgestaltige Italien. Hundert Jahre Italia Pontificia*, Berlin, 2009, p. 261-273 ; Moreau, Dominic: *Deux notes sur l'utilisation du terme *rege(i)st(r)um* pour désigner les recueils de documents pontificaux antiques*. Dans: Gioanni, Stéphane et Paolo Cammarosano (Éd.): *Les correspondances en Italie 2. Formes, styles et fonctions de l'écriture épistolaire dans les chancelleries italiennes (Ve-XVe siècle)*. Trieste-Rome 2013, p. 85-103.

	Originaux seuls		Copies seules	
	Autres types	Actes pontificaux	Autres types	Actes pontificaux
300-349	0	0	1	1
350-399	0	0	7	0
400-449	0	0	4	2
450-499	1	0	9	24
500-549	0	0	39	3
550-599	3	0	23	613
600-649	15	0	86	299
650-699	51	0	260	51
700-749	80	0	465	21
750-799	298	2	1789	131
800-849	263	0	1780	55
850-899	705	8	2431	755
900-949	446	0	2647	78
950-999	1200	9	6154	196
1000-1049	1861	9	7217	198
1050-1099	2621	101	9980	1035
1100-1149	4088	256	15820	3306
1150-1199	7668	623	27346	5676
1200-1249	8632	895	35886	5203
1250-1299	6626	486	29513	1718
Somme	34558	2389	141457	19365

Fig. 5 : Originaux et copies dans les CEMA (IV^e-XIII^e siècle).

Qu'en est-il du côté des actes conservés en original ? Sur l'ensemble des originaux inclus dans les CEMA, toutes périodes confondues, nous constatons que les documents pontificaux représentent 6,46% du corpus. C'est relativement peu, et en tout état de cause, presque deux fois moins que pour les copies, puisque les documents des pontifes constituent 12% de ce second sous-corpus. Cela signifie d'emblée que les textes pontificaux sont a priori plutôt moins bien conservés que les autres typologies diplomatiques, à l'échelle de l'ensemble de la période. Cette situation pourrait en partie s'expliquer par des facteurs endogènes : d'une part, par la quantité même des documents, et la difficulté de conserver cette masse avant les registres ; d'autre part, par la variété et l'évolution typologique de documents des pontifes, puisque leurs lettres, proportionnellement très présentes dans le haut Moyen Âge, sont massivement connues par des copies. Mais au-delà de ces biais, cette situation paraît tout de même renvoyer à des phénomènes concrets, qui apparaissent plus nettement en faisant varier les chronologies. Ainsi, avant 1050, on ne décompte que quelques documents pontificaux originaux dans les CEMA, soit environ 0,5% de l'ensemble des originaux pour cette période⁸⁰. Entre 1050 et 1299, cette proportion monte à environ 8%, soit 14 fois plus⁸¹.

Par ailleurs, si l'on considère cette fois le ratio entre originaux et copies, on note que seulement 1,14% de tous les documents pontificaux inclus dans les CEMA avant le milieu du XI^e siècle sont des

⁸⁰ Tous les actes considérés comme faux dans les bases de données incluses dans les CEMA ont été retirés des décomptes. C'est le cas en particulier pour le corpus des originaux de l'Artem. Les analyses ont été réalisées sur la version de novembre 2019 de la base. Sur la rareté des documents pontificaux originaux jusqu'au IX^e siècle, voir Noble, Thomas F. X.: Literacy and the papal government in late antiquity and the Early Middle Ages. Dans: McKitterick, Rosamond (Dir.): The uses of literacy in early mediaeval Europe. Cambridge 1990, p. 82-133, ici p. 86-87.

⁸¹ Le nombre d'originaux dans les CEMA avant 1200 (1 008) est d'ailleurs très cohérent avec celui donné par Julius von Pflugk-Harttung en 1884, qui donne dans son catalogue 963 numéros avant Innocent III, dans Von Pflugk-Harttung, Julius: Päpstliche Original-Urkunden und Scheinoriginale. Dans: Historisches Jahrbuch 5 (1884), p. 489-575.

originaux⁸², contre plus de 17,68% pour les autres types⁸³. Or, cette proportion grimpe à 12,24% entre 1050 et 1299⁸⁴, et se maintient à 19,9% pour les autres types⁸⁵. Si la *PL* joue probablement un rôle dans ces chiffres, avec encore une fois l'omniprésence des lettres dans cette collection, on note que les documents pontificaux sont particulièrement mal conservés pour le haut Moyen Âge. Une telle observation interroge l'hypothèse de la préciosité de ces actes avant le XI^e {62} siècle – à l'inverse de ce que l'on lit parfois dans la littérature –, du moins aux yeux des institutions qui auraient dû les conserver⁸⁶.

On peut, pour s'en convaincre, proposer une comparaison avec les diplômes royaux et impériaux de la collection des MGH. Dans les CEMA, elle réunit en effet un peu plus de 6 100 documents avant 1200, dont près de 2 500 antérieurs au XI^e siècle. Pour cette dernière période, les documents pontificaux regroupent environ 1 600 textes dans le corpus. Or, dans le cas des diplômes, environ 48% des documents authentiques sont conservés sous la forme d'originaux. En ce qui concerne les documents pontificaux, la moyenne sur cette même période est inférieure à 1%. L'effet n'est d'ailleurs pas strictement lié au haut Moyen Âge, puisqu'au XI^e siècle encore, 53,2% des documents impériaux sont des originaux, contre 6,62% des documents pontificaux, mais encore au XII^e siècle, avec un ratio originaux-copie de 32,8% contre 8,6%. Ces chiffres ne peuvent en outre s'expliquer par la masse supposée des textes pontificaux, puisque {63} jusqu'au XI^e siècle au moins, la production pontificale conservée reste limitée – même en incluant les lettres.

Certes, plusieurs facteurs doivent être conservés à l'esprit pour interpréter convenablement ces observations. D'une part, nous savons que pour les périodes anciennes les documents pontificaux étaient rédigés sur papyrus⁸⁷. Les difficultés de conservation propres à ce support sont connues et pourraient expliquer une partie des différences observées, en particulier car cette pratique perdure à la chancellerie pontificale jusqu'au début du XI^e siècle⁸⁸. Mais si l'on regarde attentivement les chiffres pour la seconde moitié du XI^e siècle, la proportion d'originaux pour les bulles au sein des

⁸² Soit 28 originaux contre 2 427 copies.

⁸³ Soit 4 923 originaux contre 22 912 copies.

⁸⁴ Soit 2 361 originaux contre 16 938 copies.

⁸⁵ Soit 29 635 originaux contre 118 545 copies.

⁸⁶ Voir toutefois nos remarques ci-dessous, en particulier en ce qui concerne la nature du support, le nombre et la taille des documents, qui entraînent des difficultés variables de conservation.

⁸⁷ Bresslau, Harry: *Papyrus und Pergament in der päpstlichen Kanzlei bis zur Mitte des 11. Jahrhunderts*. Ein Beitrag zur Lehre von den älteren Papsturkunden. Dans: *Mitteilungen des Instituts für Österreichische Geschichtsforschung* 9 (1888), p. 1-33 ; Omont, Henri: *Bulles pontificales sur papyrus (IX^e-XI^e siècle)*. Dans: *Bibliothèque de l'École des chartes* 65 (1904), p. 575-582 ; Bello, Carme, et Àngels Borrell: «The Papal Bulls on Papyrus: an Approach to their Conservation». *Imago temporis: medium Aevum* 7 (2013), p. 377-400, qui est centré sur les difficultés de conservation et de restauration de ces documents.

⁸⁸ Outre la fragilité du support, Henri Omont souligne à juste titre que la taille souvent exceptionnelle des documents fut probablement un facteur aggravant de leur mauvaise conservation.

CEMA est encore de 8,9%, contre 20,8% pour les autres types documentaires. L'hypothèse est donc intéressante, mais pas forcément de façon isolée.

Il y a ensuite la question de la place des lettres face aux privilèges, comme « deux mondes distincts qui se rencontrent sans se confondre », pour reprendre les mots de Laurent Morelle⁸⁹. Là encore, s'il ne fait aucun doute que la sélection très large des documents entrant dans le corpus influence nos résultats, il nous semble exagéré de dire qu'elle le conditionne entièrement. Le cas du XI^e siècle, à nouveau, est assez édifiant : alors que les chartes pontificales s'y développent fortement, le taux d'originaux demeure encore en deçà de celui obtenu sur les autres typologies diplomatiques⁹⁰. Ainsi, si ces mécanismes structurels (support, quantité, typologie) ne sont pas à exclure dans {64} l'analyse, il nous semble que l'évolution observée reste valable, en premier lieu parce que le corpus réuni est assez vaste.

Elle est en outre en partie corroborée par les observations réalisées par Harald Müller, concernant la position des bulles dans les cartulaires français aux XI^e-XIII^e siècles⁹¹. Alors que les documents pontificaux ne sont pas mis en avant dans la plupart des manuscrits anciens qu'il examine, contrairement aux actes provenant d'autres autorités, ceux-ci prennent progressivement une plus grande importance – jusqu'à parfois former la tête de certains *codices* en fin de période. Ce changement est strictement parallèle à nos observations et pourrait signifier qu'un nouveau regard a progressivement été porté sur ces textes et donc sur l'autorité des papes, qui s'affirment progressivement comme la tête et le centre de l'Église. Ainsi, le faible nombre des originaux avant la seconde moitié du XI^e siècle, mais aussi l'explosion du nombre des actes (copies comprises) après cette période, pourrait signifier qu'un changement profond s'est opéré – non pas tant dans la conservation des actes, que dans le rapport aux papes, de la part des bénéficiaires. Une hypothèse est que dans le cadre de l'évolution radicale des rapports sociaux que connaît l'Europe médiévale au cours des X^e-XIII^e siècles, où se développe considérablement le maillage topo-spatial du système et l'articulation de ses différentes échelles (donc des niveaux d'encadrement différents), le pape a

⁸⁹ Morelle, Laurent: Tournures épistolaires et usages de la lettre. L'écriture administrative de Lambert évêque d'Arras (1093/1094-1115). Dans: Sirantoine: La lettre diplomatique (voir n. 5), p. 99-114, ici 114.

⁹⁰ Sur une échelle plus limitée, Harald Müller fait des observations similaires, dans *Überlieferungsformen franko-römischer Kontakte: Zur Position der Papsturkunden in französischen Chartularen*. Dans: Klaus, Herbers et Ingo Fleisch (Dir.): Erinnerung - Niederschrift - Nutzung : Das Papsttum und die Schriftlichkeit im westeuropäischen Mittelalter. Berlin 2011, p. 195-217, ici p. 196. Pour l'auteur, la raison de ces pertes pourrait être aussi pragmatique : les changements rapides de pontifes auraient entraîné un désintérêt pour les actes plus anciens ; par ailleurs, Harald Müller mentionne que la conservation de ces actes n'était plus indispensable après l'« achèvement » (*Erledigung*) de l'affaire concernée par le document. L'hypothèse est intéressante, mais elle n'explique pas pourquoi les originaux semblent mieux conservés à partir des XI^e-XII^e siècles. Sur cette potentielle valeur transitoire des lettres et la place des documents pontificaux dans les cartulaires provençaux, voir Renault, Jean-Baptiste: La rédaction de cartulaires et la réception des documents pontificaux en Provence (fin XI^e-début XII^e siècle) : quelles interactions ?. Dans: Francia 45 (2018), p. 61-84. Ce dernier indique, à l'instar d'Harald Müller, que l'obtention de privilèges pontificaux ne joue pas systématiquement un rôle dans la forme et la rédaction des cartulaires, du moins avant le milieu du XII^e siècle.

⁹¹ Voir encore une fois l'intéressant article de Müller: *Überlieferungsformen* (voir n. 90).

radicalement gagné en importance, y compris au niveau local, en tant que référence ultime du système à son échelle la plus large⁹².

II.2. Mesurer la richesse lexicale

La richesse documentaire ne peut toutefois se résumer à la quantité des textes conservés. Une seconde manière d'aborder ce corpus est ainsi d'étudier son contenu, et tout d'abord sa richesse lexicale. Entre complexité et stéréotypicalité⁹³, il semblait intéressant de savoir à quel degré les documents pontificaux étaient riches d'un vocabulaire varié. Pour ce faire, un sous-ensemble documentaire a été sélectionné, allant de la fin du pon-**{65}**-tificat de Grégoire le Grand au début de celui d'Innocent III. Ce choix chronologique est lié à la nature des corpus des deux papes, dont la densité induit des biais statistiques qu'il est difficile de contrôler. Sur les 11 000 actes sélectionnés, entre 604 et 1198, il a été possible de mesurer la richesse du lexique lemmatisé, d'abord par demi-siècle, puis pour une vingtaine de pontificats dont la densité documentaire le permettait.

Mesurer la richesse lexicale de corpus, autrement dit la variété du vocabulaire qu'ils contiennent, n'est toutefois pas chose aisée. En premier lieu car ces derniers sont de taille variable en fonction des époques⁹⁴. Les travaux du linguiste George Kingsley Zipf et du mathématicien Benoît Mandelbrot ont montré qu'il était possible de mesurer la richesse des textes, tout en tenant compte de leur longueur variable⁹⁵. Pour ce faire, il s'agit de compter le nombre de lemmes uniques dans un

⁹² Nous nous permettons de renvoyer ici à Perreaux, Nicolas: Des structures domaniales aux territoires ecclésiiaux ? Entités spatiales et dynamique du processus de spatialisation dans les actes diplomatiques (VII^e-XIII^e siècles). Dans: Martine, Tristan ; Nowak, Jessika et Schneider, Jens (Dir.): Espaces ecclésiastiques et seigneuries laïques. Définitions, modèles et conflits en zones d'interface (IX^e-XIII^e siècle). Paris 2021, p. 55-78.

⁹³ Outre la stéréotypicalité de certaines formules pontificales déjà évoquées, l'influence massive du *Liber Diurnus* est longuement discutée dans l'historiographie. Voir Von Hans Foerster (Éd.): *Liber diurnus romanorum pontificum*. Bern 1958 ; Santifaller, Leo: *Liber Diurnus. Studien und Forschungen* (Éd. Harald Zimmermann). Stuttgart 1976 ; Frenz: *Papsturkunden* (voir n. 2), p. 50-51, indique quant à lui que le formulaire a été identifié dans 104 documents entre 595 et 1071, avec 6 reprises exactes. Ce chiffre est sans doute en deçà de la réalité, car il y a plus de 2 500 documents pontificaux dans les CEMA pour cette période. Un examen systématique des usages du formulaire, à l'aide d'outils numériques, serait donc à réaliser. On constate toutefois que les chiffres sont relativement faibles, ce qui nous laisse penser que le *Liber Diurnus* était plutôt un guide qu'un modèle, et que la normalisation de l'écrit pontifical est effectivement un phénomène plus tardif.

⁹⁴ La richesse lexicale d'un texte dépend bien entendu et en premier lieu de sa taille. Toutefois, l'accroissement du nombre de mots (ou lemmes) uniques dans un texte donné n'est pas une variable linéaire de la taille dudit document. Les textes suivent en effet des distributions statistiques de type LNRE (pour Large Number of Rare Events, l'appellation est proposée par Harald Baayen). C'est entre autres cette particularité qui rend délicate la mesure de la richesse lexicale. Voir en premier lieu : Baayen, Harald: *Word Frequency Distributions*. Dordrecht 2001 ; Evert, Stefan: *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. Stuttgart 2005.

⁹⁵ Cette méthode a déjà été appliquée dans Perreaux, Nicolas: *Langue des capitulaires et langue des chartes : richesses, circulations, spécificités*. Dans: Jussen, Bernhard et Ubl, Karl (Dir.): *Die Sprache des Rechts. Karolingische Kapitularien und historische Semantik*. A paraître (2021). On trouvera les fondements théoriques dans Zipf, George Kingsley: *Human Behaviour and the Principle of Least Effort. An Introduction to Human Ecology*. Cambridge 1949 ; Mandelbrot, Benoît: *Les objets fractals. Forme, hasard et dimension*. Paris 1975 ; Barbut, Marc: Note sur l'ajustement des distributions de Zipf-Mandelbrot en statistique textuelle. Dans: *Histoire & Mesure* 4 (1989), p. 107-119. Pour une application concrète des lois de Zipf, voir Guiraud, Pierre: *Problèmes et méthodes de la statistique linguistique*. Dordrecht 1959.

corpus donné (dans les figures qui suivent, cela correspond à l'axe des ordonnées), en le faisant lire du début à la fin par un programme (axe des abscisses). Les courbes ainsi obtenues représentent la progression de la variété lexicale au sein d'un corpus donné, à mesure qu'on le lit. Ainsi, plus ces courbes augmentent rapidement, plus le corpus examiné possède un vocabulaire riche⁹⁶. {66}

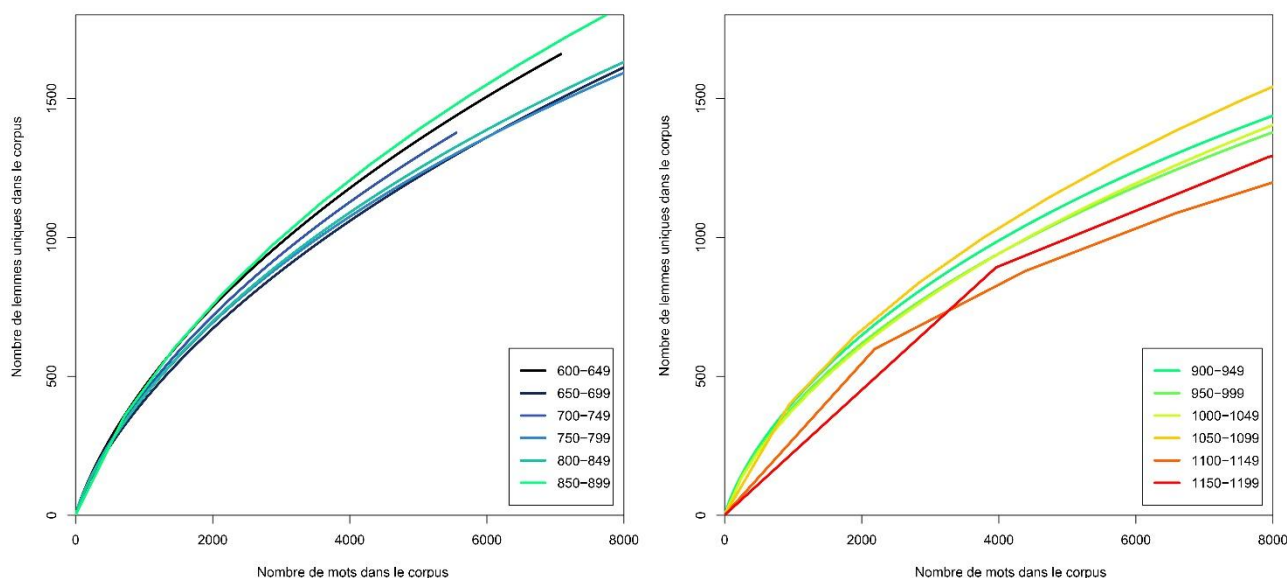


Fig. 6a et 6b : Évolution de la richesse du lexique des documents pontificaux (lettres et privilèges), par demi-siècle. Plus la courbe est haute, plus le lexique (lemmes) est varié (ZipfR / Cooc).

En procédant de la sorte pour les documents pontificaux réunis par demi-siècles (fig. 6a et 6b), on réalise que la richesse lexicale des documents est plus grande dans le haut Moyen Âge (à gauche) que pour n'importe quelle autre période examinée. Cette première observation semble en parfaite contradiction avec l'idée d'une très grande stéréotypicalité des actes de cette période, imputable en partie au *Liber Diurnus*⁹⁷. L'observation doit toutefois être envisagée à l'aune des transformations typologiques du corpus : essentiellement composé de lettres jusqu'au X^e siècle, il évolue ensuite, comme nous l'avons vu, vers un ensemble incluant au moins pour moitié des

⁹⁶ Evert, Stefan, et Marco Baroni: ZipfR: Word frequency distributions. Dans: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics. Prague 2007 ; Baayen, Harald: Analyzing Linguistic Data. A Practical Introduction to Statistics using R. Cambridge 2008, p. 165-240. Ces analyses n'auraient toutefois pas été possibles sans les bibliothèques de fonctions pour R développées par Alain Guerreau, en particulier Cooc. Celle-ci permet de manipuler statistiquement les corpus documentaires latins, en surajoutant une série d'algorithmes interactifs à l'ensemble formé par CWB et RCQP (une bibliothèque R destinée à l'intégration de CWB dans celui-ci). RCQP est développé par deux chercheurs, Bernard Desgraupes et Sylvain Loiseau : <https://cran.r-project.org/web/packages/rcqp/rcqp.pdf> (consulté le 23.11.2020).

⁹⁷ Ce qui ne signifie pas, bien entendu, que le *Liber Diurnus* ne joue pas de rôle dans la rédaction des bulles. Simplement que son usage n'entraîne pas la stéréotypicalité généralement présentée par l'historiographie pour ces périodes. D'ailleurs, le formulaire (et ses variantes) est assez important pour ne pas standardiser les actes.

privilèges. Cette évolution influence nécessairement notre analyse. Si l'on en reste pour le moment à ce niveau général, on observe le schéma suivant : parfois élevée dès le VII^e siècle, la variété lexicale des documents pontificaux diminue légèrement au VIII^e siècle et dans la première moitié du IX^e siècle. Elle augmente de nouveau nettement après 850 (cf. fig. 6a, la courbe bleu-vert, la plus haute, à gauche) – période où les lettres composent plus de 86% du corpus. Ces premières fluctuations ne peuvent être imputées à l'évolution du ratio lettres-privilèges, puisque du VII^e au IX^e siècle (inclus), il se maintient de façon stable entre 84% et 91%. Les X^e et XI^e siècles voient quant à eux s'opérer une réduction globale de la richesse des documents pontificaux (fig. 6b), même si cette chute est contrebalancée par sursaut dans la seconde moitié du XI^e siècle – sans toutefois que la variété lexicale revienne aux niveaux observables au IX^e siècle. Là encore, si l'évolution lexicale du début {67} du Xe siècle pourrait en partie s'expliquer par des changements typologiques (on passe d'environ 85% de lettres au IX^e siècle, à seulement 38% au X^e siècle), les oscillations après cette période ne peuvent se réduire à ce phénomène. En effet, le ratio de privilèges au sein du corpus est d'environ 63% dans la première moitié du X^e siècle, contre 61% dans la seconde moitié. Ce taux est par ailleurs relativement stable dans les périodes postérieures, avec 57% au XI^e siècle, 59% au XII^e siècle et 58% au XIII^e siècle. Dans ces conditions, les évolutions de la richesse lexicale au cours des X^e-XIII^e siècles ne peuvent s'expliquer par la seule transformation typologique du corpus, en tout cas pas par l'évolution du ratio entre lettres et privilèges. Ainsi, il s'opère au XII^e siècle à nouveau un très net recul en matière de variété lexicale. Les documents de la première moitié du XII^e siècle sont environ un tiers moins variés lexicalement que ceux de la première moitié du XI^e siècle. Cette réduction n'est toutefois pas linéaire, puisque certains pontificats correspondent à un sursaut de variété lexicale, ainsi que le montrent des analyses plus précises⁹⁸. {68}

⁹⁸ A l'inverse, les deux moitiés du XII^e siècle connaissent des ruptures à la baisse, bien visibles sur la fig. 6b. Alors que l'on progresse chronologiquement, on observe en effet une forme de « décrochage lexical », bien visible sur les deux courbes correspondantes.

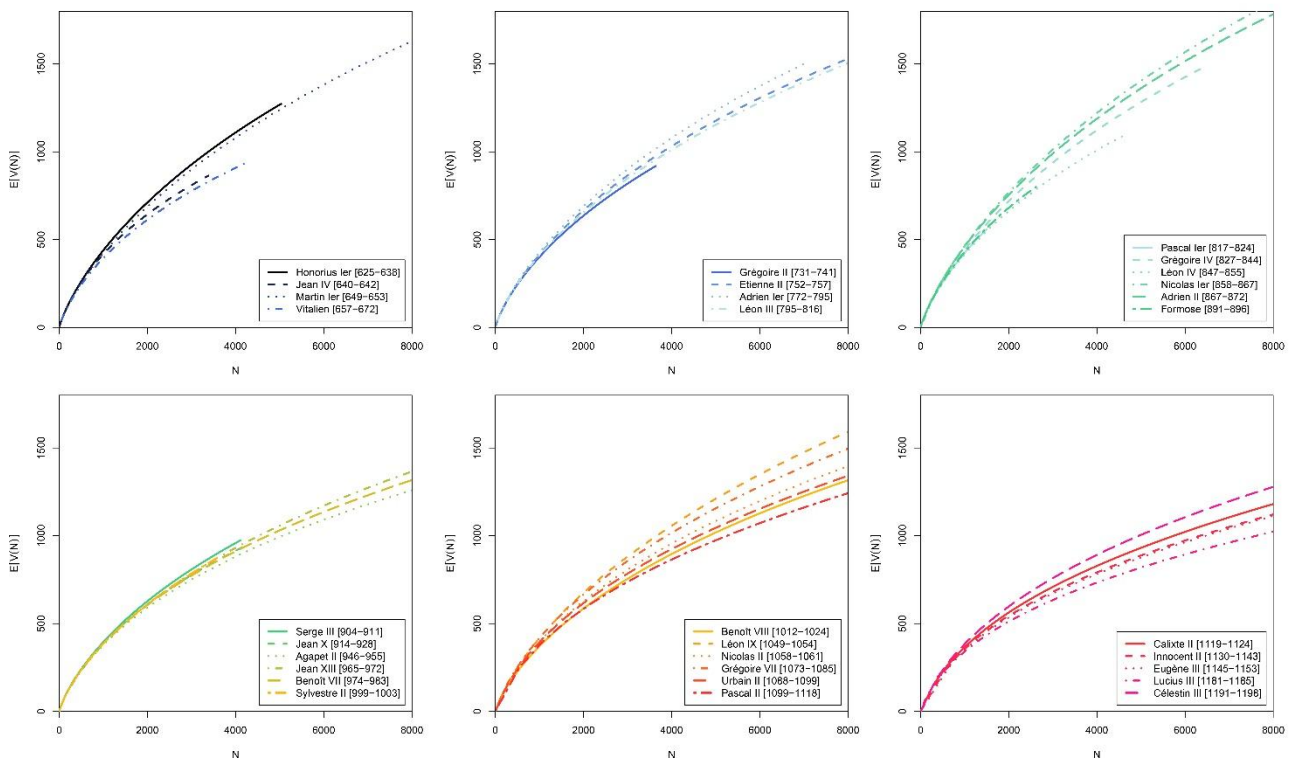


Fig. 7 : Évolution de la richesse du lexique des documents pontificaux (lettres et privilèges), par pape (VII^e-XII^e siècles). Plus la courbe est haute, plus le lexique (lemmes) se révèle être varié (ZipfR / Cooc).

Les résultats précédents invitaient à poursuivre cette exploration de la variété lexicale des actes, cette fois à partir de quelques pontificats pour lesquels il était possible de réunir une documentation suffisante, soit entre quatre et six pour chaque siècle (fig. 7). Ils permettent de préciser les observations précédentes. Au VII^e siècle, la richesse lexicale est plutôt élevée, mais elle semble aussi très contrastée d'un pontife à l'autre, par exemple entre Honorius I^{er} et Martin I^{er}. La rareté documentaire invite certes à considérer certaines de ces mesures avec prudence⁹⁹. Mais pour des corpus comportant une vingtaine de textes, comme ceux d'Honorius I^{er} ou de Martin I^{er}, on constate que la variété des lemmes employés est plutôt élevée. Au VIII^e siècle, cette tendance scripturaire s'affirme, la richesse lexicale devenant plus homogène d'un pontificat à l'autre. Pour les quatre papes retenus dans ce siècle (Grégoire II, Étienne II, Adrien I^{er} et Léon III), les mesures sont non seulement

⁹⁹ Parmi les pontifes retenus pour ce siècle, Jean IV est certainement le plus problématique (voir Morelle, Laurent: La liberté de Luxeuil et son expression diplomatique : à propos d'une charte épiscopale absente et d'un privilège pontifical encombrant (Jean IV, 640-642). Dans: Sébastien Bully, Alain Dubreucq et Aurélie Bully (Dir.), *Colomban et son influence : moines et monastères du haut Moyen Âge en Europe*. Rennes 2018, p. 239-260). Plusieurs des actes qui lui sont attribués sont en effet des faux (par exemple les actes en faveur de l'abbaye de Remiremont, édités dans Bridot, Jean (Éd.): *Chartes de l'abbaye de Remiremont des origines à 1231*. Nancy-Turnhout 1997, n° 2 et 3, le dernier correspondant document à l'Artem n° 2478 ; ou encore les documents examinés par Laurent Morelle). Le corpus est ainsi extrêmement limité pour ce pontife, avec 5 documents seulement. En revanche, Honorius I^{er}, Martin I^{er} et Vitalien offrent des corpus plus substantiels (en particulier de lettres pastorales pour Martin I^{er}), qui confirment d'une part le contraste qui prévaut dans les usages lexicaux à cette période, d'autre part la grande richesse relative des corpus conservés.

cohérentes mais aussi élevées¹⁰⁰. Au IX^e siècle, d'importantes disparités réapparaissent d'un pontife à l'autre. Si les actes de Léon IV sont peu variés lexicalement¹⁰¹, ceux de Nicolas I^{er} ou de Adrien II correspondent en revanche à la plus haute richesse rencontrée dans l'ensemble du corpus examiné, toutes périodes {69} confondues. Le X^e siècle est ensuite très cohérent d'un pontificat à l'autre, avec peu de variation dans la richesse lexicale, mais il correspond aussi à une chute de cette richesse (environ 25% par rapport au siècle précédent). Alors que six pontifes ont été retenus pour cette période (de Serge III à Sylvestre II), on note que les scores obtenus sont très homogènes. Cette évolution était déjà visible lors de la lecture par siècle (fig. 6b). La tendance se poursuit au XI^e siècle, mais de façon nuancée : la chute de la richesse lexicale est alors contrebalancée par l'existence d'une forte variation entre les pontifes – ce qui oppose, sur ce dernier plan, ce siècle et le précédent. Le lexique se fait en effet beaucoup plus varié au milieu du XI^e siècle, sous Léon IX, puis Nicolas II et Grégoire VII, avant de se concentrer à nouveau dans son dernier tiers, sous Urbain II et surtout Pascal II¹⁰². Le XII^e siècle, enfin, entérine la tendance à la baisse déjà évoquée, même si une certaine variabilité existe, comme au siècle précédent¹⁰³.

II.3. *Interprétation et comparaisons*

Comment expliquer ces changements ? Différentes hypothèses de travail peuvent être proposées, les principales étant liées au degré de standardisation des formulaires et des modèles pontificaux (à travers l'usage, ou non, des préambules et des clauses comminatoires, du *cursus*, de créations conceptuelles, etc.), mais aussi l'homogénéité des sujets traités par les documents. Comme souvent, les niveaux lexicaux, sémantiques et sociaux s'entrecroisent dans un enchevêtrement impossible à démêler. Distinguer ces strates paraît d'autant plus illusoire qu'elles forment en fait un tout, à la fois articulé et indissociable¹⁰⁴. Énoncé ainsi, le problème revient donc à évoquer l'hypothèse d'une extension, puis d'une formalisation progressive de l'institution pontificale, dont le lexique serait le révélateur¹⁰⁵. Il nous semble que c'est cela que les graphiques permettent d'observer : les thèmes évoqués par les papes (ou ceux sur lesquels on les sollicite, du moins à partir d'une période

¹⁰⁰ Après la lecture de 8 000 mots dans leurs corpus respectifs, les papes Étienne II, Adrien I^{er} et Léon III dépassent les 1 400 lemmes uniques.

¹⁰¹ Le corpus est par ailleurs très limité pour ce pontife.

¹⁰² Concernant le lexique d'Urbain II, nous renvoyons à Becker, Alfons: *Papst Urban II : 1088-1099*. Stuttgart 1964-2012, en particulier les tomes 2 et 3.

¹⁰³ Les deux extrêmes repérés étant pourtant Lucius III et son successeur, Célestin III.

¹⁰⁴ L'exemple classique est celui de Klemperer, Victor: *LTI, la langue du III^e Reich*. Paris 1996 (première édition Leipzig 1975).

¹⁰⁵ Voir les remarques concordantes de Grévin, Benoît: *Les mystères rhétoriques de l'État médiéval. L'écriture du pouvoir en Europe occidentale (XIII^e-XV^e siècle)*. Dans: *Annales HSS* 63^e année (2008), p. 271-300, autour de la standardisation de l'*ars dictaminis* au XIII^e siècle.

avancée¹⁰⁶) et les réponses que l'institution fournit sont en effet plus ou moins standardisées, plus ou moins focalisées, en fonction des chronologies¹⁰⁷.

{70} On aurait ainsi une grande variabilité (et peut-être même une incertitude quant au rôle des papes) aux VII^e-VIII^e siècles, suivie d'une implication plus grande des pontifes dans différents domaines au IX^e siècle. La tendance lexicale suit donc relativement bien ce que nous savons du rôle concret et du rôle perçu des pontifes pour ces siècles, avec des reculs et des contradictions, mais aussi une évolution globale vers une plus grande importance des papes – qui s'affirme très nettement à l'époque carolingienne. Les X^e-XII^e siècles seraient quant à eux des périodes de progressive régulation des thèmes traités par l'institution¹⁰⁸, ce qui n'empêche pas l'intégration de nouveaux sujets (et donc une plus grande variabilité lexicale) au cours de certains pontificats, en particulier au XI^e siècle. Là encore, on ne peut s'empêcher de trouver des correspondances socio-historiques à nos observations : le développement du rôle et de la centralité du pontife se poursuit au cours de ces siècles, avec toutefois des tendances plus nettes à l'auto-affirmation sous certains pontificats, par exemple au milieu du XI^e siècle. Ces hypothèses socio-sémantiques n'excluent pas, bien entendu, l'existence de facteurs linguistiques et diplomatiques¹⁰⁹. Ces multiples dimensions se combinent et forment un ensemble cohérent : c'est parce que l'institution et le regard que l'on porte sur elle évolue, que son vocabulaire et son « style » se transforment. Une autre hypothèse explicative consisterait à énoncer que les lettres conservées pour le haut Moyen Âge le furent parce qu'elles étaient exceptionnelles : elles auraient ainsi paru singulières, importantes, aux yeux des récepteurs/bénéficiaires. C'est cela qui expliquerait la richesse des thèmes, et donc du lexique, dans les plus anciens documents conservés pour les pontifes. Cette lecture n'est pas à exclure, mais, par essence, sa vérifiabilité est faible. Une hypothèse plus probable est celle d'une « bureaucratisation » de l'institution pontificale, concomitante avec une évolution du rôle et du regard porté sur celle-ci¹¹⁰.

¹⁰⁶ Sur ces questions, nous renvoyons aux travaux de Jochen Johrendt, en particulier *Papsttum und Landeskirchen im Spiegel der päpstlichen Urkunden* (896-1046). Hanovre 2004 ; ainsi qu'à Johrendt, Jochen et Harald Müller (Dir.): *Römisches Zentrum und kirchliche Peripherie: das universale Papsttum als Bezugspunkt der Kirchen von den Reformpäpsten bis zu Innozenz III.* Berlin 2009.

¹⁰⁷ Si les ouvrages de synthèses sont précieux (outre les références déjà mentionnées note 66, dont Herbers, Frenz et Tanase, ainsi que Knowles, David et Obolensky, Dmitrij Dmitrievič (Éd.): *Nouvelle histoire de l'Église. Tome 2 : Le Moyen Âge.* Paris 1968 ; Hilaire, Yves-Marie (Dir.): *Histoire de la papauté. 2000 ans de mission et de tribulations.* Paris 1996), ils se focalisent essentiellement sur les querelles temporelles autour des pontifes, ainsi que leurs relations avec les autres dominants de leurs époques (évêques, empereurs, rois, princes, abbés, etc.). On trouve ainsi une multitude d'histoires ou dictionnaires de la papauté, qui résument de façon plus ou moins efficaces des séries d'événements. Si ces textes ont leur intérêt, ils n'éclaircissent certainement pas le sens théologique et sociologique du pape, considéré comme la tête du système ecclésial. On verra cependant Ullmann, Walter: *Growth of Papal Government in the Middle Ages. A Study in the Ideological Relation of Clerical to Lay Power.* Londres 1962 ; Congar, Yves: *Le pape patriarche d'Occident* (voir n. 12) ; D'Alberto: *Papauté* (voir n. 9) ; Demacopoulos.: *The Invention of Peter* (voir n. 10) ; Paravicini Bagliani, Agostino: *Le corps du pape.* Paris 1997 ; Id.: *Il potere del papa. Corporeità, autorappresentazione, simboli.* Firenze 2009.

¹⁰⁸ Fossier, Arnaud: *L'exemple ou la norme ? De l'art d'administrer par formulaire (XIII^e-XIV^e siècles).* Dans: Giavarini, Laurence (Dir.): *Pouvoir des formes, écriture des normes.* Dijon 2017, p. 21-40.

¹⁰⁹ En particulier, il faudrait creuser la question des usages scribaux et de la formation de ces derniers.

¹¹⁰ Sur le paradigme de la « bureaucratisation », voir Fossier, Arnaud ; Petitjean, Johann et Revest, Clémence: *Introduction. Pour une histoire pratique de l'administration.* Dans: Fossier, Petitjean et Revest (Dir.): *Écritures grises* (voir

Cette standardisation documentaire, partiellement liée à l'évolution typologique, est sans doute ce qui explique le mieux ces variations dans la richesse des textes pontificaux. C'est parce que le pape est régulièrement consulté, que lui et sa chancellerie produisent des documents fortement standardisés, autour de thèmes récurrents – et c'est aussi, de façon circulaire, parce qu'ils sont perçus comme des références définitives, qu'ils sont consultés, dans une société où la hiérarchisation ecclésiastique s'est fortement accentuée et où l'articulation verticale des clercs s'est considérablement développée. {72}

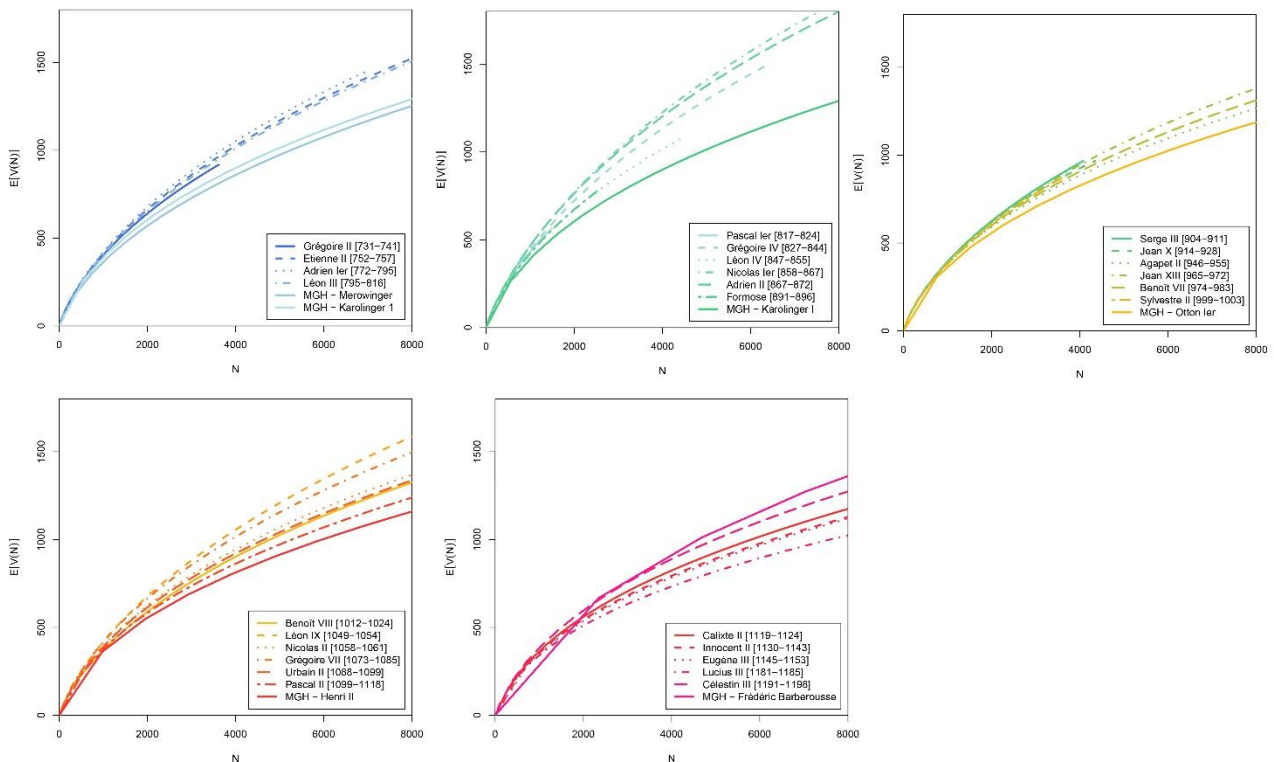


Fig. 8 : Comparaison de l'évolution de la richesse du lexique des documents pontificaux (lettres et privilèges) et des diplômes royaux/impériaux (VIII^e-XII^e siècles). Plus la courbe est haute, plus le lexique (lemmes) se révèle être varié (ZipfR / Cooc).

Ces hypothèses d'un lien étroit entre richesse lexicale et variété des thèmes abordés par l'institution peuvent par ailleurs être étayées par une série d'expériences complémentaires (fig. 8). La

note 49), p. 8-11 ; Fossier: Innocent III (voir note 26) ; ainsi que, dans le même volume, Bertrand, Paul et Chastang, Pierre: Les temps des écritures grises. Formation et temporalités du gouvernement par l'écrit (v. 1080-v. 1350), p. 29-65. Il est bien entendu directement lié à l'évolution des pratiques documentaires au Moyen Âge central : Bertrand: Les écritures ordinaires (voir note 16) ; Maire-Vigueur, Jean-Claude: Révolution documentaire et révolution scripturaire : le cas de l'Italie médiévale. Dans: Bibliothèque de l'École des chartes 153 (1995), p. 177-185. Plus généralement, il s'agirait de comprendre comme cette « bureaucratisation » de l'institution préfigure la dynamique scripturaire de toute l'Europe en la matière. Sur ces questions, outre les références mentionnées, nous renvoyons au classique de Clanchy, Michael T.: From Memory to Written Record. England 1066-1307. London 1979.

comparaison des graphiques précédents à ceux obtenus sur les corpus des diplômes royaux et impériaux offrent en effet différentes pistes de réflexion¹¹¹. Nous avons ainsi retenu une série d'éditions issues des MGH : les actes authentiques des mérovingiens, puis ceux de Pépin et de Charlemagne, d'Otton Ier, d'Henri II et enfin de Frédéric Barberousse¹¹². Ces points de comparaison permettent d'observer un phénomène intéressant : jusqu'au XI^e siècle inclus, les documents pontificaux sont systématiquement plus riches lexicalement que les diplômes. Parallèlement, les diplômes sont d'une richesse moins variable d'un souverain à l'autre, depuis les rois mérovingiens à Henri II, malgré une augmentation de cette variété entre la fin du VIII^e et le milieu du IX^e siècle¹¹³. Cela n'exclut pas, bien entendu, d'importants changements lexicaux dans ce dernier corpus, mais indique simplement qu'un degré de standardisation plus élevé se maintient dans les actes de ces souverains, pourtant fort éloignés. Mais avec Frédéric I^{er}, la richesse lexicale des actes impériaux dépasse largement celle des papes. Ce phénomène montre au moins une chose : que la diminution de la variété lexicale des documents pontificaux n'est pas seulement imputable à un mouvement général et uniforme de simplification des actes diplomatiques, qui serait réductible à l'éviction des préambules et des clauses comminatoires, ou encore à l'usage, ou non, de tel ou tel formulaire.

{73} III. Évolutions et ruptures au sein du lexique pontifical

III.1. *Lettres et privilèges (VII^e-XII^e siècles)*

Comment aller plus loin et cerner les tendances, les ruptures et les rythmes dans l'écriture pontificale ? La littérature propose différents éléments de réponse, mais souvent à partir d'un point de vue précis, lié par exemple à un caractère externe, à une formule ou encore à un formulaire ponctuel¹¹⁴. La méthode proposée ici consiste à analyser globalement le lexique lemmatisé des actes,

¹¹¹ Sur les circulations entre les actes royaux et pontificaux, nous renvoyons, pour la fin du Moyen Âge, à l'important travail de Grévin, Benoît: *Rhétorique du pouvoir médiéval. Les Lettres de Pierre de la Vigne et la formation du langage politique européen (XIII^e-XV^e siècles)*. Rome 2008. Pour la circulation entre les documents pontificaux et épiscopaux, on pense à l'étude Guyotjeannin, Olivier: *Traces d'influence pontificale dans les actes épiscopaux et royaux français (XIII^e-XV^e siècle)*. Dans: Herde, Peter et Jakobs, Hermann (Éd.): *Papsturkunde eine europäisches Urkundenwesen : Studien zu ihrer formalen und rechtlichen Kohärenz vom 11. bis 15. Jahrhundert (Archiv für Diplomatik, Schriftgeschichte, Siegel – und Wappenkunde, Beihefte 7)*. Cologne-Vienne-Weimar 1999, p. 337-362 ; Id.: *Écrire en chancellerie*. Dans: Zimmermann, Michel (Dir.): *Auctor et auctoritas. Invention et conformisme dans l'écriture médiévale*. Paris 2001, p. 17-35.

¹¹² Louis le Pieux aurait aussi été un bon candidat pour ces comparaisons. Malheureusement, nous ne disposons pas de l'édition récente des MGH dans les CEMA.

¹¹³ Voir nos observations dans Perreaux: *Langue des capitulaires* (voir note 95). Cette hausse est toutefois beaucoup moins importante que les variations observées au sein du corpus pontifical, ce qui tient, à cette échelle, à minimiser les évolutions dans les diplômes (qui est cependant réelle).

¹¹⁴ Parmi les éléments de rupture les plus souvent évoqués, on trouve en particulier la « Réforme », avec les pontificats de Léon IX (cf. Frenz: *Papsturkunden* (voir n. 2), p. 6) et de Grégoire VII, mais aussi l'avènement des registres systématiquement conservés, avec Innocent III. Sur le *Liber Diurnus*, voir les références mentionnées note 93.

afin d'observer les tendances scripturaires de la chancellerie pontificale. En pratique, elle consiste à extraire une liste d'environ 20 000 bi-lemmes du corpus¹¹⁵, puis à calculer la fréquence de ceux-ci pour les textes de chaque pontificat possédant un nombre suffisant d'actes (au moins 5), entre la mort de Grégoire le Grand et l'avènement d'Innocent III¹¹⁶. L'analyse factorielle des correspondances (AFC) produite à partir du tableau permet ainsi de mesurer l'évolution lexicale d'un pontife à l'autre, du début du VII^e à la fin du XII^e siècle¹¹⁷. Sa lecture est {74} relativement simple : plus le lexique des actes de deux papes présente de points communs, plus ces papes figurent à proximité l'un de l'autre sur le plan factoriel ; plus ce lexique est éloigné, plus ils s'éloignent sur l'AFC. Pour cette première expérience, nous avons fait le choix de retenir la totalité du corpus des documents pontificaux, autrement dit aussi bien les lettres que les privilèges.

Concernant Grégoire VII, voir Hoffmann, Hartmut: Zum Register und zu den Briefen Papst Gregors VII. Dans: Deutsches Archiv für Erforschung des Mittelalters 32 (1976), p. 86-130.

¹¹⁵ Autrement dit des groupes de deux mots lemmatisés, par exemple « *et in* », « *notum sit* », « *actum est* », etc.

¹¹⁶ La limite de cinq actes nous a semblé être minimale pour faire des calculs. Du reste, entre cinq et dix actes par pontife, les hypothèses restent nécessairement sujettes à caution, car elles sont alors très liées aux spécificités des rares documents conservés. Au-delà de vingt bulles, on peut considérer que les analyses sont relativement robustes. La liste des papes pour lesquels nous avons réuni plus de dix textes entre le VII^e et le XII^e siècle est la suivante, le nombre d'actes étant indiqué entre parenthèses : Honorius I (19), Jean IV (7), Martin I (20), Vitalien (13), Agathon (6), Léon II (6), Constantin (5), Grégoire III (9), Zacharie (7), Étienne II (19), Paul I^{er} (38), Étienne III (6), Adrien I^{er} (65), Léon III (38), Grégoire IV (8), Léon IV (45), Benoît III (6), Nicolas I (186), Adrien II (41), Jean VIII (412), Marin I^{er} (6), Étienne V (40), Formose (10), Étienne VI (5), Jean IX (7), Serge III (10), Jean X (16), Jean XI (10), Léon VII (14), Marin II (13), Agapet II (27), Jean XII (25), Jean XIII (41), Benoît VI (7), Benoît VII (42), Jean XV (29), Grégoire V (30), Sylvestre II (20), Jean XVIII (21), Serge IV (22), Benoît VIII (66), Jean XIX (32), Benoît IX (14), Grégoire VI (5), Clément II (12), Léon IX (160), Victor II (25), Étienne IX (11), Nicolas II (53), Grégoire VII (179), Clément III (antipape) (11), Urbain II (465), Gélase II (40), Calixte II (423), Honorius II (180), Pascal II (860), Anaclet II (antipape) (53), Innocent II (902), Célestin II (62), Lucius II (30), Eugène III (1005), Anastase IV (141), Adrien IV (485), Victor IV (antipape) (30), Alexandre II (246), Calixte III (antipape) (9), Alexandre III (2706), Lucius III (787), Grégoire VIII (40), Urbain III (339), Clément III (330), Célestin III (434). Ces chiffres peuvent intégrer certains faux, soit connus soit ignorés de l'historiographie, lorsque l'information n'était pas disponible dans les CEMA (c'est en revanche le cas pour l'Artem ou *Diplomata Belgica* – même si ces indications concernant l'authenticité sont parfois sujettes à caution). L'intégration systématique des métadonnées concernant l'authenticité des actes inclus au sein de la base est une perspective essentielle, mais de longue haleine. Du reste, nous pensons qu'avec une telle quantité d'actes, cela ne change qu'à la marge les résultats – la liste ci-dessus permettant précisément d'être attentif à ce type de biais, en particulier lorsque l'effectif de textes pour un pontife est faible.

¹¹⁷ Très singulier par sa taille et en partie par son contenu, le corpus de Grégoire le Grand a été volontairement exclu des analyses qui suivent.

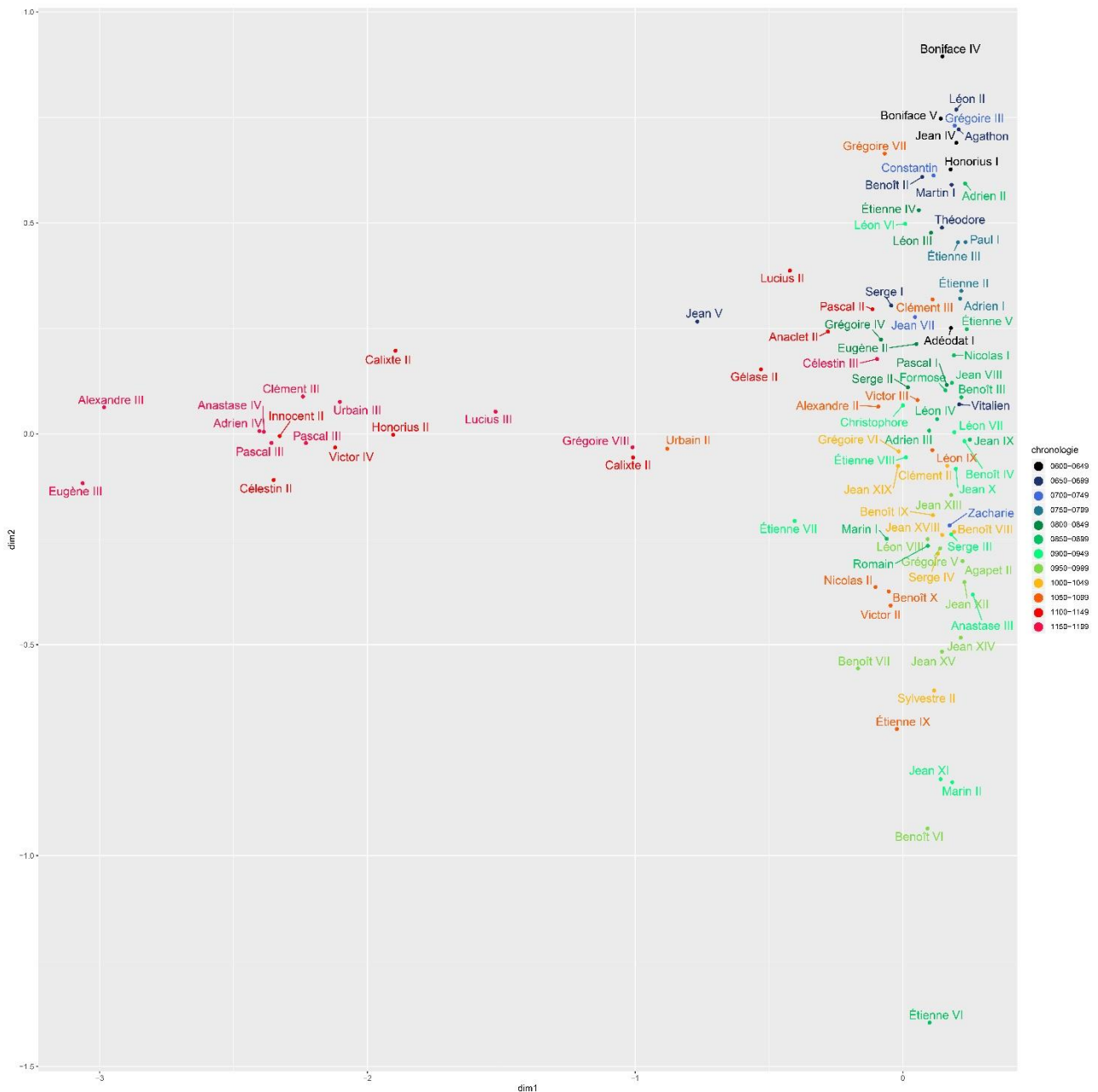


Fig. 9 : Analyse factorielle de l'ensemble du lexique pontifical (20 000 bi-lemmes, axes factoriels 1-2), VII^e-XII^e siècles. Chaque pape correspond à un point. L'analyse inclut la totalité des documents pontificaux, depuis les lettres jusqu'aux privilèges.

{75} Cette analyse initiale montre qu'en dépit de recouvrements indiscutables, différents groupes chronologiques se distinguent à partir du lexique (fig. 9). La rupture principale apparaît en effet sur l'axe 1 de l'AFC : elle oppose les actes issus des pontificats du XII^e siècle¹¹⁸, voire peut-être

¹¹⁸ Tout à gauche de la carte factorielle, on retrouve Alexandre III, Eugène III, Anastase IV, Adrien IV, Pascal II, Célestin II, Clément III, Innocent II, Urbain II, Calixte II, Honorius II, Lucius II ; puis un peu plus au centre, Grégoire VIII, Calixte II et Urbain II ; enfin à la charnière du groupe réparti sur l'axe 2, à droite, Gélase II, Anaclet II, Pascal II, Lucius II, Célestin III.

de la fin du XI^e siècle, au reste de la production papale. La tendance est forte, mais pas absolue : des pontifes comme Célestin III, Pascal II, Lucius II, Anaclet II, voire Gélase II, qui appartiennent tous au XII^e siècle, paraissent se mêler aux pontificats des siècles antérieurs, sans distinction majeure. C'est en analysant plus en profondeur l'AFC que cette contradiction initiale s'éclaire. En effet, l'axe 2 de la carte factorielle n'est pas inintelligible, en dépit d'une certaine confusion. On note qu'une part importante des papes des VII^e-VIII^e siècles se massent en haut à droite de l'analyse, de même pour ceux du IX^e siècle, un peu plus bas, vers le centre de l'axe. Un {76} second graphique, présentant la même AFC mais cette fois selon un découpage par siècle, permet d'éclaircir cette tendance (fig. 8b).

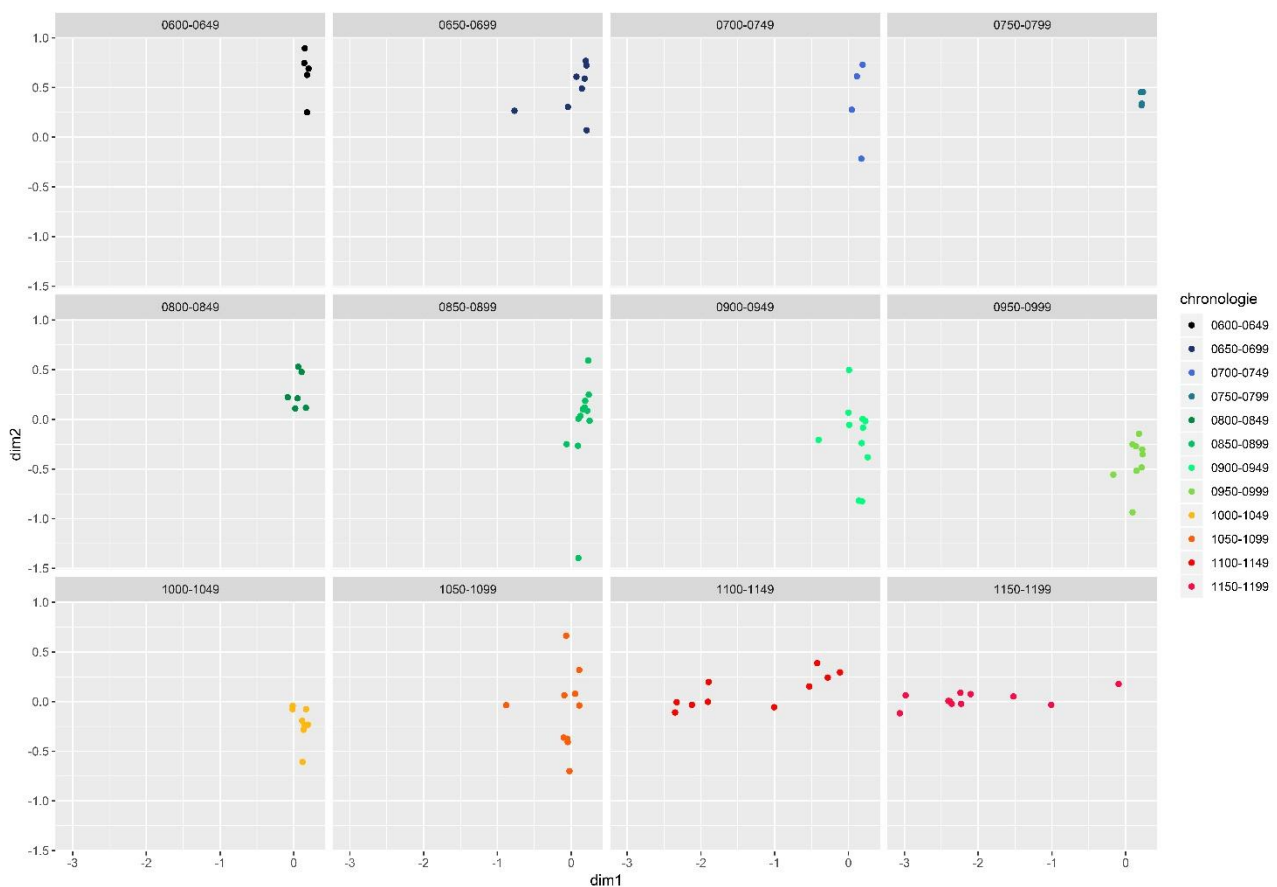


Fig. 10 : Analyse factorielle de l'ensemble du lexique pontifical (20 000 bi-lemmes, axes factoriels 1-2), VII^e-XII^e siècles – découpage par siècle. Chaque pape correspond à un point. L'analyse inclut la totalité des documents pontificaux, depuis les lettres jusqu'aux privilèges.

On y observe de nouveau la forte spécificité du XII^e siècle, qui se distingue sur l'axe 1 de l'AFC. Mais c'est cependant pour la compréhension de l'axe 2 que ce nouveau graphique apporte le plus : de haut en bas, il correspond en effet à un glissement progressif du lexique du VII^e au milieu du XI^e siècle. Autrement dit, il y a bien une évolution du lexique sur cette première période, mais elle est en partie distincte de celle qui s'opère au XII^e siècle. La position de la seconde moitié du XI^e siècle

le confirme indirectement : les points pour les papes de cette période s'étalent sur la quasi-totalité de l'axe 2, sauf Urbain II, qui se situe plus à gauche sur l'axe 1. Ainsi, c'est précisément parce qu'il existe un profond changement entre la seconde moitié du XI^e siècle et le début du XII^e siècle, que les papes de la période 1050-1099 se trouvent dans une position incertaine. Pour le dire autrement : si l'on considère simultanément les lettres et les privilèges, il existe deux populations de documents pontificaux¹¹⁹. Au plan historique, en considérant les deux graphiques simultanément, on dégage ainsi un double mouvement : d'une part, un glissement progressif du lexique pontifical des VII^e au X^e, voire XI^e siècle, puis une rupture très nette au XII^e siècle, au cours duquel se dessine un second mouvement.

Toutefois cette analyse pose une série de difficultés, la première étant évidemment cette coprésence des deux « populations de documents pontificaux ». Elle peut s'expliquer de différentes façons, d'ailleurs très complémentaires : par une rupture lexicale, mais aussi par une évolution typologique – des lettres aux privilèges. C'est dans cette perspective que nous avons décidé de mener une seconde expérience, cette fois uniquement à partir des privilèges.

III.2. *Privilèges seuls (VII^e-XIV^e siècles)*

L'existence de populations statistiques hétérogènes au sein d'une même analyse, autrement dit ici d'au moins deux groupes documentaires plus ou moins discrets, est souvent la cause d'une moindre lisibilité des analyses factorielles. Afin de préciser les tendances observées, nous avons donc choisi d'affiner le corpus initial, en retirant tout ce qui pouvait potentiellement le parasiter typologiquement. De la *PL*, il n'a donc été conservé que les actes présentés dans l'analyse comme étant des « Privilèges » ou des « Confirmations »¹²⁰. De la même façon, les documents présentés comme des « lettres » dans la base *Aposcripta* ont été retirées¹²¹. Enfin, parce que la rupture principale semblait se dérouler au XII^e siècle, nous avons étendu chronologiquement le corpus, en allant jusqu'au pontificat de Boniface VIII inclus (soit 1303)¹²². L'ensemble est ainsi composé d'un peu plus de 11 000 bulles.

¹¹⁹ Le terme population est entendu au sens statistique. On trouvera sur cette question des réflexions intéressantes dans Barbut, Marc: Des bons et des moins bons usages des distributions parétiennes en analyse des données. *Histoire & Mesure* 3-1 (1988), p. 111-128 ; Guerreau, Alain: *Statistique pour historiens*. Paris 2004, en particulier p. 25-72.

¹²⁰ À l'inverse, les lettres y sont relativement faciles à repérer, puisqu'elles débutent le plus souvent par une courte analyse, du type « *Epistola prima...* », « *Epistola ad...* », « *Ad...* ». On trouve ainsi le corpus des confirmations et privilèges en creux.

¹²¹ Dans le cas de cette base, une partie des métadonnées étaient disponibles. Nous avons poursuivi les tris.

¹²² Cette opération vise à repérer si la rupture du XII^e siècle est liée à un basculement typologique ou à une évolution lexicale – sans toutefois que les deux hypothèses s'excluent totalement (le changement lexical pouvant être inter- ou intra-typologique).

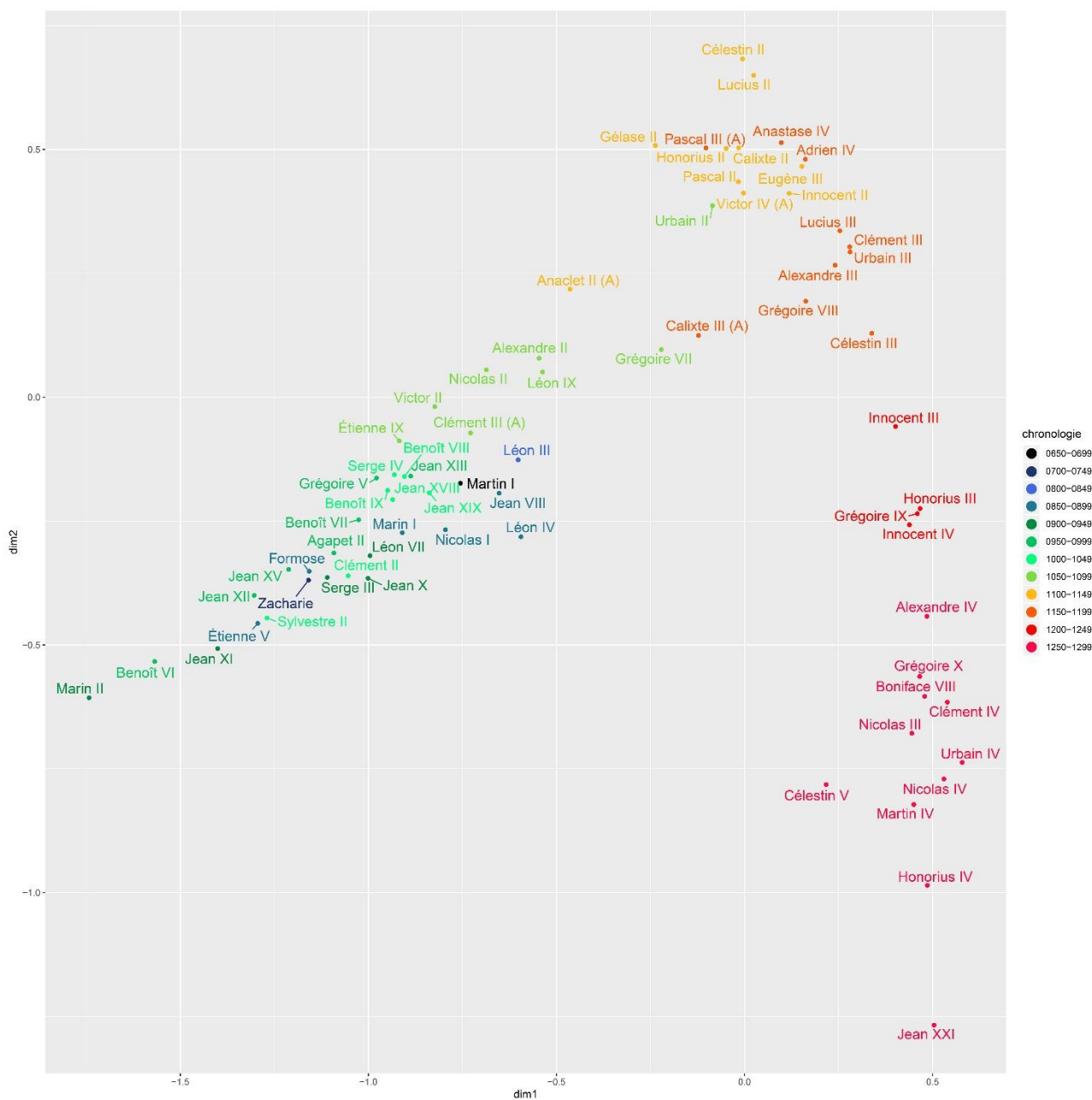


Fig. 11 : Analyse factorielle de l'ensemble du lexique pontifical (20 000 bi-lemmes, axes factoriels 1-2), milieu du VII^e-XII^e siècle – découpage par siècle. Seuls les privilèges ont été conservés. Chaque pape correspond à un point.

Une fois la typologie documentaire resserrée, on constate sur la nouvelle analyse que les groupes formés à partir du lexique sont beaucoup plus francs. À gauche sur l'axe 1, se distinguent les actes issus des pontificats des VII^e-X^e siècles, qui paraissent plutôt cohérents au plan lexical¹²³. À la

¹²³ Les analyses complémentaires qui suivent, portant uniquement sur les privilèges, montrent toutefois qu'il y a là aussi un glissement progressif. Par ailleurs, la relative cohérence des documents du haut Moyen Âge dans la présente expérience est liée au fait que celle-ci intègre à la fois des lettres et des privilèges, sur une chronologie très longue – ce qui a tendance

limite de ce groupe, se distingue nettement les papes de la seconde moitié du XI^e siècle, qui forment en quelque sorte un sous-groupe « tampon », entre ceux du haut Moyen Âge et ceux du premier XII^e siècle. Cependant, si Étienne IX ou Victor II paraissent être encore fortement liés aux actes du X^e siècle (Jean XIII, Grégoire V) ou du premier XI^e siècle (Benoît VIII, Serge IV, Jean XVIII), des pontifes comme Grégoire VII et surtout Urbain II appartiennent pleinement à la logique du XII^e siècle. Mais cette tendance ne dépend pas seulement de la chronologie, puisqu'Alexandre II [† 1073] apparaît comme plus proche des actes du haut Moyen Âge que Grégoire VII [† 1085], qui est pourtant son successeur. Cette observation montre que si l'évolution est chronologique, elle dépend aussi des choix et des actions des pontifes, de leur rapport au changement et à la tradition (scripturaire, sémantique, ou même sociale).

En haut à droite, se forme ensuite un second groupe, lui-même composé de deux sous-ensembles. L'écriture pontificale présente là une forte rupture, déjà relevée à partir de la fig. 9. Cette évolution est claire pour les pontificats de Léon IX, Nicolas II, Alexandre II et Grégoire VII (cf. le groupe 2a, fig. 10), mais c'est en fait Urbain II qui montre un fort renouvellement lexical, car il va se placer à la charnière du groupe contenant les pontificats de la seconde moitié du XI^e et du XII^e siècles – ce qui est toutefois cohérent chronologiquement.

Le troisième groupe, en bas à droite, est lui aussi composé de deux sous-ensembles – même si, pris ensemble, ils forment une seconde rupture lexicale. Comme on pouvait s'y attendre, c'est Innocent III qui l'inaugure. Ce sous-groupe (3a) conserve ensuite une forte cohérence lexicale jusqu'à Alexandre IV, qui ouvre lui aussi une nouvelle phase (3b), couvrant quant à elle la seconde moitié du XIII^e siècle. À l'extrême fin de notre chronologie, il semble que Jean XXI inaugure une nouvelle tranche d'évolution lexico-sémantique, sans que nous puissions véritablement préciser les enjeux à ce stade. En définitive, l'ensemble des analyses réalisées (dont seulement une partie est présentée ici) permet de proposer un modèle pour l'évolution globale de des privi-~~{78}~~-lèges pontificaux, composée d'au moins trois grandes phases : a. VII^e-X^e siècles ; b. XI^e-XII^e siècles ; c. XIII^e-XIV^e siècles. Cette évolution n'est toutefois pas parfaitement linéaire, et certains pontifes infléchissent plus nettement la tendance chrono-lexicale que d'autres. En outre, on peut aussi noter que les actes des malheureux candidats au siège apostolique, parfois dits « antipapes », sont régulièrement en retard ~~{79}~~ sur leur « époque lexicale »¹²⁴. C'est le cas par exemple de Pascal III ou de Calixte III, même si ces différences sont subtiles... et paradoxalement téléologiques.

à écraser certains phénomènes (l'évolution lexicale plus lente du haut Moyen Âge ; la relative cohérence des lettres) au profit d'autres plus puissants (l'évolution des X^e-XII^e siècles).

¹²⁴ Nous renvoyons en dernier lieu à Müller, Harald et Hotz, Brigitte (Dir.): *Gegenpäpste. Ein unerwünschtes mittelalterliches Phänomen.* Wien/Köln/Weimar 2012 ; Schönfeld, Benjamin: *Die Urkunden der Gegenpäpste : zur Normierung der römischen Kanzleigewohnheiten im 11. und beginnenden 12. Jahrhundert.* Wien/Köln/Weimar 2018.

III.3. *Situation spécifique du haut Moyen Âge et dynamique globale*

Concernant les VII^e-XI^e siècles, il s'agissait de réaliser une analyse complémentaire, permettant de mieux observer l'existence (ou non) d'une forte évolution lexicale des documents pontificaux au cours de cette période, toujours à partir des seuls privilèges (fig. 12). Elle révèle une autre évolution, plus subtile. Si les documents des VII^e-VIII^e siècles ne s'insèrent pas toujours dans une dynamique lexicale tranchée¹²⁵, les pontifes du IX^e siècle se regroupent assez nettement en bas à gauche de l'analyse (Léon IV, Nicolas I^{er}, Jean VIII, Marin I^{er}, Étienne V). Ils côtoient les papes de la première moitié du X^e siècle (Jean X, Jean XI, Léon VII), même si, là encore, quelques hésitations lexicales surviennent (Formose, Serge III). L'évolution se poursuit par la suite, puisque les actes de la seconde moitié du X^e siècle (Marin II, Agapet II, Jean XII, Benoît VI, Benoît VII, Jean XV, Grégoire V, Sylvestre II) et de la première moitié du XI^e siècle (Jean XVIII, Serge IV, Benoît VIII, Jean XIX, Benoît IX, Clément II) se regroupent. La rupture principale est, là encore, posée dans la seconde moitié du XI^e siècle, où les pontifes se démarquent nettement de leurs prédécesseurs. Ce mouvement s'opère peut-être en deux temps, avec d'une part la séquence allant de Léon IX à Alexandre II, puis celle qui s'ouvre avec Grégoire VII. La présente analyse concorde donc avec les résultats précédents (fig. 9, 10 et 11), tout en les complétant. {80}

¹²⁵ Cela peut signifier plusieurs choses : soit le nombre de faux/d'interpolations est trop élevé ; soit ces actes sont mal attribués ; soit ils appartiennent à une autre catégorie typologique – sans que ces trois hypothèses soient exclusives. Du reste, la faiblesse des échantillons pourrait expliquer la difficulté d'une analyse stylométrique pour ces siècles, du moins lorsqu'elle se fonde comme ici seulement sur les privilèges.

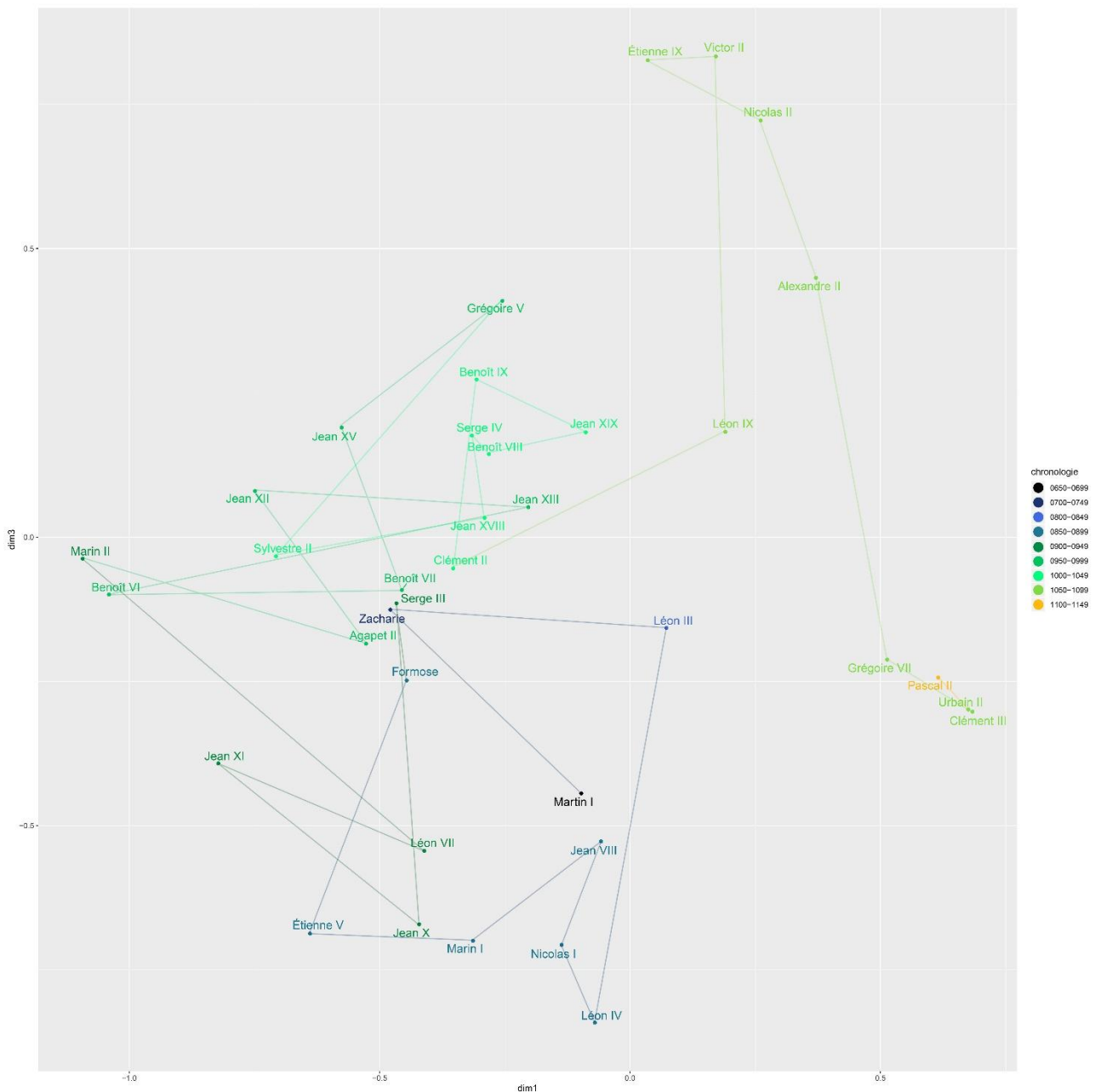


Fig. 12 : Analyse factorielle de l'ensemble du lexique pontifical (20 000 bi-lemmes, axes factoriels 1-3), 649-1118 – découpage par siècle. Seuls les privilèges ont été conservés. Les papes n'ayant pas un nombre d'actes suffisants ont été écartés de l'analyse. Chaque pape correspond à un point. Les points représentant les pontifes ont été reliés par ordre chronologique.

En l'état, nous pensons que la situation singulière des VII^e-VIII^e siècles est imputable à plusieurs facteurs¹²⁶, sur lesquels il nous faut dire un mot. La difficulté que ces pontifes ont à s'insérer dans un schéma lexical précis n'est sans doute pas seulement liée à l'existence d'une grande proportion de lettres dans leurs corpus – puisque dans la présente expérience, les lettres sont retirées

¹²⁶ Voir note 125.

(fig. 11)¹²⁷. Même s'il semble que cette originalité des documents pontificaux du très haut Moyen Âge s'accroît lorsqu'on ne retire pas les lettres de l'analyse (fig. 9 et 10). S'il paraît difficile de s'avancer, en particulier parce que la perte d'hypothétiques registres, la mauvaise conservation des papyrus et les sélections probables par les bénéficiaires ou les compilateurs ont sans doute joué un rôle dans cette situation, on observe tout de même que les lettres et, plus généralement, les documents dont le contenu lexical se signale par son originalité, prennent une part importante dans la typologie de bulles de ces siècles. S'agirait-il donc d'un biais de sélection, par lequel les actes les plus « hors-normes » auraient été préservés ? {81} Nous pensons plutôt que ce phénomène peut être rapproché des observations précédentes quant à la répartition chronologique des actes et la progressive standardisation du lexique. En effet, si les documents pontificaux du très haut Moyen Âge nous sont probablement parvenus que de façon lacunaire, on note des différences extrêmes, en termes de nombre de documents, d'un pontife à l'autre – nombre que l'on peut souvent corrélérer avec l'activité des pontifes, que l'on connaît par d'autres corpus textuels, en particulier narratifs¹²⁸. Il se pourrait donc que cette « inégalité de la production pontificale » soit aussi liée à des phénomènes proprement historiques, autrement dit à l'implication et à la perception du rôle de ces pontifes dans la société ecclésiale¹²⁹. Cela corrobore aussi les observations concernant la standardisation du lexique, faible pour ces hautes époques, et surtout, en première approximation, variable d'un pontife à l'autre. L'apparition d'une tendance chronologique plus nette à partir du IX^e siècle (en particulier de la seconde moitié), qui s'inscrit dans une évolution qui s'étire de ce siècle au milieu du XI^e siècle, signe probablement une première mise en forme de l'institution. Si les actes montrent des évolutions fréquentielles et lexicales, c'est probablement parce que le pontife lui-même se voit alors associé à des fonctions plus précises et mieux articulées avec le reste de la société médiévale, en tant que tête de l'*ecclesia* – en particulier dans le cadre de la dynamique carolingienne.

¹²⁷ On constate toutefois que l'inclusion des lettres dans les figures 9 et 10 brouillait la dynamique lexicale du haut Moyen Âge. Autrement dit, on peut supposer qu'en dépit d'une frontière floue entre les deux genres, il existe des évolutions scripturaires différenciées ou au moins complémentaires entre lettres et privilèges. L'hypothèse la plus probable, qu'il faudrait explorer, est celle d'une évolution à deux vitesses.

¹²⁸ C'est en particulier le cas à l'époque carolingienne, avec par exemple Nicolas I^{er} ou encore Jean VIII. Sur le rôle de la période carolingienne dans la formation du discours sur une fonction pontificale éternelle, voir à nouveau Demacopoulos: *The Invention of Peter* (voir n. 10) ; ainsi que Bougard, François: *Petitor et medius* : le rôle de la papauté dans les relations internationales de Grégoire le Grand à Jean VIII. Dans: *Le relazioni internazionali nell'alto medioevo*. Spolète 2011 (Settimane di studio del Centro italiano di studi sull'alto medioevo 58), p. 299-340 ; Bon, Bruno et Bougard, François: *Le Liber Pontificalis* et ses auteurs au IX^e siècle. Enquête stylo-métrique. Dans: Herbers, Klaus et Simperl, Matthias (Dir.): *Das Buch der Päpste : Der Liber Pontificalis*. Ein Schlüsseldokument zur europäischen Geschichte. Freiburg im Breisgau 2020 (Römische Quartalschrift Supplementbände 67), p.286-303, 2020, qui constitue une avancée importante en matière d'analyse des réécritures dans le *Liber pontificalis* ; McKitterick, Rosamond: *Rome and the Invention of the Papacy. The Liber Pontificalis*. Cambridge 2020 ; Costambeys, Marios, et Leyser, Conrad: *To be the neighbour of St Stephen: patronage, martyr cult, and Roman monasteries, c. 600-c. 900*. Dans: Cooper, Kate Mason, et Hillner, Julia (Dir.): *Religion, dynasty and patronage in early Christian Rome: 300-900*. Cambridge 2007, p. 262-287.

¹²⁹ Encore une fois, en la matière, les travaux nous semblent relativement rares. Voir cependant ceux mentionnés note 107.

Pour aller plus loin dans ces hypothèses, il est possible de produire un autre graphique, indiquant cette fois la dynamique de l'évolution lexicale des documents pontificaux. Pour ce faire, on mesure tout simplement la distance du lexique d'un pontife au lexique qui le précède immédiatement. Cette modélisation est complémentaire de celle concernant la richesse lexicale, mais ne dit pas la même chose : le lexique employé par un pontife peut être très standardisé (*i.e.* peu varié), tout en se démarquant fortement de celui de son prédécesseur (fig. 13)¹³⁰. {82}

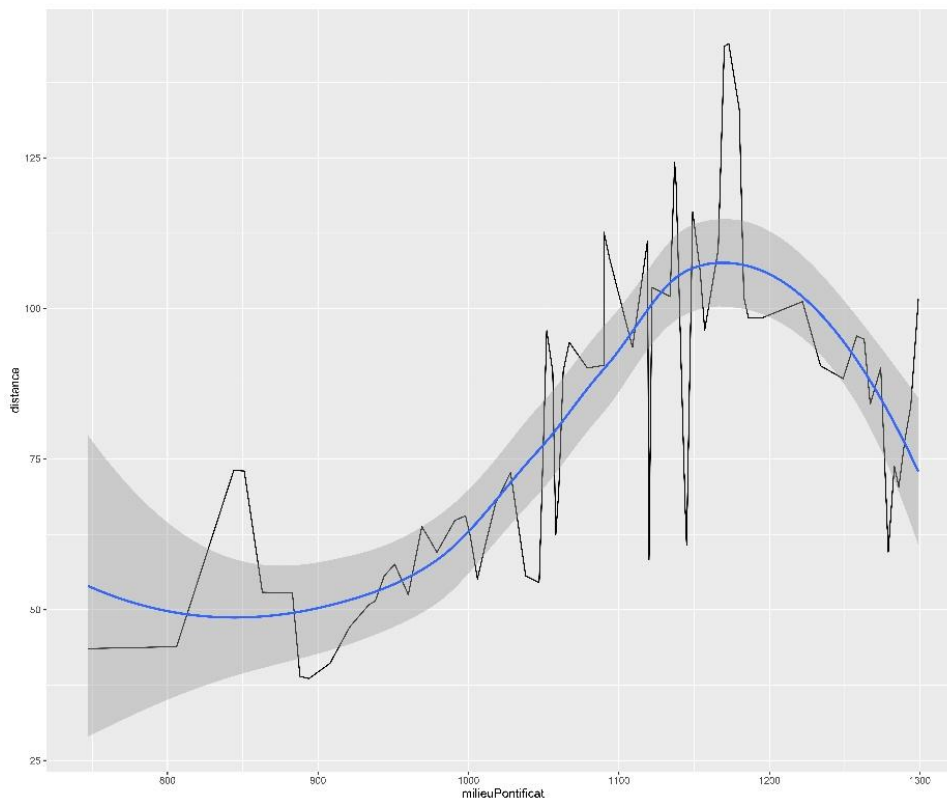


Fig. 13 : Évolution de la distance lexicale entre les actes d'un pontife donné et son prédécesseur immédiat, par ordre chronologique. Plus la courbe est haute, plus la vitesse d'évolution du lexique d'un pontife à l'autre est rapide (ce qui n'est pas contradictoire avec la concentration/standardisation lexicale observée précédemment).

Ce dernier graphique montre des tendances cohérentes avec les précédentes, mais cependant complémentaires. On note que dans une première phase, assez longue, qui s'étend jusqu'au début du XI^e siècle, le lexique pontifical (en dépit de sa grande richesse) connaît une évolution relativement lente d'un pontife à l'autre. Cette tendance globale est contrebalancée par des poussées évolutives, en particulier au milieu du IX^e siècle, au cours desquelles les lemmes employés (donc les thèmes évoqués dans les actes) changent très fortement. Passé l'An Mil, on observe un profond changement

¹³⁰ Ce graphique a été réalisé sur une suggestion de Sébastien de Valeriola, et avec son concours. Nous le remercions vivement pour son aide.

dans le lexique d'un pape à l'autre, et le phénomène paraît s'accroître entre 1020 et 1050¹³¹. Cette accélération est d'autant plus surprenante qu'elle s'associe à la standardisation du lexique déjà observée (fig. 6a, 6b, 7 et 8). Enfin, lors d'une troisième phase, on observe vers 1200 que l'évolution du vocabulaire d'un pontife à l'autre a tendance à {83} diminuer fortement. Ces dynamiques ne sont en outre pas liées au nombre de pontifes par siècle¹³².

En combinant ces informations aux précédentes, on obtient une idée de l'évolution générale de l'écriture pontificale au cours des VII^e-XIII^e siècles, complémentaire de l'historiographie – qui se fonde quant à elle sur d'autres critères, le plus souvent diplomatiques. Les analyses font ainsi apparaître quatre périodes bien distinctes, allant tout d'abord jusqu'au VIII^e siècle (inclus) ; puis du IX^e au milieu du XI^e siècle ; du milieu du XI^e à la fin du XII^e siècle ; puis enfin les XIII^e-XIV^e siècles. Elles correspondent assez bien avec ce que nous savons de la construction du pouvoir, du rôle et de l'image pontificale, tout en permettant d'émettre des hypothèses complémentaires.

IV. La construction variable de l'identité pontificale

Notre analyse ne pouvait en rester là. La possibilité d'interroger globalement cet ensemble rend envisageable des enquêtes sémantiques et diplomatiques complexes. Il va sans dire que nous ne pourrions que les effleurer ici. Une première possibilité consiste à regrouper les pontifes selon les quatre grandes phases identifiées lors des analyses précédentes, afin d'étudier les lemmes caractéristiques des différentes périodes. De nombreuses méthodes sont ici envisageables, en particulier l'application de coefficients permettant de faire ressortir les mots clés¹³³, mais aussi les analyses multivariées. Dans ce second cas, nous avons regroupé les textes selon les phases identifiées, en choisissant un échantillon aléatoire de textes pour que chacun des quatre sous-groupes contienne un nombre de mots équivalent¹³⁴. Sur la totalité de cet échantillon, près de 9 300 lemmes ont été retenus pour l'AFC – dont seulement 200 apparaissent toutefois sur le graphique, afin d'en faciliter

¹³¹ Entre 980 et 1050, certains pontifes se distinguent par un lexique relativement différent de son prédécesseur. Mais cette évolution n'est pas linéaire, et se fait plutôt par à-coups – tandis que lors de la phase suivante, l'évolution est continue.

¹³² En effet, on peut faire l'hypothèse qu'un lexique évolue plus lors d'un pontificat de 30 ans que lors d'un pontificat de 5 ans. Cependant, si l'on prend les choses à l'échelle globale, le nombre de pontifes par siècle est grossièrement équivalent du VII^e au XIII^e siècle. Autrement dit, le nombre de papes par siècle ne joue pas un rôle majeur dans l'évolution observée.

¹³³ On parle alors d'« extraction de mots-clés » ou de *topic modeling*. Parmi les méthodes possibles, la pondération des effectifs d'occurrences par des coefficients est assez efficace, par exemple le TF-IDF.

¹³⁴ Réunir des sous-groupes possédant un nombre de mots équivalent est fondamental, car les distributions lexicales sont particulièrement sensibles aux effets de taille. Pour avoir tenté une analyse sans ce paramètre, on peut affirmer que la difficulté se confirme empiriquement – l'absence d'échantillons de taille comparable rendant les résultats beaucoup moins robustes et de ce fait, difficiles à analyser.

la lecture. Ils ont été retenus en fonction de leur contribution à la distribution de l'AFC, autrement dit leur importance dans la spécificité de chaque période¹³⁵. {84}



Fig. 14 : Les lemmes contribuant le plus fortement à l'évolution du lexique pontifical (lettres et privilèges), selon les quatre périodes identifiées précédemment (en rouge : VII^e-VIII^e s. ; IX^e-milieu du XI^e s. ; milieu du XI^e s.-XII^e s. ; XIII^e-XIV^e s.). Axes factoriels 1-2 (80,8% de l'information résumée). Le gradient de vert à orange indique la contribution de chaque terme à l'AFC.

¹³⁵ Certains facteurs (on parle aussi de « modalités » ; ici, ce sont les lemmes) jouent plus fortement que d'autres dans la structure d'une analyse factorielle, parce qu'ils contribuent de manière décisive aux oppositions et aux complémentarités dans les données analysées (ici les papes et leurs lexiques). Plus cette contribution est forte, plus le facteur/lemme joue un rôle important. Dans la figure 14, les lemmes ayant les contributions les plus élevées apparaissent en jaune/orange.

Combinée à une analyse plus qualitative, fondée sur l'étude des coefficients, l'AFC permet de distinguer plusieurs éléments concernant l'évolution du discours pontifical. La spécificité des textes des VII^e-VIII^e siècles, évoquée tout au long de l'article, est sans doute ce qui ressort en premier lieu¹³⁶. Des lemmes comme *inconfusus*, *eximietas*, *ope-~~{85}~~-ramentum*, *compater*, *synodaliter*, *subsistentia*, *liberator*, *precellentia*, ou encore *unitio* ne se rencontrent quasiment que dans les documents de cette période, du moins au sein du corpus. Plus généralement, on note que ressortent sur l'AFC un grand nombre de termes liés à l'orthodoxie et à l'unité (*hereticus*, *orthodoxus*, *unus*, *unanimiter*), l'erreur (*error*, *peccatum*), au couple Créateur/créatures (*Deus*, *mens*, *spiritus*, *caro*, *cor*, *imago*, *natura*, *humanus*, *celum*, *operatio*, *divinitas*), et plus globalement à la juste doctrine et aux qualités chrétiennes (*excellencia*, *fides*, *dogma*, *perfectus*, *doctrina*, *bonitas*, *humanitas*, *gaudium*, *humilis*, *veritas*). Pour les textes conservés de cette période, il nous semble que le discours insiste particulièrement donc sur le rôle moral, apologétique et pastoral du pontife¹³⁷. C'est aussi l'évêque de Rome qui ressort, avec des termes comme *metropolis* ou *urbs*.

En opposition sur l'axe 2, mais toujours à gauche (axe 1) sur l'analyse, se trouve la période couvrant le IX^e, le X^e et la première moitié du XI^e siècle. Différents termes sont plus spécifiques à cette phase, en lien avec le développement des privilèges (bien que l'activité épistolaire soit aussi présente, mais de manière variable) : le pontife s'affirme alors comme une autorité dont le pouvoir est lié à l'écrit (*preceptum*, *decretum*, *epistola*, *scrinarius*, *synodus*), qui traite avec les autres puissants (*imperator*, *princeps*) et condamne (*damno*, *anathema*). Le champ sémantique du temps et de l'organisation spatiale se développe aussi, avec des lemmes comme *semper*, *successor*, *perpetuus*, mais aussi *fundus*, *rivus*, *vallis*, *campus*, *pons*, *adiacentia*¹³⁸. Nous rejoignons donc les hypothèses émises par Kristina Sessa, sur l'affirmation progressive du pontife comme « juste administrateur » des biens de l'Église, cependant avec un décalage chronologique¹³⁹. L'autorité morale de la précédente phase s'est donc transférée à une première forme d'autorité temporelle beaucoup plus nette, en lien avec l'implication grandissante des pontifes dans le monde.

La suite de l'analyse montre que les périodes couvrant la seconde moitié du XI^e siècle et le XII^e siècle d'une part, puis les XIII^e-XIV^e siècles d'autre part (à droite sur l'axe 1), sont relativement

¹³⁶ Elle est nettement plus forte si l'on ne réalise pas d'échantillonnage – autrement dit que l'on conserve tous les actes pour toutes les périodes. Dans ces conditions, les mots rares et moyennement fréquents ressortent. Il en résulte des AFC sur lesquelles l'opposition majoritaire est « VII^e-VIII^e s. » vs les trois autres périodes.

¹³⁷ L'AFC impliquant les 400 lemmes contribuant le plus à l'analyse concorde parfaitement avec cette première analyse – que ce soit en matière de qualités chrétiennes ou de vision du monde (avec des lemmes comme *anima*, *sermo*, *potentia*, *voluntas*, *heresis*, *leatitia*, *mysterium*, etc.).

¹³⁸ Ce lexique apparaissant dans des formules stéréotypées, mais aussi des descriptions de terres.

¹³⁹ Kristina Sessa, *The Formation of Papal Authority in Late Antique Italy. Roman Bishops and the Domestic Sphere*. Cambridge 2012.

proches si on les oppose aux périodes antérieures¹⁴⁰. Cela confirme la rupture que représente cette phase, envisagée globalement, autrement dit l'évolution lexicale qui s'y déroule (sans que cela présage de la standardisation lexicale qui, comme nous l'avons vu, est alors plus forte et en accélération). Au plan sémantique, les termes qui ressortent relèvent du temps (*annus, pontificatus, presens, idus, {86} calende, secularis*), de l'autorité dans les jugements (*iustus, ius, auctoritas, excommunicatio, indulgentia, supplicatio, interdictum, sententia, protectio*), de l'organisation spatiale et des entités englobantes (*diocesis, districtus, diocesanus*)¹⁴¹, mais aussi et surtout de l'institution ecclésiale, ainsi que de la hiérarchie dans l'Église (*ecclesia, cardinalis, ecclesiasticus, episcopus, abbas, archiepiscopus, archidiaconus, decanus, officium*), et de la parenté spirituelle (*frater, filius*). Ce mouvement passe globalement par la question de l'écrit, qui apparaît dans de nouvelles modalités (*pagina, signum, mando, supplicatio*). La transformation sémantique du discours pontifical est particulièrement nette, avec une insistance forte sur le rôle « capital » (au sens de *caput*) et englobant des papes, tête(s) de l'Église dans le siècle et sur terre, qui jugent et ordonnent (dans tous les sens de ce dernier terme)¹⁴². La dynamique quantitative exponentielle dans la production des actes s'accompagne ainsi d'une mutation de l'institution, de la vision qu'elle propose et de la perception que l'on en a. Ce qui nous semble particulièrement intéressant, c'est que cette dimension englobante est peu présente dans les documents antérieurs, qui se focalisent plutôt sur d'autres niveaux – en particulier la dimension pastorale.

Il ne convient toutefois pas d'assimiler complètement la phase 1050-1199 et 1200-1399, qui s'opposent sur l'axe 2¹⁴³. La première semble plutôt centrée sur des questions d'autorité et d'affirmation du rôle de l'*ecclesia*¹⁴⁴, tandis que la seconde défend l'importance des entités spatiales contrôlées par l'institution et la hiérarchie ecclésiale, donc l'universalisme¹⁴⁵. On comprend en définitive qu'à travers ce mouvement sémantique, qui va du VII^e au XIV^e siècle (inclus), c'est en fait l'ecclésiatisation du social que l'on observe, du point de vue de la figure qui en devient progressivement la tête institutionnelle : l'évêque de Rome.

¹⁴⁰ Paradoxalement, cela n'exclut pas la création d'un très grand nombre de néologismes, qui caractérisent aussi ces phases, surtout aux XII^e-XIII^e siècles.

¹⁴¹ Ce qui corrobore nos observations dans Perreaux: Des structures domaniales aux territoires ecclésiaux ? (voir note 92).

¹⁴² Nos observations rejoignent donc ici celles de Anheim, Étienne: Clément VI au travail. Lire, écrire, prêcher au XIV^e siècle. Paris 2014.

¹⁴³ Celui-ci est moins significatif que l'axe 1, mais résume tout de même presque 25% de l'information du tableau initial.

¹⁴⁴ Ladner, Gerhart B.: The Concept of Ecclesia and Christianitas and their Relation to the Idea of Papal Plenitudo Potestatis from Gregory VII to Boniface VIII. Dans: *Miscellanea Historiae Pontificiae* 18 (1954), p. 49-77.

¹⁴⁵ Voir ici les importantes remarques de Théry, Julien: Le gouvernement romain de la Chrétienté autour de 1206 : Innocent III et les débuts de la théocratie pontificale. Dans: *Mémoire dominicaine* 21 (2007 - Dominique avant les dominicains), p. 33-37 ; Id., Le triomphe de la théocratie pontificale du III^e concile du Latran au pontificat de Boniface VIII (1179-1303). De Cevins, Marie-Madeleine et Matz, Jean-Michel (Dir.): Structures et dynamiques religieuses dans les sociétés de l'Occident latin (1179-1449). Rennes 2010, p. 17-31 ; ainsi que Théry: Le Gouvernement pontifical (voir note 35).

{87} En guise de conclusion

Après ces différentes analyses, il convient de faire la part des confirmations historiographiques et des hypothèses nouvelles qu'ont permis de dégager cette approche numérique. Rappelons toutefois quelques préalables méthodologiques : en combinant les bases de données diplomatiques existantes, il est possible d'acquérir dès aujourd'hui un corpus étendu et représentatif de documents pontificaux, lettres et privilèges confondus, pour les VII^e-XIII^e siècles. Notre hypothèse de départ était en effet qu'un grand nombre de ces textes avaient été édités soit dans des collections antérieures, soit dans des éditions de cartulaires, qui ont été numérisées. La prise en compte de la *PL* (souvent sous-estimée, alors que sa richesse reste immense et incomparable), mais aussi de tous les actes présents dans les *CEMA*, via des outils algorithmiques (*Machine Learning*), a permis d'identifier un très vaste ensemble documentaire composé de presque 30 000 documents pontificaux¹⁴⁶. Indirectement, l'approche confirme que les outils de classification empruntés à la fouille de données (*text mining*) sont parfaitement à même de repérer ces textes dans un corpus diplomatique généraliste à partir d'un entraînement approprié – confirmant indirectement la spécificité de leur lexique. Dans nos différentes expériences, nous avons fait le choix de retenir la totalité de ces documents, tout en faisant régulièrement varier les échelles typologiques (lettres plus privilèges, lettres vs. privilèges) et chronologiques. Cette approche alternant entre lectures globales et particulières permet de contrôler certains biais majeurs de cette documentation – qui sont de fait quasi-incontournables si l'on se limite à une échelle unique. Parallèlement, une réflexion sur le traitement des données chronologiques et éditoriales (doublons, actes abrégés) a permis de proposer différentes méthodes permettant de résoudre différents problèmes et d'arriver à un corpus, certes imparfait, mais exploitable statistiquement¹⁴⁷.

{88} Concernant la typologie documentaire, il faudrait toutefois aller plus loin et voir sur quels critères lexicaux il est possible de distinguer entre privilèges et lettres. Il y a *de facto* des

¹⁴⁶ Nous ne pouvons donc qu'être étonnés de la quantité d'actes déjà disponibles *au format numérique*, ce qui renforce l'intérêt des grands corpus comme les *CEMA* et *Aposcripta*. Parallèlement, nous avons pu constater que les doublons n'étaient pas si répandus qu'on pouvait le croire, mais proviennent essentiellement des corpus d'originaux, qui viennent (utilement) redoubler les éditions existantes. Cela ne contredit pas l'idée exprimée plus haut selon laquelle « la tradition imprimée des documents pontificaux est souvent l'histoire d'une multi-édition », mais indique simplement que (parce que la numérisation intégrale de toutes les éditions diplomatiques n'est pas encore d'actualité) les recouvrements seront inévitablement de plus en plus nombreux, à mesure que l'on ajoute des cartulaires aux *CEMA*. L'objectif du corpus est en effet de continuer à croître, en particulier en recourant à la reconnaissance optique de caractères (optique présentée dans notre thèse : Perreaux, Nicolas: *L'écriture du monde. Dynamique, perception, catégorisation du mundus au Moyen Âge (VII^e-XIII^e siècles)*. Recherches à partir des bases de données numérisées. Dijon 2014, en ligne : <https://halshs.archives-ouvertes.fr/tel-03084322> (l'édition du travail est actuellement en cours dans la *Collection d'études médiévales de Nice*, chez Brepols).

¹⁴⁷ Le manque de métadonnée est l'un point faible des corpus diplomatiques actuels, et il faut améliorer les passerelles entre les corpus. C'est évidemment l'un des objectifs des *CEMA*, qui regroupent d'ores et déjà la plupart des métadonnées antérieures – d'autres pouvant par exemple être générées par les algorithmes d'intelligence artificielle.

différences entre ces « genres », puisque l'exclusion des lettres, dans certaines de nos analyses, permet d'entrevoir des résultats sensiblement différents (mais néanmoins complémentaires, cf. fig. 9-10 vs. 11-12). Cette situation est probablement plus liée au fait que certains documents étiquetés comme des « lettres » dans les éditions concernent de facto quasiment les mêmes thèmes que les privilèges – tandis que d'autres relèvent clairement du genre épistolaire, au sens fort, c'est-à-dire concerne des thèmes théologiques, didactiques et pastoraux. Il semble ainsi clair que la coupure n'est pas à faire (du moins pas uniquement) entre « lettres » et privilèges », mais plutôt au sein des lettres elles-mêmes. En outre, l'homogénéité de ces catégories fluctuent dans le temps, comme nous l'avons montré à plusieurs reprises – avec des évolutions fortes, en la matière, à l'époque carolingienne. Cela signifie que le découpage du corpus devra par la suite se faire à l'aide de critères lexico-sémantiques et non diplomatiques (du moins pas exclusivement), eux-mêmes évolutifs dans le temps. On arriverait ainsi à une typologie chrono-sémantique, complémentaire voire différente des typologies traditionnelles.

Pour revenir à un niveau plus général, le tableau suivant synthétise nos principales observations sur le corpus, à partir de différents points de vue adoptés : le nombre et la production des documents ; la conservation des originaux ; la richesse du lexique ; la rapidité de l'évolution de ce même lexique ; la typologie dominante, entre lettres et privilèges ; enfin, les tendances sémantiques :

	VII ^e -VIII ^e s.	IX ^e -milieu du XI ^e s.	milieu du XI ^e - fin du XII ^e s.	XIII ^e -XIV ^e s.
Nombre de documents	Faible	Hausses inégales, plutôt faible	Forte hausse continue, nombreux	Forte hausse, très nombreux
Proportion d'originaux	Faible	Faible	En hausse	En hausse
Richesse lexicale	Forte, mais variable	Très forte	Forte, mais variable	Moyenne
Rapidité de l'évolution du lexique	Lente, mais variable	Lente, en hausse vers la fin	Très forte	En baisse
Typologie dominante	Lettres	Lettres, puis privilèges (Xe s.)	Privilèges	Privilèges, puis différentes évolutions
Sémantique	Discours pastoral	Un « juste administrateur »	Un arbitre, à la tête de l'Ecclesia	Une théocratie pontificale

Fig. 15 : Synthèse des observations, hors analyse sémantique.

En ce qui concerne la production des actes, nous avons été frappé par la rupture que constitue le XI^e siècle. Elle est certes bien connue de l'historiographie, à travers la question de la perte de registres ou pseudo-registres, mais aussi des originaux, souvent articulée à la fragilité des supports¹⁴⁸,

¹⁴⁸ Là encore, si cet aspect veut sa part, il est impossible de ramener l'équation à la seule question des *papyrii*. En effet, la conservation des documents pontificaux originaux reste médiocre, y compris au-delà du XI^e siècle, alors que ce support est délaissé puis abandonné. Cette conservation par le biais de copies peut toutefois faire l'objet de différentes hypothèses. En définitive, elle soulève le problème de la sémantique de la copie et de l'action de copier dans l'Europe médiévale. Les historiens présupposent d'emblée qu'un original avait plus de valeur, parce que cela correspond à des critères ecclésiastiques et plus généralement éditoriaux (mais encore « romantiques »). Mais nous ne pouvons affirmer qu'il en allait systématiquement de même pour les médiévaux. Sur ce thème, une intéressante enquête pourrait être ouverte sur la valeur de la copie, au-delà des travaux sur la dimension matérielle de celle-ci.

à la taille et au nombre des documents¹⁴⁹. Toutefois, ces différentes hypothèses ne peuvent à elles seules expliquer le bond dans la production documentaire qui s'observe autour du pontificat de Léon IX. Nous y verrions plus volontiers une manifestation du changement spatial, social et sémantique dans les rapports qu'entretiennent les pouvoirs locaux de l'Europe médiévale aux pontifes. L'analyse lexicale confirme en effet une évolution considérable sur l'ensemble de la période envisagée, à travers différentes phases, allant du pape comme producteur de {89} lettres centrées sur le respect des qualités chrétiennes et la pastorale, à un pontife qui arbitre, tête dominante et universelle de l'*ecclesia*¹⁵⁰. Ce qui change, en outre, c'est non seulement le contenu de la documentation, mais aussi sa typologie : essentiellement composée de lettres dans le haut Moyen Âge, les privilèges se développent lentement à partir de l'époque carolingienne puis au X^e siècle, allant jusqu'à composer la majeure partie du corpus au XI^e-XII^e siècles¹⁵¹. Nous pensons qu'il faut relier ce phénomène à ce que Joseph Morsel nomme l'« enjeu de la réponse » : le pape n'écrit plus seulement des missives enjoignant à la régulation, mais il est progressivement sollicité par les bénéficiaires¹⁵². Alors que dans le premier cas les missives imposent un discours qui n'attend pas nécessairement de réponse, dans le second, il s'agit d'un dialogue entre un requérant et un arbitre, un protecteur. Les évolutions de la richesse lexicale vont aussi dans ce sens : alors que les pontifes traitent jusqu'au IX^e siècle inclus de sujets très variés (en lien avec des questions théologiques et pastorales donc, puis avec la question de la « juste administration »), le lexique tend à se concentrer progressivement à par-{90}-tir du X^e siècle, pour se standardiser ensuite fortement au XII^e siècle¹⁵³. Les périodes carolingienne et post-carolingienne apparaissent ainsi, à bien des égards, comme des lieux d'expérimentations pour le système pontifical et sa chancellerie en devenir. Cette concentration progressive n'empêche pas, par ailleurs, une évolution de plus en plus rapide du lexique lui-même entre 880 et 1180¹⁵⁴, signe des transformations rapide du système pontifical (et plus largement, ecclésial).

¹⁴⁹ Ce dernier argument ne pouvant bien entendu pas s'appliquer au haut Moyen Âge, puisque, précisément, les documents pontificaux y sont rares – en particulier si l'on exclut les lettres de l'analyse.

¹⁵⁰ Le terme étant ici entendu au sens de « système ecclésial » et non pas seulement d'Église.

¹⁵¹ Il y a un basculement du ratio lettres/privilèges au X^e siècle, avec, avant cette date une proportion très élevée de lettres, mais stable du VII^e au IX^e siècle ; puis après cette date, une proportion très élevée de privilèges, du XI^e au XIII^e siècle.

¹⁵² Morsel, Joseph: Communication et domination sociale en Franconie à la fin du Moyen Âge : l'enjeu de la réponse. Dans: Boucheron, Patrick et Nicolas Offenstadt (Dir.): L'espace public au Moyen Âge. Débats autour de Jürgen Habermas. Paris 2011, p. 353-365 (une version longue de l'article est disponible en ligne : <https://hal-paris1.archives-ouvertes.fr/halshs-00388554/>). Concernant la perception des bénéficiaires d'actes pontificaux, nous renvoyons aux travaux de Jochen Johrendt mentionnés note 106.

¹⁵³ Cette richesse relative du lexique pontifical dans le haut Moyen Âge n'était pas nécessairement attendue, le maximum étant atteint à l'époque carolingienne. Là encore, l'évolution typologique du corpus joue probablement un rôle, les lettres étant a priori plus variées que les privilèges (encore faudrait-il le montrer et cela doit dépendre des périodes), mais n'explique pas tout. En outre, si l'étiage lexical a été observé au XII^e siècle, nous faisons l'hypothèse de travail que la richesse lexicale pourrait repartir à la hausse aux XIII^e-XIV^e siècles, dans un contexte de création conceptuelle appuyant les efforts des pontifes et de la curie.

¹⁵⁴ Avec une première phase de changement rapide à l'époque carolingienne, plutôt au début du IX^e siècle.

Ainsi, alors que le pape n'est plus tout à fait un être inaccessible et distant, y compris pour les acteurs locaux ou régionaux comme les abbayes, les chapitres et les évêques, ces bénéficiaires se mettent à conserver de manière beaucoup plus efficace les documents¹⁵⁵. Ce qui intéresse alors, pensons-nous, ce n'est plus seulement la fonction théologique du pape, mais son rôle de régulateur, de plus en plus universel, qui se développe évidemment dans le contexte des exemptions, mais aussi des voyages pontificaux. D'importantes évolutions lexicales s'opèrent en effet au moment même où se développe l'itinérance des papes, par exemple avec Urbain II, qui marque une évolution nette. Considérés simultanément, ces différents indices renforcent l'idée que ce qui a changé, c'est non seulement l'image que le pape donne de lui-même, mais aussi l'image qu'on en perçoit – dans une chrétienté désormais « encellulée », c'est-à-dire capable de s'extraire (pour les dominants) de l'échelle locale, parce que celle-ci est maîtrisée, en particulier via un semis d'églises, de lieux et de personnes. La documentation, ici examinée comme un tout, est non seulement un révélateur de ces évolutions, situées à mi-chemin entre les représentations et les dynamiques sociales, hiérarchiques, mais aussi un catalyseur. En produisant en masse des titres pour tous ceux qui la sollicite, la tête de l'Église s'impose, surtout à partir des XII^e-XIII^e siècles, comme le pivot du système. Or, ce changement n'a pu être opéré que parce que l'Europe médiévale avait atteint un certain stade d'organisation socio-spatiale, en particulier à l'échelle locale.

Nicolas Perreaux
Université de Paris 1 Panthéon-Sorbonne
UMR 8589 LaMOP
LabEx HASTEC (ANR-10-LABX-85)

¹⁵⁵ Nous avons constaté que le corpus pontifical était assez mal conservé sous la forme d'originaux, y compris pour les seuls privilèges, et y compris pour les X^e-XIII^e siècles. En comparaison avec d'autres typologies, en particulier les diplômes royaux et impériaux, le fossé est énorme.