

RENDRE COMPTE DES USAGES BIBLIQUES D'UN AUTEUR : L'EXEMPLE DE BERNARD DE CLAIRVAUX.

I. – INTRODUCTION

Comment rendre compte du rapport de saint Bernard à la Bible ? Comment faire, à partir de la mosaïque des quelque 38.500¹ insertions bibliques ciselées, qui justifient chacune une analyse littéraire, théologique, pour dégager une vue synthétique qui donne accès aux comparaisons synchroniques et diachroniques, aux enjeux d'éventuelles évolutions ?

Pour tenter de répondre à ces questions, est née en 2016 à l'Institut des Sources Chrétiennes, en collaboration avec l'équipe de textométrie de l'ENS de Lyon² et la LLT-A de Brepols³, et dans le prolongement du projet BIBLINDEX⁴, une entreprise spécifiquement consacrée au traitement du matériau biblique des œuvres de Bernard de Clairvaux. Elle repose sur la construction d'un index numérique à vocation exhaustive, qui permette des interrogations multiples, aussi bien sur les références chiffrées des versets que sur le texte latin de Bernard⁵. Tout

¹ Ce calcul repose sur les apparats scripturaires des volumes parus de la collection Sources Chrétiennes, et sur une estimation de ceux encore en préparation.

² <http://textometrie.ens-lyon.fr/>.

³ *Library of Latin Texts*, www.brepolis.net.

⁴ Index en ligne des références scripturaires chez les Pères de l'Église : <http://www.biblindex.info>. Pour de plus amples informations, voir par exemple L. MELLERIN, « BIBLINDEX, index en ligne des références scripturaires chez les Pères de l'Église : état des lieux et perspectives », dans *Écriture et tradition chez les Pères de l'Église*, M. Stavrou et al. éd., *Cahiers de Biblia Patristica* 17, Strasbourg, 2017, 180 p., p. 145-170 (pre-print : halshs-01176139).

⁵ Est mise en place une chaîne éditoriale complète, depuis les textes bruts fournis par Brepols, ou depuis la version numérique des textes déjà édités dans Sources Chrétiennes, jusqu'à un ensemble de fichiers XML conformes à la TEI comportant délimitation et typage fin de toutes les occurrences. Ces fichiers-pivots seront exploités diversement sur le web, via des feuilles de transformation XSL et des liens avec la base de données de références de BIBLINDEX. Cette chaîne, aujourd'hui pratiquement finalisée grâce au travail d'Élysabeth Hue-Gay, assistée pendant quelques mois par Lou Cécille et Charles Bourdot, repose sur les outils Métopes

comme BIBLINDEX dans son ensemble, ce projet sur corpus spécifique entend constituer le maillon manquant entre des outils textuels existants – en l’occurrence la base de données des textes latins de Bernard dans la LLT-A – et les index scripturaires disponibles – pour Bernard, ceux des *Sancti Bernardi Opera* et des Sources Chrétiennes. Mais l’outil numérique ne se conçoit qu’en interaction permanente avec des parcours de recherche : si l’accès aux données brutes est un point de départ, la conception de l’index, la structuration des fichiers, le choix des attributs, sont informés par les requêtes envisagées, et les questionnements sont à leur tour enrichis, modifiés, par ce que donnent à voir les extractions.

Envisageant ici le corpus bernardin non seulement pour lui-même, mais aussi en tant qu’archétype, du fait de la nature particulièrement topique de ses données – profusion et diversité des occurrences, complexité et variété de leurs modalités d’insertion ; connaissance partielle mais importante des métadonnées chronologiques –, nous proposons un aperçu des principales étapes méthodologiques d’une enquête sur la Bible d’un auteur.

II. – LA PREPARATION DU CORPUS

A. *Le choix du référentiel biblique*

En premier lieu, il faut choisir l’édition moderne du texte biblique dont la numérotation sera utilisée, et par rapport à laquelle seront notées les éventuelles divergences textuelles. Chez Bernard, les citations proviennent massivement de la Vulgate, l’édition Weber-Gryson⁶ sera donc prise comme référentiel : facilement consultable, son texte est disponible en ligne sur le site de BIBLINDEX⁷.

B. *Les délimitations d’unités textuelles*

Ensuite, la segmentation souhaitée des unités textuelles doit être définie, pour la Bible comme pour les œuvres de Bernard.

1) *Les textes bibliques cités*

Pour le confort des usagers modernes, il est indispensable de conserver les divisions usuelles, en livres, chapitres, versets, même si, encore à l’époque de Bernard, elles ne sont pas pertinentes. Chez lui, les citations plus longues que le verset sont extrêmement rares ; en revanche, les citations complètes de versets le

développés par le Pôle Documentaire de la MSH de Caen. Sa documentation complète est consultable sur demande.

⁶ *Biblia sacra juxta vulgatam versionem*, R. Weber, B. Fischer, J. Gribomont *et al.* éd., Stuttgart, 1969, 1994⁴, 2007⁵.

⁷ Voir la section « Outils bibliques » du site. À propos des référentiels, nous renvoyons à L. MELLERIN, « Methodological Issues in BIBLINDEX, An Online Index of Biblical Quotations in Early Christian Literature », dans *Biblical Quotations in Patristic Texts*, M. Vinzent, L. Mellerin, H.A.G. Houghton éd., *Studia Patristica* LIV (Papers presented at the Sixteenth International Conference on Patristic Studies held in Oxford 2011), Leuven-Paris-Walpole, MA, 2013, 129 p., p. 11-32, ici p. 15-18 (= MELLERIN, « Methodological Issues »).

sont presque tout autant, puisqu'une estimation par sondage montre que ses citations littérales font en moyenne 8 mots, ce qui est bien inférieur à la longueur moyenne d'un verset. Il faudra donc croiser questionnement du corpus par nombre d'occurrences – une occurrence pouvant renvoyer à plusieurs lieux bibliques –, par nombre de références scripturaires et par nombre de mots : on peut par exemple affirmer à la fois que Bernard cite tous les livres de l'Ancien Testament, et qu'il ne cite qu'1/15 de ses mots ; ou qu'il cite 33 fois *Malachie* 4, 2 et qu'il utilise seulement l'expression « *sol iustitiae* », etc. Par ailleurs, selon les études, seront à définir les regroupements pertinents de textes bibliques. Par exemple, chez Bernard qui cite très peu les livres historiques mais énormément les épîtres pauliniennes, il faudra regrouper les premiers, mais décomposer les secondes.

2) *Les œuvres citantes*

À l'échelle de BIBINDEX, qui regroupe nombre d'œuvres fragmentaires, partiellement citées, conservées dans des langues différentes, avec des doublons, des recoupements, des problèmes d'attribution, établir une nomenclature des œuvres est extrêmement complexe. Dans le cas d'un auteur unique, et dont l'œuvre est bien connue et transmise comme celle de Bernard, la difficulté est certes moindre ; cependant des décisions sont à prendre qui conditionnent toute la suite des analyses. Se posent la question de l'authenticité, ou du moins du degré d'implication de Bernard dans la rédaction finale⁸ ; celle de la prise en compte des variantes de la tradition manuscrite recensées dans les apparats critiques, des regroupements que la tradition a imposés. Si la plupart des textes ont un genre littéraire clairement déterminé, quelques-uns résistent au classement. Reste enfin la question de la granularité à adopter.

C. La préparation de métadonnées

1) *Le décompte des mots (et des versets)*

Il convient d'évaluer systématiquement le nombre de mots – et, pour la Bible, de versets – contenus dans chacune des unités textuelles à distinguer. En effet, l'une des principales pondérations des résultats de fréquence sera la longueur relative des œuvres et des livres bibliques.

2) *La détermination d'une chronologie*

Rappelons également la nécessité d'établir la chronologie la plus solide possible pour étudier les évolutions des phénomènes citationnels⁹. Certes,

⁸ Sur ces questions d'authenticité, voir l'étude récente de J. DE GUSSEM, « Bernard of Clairvaux and Nicholas of Montiéramey: Tracing the Secretarial Trail with Computational Stylistics », *Speculum* 92/S1, 2017, p. 190-225.

⁹ Dans les exemples qui suivent pour Bernard, nous nous sommes fondés sur la chronologie donnée par Guy Lobrion au début du volume collectif *Bernard de Clairvaux : histoire*,

l'approche chronologique laisse de côté les œuvres non datables : dans le cas de Bernard, il s'agit d'environ 43% des mots¹⁰, de 67% des occurrences scripturaires¹¹, situées en particulier dans sa prédication qui représente les ¾ de l'œuvre totale ; mais elle fournit tout de même des indications significatives sur le degré de familiarité de Bernard avec un livre donné à une période donnée.

3) *Les corpus des autres auteurs*

Un pôle d'auteurs contemporains (Guillaume de Saint-Thierry, Aelred de Rievaulx, Pierre le Vénéral, Hugues de Saint-Victor) qui servirait d'étalon comparatif est indispensable. Malheureusement, pour l'obtenir, il faudrait déjà avoir fait un travail analogue sur ces autres auteurs... Dans un premier temps, on pourra comparer les usages de Bernard à ceux de l'ensemble des textes patristiques aujourd'hui recensés dans BIBINDEX.

III. – LE TRAVAIL SUR LES OCCURRENCES SCRIPTURAIRES

A. *Le repérage*

Après ces indispensables préalables vient le repérage des passages scripturaires. Il est toujours possible de recourir aux concordances verbales¹² ; par ailleurs, les outils de détection automatique des similarités textuelles, qui s'améliorent pour les langues anciennes, fournissent une base de travail. Cependant, ils sont fondamentalement pris en défaut : ils donnent des séquences de lemmes, voire de caractères communs ; mais les résultats pertinents sont perdus au milieu d'un bruit considérable. Le champ des allusions ténues, progressivement préparées, des paraphrases demeure inexploré ; ces outils peinent à délimiter correctement les citations en l'absence de marqueurs énonciatifs nets¹³. Nous sommes là au cœur d'un paradoxe qui sous-tend tout

mentalités, spiritualité (SC 380, Paris, 1990, p. 32-41) et les introductions et notes des volumes de Sources Chrétiennes.

¹⁰ Pour ce calcul, nous avons estimé à 15% du total des *Lettres* (sans *Ep* 42, 77 et 190) les lettres non datées. Nous arrivons à 364.807 mots sur 840.909.

¹¹ Le corpus datable comporte 11.638 occurrences bibliques d'après G. HENDRIX, *Index Biblicus in Opera Omnia Sancti Bernardi, Sancti Bernardi Opera IX*, Louvain, 1998 (= HENDRIX, *Index Biblicus*).

¹² Pour la Bible latine, F. P. DUTRIPON, *Vulgatae editionis Bibliorum sacrorum concordantiae*, Paris, 1844 ; B. FISCHER, *Novae Concordantiae Bibliorum Sacrorum Iuxta Vulgatam*, Stuttgart-Bad Cannstatt, 1977 ; E. PEULTIER, L. ÉTIENNE, L. GANTOIS, *Concordantiarum universae Scripturae sacrae Thesaurus*, Paris, 1897.

¹³ Voir les tests opérés avec TRACER, partiellement sur le corpus bernardin, dans l'article de M. MORITZ, A. WIEDERHOLD, B. PAVLEK, Y. BIZZONI, M. BÜCHLER, « Non-Literal Text Reuse in Historical Texts: An Approach to Identify Reuse Transformations and its Application to Bible Reuse », dans *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, November 1-5, 2016, p. 1849–1859 [<https://aclweb.org/anthology/D16-1190>].

BIBLINDEX : il faut rendre compte d'une relation binaire texte cité-texte citant, mais le texte citant ne laisse bien souvent transparaître l'objet cité qu'à travers le prisme d'une réappropriation mémorielle, d'une réécriture qui le transforme en bien propre de celui qui cite.

Que faire alors ? Adopter la méthode radicale d'Anne-Marie La Bonnardière qui, interrogée sur son étude de la Bible d'Augustin, répondait : « C'est très simple. Je lis la Bible ; puis je lis Augustin ; et je relis la Bible, et je relis Augustin ! » ? Il s'agirait de pratiquer une sorte de *lectio* à deux niveaux – ruminant de la Bible, ruminant du texte de Bernard –, pour faire l'expérience en soi-même de l'intrication des textes, dans un va-et-vient incessant, successif, mais dévoilant ponctuellement des connexions par le biais de la mémoire. Cette participation existentielle à l'imprégnation biblique de Bernard est une voie fructueuse, en ce qu'elle déjoue les pièges du cercle herméneutique : il faut s'être pénétré en profondeur de toute l'œuvre pour déceler dans une allusion ténue les harmoniques spirituelles et théologiques qui constituent son arrière-plan ; pour autant, il faut aussi scruter une à une les occurrences où tel verset est tissé dans la phrase bernardine pour voir se dessiner la structure à laquelle il contribue. Dans l'approche d'Anne-Marie La Bonnardière, tout comme, paradoxalement, dans celle froidement technique de TXM (cf. *infra*), il n'y a plus mise en regard de deux corpus textuels distincts, la Bible d'un côté, le texte de Bernard de l'autre, mais constitution d'un corpus englobant, mental ou numérique, qui donne à voir tous les échos d'une thématique, d'un épisode, d'une formulation dans des réseaux de passages textuels – ce que les notes marginales des Bibles indiquent sous forme de renvois intrabibliques, et ce que la spécialiste d'Augustin appelle « orchestrations scripturaires¹⁴ » lorsque ce réseau se tisse aussi dans l'œuvre citante. Le va-et-vient nécessaire entre le tout de l'œuvre et son détail est une constante dans les études de corpus littéraire, mais il est rendu plus complexe chez les auteurs imprégnés d'une source elle-même formée par des auteurs imprégnés d'une source. Il y a là un point aveugle de la méthodologie : Georges Valayer, collaborateur infatigable du CADP, se définissait comme « chercheur de truffes¹⁵ » ; Jean Figuet, explorateur du corpus bernardin, parlait de « recettes de cuisine » pour définir ses analyses d'usages bibliques chez Bernard. Et finalement, le paramétrage sans cesse recommencé de la taille des segmentations, de l'usage de la lemmatisation dans TRACER relève du même empirisme. Cette première étape de la recherche est forcément itérative. Pour la rendre la plus exhaustive possible, il convient de faire se croiser sur le texte les regards

¹⁴ Chaque verset de la *Biblia Augustiniana* est présenté avec son « orchestration scripturaire ».

¹⁵ Cf. des extraits de sa correspondance donnés dans l'article [<https://biblindex.hypotheses.org/1173>].

successifs de personnes dont la culture biblique est complémentaire, et, une fois entraîné le détecteur biblique humain par la fréquentation des textes, d'affiner par le recours assez systématique aux éditions numériques de la Vulgate.

B. La caractérisation des occurrences scripturaires

Voilà donc repérées les connexions entre texte biblique et texte patristique : il s'agit alors de les décrire de la façon la plus homogène et juste possible.

1) *L'expression des sources*

À cette fin, les analystes doivent disposer d'une typologie unifiée et constante dans le temps, telle que donnée dans les directives de *Biblia Patristica* ou BIBLINDEX¹⁶. Nous avons exposé en détail dans un autre article les caractérisations prévues dans BIBLINDEX et le projet « Bible de Bernard », en particulier l'encodage TEI pour en rendre compte¹⁷, nous nous permettons d'y renvoyer ici.

Que ce soit dans les *SBO*, les *SC* ou l'encodage TEI, la première exigence est de trouver un mode d'expression générique des occurrences qui colle de la moins mauvaise façon possible au texte utilisé par l'auteur. Dans le cas simple, il y a une référence par occurrence ; mais souvent il y en a plusieurs, pour dire aussi bien la fusion des textes dans leur réemploi, que l'ambiguïté du référentiel biblique lui-même, dans les cas de citations intrabibliques ou de passages synoptiques. C'est le même procédé, l'ajout d'une référence, qui sert pour décrire une combinaison de sources, donc un acte citationnel complexe qui engage la liberté de l'auteur, et le caractère indécidable d'une alternative entre un passage biblique et un autre qui lui est similaire. De la même façon, les renvois à d'autres sources que les manuscrits bibliques sont à la fois l'indication d'une altération de la source biblique et un renseignement sur l'acte citationnel, puisque, sous-jacente à ces modifications, il y a une réminiscence, un fonctionnement de la mémoire qui obéit à d'autres impératifs que la restitution exacte d'un texte reçu figé dans une lettre. Dans les *SBO* comme dans *SC*, les biais par lesquels le texte a pu arriver jusqu'à Bernard – un texte patristique ; la Règle de saint Benoît ; la Liturgie – sont exprimés par une mention entre parenthèses dans les notes. La même information, dans l'encodage TEI, peut être considérée comme un attribut de la « cible », c'est-à-dire du référentiel biblique.

¹⁶ J. ALLENBACH (dir.), *Étapes, moyens et méthode d'analyse pour la constitution du Fichier microphotographique des citations de l'Écriture chez les Pères*, Faculté de théologie protestante, CADP, Strasbourg, 1967 ; MELLERIN, « Methodological Issues ».

¹⁷ E. HUE-GAY, L. MELLERIN, E. MORLOCK, « TEI-encoding of text reuses in the BIBLINDEX Project », dans *Computer-aided Processing of Intertextuality in Ancient Languages*, M. Büchler, L. Mellerin (éd.), *Journal on Data Mining and Digital Humanities*, special issue, 2017 [consultable en ligne, <https://jdmdh.episciences.org/paper/view/id/3989>].

On repère pour chaque support des constantes : l'exigence d'un système bref de références canoniques légèrement typées ; la définition d'une catégorie simple qui rassemble les occurrences les plus littérales. La nécessité de jouer sur la complémentarité entre référence canonique et texte est prégnante. Il faut des indications systématiques, sous forme de sigles ou d'attributs, qui réduisent au maximum la part de l'interprétation et servent ensuite de critère de tri ou de sélection dans des recherches ; mais les frontières entre catégories sont trop subtiles pour que cela suffise à décrire avec exactitude l'occurrence.

2) *La représentation de l'acte citationnel*

Pour représenter avec finesse l'acte citationnel, au-delà du repérage et du typage normé des sources, il convient de mener une enquête plus fine sur les modalités d'insertion de l'occurrence biblique dans le discours, qui relève de l'encodage des mots¹⁸. Ici le recours au fichier pivot XML permet d'aller plus loin que l'édition papier, en dépassant la dissociation entre le texte observé et sa description. Le fichier XML peut servir à générer plusieurs visualisations superposables et/ou successives des phénomènes, qui correspondent à autant de parcours de lectures. Il est libéré des limitations propres à une édition qui doit tout exprimer en même temps et rester humainement lisible. Ainsi, il peut décrire plusieurs couches rédactionnelles dans la même unité textuelle. La prise en compte de l'apparat critique des *SBO* pour identifier les corrections des textes bibliques, les modifications en cours de rédaction, l'impact des harmonisations de copistes donnera par exemple une idée des transformations au cours du temps, volontaires ou non, du texte biblique de Bernard.

Outre la systématisation des relevés, qui facilite la production de résultats homogènes, il vise à un repérage des indices objectifs de l'acte citationnel. Nous analyserons la présence ou l'absence de formules introductives, explicites ou non, du type *Apostolus*, *Iob*, *Propheta dicit*, etc. C'est un champ d'étude *a priori* fécond car, comme le note Jean Figuet, il semblerait que Bernard n'emploie pas les mêmes formules pour les versets qui lui sont spécialement chers. On ne trouverait pas pour eux des formulations neutres, du type « L'Écriture dit », mais des formulations plus personnalisées¹⁹. Seront aussi à analyser des erreurs

¹⁸ Le texte citant est encodé avec la granularité du mot, la cible biblique n'est pas encodée à une granularité inférieure à la partie de verset, dans la mesure où l'expression canonique du verset permet de montrer à l'utilisateur le texte de ce verset en regard du texte citant. Cette complémentarité entre expression de la référence canonique et analyse textuelle permet de contourner une difficulté : l'impossibilité de faire correspondre mot à mot texte citant et texte biblique en cas de reformulation.

¹⁹ J. FIGUET, « La Bible de saint Bernard : données et ouvertures », dans *Bernard de Clairvaux, Histoire, mentalités, spiritualité, Colloque de Lyon-Cîteaux-Dijon*, SC 380, Paris, 1992, p. 237-270 (= FIGUET, « Bible »), ici p. 242.

d'attributions²⁰, ou encore les « prélibations », ces anticipations de citations manifestées par quelques mots-cheville ou avant-coureurs.

Quant aux modifications morphosyntaxiques qui rendent possible l'insertion du texte biblique dans la phrase, qu'elles soient encodées en détail – ce qui *a priori* n'est pas l'option que nous avons retenue – ou aisément discernables par des visualisations adaptées, elles pourront être observées et caractérisées. L'ambition est de donner aux intuitions de J. Figuet une assise méthodologique pour les confirmer, infirmer ou compléter. Il pointait ainsi des modifications provenant d'habitudes de pensée – par ex. des verbes au pluriel passant au singulier²¹ – ; d'autres justifiées par un souci stylistique²², etc. Ajoutons que Bernard est un adepte des jeux de mots et de sonorités, aussi dans ses citations bibliques²³.

Argument d'autorité, exhortation, ornement, clin d'œil... Grâce à la présence d'attributs réguliers, on interrogera à l'échelle du corpus des phénomènes qui semblaient caractérisés par leur originalité contextuelle ; tous phénomènes qui ont fait l'objet d'études ponctuelles, dont l'idée est née au fil de la lecture. Sans doute aussi pourra-t-on observer des changements en fonction des genres littéraires ou de la chronologie des œuvres.

Ajoutons un domaine que l'intégration des références scripturaires au fil d'un fichier XML va rendre plus balisable, en autorisant le croisement des analyses lexicales par mots et des repérages par références : la description de l'imprégnation biblique des auteurs citants. Dans le cas de Bernard, la frontière entre texte biblique et non biblique est spécialement difficile à déterminer. Confronté aux questions méthodologiques pour élaborer les index des *Sancti Bernardi Opera*, Jean Leclercq décrit des problèmes insolubles à partir de quelques exemples, nous en relevons un : une majorité des 150 occurrences du verbe *abeo* est biblique. Faut-il les classer toutes dans l'index du vocabulaire de Bernard, ou bien isoler celles qui proviennent des versets cités qui le contiennent²⁴ ? Mais lorsque Bernard emploie ce verbe hors des versets déterminants pour sa familiarité avec lui, s'agit-il de réminiscence biblique très implicite ? La signification spirituelle sous-jacente est-elle présente ? On ne peut comprendre les nuances de la plupart des termes employés par Bernard qu'en rapport avec un arrière-plan biblique, plus ou moins explicite. À terme, chaque

²⁰ FIGUET, « Bible », p. 265.

²¹ FIGUET, « Bible », p. 252.

²² FIGUET, « Bible », p. 253.

²³ FIGUET, « Bible », p. 254. L'ouvrage de référence sur ces questions est D. SABERSKY-BASCHO, *Studien zur Paronomasie bei Bernhard von Clairvaux*, Fribourg, 1979.

²⁴ Voir LECLERCQ, « Problèmes des *Indices* de S. Bernard », dans *Recueil d'études sur saint Bernard et ses écrits*, 5 t., Rome 1962-1992, t. IV, p. 385-400, ici p. 395.

occurrence biblique de Bernard sera encodée en descendant jusqu'au niveau du mot, avec lemmatisation par TreeTagger²⁵. Cet encodage rendra possible de caractériser chaque emploi d'un verbe comme *abeo* en fonction de son contexte. Il sera envisageable d'afficher tous les emplois, classés par type, par référence scripturaire où le terme figure ; ce type d'interrogation n'est possible ni à partir d'index scripturaires seuls, ni à partir de bases textuelles seules, c'est le croisement des deux qui est fructueux. Les graphiques *infra* montrent ce que le projet « Bible de Bernard » permet d'obtenir aujourd'hui dans l'interface TXM : d'abord une liste des occurrences du lemme « *abeo* » classées en fonction de leur source, biblique ou non, puis un exemple de retour au texte à partir de la première entrée de cette liste.

src, ref	Contexte gauche	Pivot	Contexte droit
[Vg.Ct.1.7], Sermo 34, ch. 1	ignoras te o pulchra inter mulieres egredere et	abi	post greges sodalium tuorum et pasce haedos tuos iuxta tabernacula pastorum Olim
[Vg.Ct.1.7], Sermo 35, ch. 3	Et vide quia non simpliciter egredere inquit sed	abi	post greges sodalium tuorum et pasce haedos tuos in quo ut mihi
[Vg.Ct.2.11], Sermo 33, ch. 6	certe si hoc magis probas sola tunc hiems	abii	et recessit Hunc locum inquit tantae claritatis et pacis et plenitudinis indica
[Vg.Ct.2.11], Sermo 58, ch. 8	postquam ille infidelitatis imber aliquantulum vel ex parte	abii	et recessit secuta mox est pluvia voluntaria quam segregavit Deus hereditati suae
[Vg.Ct.2.11][Vg.Ct.2.12], Sermo 58, ch. 11	vobis profecto de pluribus – hic iam imber	abii	et recessit nam et flores apparent indices pluviae suavioris Habet et aestas
[Vg.Ct.2.11][Vg.Ct.2.12], Sermo 58, ch. 12	gestetis in manibus Ergo si hiems transit imber	abii	et recessit si demum flores apparuerunt in terra nostra et subinde quaedam
[Vg.Ct.2.11][Vg.Ct.2.12], Sermo 58, ch. 2	necabat vel seri vetabat is inquam imber excurrit	abii	et recessit flores apparuerunt in terra nostra vemalem profecto temperiem adesse signantes
[Vg.lis.50.5], Sermo 28, ch. 6	mihī aurem et ego non contradico retrorsum non	abii	Ubi et voluntariae habes oboedientiae formam et longanimitatis exemplum Qui enim non
auteur, Sermo 17, ch. 2	nescimus qua hora Spiritus venturus sit seu iterum	abilturus	It et redit Spiritus et qui stat eo tenente deserente cadat necesse
auteur, Sermo 26, ch. 8	es et haec omnia simul Tecum omnes pariter	abiere	deliciae et laetitiae meae iam curae imiunt iam molestiae hinc inde pulsant
auteur, Sermo 26, ch. 8	angustiae unidique solum me reperunt solae mihi te	abeunte	remanserunt solus sub sarcina gemo Aut ponere aut opprimere necesse est quia
auteur, Sermo 28, ch. 6	non contradicit spontaneus est et qui retro non	abii	perseverat Utrumque necessarium quoniam hilarem datorem diligit Deus et qui perseveraverit usque
auteur, Sermo 35, ch. 3	non beneficit Il Quomodo post gregem pecudum homo	abit	Et vide quia non simpliciter egredere inquit sed abi post greges sodalium
auteur, Sermo 35, ch. 3	non saltem inter greges remanere permittitur sed post	abire	iubetur « Quomodo » inquis Quomodo legis Homo cum in honore esset
auteur, Sermo 35, ch. 7	secundo ignorare adiecerit Egreditur itaque homo et solitarius	abit	post greges sodalium suorum cum solus in inferno inferiori retruditur An non
auteur, Sermo 49, ch. 1	votis dulci admodum familiarique colloquio cum dilecto illo	abeunte	sponsa regreditur ad adolescentulas aspectu ita ipsius affatuque reflecta atque accensa quatenus
auteur, Sermo 54, ch. 9	pavidus Time ergo cum arriserit gratia time cum	abierit	time cum denuo revertetur et hoc est semper pavidum esse Succedant vicissim
auteur, Sermo 74, ch. 1	idque non longe ante quippe qui dum adhuc	abiret	revocari videtur Interpestiva revocato magni unius amoris magnae alterius amabilitatis indicium est
auteur, Sermo 74, ch. 1	forte hae lacrimae pariter videbuntur Revertere ait Bene	Abibat	revocatur Quis mihi huius reseret mutabilitatis sacramentum Quis mihi digne explicet ire
auteur, Sermo 74, ch. 2	libebit Restat igitur ut absentem studioso requirat revocat	abeuntem	ita ergo revocatur Verbum et revocatur desiderio animae sed eius animae cui
auteur, Sermo 74, ch. 2	Denique desiderium pauperum inquit exaudivit Dominus Verbo igitur	abente	una interim et continua animae vox continuum desiderium eius tamquam unum continuumque
auteur, Sermo 74, ch. 3	devota sibi anima sedulo actitare Praeteritis teneri vult	abiens	revocari Neque enim hoc irrevocabile Verbum it et redit pro beneplacito suo
auteur, Sermo 74, ch. 5	quidem Nam unde in animam meam venerit quove	abierit	denuo eam dimittens sed et qua introierit vel exierit etiam nunc ignorare
auteur, Sermo 74, ch. 7	a me nec cessabo clamitare quasi post tergum	abeuntis	ardenti desiderio cordis ut redeat et reddat mihi laetitiam salutaris sui reddat

Illustration 1. Liste des apparitions du lemme *abeo* dans les Sermons sur le Cantique, classées par source.

I. QUOMODO GRANDIA PETENTES AD GRADUM HUMILITATIS REVOCANTUR.

1 Si ignoras te, o pulchra inter mulieres, egredere, et abi post greges sodalium tuorum, et pasce haedos tuos iuxta tabernacula pastorum [+][9]. Olim sanctus Moyses, quando multum praesumebat de familiaritate et gratia quam invenerat apud Deum, aspirabat ad quandam visionem magnam, ita ut diceret Deo: Si inveni gratiam in oculis tuis, ostende mihi teipsum [+][9]. Accepit autem pro ea visionem longe inferioriorem, ex qua tamen ad ipsam, quam volebat, posset aliquando pervenire. Filii quoque Zebedaei in simplicitate cordis sui ambulantes [+][9], magnum aliquid et ipsi ausi sunt, sed ad gradum nihilominus sunt redacti, per quem fuerat ascendendum [+][9]. Ita et modo sponsa, quoniam rem grandem postulare [+][9] videtur, reprimitur sane austeriori responsione, sed plane utili et fideli. Oportet namque humiliter sentire de se nitentem ad altiora: ne, dum supra se attollitur, cadat a se, nisi in se firmiter per veram humilitatem fuerit solidatus. Et quia nisi humilitatis merito maxima minime obtinentur, propterea qui provehendus est correptione humiliatur, humilitate meretur. Tu ergo, cum te humiliari videris, habeto id signum in bonum [+][9] omnino argumentum gratiae propinquantis. Nam sicut ante ruinam exaltatur cor [+][9], ita ante exaltationem humiliatur. Sane utrumque legis, Dominum scilicet et superbis resistere, et humilibus dare gratiam [+][9]. Nonne denique servum suum Iob [+][9], cum post insigne triumphum, tantam et tam probatam ipsius patientiam larga remunerandam benedictione censeret, prius in multis et districtis percunctationibus humiliare curavit, et sic parare viam benedictioni?

a. Ct. 1. 7 #

Requête: [lemma="abeo"]

Clés de tri: #1 Références #2 Aucun #3 Aucun #4 Aucun Tri

src, ref	Contexte gauche	Pivot	Contexte droit
[Vg.Ct.1.7], Sermo 34, ch. 1	ignoras te o pulchra inter mulieres egredere et	abi	post greges sodalium tuorum et pasce haedos tuos iuxta tabernacula pastorum Olim
[Vg.Ct.1.7], Sermo 35, ch. 3	Et vide quia non simpliciter egredere inquit sed	abi	post greges sodalium tuorum et pasce haedos tuos in quo ut mihi

Illustration 2. Retour au texte de Sct 34, 1 à partir d'une occurrence du lemme *abeo*.

²⁵ Le lemmatiseur de latin de TreeTagger a été entraîné par G. Bandolini sur le corpus (classique) de Perseus, mais il est rarement pris en défaut sur le texte de Bernard.

IV. – L'INTERROGATION DU CORPUS

Reste maintenant à accéder à l'ensemble des éléments patiemment décrits, pour en tirer des informations au niveau macroscopique. L'encodage fin n'est d'une certaine façon, localement, qu'une traduction en XML de méthodologies d'analyses littéraires très classiques. Il prend sens lorsqu'on peut observer des régularités ou des évolutions à l'échelle d'un corpus, phénomènes indécélables dans une lecture myope.

A. Le choix des outils

Pour ce faire, nous devons recourir à des outils exploitant des sources statistiques aussi bien chiffrées que textuelles, et offrant des visualisations susceptibles de faire jaillir de l'information supplémentaire. Pour les approches quantitatives, menées à partir des références exprimées canoniquement, nous nous servons pour l'heure du tableur Excel ; pour celles qui combinent quantitatif et qualitatif, nous avons trouvé en TXM, puissant logiciel de textométrie développé depuis 2007 à l'École Normale Supérieure de Lyon²⁶, dont la vocation est d'appliquer des méthodes lexicométriques à des corpus enrichis et structurés, un outil très performant. TXM réalise des analyses qualitatives à l'aide du langage CQL : des concordances – recherche autour d'un mot ou d'une référence des portions de textes apparaissant avant ou après – ; des analyses de fréquences – de formes, de références, de structures textuelles, de caractéristiques morphosyntaxiques en fonction de l'étiquetage des mots du corpus ; des graphiques de progression mettant en évidence l'évolution chronologique de l'emploi d'un mot ou d'une référence. D'autres outils d'analyse quantitative sont basés sur des modules R : calcul de spécificités représentées en histogrammes ; recherches de cooccurrences de mots ou de versets, avec coefficient de proximité paramétrable ; analyse factorielle des correspondances ; classification ascendante hiérarchique. En outre, et c'est pour nous fondamental, TXM permet un retour au texte source, offrant une visualisation du contexte chaque fois que c'est nécessaire, par navigation hypertextuelle au sein d'éditions enrichies en HTML, issues de l'encodage TEI fourni.

B. Le choix des questionnements

Le fonctionnement des outils est conditionné par la pertinence des questions de recherche posées. Nous donnons ici une liste indicative de celles que nous souhaitons adresser au corpus de Bernard et qui pourraient être appliquées à d'autres corpus :

²⁶ Voir S. HEIDEN, J.-P. MAGUE, B. PINCEMIN, « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », dans *Proceedings of 10th International Conference on the Statistical Analysis of Textual Data*, I. C. Sergio Bolasco (éd.), Journées internationales d'Analyse statistique des Données Textuelles (JADT), vol. 2, Rome, 2010, p. 1021-1032 [<https://halshs.archives-ouvertes.fr/halshs-00549779>].

1) *Description du corpus d'occurrences*

- nombre total d'occurrences scripturaires ;
- rapport Ancien/Nouveau Testament ;
- longueur totale en mots des passages issus de la Bible, rapporté au nombre de mots de l'œuvre ; mêmes calculs à l'échelle de groupements de livres et des livres bibliques ; affinage et, sur des ensembles bibliques sélectionnés, en fonction des résultats obtenus, calculs à granularité plus fine ; mêmes calculs en étudiant des sous-corpus de l'œuvre, à géométrie variable ;
- liste des livres absents du corpus, éventuellement liste de passages significatifs absents ; place des deutérocanoniques, des apocryphes ;
- part (numérique et avec le détail des versets) de chaque livre biblique utilisé par l'auteur citant ; reconstitution linéaire de la Bible d'un auteur, sur tout, sur un livre, sur un chapitre : visualisation des passages absents ou faiblement cités, et inversement des pics incitant à un retour ciblé au texte ;
- établissement de classements : les passages les plus cités, les moins cités, là encore avec une granularité modulable ;
- affinage éventuellement, en fonction des résultats, mêmes calculs avec des granularités plus fines ;
- reprise des acquis précédents en faisant jouer le critère du genre littéraire ; des thèmes ou mots-clefs ; des destinataires ;
- répartition chronologique des occurrences ;
- croisement des critères de genre et de date.

2) *Description des versions bibliques utilisées*

L'analyse fine de la lettre des citations bibliques littérales, voire la teneur de certaines allusions, nous renseigne sur l'état du texte biblique à la disposition de l'auteur citant.

3) Description de la manière de citer

- calcul d'un taux de références bibliques et/ou de mots bibliques sur des extensions paramétrables de corpus ;
- rapport entre le nombre d'occurrences d'une section et le nombre de ses parties citées, par exemple un livre et les versets qui le composent ;
- repérage des cooccurrences de versets et/ou de mots bibliques, selon des critères d'éloignement et d'exactitude paramétrables ;
- classement des occurrences en fonction de leur longueur ; de la présence ou non de mots introductifs ; de leur caractère explicite ou implicite ; de leur caractère plus ou moins allusif, de leur typage comme lemme ou reprise de lemme ;
- révision des résultats obtenus en faisant jouer les critères de genre, de thème, de date, de longueur des textes ;
- étude en fonction des modalités d'insertion dans la phrase ;
- analyse des contextes, du rôle sémantique des textes bibliques.

Dans le cas d'un corpus à auteurs multiples, on effectuera des comparaisons. Par ailleurs l'approche géographique et chronologique, ou par langue rédactionnelle, prendra tout son sens.

Dans la mesure où chacune de ces interrogations du corpus est paramétrable, les possibilités sont à peu près infinies. Si l'on peut dans un premier temps tenter l'investigation la plus objective possible, en observant à grande échelle l'ensemble des phénomènes, il faudra dans un second temps restreindre le champ d'analyse et choisir des parcours d'enquêtes spécifiques. Dans la constitution des données qui seront le fondement des analyses, quatre points d'attention sont à noter : la définition de seuils ; le choix des granularités²⁷ ; le choix des

²⁷ Prenons un exemple sur le corpus de Bernard : d'après HENDRIX, *Index biblicus*, Bernard cite 1469 fois à la *Genèse*, cela constitue 11,9 % de ses citations de l'Ancien Testament. Or la *Genèse* représente 5% de l'Ancien Testament en nombre de mots, donc Bernard la cite environ deux fois plus qu'on ne pourrait l'attendre – si tant est que la longueur des livres bibliques puisse être le seul critère de pondération. Mais si l'on ne prend en compte que les chapitres 1-4 de la *Genèse*, que Bernard cite 571 fois, ce sont 4,6 % des citations de l'Ancien Testament, alors que ces chapitres constituent 0,3 % de l'Ancien Testament en nombre de mots, donc l'intérêt spécifique de Bernard est bien plus remarquable (15 fois plus que la « norme »), ce qui passait inaperçu avec une granularité à l'échelle du livre. Les biais sont encore très sensibles à l'échelle du verset : Bernard cite 61 fois le livre de *Baruch*, et 23 versets différents. Si on calcule la moyenne d'emploi de chacun de ces versets, on arrive à 2,9 fois par verset. Mais en réalité,

pondérations : longueur des textes, genre, etc. ; le choix des points de comparaison.

V. – LA VISUALISATION DES RESULTATS

Il faut enfin rendre partageables les résultats. Nous donnons ci-dessous quelques exemples de visualisations déjà obtenues à partir du corpus de Bernard.

A. La visualisation des données statistiques

1) Les tableaux

La forme la plus basique des visualisations statistiques est le tableau de chiffres.

	Ancien Testament	Nouveau Testament
% occurrences ²⁸	46,23	53,77
% versets cités / total des citations ²⁹	56,65	43,35
% versets cités / total des versets du corpus biblique ³⁰	18,67	50,07
nb moyen d'apparitions d'un verset cité	3,07 ³¹	4,67

Dans celui présenté ci-dessus, on voit qu'il y a dans l'œuvre de Bernard nettement plus d'occurrences du Nouveau Testament que de l'Ancien ; en nombre de versets cependant, ce rapport s'inverse, puisque le Nouveau Testament ne représente que 43,3% des versets bibliques cités. L'ensemble des chiffres donnés ci-dessus vont cependant dans le même sens : le Nouveau Testament est très nettement plus utilisé que l'Ancien par Bernard : des occurrences plus nombreuses, une part du Nouveau Testament citée proportionnellement presque 3 fois plus grande que celle de l'Ancien, un nombre moyen d'occurrences par verset du Nouveau Testament présent dans le corpus 1,52 fois plus élevé que pour l'Ancien. Cette conclusion est incontestable, et Bernard se situe ici au-dessus de la moyenne patristique (48/52), mais il a fallu l'ensemble de ces approches pour le montrer.

Cependant la présentation en tableau ne fait pas ressortir les chiffres les plus significatifs. On leur préférera souvent les graphiques, qui offrent des résumés visuels, facilitant au lecteur le repérage d'un sens. Nous en avons principalement

Ba 3, 38 (*post haec in terris visus est et cum hominibus conversatus est*), appliqué au premier avènement du Seigneur, l'Incarnation, est cité 28 fois et fournit à lui seul presque la moitié des occurrences.

²⁸ (Nombre d'occurrences venant de l'AT ou du NT/nombre total d'occurrences) x100.

²⁹ (Nombre de versets différents venant de l'AT ou du NT/nombre total de versets différents cités) x100.

³⁰ (Nombre de versets cités / nombre de versets de l'AT ou du NT) x100.

³¹ Nombre d'occurrences de l'AT ou du NT / nombre de versets de l'AT ou du NT.

réalisé quatre sortes : les diagrammes sectoriels, les histogrammes, les courbes et les radars³².

2) Les diagrammes sectoriels

Les diagrammes sectoriels – tout comme les diagrammes en aires empilées cumulatifs – sont bien adaptés à la visualisation de sous-corpus pas trop nombreux dans un corpus total symbolisé par le disque complet.

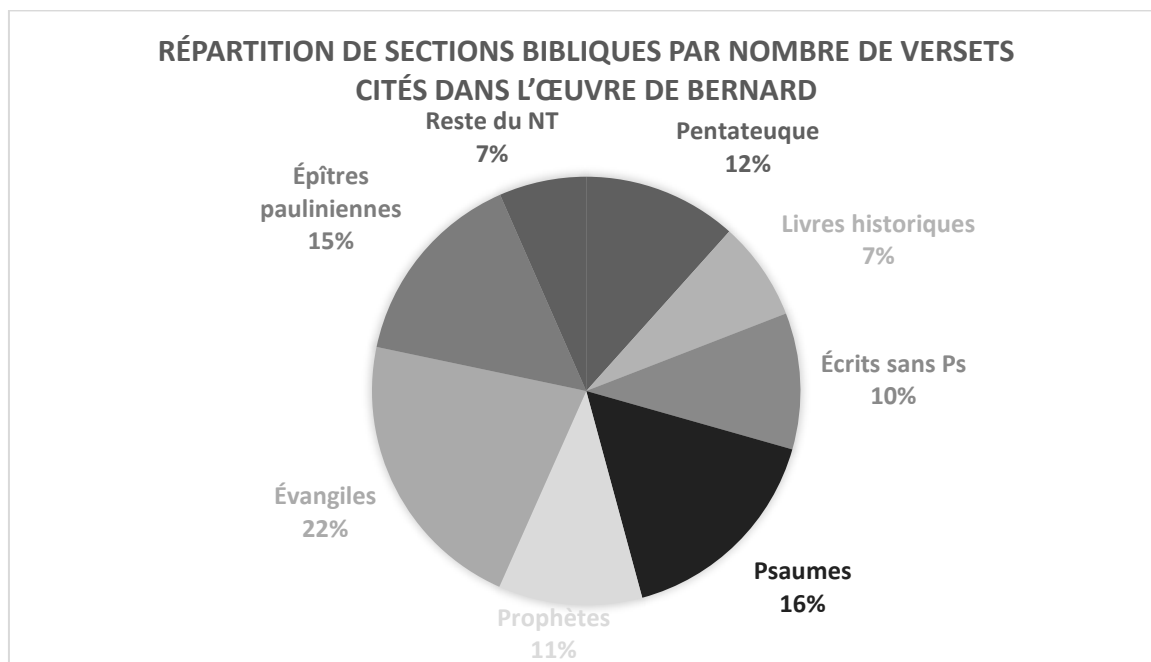
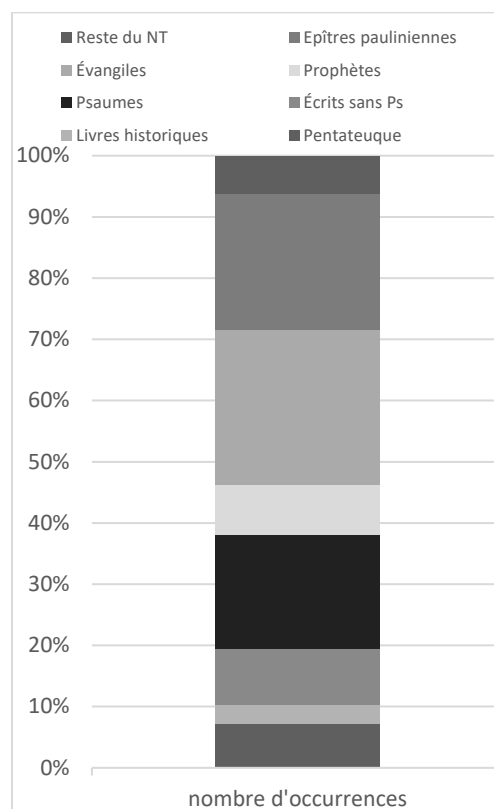
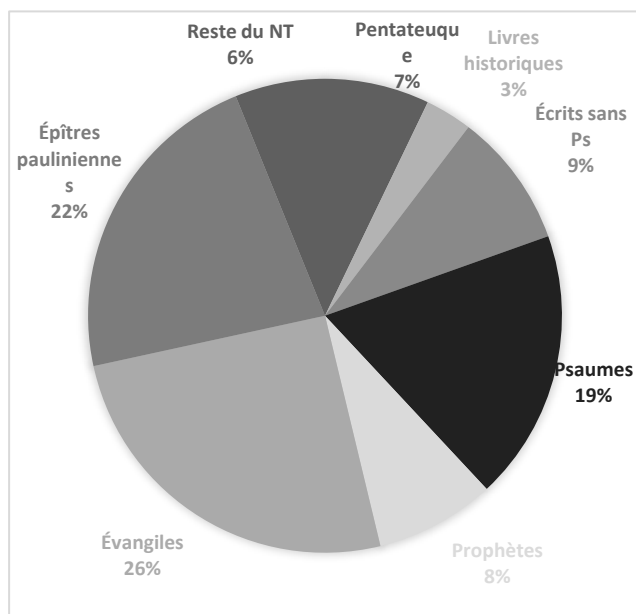


Illustration 3. Répartition de sections bibliques par nombre de versets dans l'œuvre de Bernard.

³² Nous appuyons ici tous nos exemples sur des chiffres réels : index scripturaire des *Sancti Bernardi Opera* et/ou corpus textuel des *Sermons sur le Cantique* tel qu'importé dans TXM dans sa version 1 (octobre 2017).



Illustrations 4 et 5. Répartition de sections bibliques par nombre d'occurrences dans l'œuvre de Bernard.

3) Les « histogrammes »

La comparaison de diagrammes en aires empilées cumulatifs sur des sous-corpus présente l'avantage de pouvoir intégrer une estimation de la répartition sur le total, qui fait office de moyenne normative. Par le classement en ordre croissant ou décroissant, la position de chaque section d'œuvre par rapport à la moyenne est immédiatement visualisable. On en a une illustration dans les deux graphiques suivants. Le premier met en évidence une grande homogénéité dans la distribution des sections bibliques principales dans les différents groupements d'œuvres de Bernard, mais aussi la présence d'un élément atypique, à savoir les *Sentences* et les *Paraboles*, qui comportent une part beaucoup plus importante de références au Pentateuque et au reste de l'Ancien Testament que l'ensemble, mais beaucoup moins de renvois aux épîtres de Paul.

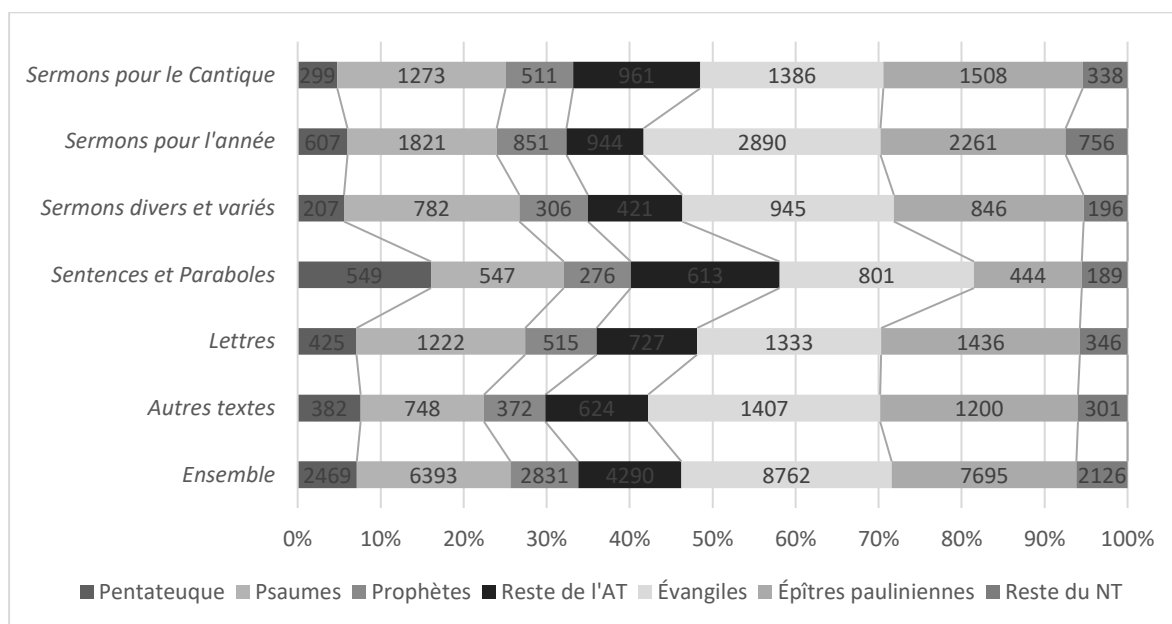


Illustration 6. Répartition en pourcentage d'occurrences des principales sections bibliques utilisées, classées par groupements d'œuvres de Bernard.

En isolant l'élément atypique, on met mieux en évidence des phénomènes plus discrets concernant les autres groupes d'œuvres, d'où le second graphique, après élimination des *Sentences* et des *Paraboles*. On constate alors que les renvois aux *Psaumes* sont notablement moins fréquentes dans les traités et opuscules, qui citent, tout comme les *Sermons pour l'Année*, davantage les Évangiles que la moyenne ; cette observation n'est pas valable pour les *Lettres*.

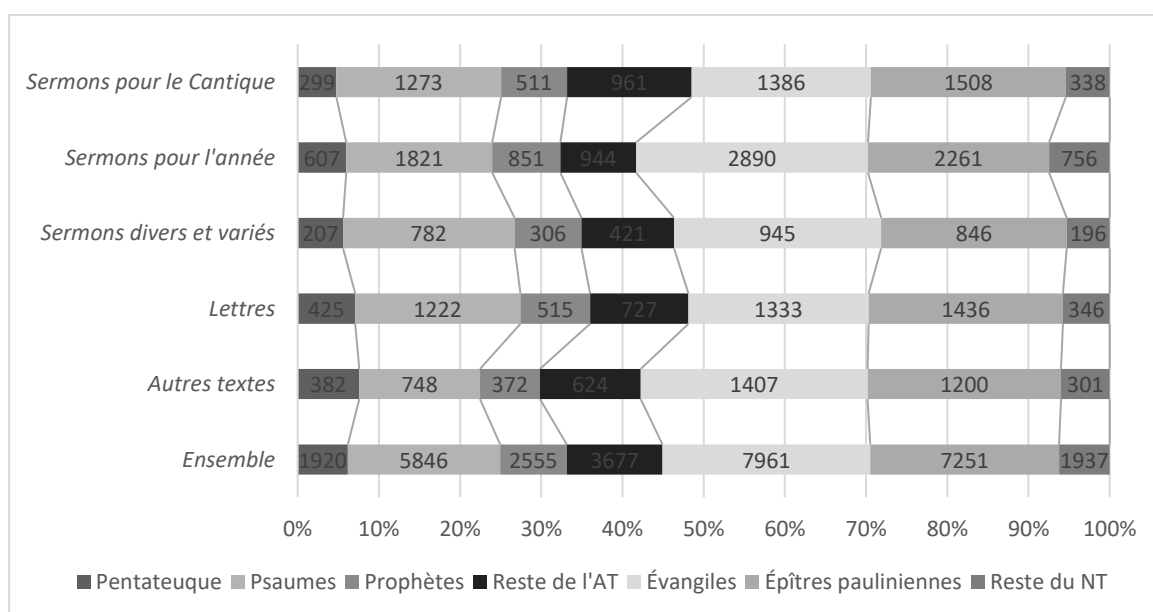


Illustration 7. Même graphique, sans les *Sentences* et les *Paraboles*.

Lorsque le rapport au tout n'est pas essentiel, il est préférable d'utiliser les autres histogrammes, à colonnes unique ou multiples pour une variable donnée. Plus lisibles pour l'œil humain que les secteurs, ils sont pratiques dès que les sous-corpus sont nombreux.

Ils peuvent s’appréhender de gauche à droite, dans le mouvement de la lecture. Si l’axe des ordonnées représente toujours dans nos études un nombre d’occurrences, les variables sur l’axe des abscisses, presque toujours discrètes, sont le plus souvent non numériques, comme les noms de sections bibliques, qui elles aussi peuvent se dérouler dans l’ordre du Livre. Nous illustrons ce principe – en hauteur faute de place en largeur – dans le diagramme récapitulant les usages de livres bibliques dans l’œuvre de Bernard donné *infra*. En faisant implicitement de l’axe des abscisses une flèche temporelle, un paramètre chronologique peut intervenir.

Dans les interfaces TXM du projet « Bible de Bernard », les histogrammes simples donnent à voir les spécificités, c’est-à-dire les calculs de probabilité d’apparition d’un mot ou d’un lemme, par rapport à une probabilité neutre, normative. Est alors utilisée aussi la partie négative de l’axe des ordonnées. Nous avons choisi les exemples de mots très courants, *anima* et *spiritus* : l’absence de régularité dans les usages que fait Bernard est clairement manifestée. Et si les valeurs positives élevées permettent de mesurer la sur-représentation d’un terme, souvent perceptible à la lecture – ainsi, *anima* dans les *Sermons* 30, 45, 80-82, 84-85 – ; les valeurs négatives en revanche rendent compte d’un phénomène très difficilement perceptible : la sous-représentation d’un terme – pour *anima*, dans les *Sermons* 13 et 72. Le repérage des mots « bibliques » est également facilité : dans notre exemple, le lemme *osculum* apparaît très majoritairement dans les *Sermons* 2-4 et 7-9, avec un pic dans le *Sermon* 8. Ce type de graphique peut être très utile pour repérer les échos d’un verset sous forme de réminiscence d’un lemme-clef, ou bien, sur une étendue textuelle plus restreinte, pour examiner comment l’introduction du verset se fait progressivement.



Illustration 8. Distribution des occurrences du lemme *anima* dans les SCT.

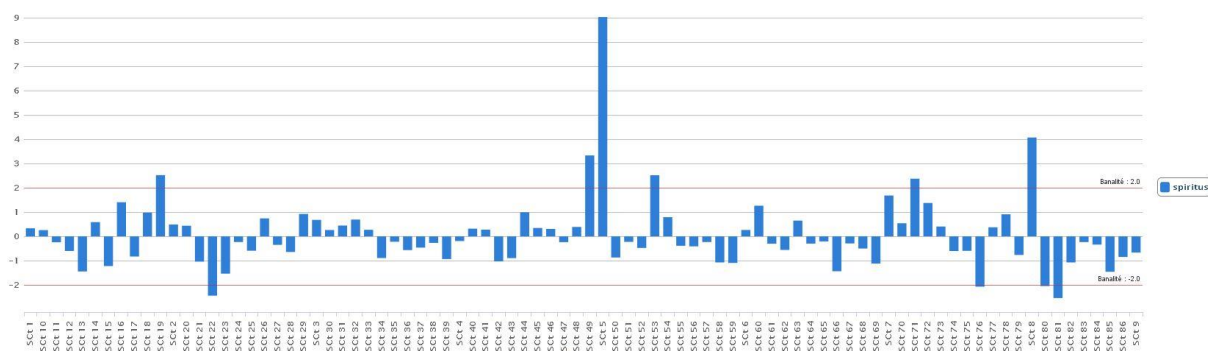


Illustration 9. Distribution des occurrences du lemme spiritus dans les SCT³³.

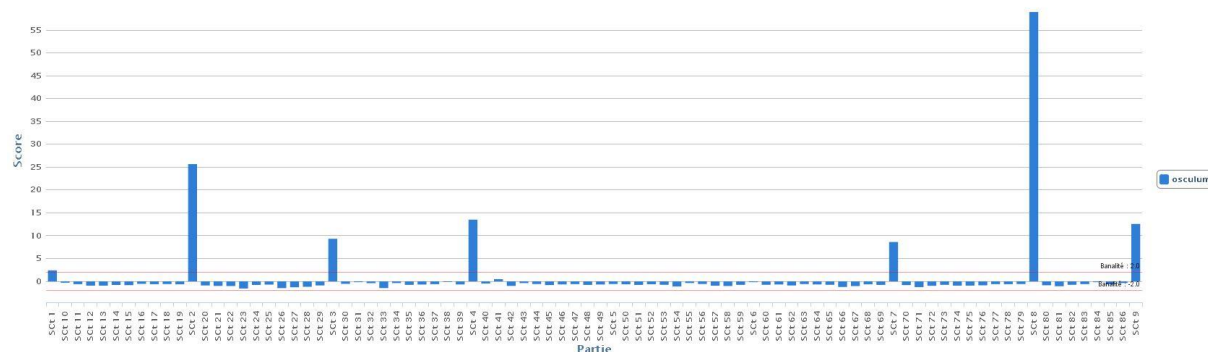


Illustration 10. Distribution des occurrences du lemme osculum dans les SCT³⁴.

Dans les cas que nous venons d’illustrer, les colonnes sont uniques pour une valeur donnée en abscisse. Mais la possibilité d’afficher plusieurs colonnes côte à côte pour une même valeur, ou de compléter le graphique par une courbe de fréquence, enrichit les visualisations possibles.

Cette première tentative pour décrire chronologiquement la Bible de Bernard en est une illustration.

³³ Attention, ce sont les *Sermons* 5 et 8 qui présentent des pics, l’ordre des bâtons n’est pas exactement celui des *Sermons* car TXM classe ici les nombres de façon alphanumérique.

³⁴ Même remarque.

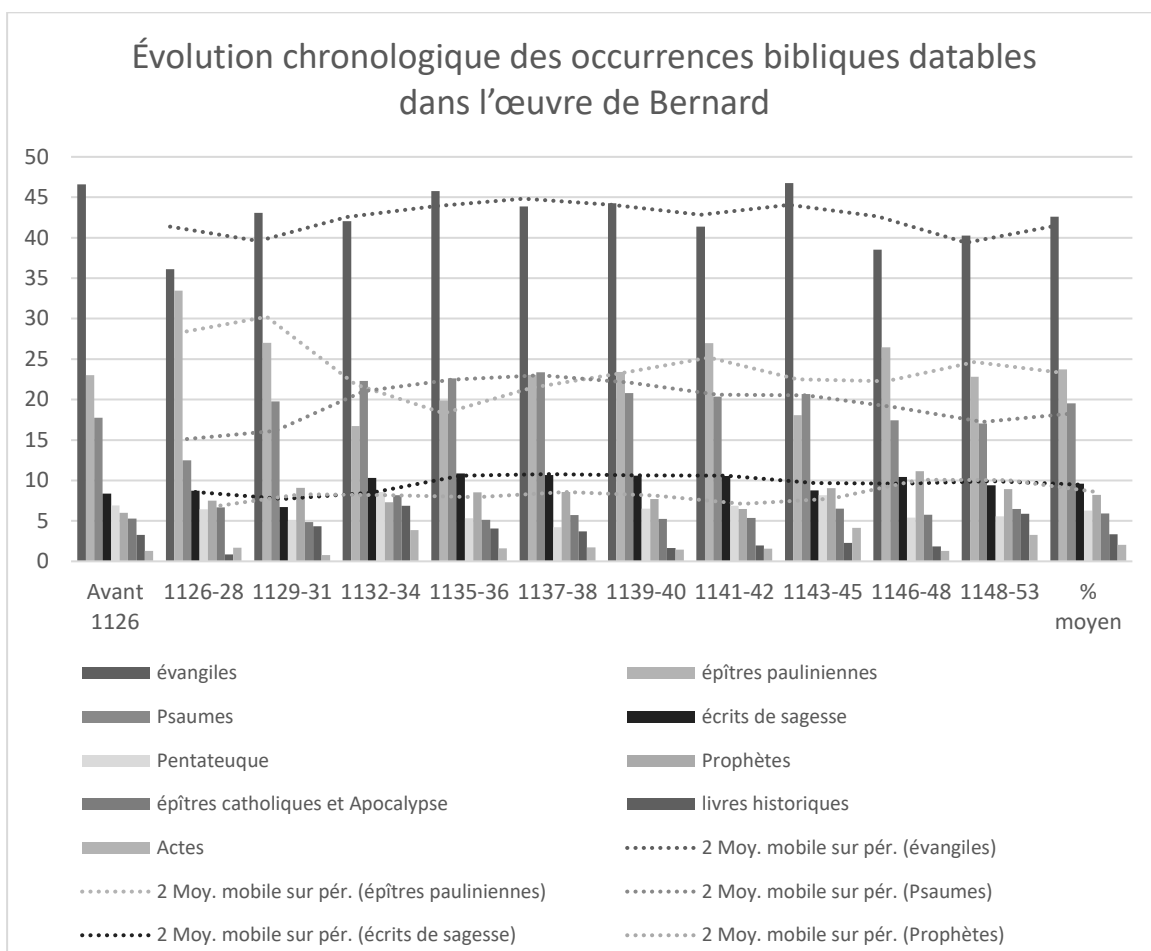


Illustration 11. Évolution chronologique des occurrences bibliques datables dans l'œuvre de Bernard.

On peut étudier chaque variable prise séparément, et constater par exemple le reflux significatif des *Psaumes* et des épîtres pauliniennes à la fin de la vie de Bernard. Mais le graphique donne aussi un aperçu du profil de chaque période.

Pour prendre un autre exemple, si nous poursuivons notre étude du lemme *osculum*, il est possible d'interroger les termes cooccurrents dans les *Sermons* qui ont été repérés comme pertinents. Le graphique ci-dessous montre les liens entre *osculum* et *pes*, *manus*, *os* dans les *Sermons* qui commentent Ct 1, 1, *osculatur me osculo oris sui*.

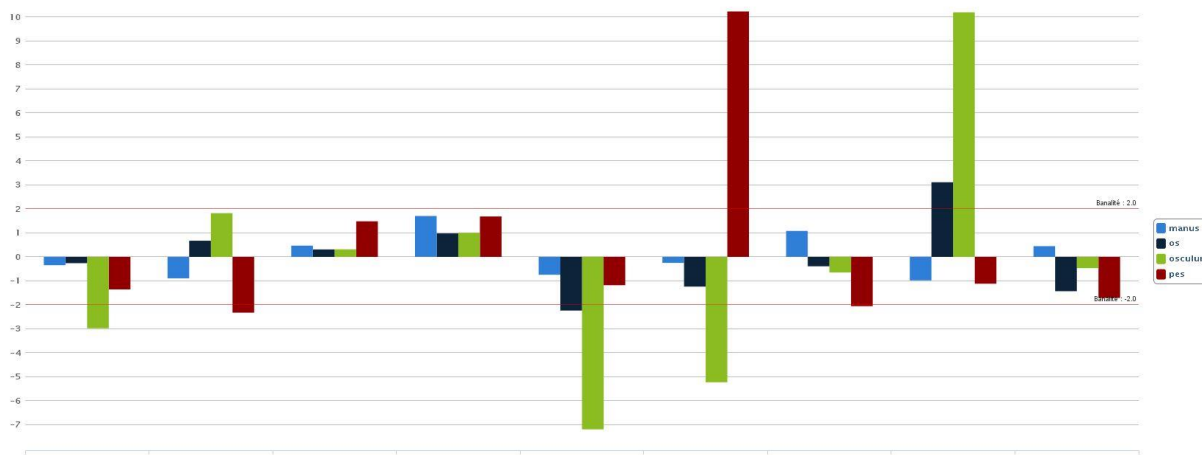


Illustration 12. Distribution des occurrences des lemmes osculum, manus, os, pes dans les SCt 1-9.

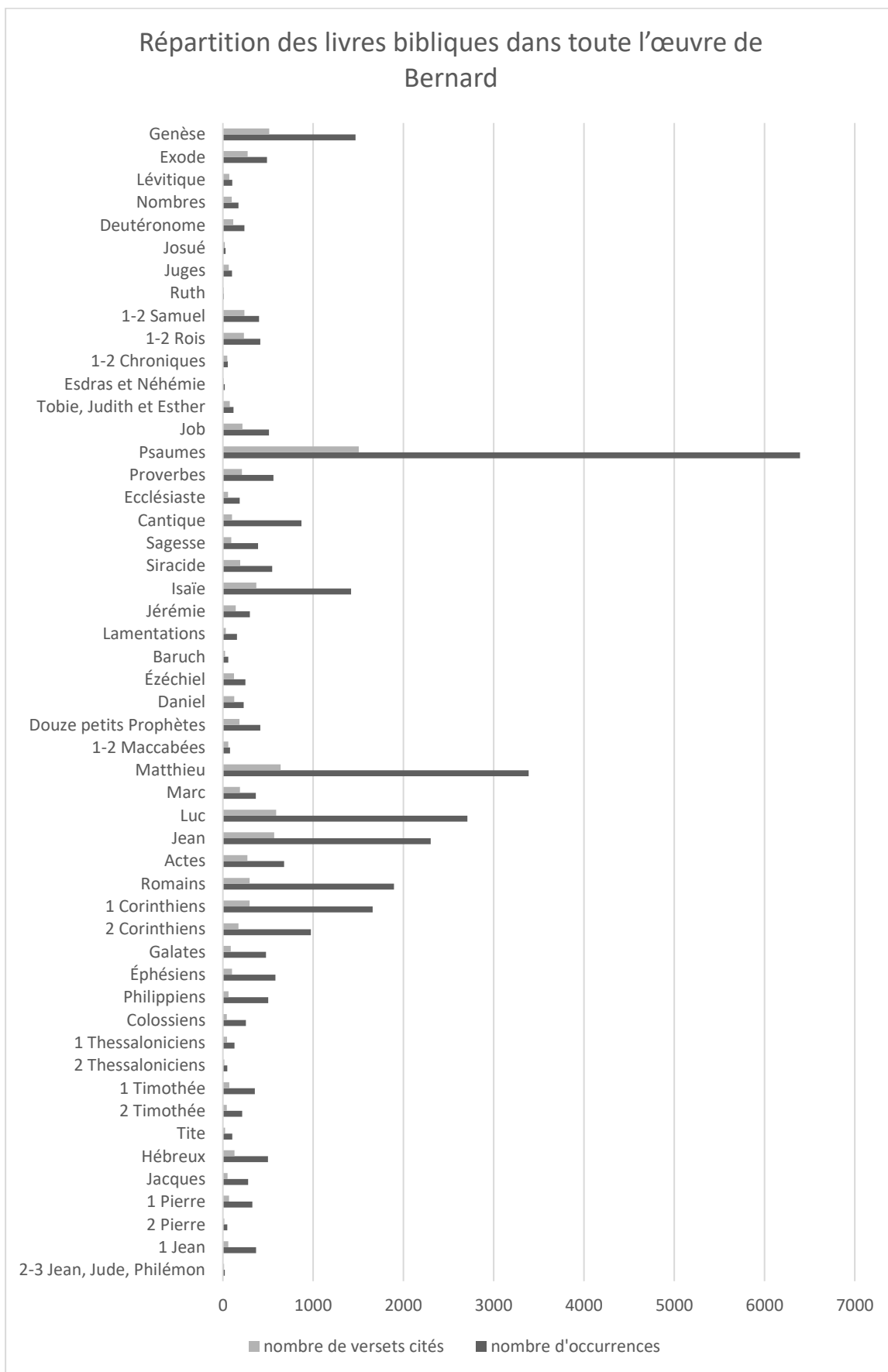
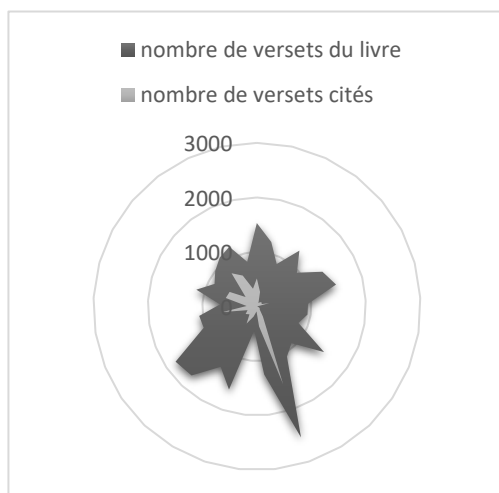


Illustration 13. Répartition des livres bibliques dans l'œuvre de Bernard

4) *Les diagrammes en radar*

Tout comme les histogrammes groupés avec référence à une moyenne, les radars mettent en évidence la différence entre des faits et une norme ; ils sont parlants lorsqu'on fait varier deux paramètres en même temps, comme la longueur du texte et le nombre de citations. Leur avantage est leur caractère compact, leur inconvénient une lisibilité moins intuitive que celle de l'histogramme.

Illustration 14. Représentation en radar de la Bible de Bernard.



En voici un exemple. Est représentée ci-contre la part des versets bibliques (en gris clair) cités par Bernard. L'inclusion d'une forme dans l'autre donne une idée des proportions : Bernard cite certes énormément la Bible, mais, et de très loin, pas toute la Bible (ici la forme englobante). En agrandissant proportionnellement la Bible de Bernard, pour gagner en lisibilité, nous pouvons comparer sa forme à celle de la Bible complète, proportionnelle aux longueurs respectives des

livres en versets. Si Bernard citait tous les livres de façon équilibrée, la forme de sa Bible serait une reproduction en miniature de celle de la Bible complète. Très nettement, ce n'est pas le cas. S'il suit le pic des *Psaumes*, il ne suit pas du tout celui des livres historiques ; il cite Isaïe ou les épîtres pauliniennes beaucoup plus que proportionnellement.

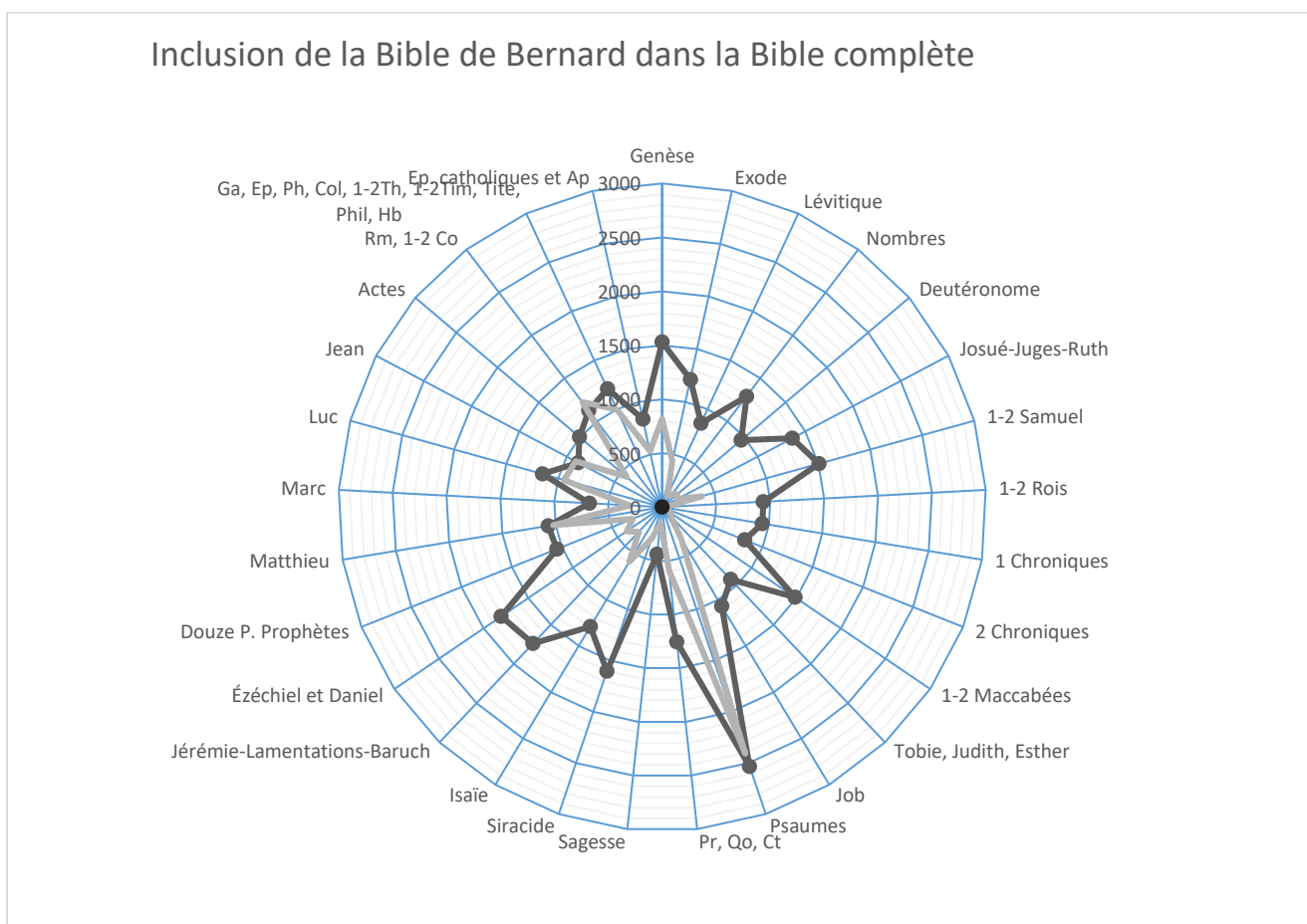


Illustration 15. Représentation en radar de la Bible de Bernard (zoom).

5) Les courbes

Quant aux courbes – ou tout au moins les successions de points reliés –, elles sont particulièrement appropriées pour les progressions chronologiques. Il est également possible de faire figurer en abscisse un nombre de versets par occurrence, ou un nombre d'occurrences par versets ; ou encore des numéros de versets, pourvu que le passage étudié ne soit pas trop long.

Pour illustrer notre propos, nous donnons quelques extraits d'une première exploitation de nos informations sur les œuvres datables de Bernard.

Les *Psaumes* sont de loin le livre qu'il cite le plus. La courbe montre cependant que leur usage n'est pas aussi immédiatement prégnant qu'on pourrait le penser chez un auteur monastique. C'est seulement après 1132 qu'ils se déploient le plus abondamment, pour refluer en proportion après 1143, le bagage biblique de Bernard s'étant diversifié.

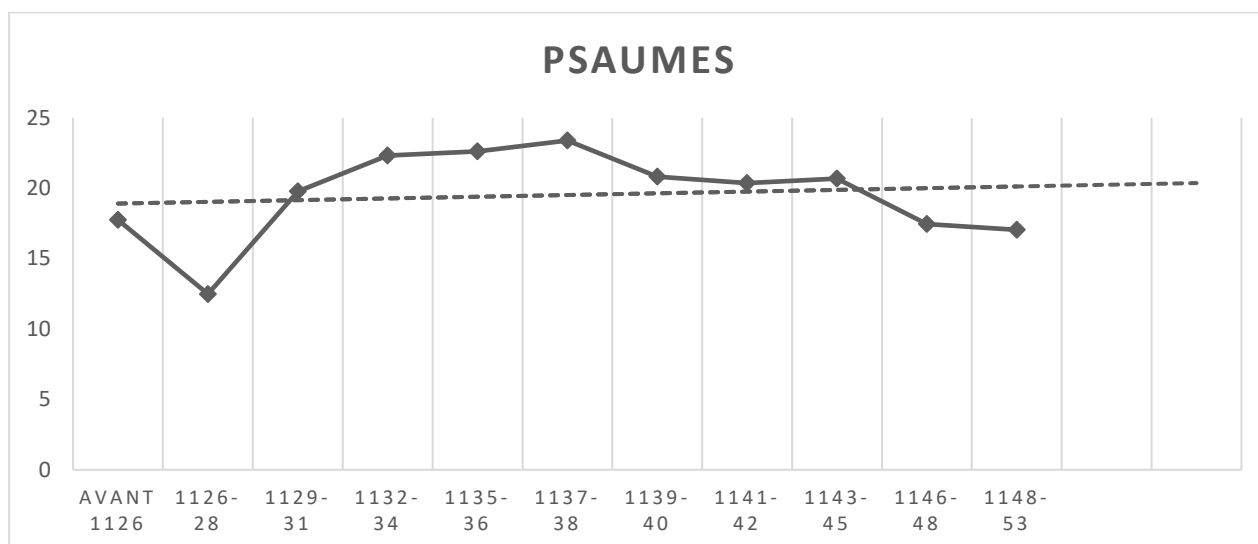


Illustration 16. Évolution en % des citations des Psaumes dans les œuvres datables de Bernard.

En ce qui concerne les évangiles, si la nette sous-représentation de *Marc*, cohérente avec le peu de goût de Bernard pour les récits et classique dans les écrits patristiques, est constante, les proportions respectives des trois autres livres varient considérablement. Chaque rupture des courbes méritera explication, mais l'orientation des linéaires est claire : aux environs de 1140, Bernard délaisse *Matthieu* et *Luc* au profit de *Jean*. De la même façon, même si la tendance est moins marquée, les usages de *2 Corinthiens* tendent à augmenter à peu près à la même période, tandis que les recours à *Romains* et *1 Corinthiens* diminuent. L'illustration 19 montre la nette prépondérance de la *Genèse* et d'*Isaïe* parmi les autres livres vétérotestamentaires, là encore un phénomène classique chez les Pères, et qui va en s'accroissant au fil de la vie de Bernard. Les échos de l'*Exode*, abondants avant 1136, diminuent considérablement ensuite. Quant aux écrits sapientiaux, indépendamment de la fulgurante et évidente montée du *Cantique* après 1136, leur répartition par livre est loin d'être homogène ; il faudra sans doute croiser ces graphiques avec d'autres paramètres – genre, destinataire – pour comprendre les évolutions bernardines.

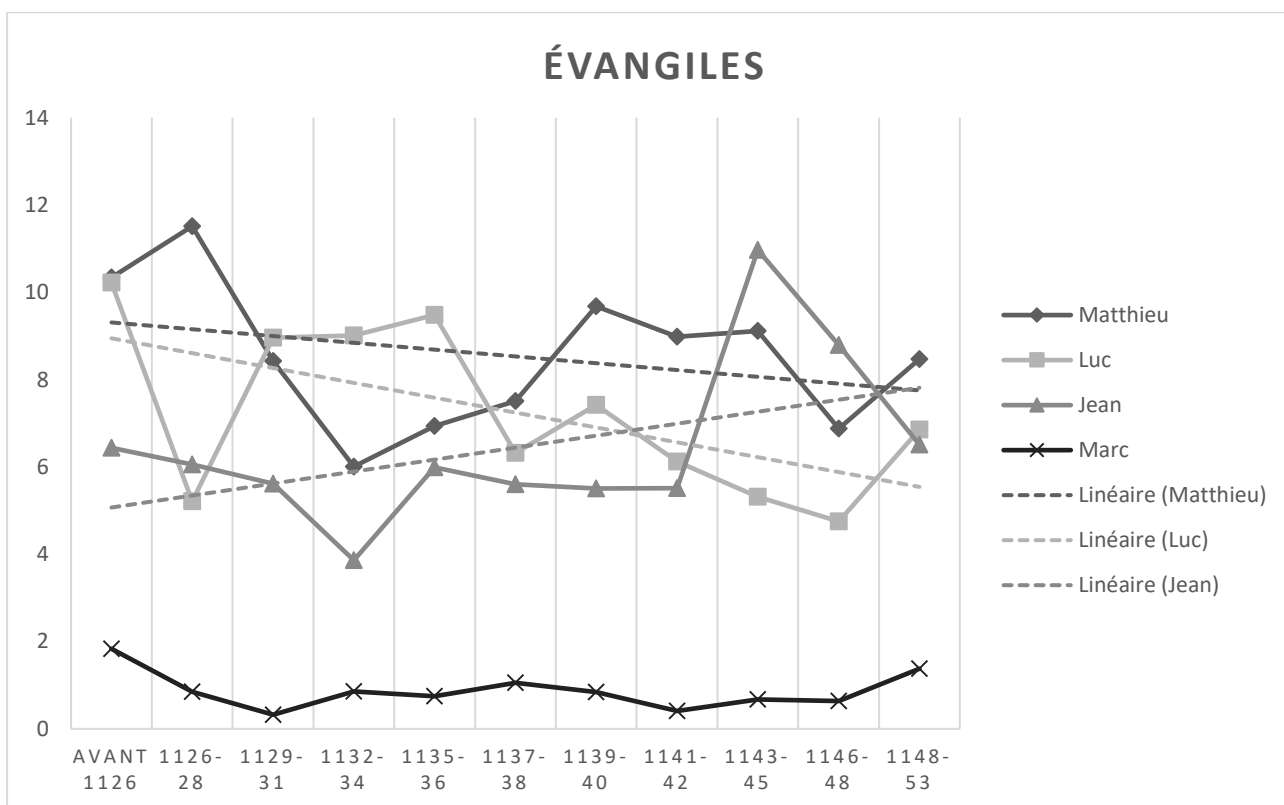


Illustration 17. Évolution en % des citations des Évangiles dans les œuvres datables de Bernard.

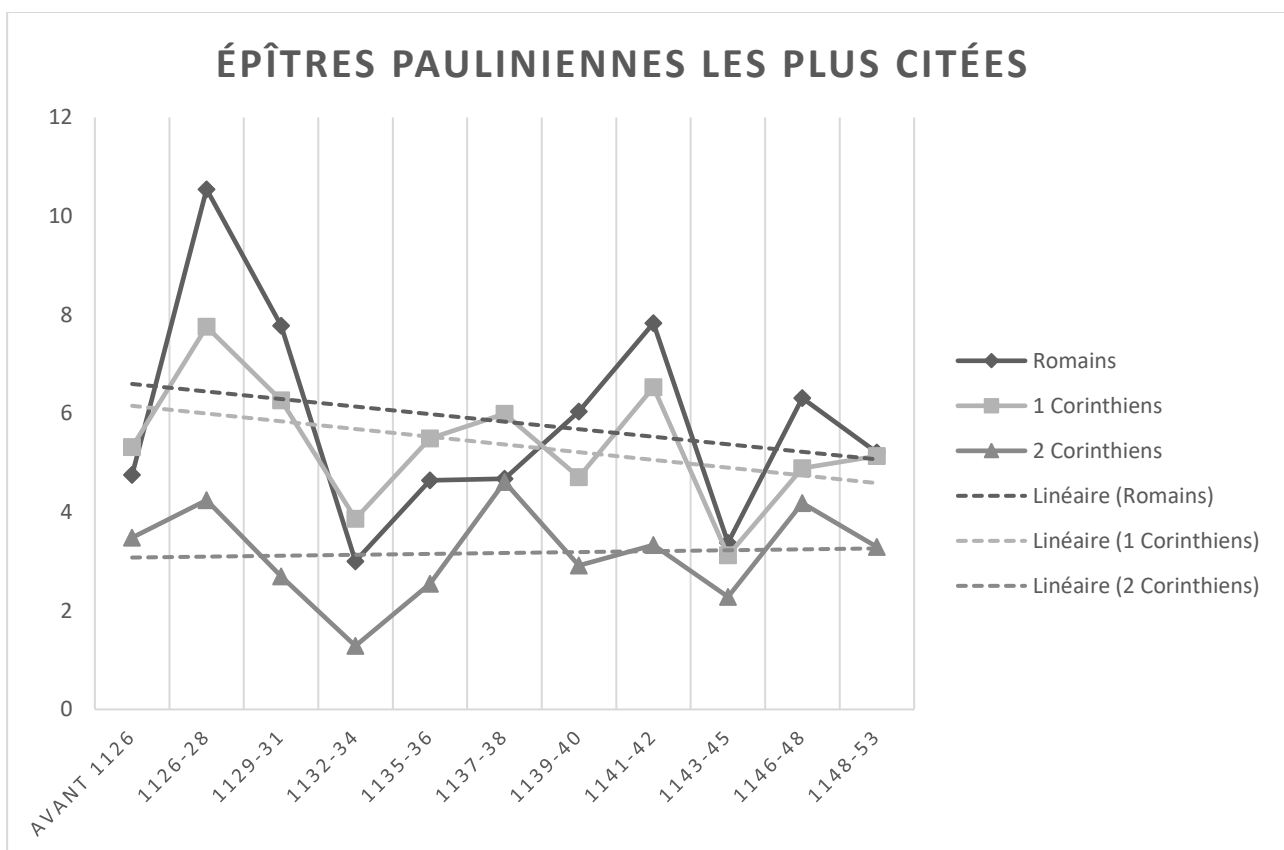


Illustration 18. Évolution en % des citations des Épîtres pauliniennes les plus citées dans les œuvres datables de Bernard.

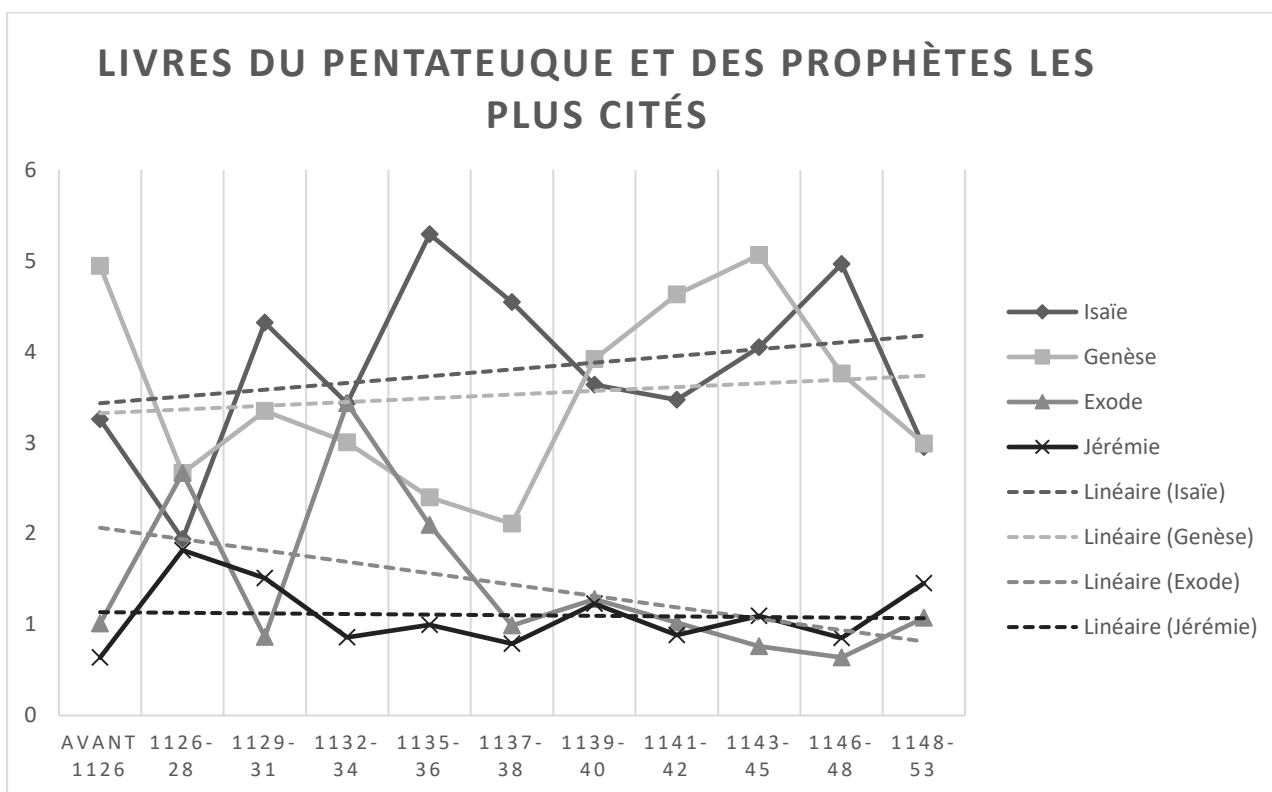


Illustration 19. Évolution en % des citations du Pentateuque et des Prophètes très cités dans les œuvres datables de Bernard.

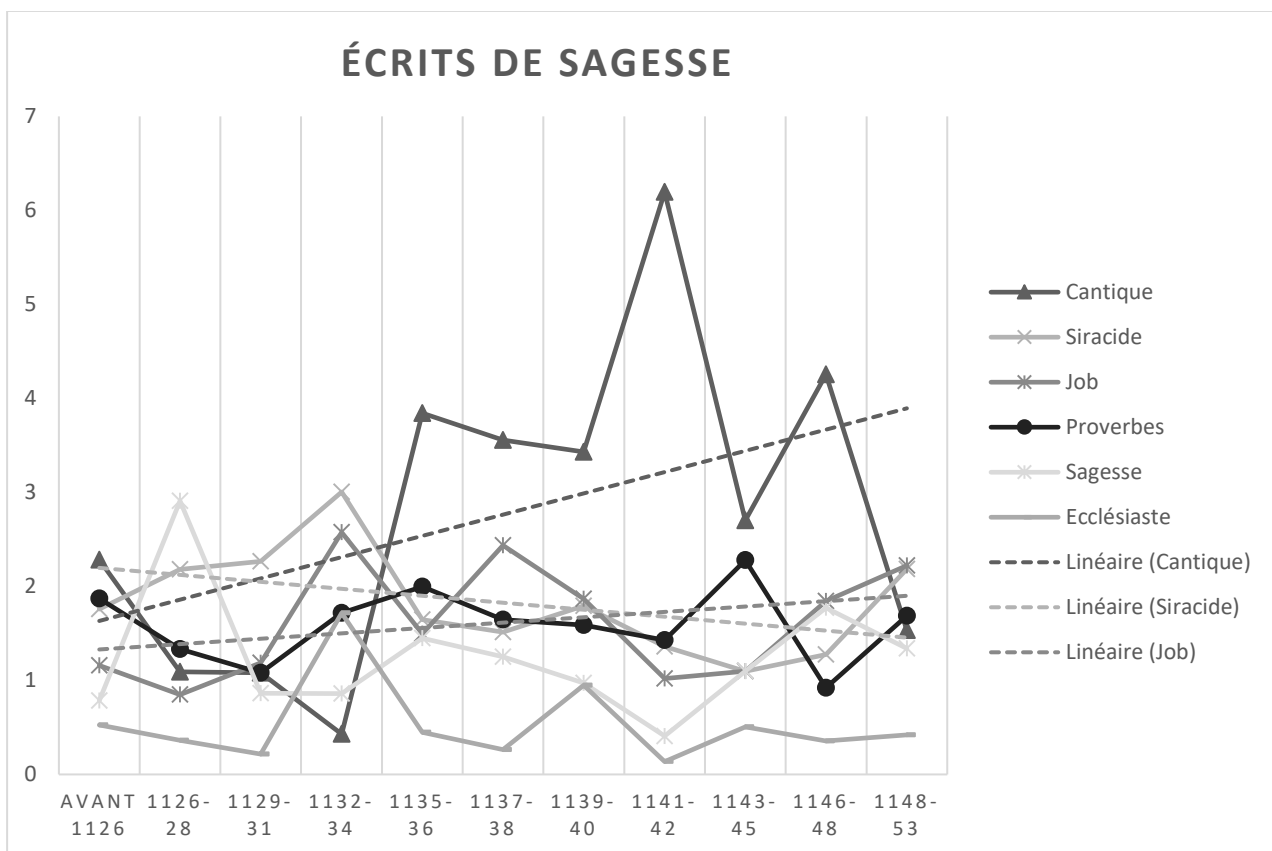


Illustration 20. Évolution en % des citations des écrits de sagesse dans les œuvres datables de Bernard.

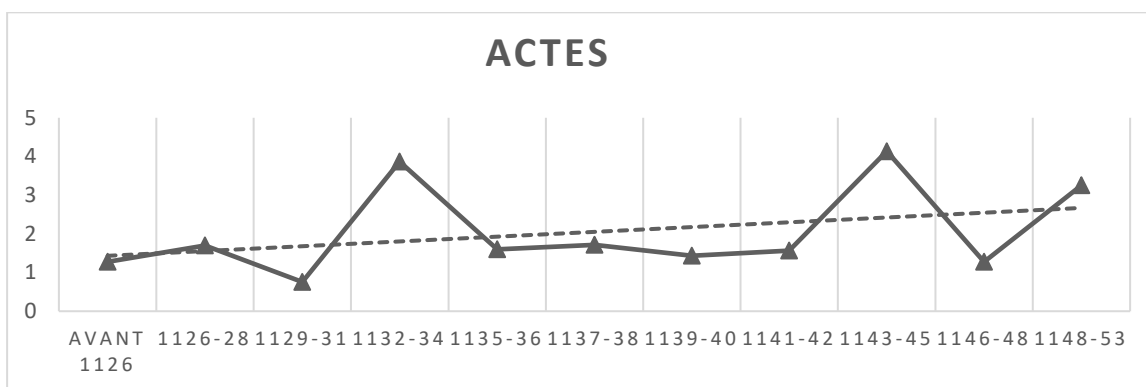


Illustration 21. Évolution en % des citations des Actes des Apôtres dans les œuvres datables de Bernard.

Dans TXM, nous avons la possibilité d'effectuer des progressions, c'est-à-dire des représentations cumulées d'effectifs. Une progression est nécessairement une courbe ascendante, puisqu'elle se construit par addition des mots bibliques au fur et à mesure qu'on avance dans le corpus, l'axe des abscisses donnant le nombre de mots. Dans l'exemple ci-dessous, les endroits où la courbe est horizontale correspondent aux passages du corpus où les Actes ne sont pas cités, ceux où elle est verticale au contraire aux passages où ils apparaissent le plus.

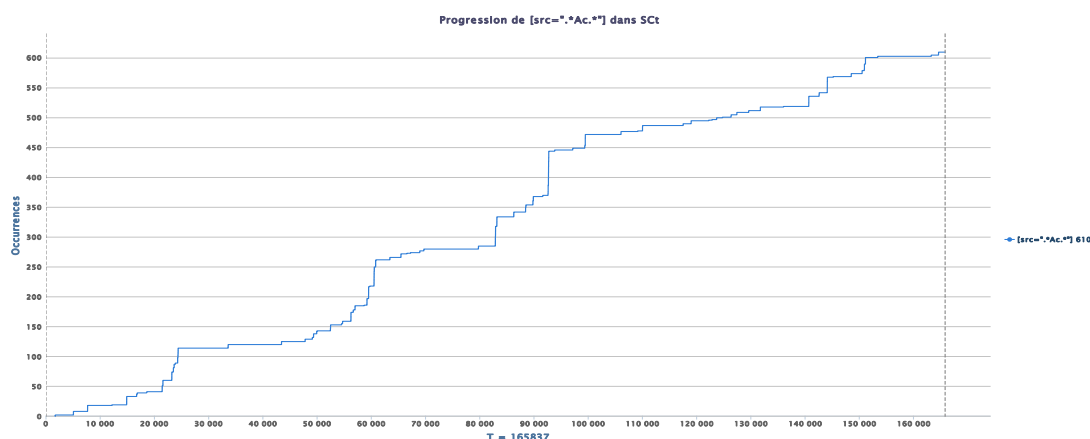


Illustration 22. Évolution des références aux Actes des Apôtres dans les SCt.

En zoomant sur le graphique, nous pouvons lire le détail des lieux intéressants.

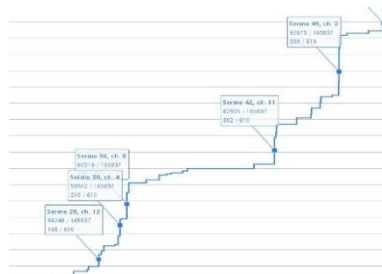


Illustration 23. Détail de la progression des Actes des Apôtres dans les SCt.

6) Conclusion

Sur ces questions de visualisation, nous renvoyons à une abondante littérature, en plein développement³⁵. Outre les outils de Google Charts ou de la suite Microsoft Office, on consultera également la documentation des fonctions graphiques du logiciel R. Il nous faut encore trouver une méthode mathématique pour construire des diagrammes concentriques de cooccurrences de versets autour d'un pivot, du type de celui que nous avons réalisé, de façon entièrement manuelle, pour Irénée autour d'Ep 1, 10³⁶. Ce travail extrêmement long doit pouvoir être fait à partir de n'importe quel verset, d'autant que nous avons toutes les informations nécessaires dans les relevés des Sources Chrétiennes³⁷ et que les concordances dans TXM, dont nous allons traiter dans la section suivante, facilitent grandement le repérage des cooccurrences.

B. La visualisation textuelle

Les graphiques sont parfois des résultats en eux-mêmes, mais souvent, ils sont seulement des moyens pour repérer les lieux du corpus qui mériteraient examen plus approfondi. Aussi le projet s'intéresse-t-il fortement aux possibilités de retour au texte.

Voici un aperçu des fonctionnalités de visualisation textuelle déjà offertes par TXM sur notre corpus de test des *Sermons sur le Cantique*³⁸. La table lexicale donne toutes les formes avec leur fréquence. L'index permet d'y opérer des sélections, puis d'avoir un retour au texte.

1) Croisement de l'analyse textuelle et des références

À partir de nos données textuelles liées à des références, nous pouvons croiser deux types d'interrogation : la recherche d'un taux scripturaire, calculé comme

³⁵ Citons entre autres : W. S. CLEVELAND, *Visualizing Data* et *The Elements of Graphing Data*, Summit (NJ), 1993 et 1994 ; M. FRIENDLY, *Visualizing Categorical Data*, Assam (India), 2000 ; L. WILKINSON, *The Grammar of Graphics*, Springer Verlag, Berlin 2005² ; P. MURRELL, *R Graphics*, Abingdon, 2011².

³⁶ Cf. L. MELLERIN, « Étude des usages bibliques d'Irénée de Lyon à l'aide de BIBINDEX », dans *Irénée de Lyon et le début de la Bible chrétienne, Actes de la journée du 1. VII. 2014 à Lyon*, A. Bastit-Kalinowska, J. Verheyden (éd.), Collection « *Instrumenta Patristica et Mediaevalia* » 77, Turnhout, 2017, p. 35-62, ici p. 56.

³⁷ L'indication des numéros de paragraphes, complétée par la lettre d'appel de l'apparat scripturaire et un numéro d'ordre de chaque occurrence dans une œuvre donnée, rend possible de définir des critères objectifs pour étudier la proximité relative d'un ensemble de versets.

³⁸ Nous ne traitons ici que de TXM, mais il existe d'autres formes de visualisation textuelle qu'il nous faudra explorer. Voir notamment M. EDER, J. RYBICKI, M. KESTEMONT, « Stylometry with R: A Package for Computational Text Analysis », *R Journal* 8, 2016, p. 107-121 ; S. JÄNICKE, G. FRANZINI, C. FAISAL, G. SCHEUERMANN, « Visual Text Analysis in Digital Humanities », *Computer Graphics Forum* 35, 2016, doi:10.1111/cgf.12873, remarquable synthèse des possibilités actuelles.

nombre de mots provenant de telle section biblique rapporté au nombre de mots de l'œuvre ; le nombre d'occurrences dans un corpus d'une référence à cette section. Dans l'exemple des illustrations qui suivent, on constate que Bernard cite 8 fois le *Psaume 32* dans les *Sermons sur le Cantique*, et que le *Psaume 32* représente 43 mots dans l'ensemble du texte des *Sermons sur le Cantique*.

src, ref	Contexte gauche	Pivot	Contexte droit
[Vg.Ps.32.15], Sermo 9, ch. 4	vestra omnia norit scrutans corda et renes Deus	qui finxit singillatim corda vestra et intelligit omnia opera vestra	Sponsa ergo sp
[Vg.Ps.32.9], Sermo 20, ch. 2	mundi fabrica tantum fatigationis auctor assumpsit illa denique	dixit et facta sunt mandavit et creata sunt	At vero hic et in c
[Vg.Ps.32.5], Sermo 22, ch. 8	enim in omnem terram exivit odor vitae quoniam	misericordia Domini plena est terra	et miseraciones
[Vg.Ps.32.9][Vg.Ps.148.5], Sermo 27, ch. 9	quod utique non quomodo ceteros dixit tantum ut	fieret	sed pugnavit ut :
[Vg.Ps.32.12], Sermo 27, ch. 9	dabit mihi saltem vestigiis adhaerere sanctae cuiuspiam animae	quam elegit in hereditatem sibi	Tamen si dignet
[Vg.Ps.32.5], Sermo 54, ch. 3	multiplicavit locupletare eam Terram visitavit non aerem quia	misericordia Domini plena est terra	Denique operat
[Vg.Ps.32.19], Sermo 71, ch. 4	domesticorum curam gerit maxime in diebus malis ut	alat eos in fame	cibans illos pan
[Vg.Ps.32.6], Sermo 85, ch. 3	per se non potuit Non opinor Quid enim	Verbo Domini caeli firmati sunt	et terra stabit sir

Illustration 24. Les 8 occurrences du Ps 32 dans les Sermons sur le Cantique³⁹.

src, ref	Contexte gauche	Pivot	Contexte droit
[Vg.Ps.32.15], Sermo 9, ch. 4	vestra omnia norit scrutans corda et renes Deus	qui	finxit singillatim corda vestra et intelligit omnia ope
[Vg.Ps.32.15], Sermo 9, ch. 4	omnia norit scrutans corda et renes Deus qui	finxit	singillatim corda vestra et intelligit omnia opera ve
[Vg.Ps.32.15], Sermo 9, ch. 4	norit scrutans corda et renes Deus qui finxit	singillatim	corda vestra et intelligit omnia opera vestra Spons
[Vg.Ps.32.15], Sermo 9, ch. 4	scrutans corda et renes Deus qui finxit singillatim	corda	vestra et intelligit omnia opera vestra Sponsa ergo
[Vg.Ps.32.15], Sermo 9, ch. 4	corda et renes Deus qui finxit singillatim corda	vestra	et intelligit omnia opera vestra Sponsa ergo spons
[Vg.Ps.32.15], Sermo 9, ch. 4	et renes Deus qui finxit singillatim corda vestra	et	intelligit omnia opera vestra Sponsa ergo sponsu
[Vg.Ps.32.15], Sermo 9, ch. 4	renes Deus qui finxit singillatim corda vestra et	intelligit	omnia opera vestra Sponsa ergo sponsum adesse
[Vg.Ps.32.15], Sermo 9, ch. 4	Deus qui finxit singillatim corda vestra et intelligit	omnia	opera vestra Sponsa ergo sponsum adesse persi-
[Vg.Ps.32.15], Sermo 9, ch. 4	qui finxit singillatim corda vestra et intelligit omnia	opera	vestra Sponsa ergo sponsum adesse persentien-
[Vg.Ps.32.15], Sermo 9, ch. 4	finxit singillatim corda vestra et intelligit omnia opera	vestra	Sponsa ergo sponsum adesse persentienis subs
[Vg.Ps.32.9], Sermo 20, ch. 2	mundi fabrica tantum fatigationis auctor assumpsit illa denique	dixit	et facta sunt mandavit et creata sunt At vero hic et i
[Vg.Ps.32.9], Sermo 20, ch. 2	fabrica tantum fatigationis auctor assumpsit illa denique dixit	et	facta sunt mandavit et creata sunt At vero hic et in c
[Vg.Ps.32.9], Sermo 20, ch. 2	tantum fatigationis auctor assumpsit illa denique dixit et	facta	sunt mandavit et creata sunt At vero hic et in dictis
[Vg.Ps.32.9], Sermo 20, ch. 2	fatigationis auctor assumpsit illa denique dixit et facta	sunt	mandavit et creata sunt At vero hic et in dictis suis
[Vg.Ps.32.9], Sermo 20, ch. 2	auctor assumpsit illa denique dixit et facta sunt	mandavit	et creata sunt At vero hic et in dictis suis sustinuit c
[Vg.Ps.32.9], Sermo 20, ch. 2	assumpsit illa denique dixit et facta sunt mandavit	et	creata sunt At vero hic et in dictis suis sustinuit cor
[Vg.Ps.32.9], Sermo 20, ch. 2	illa denique dixit et facta sunt mandavit et	creata	sunt At vero hic et in dictis suis sustinuit contradict
[Vg.Ps.32.9], Sermo 20, ch. 2	denique dixit et facta sunt mandavit et creata	sunt	At vero hic et in dictis suis sustinuit contradictores

Illustration 25. Les 43 mots empruntés au Ps 32 dans les Sermons sur le Cantique.

2) Recherche de cooccurrences

L'une des fonctionnalités les plus prometteuses est la recherche de cooccurrences. Il est possible d'observer systématiquement ce qui se trouve autour d'un verset donné.

³⁹ Actuellement, seul le dernier mot de chaque allusion a été encodé.

VI. QUADRUPLICITER QUOMODO SPONSAM DECOLORAVERIT SOL.

13Unde et addit sponsa : Nolite me considerare quod fusca sim , quia decoloravit me sol [+][9], hoc est : Nolite me notare quasi deformem , quia cernitis pro ingruente persecutione minus florentem , minus secundum gloriam saeculi coloratam . Quid exprobratis ingredinem , quam fervor persecutionis , non conversationis pudor invenit ? Vel solem dicit zelum iustitiae , quo accenditur et accingitur adversus malignantes [+][9], dicens Deo : Zelus domus tuae comedit me [+][6]; et illud : Tabescere me fecit zelus meus , quia obliiti sunt verba tua inimici mei [+][6]; illud quoque : Defectio tenuit me pro peccatoribus derelinquentibus legem tuam [+][6]; item : Nomen quo oderunt te , Domine , oderam illos , et super inimicos tuos tabescebam [+][6]? Etiam illud Sapientis caute observat : Filiae , ait , tibi sunt ? Noli ostendere laetum vultum ad ipsas [+][6], ut scilicet remissis et mollibus et fugitantibus disciplinam , non candorem serenitatis , sed obscurum severitatis exhibeat . Vel decolorari a sole , est ignescere caritate fraterna , **fieri eum flentibus , gaudere cum gaudentibus** [+][1] Vg.Rm.12.15[9], **cum infirmantibus infirmari , ut ad scandala singulorum** [+][1] Vg.1.Co.11.29[9]. **Vel sic Sol iustitiae** [+][1] Vg.Mt.4.2[9] decoloravit me Christus , cuius amore languens [+][1] Vg.Ch.2.3[9]. Languor iste coloris quaedam exterminatio est , et defectus in desiderio animae [+][9]; unde et dicit : Memor fui Dei , et delectatus sum , et exercitatus sum , et defecit spiritus meus [+][9]. Ergo , instar urentis solis , desiderii ardor peregrinatem in corpore [+][9] decolorat , dum vultu gloriae [+][9] inhiantem , impatientem facit repulsa et excruciat amantem dilatio . Quis nostrum ita sancto amore ardet , ut desiderio videndi Christum omnem colorem praesentis gloriae laetitiaeque fastidiat et deponat , illa ei prophetica voce contestans : *Et diem hominis non desideravi , tu scis* [+][9]. Item cum Sancto David : *Renuit consolari anima mea* [+][9], id est praesentium bonorum inani laetitia despicit colorari . Vel certe decoloravit me sol [+][9], sui nimirum comparatione splendoris , dum appropians illi , ex eo me obscuram deprehendo , nigram invenio , foedam despicio . Ceterum alias quidem formosa sum : quid fuscam dicitis solus Solis pulchritudini succubentem ? At sensui priori videntur magis assentire ea quae sequuntur . Adiciens siquidem : *Filii matris mae pugnaverunt contra me* [+][9], persecutionem se esse passam significat . Sed hinc aliud sermonis principum ordiemur , quoniam sufficere hac vice possunt , quae accepimus de gloria sponsi Ecclesiae , dono ipsius , qui est *Deus benedictus in saecula . Amen* [+][9].

Illustration 26. Contextes scripturaires de MI 4, 2 avec un exemple de retour au contexte en SCT 28, 13.

Il est également possible de ne chercher que des associations de versets, ou de livres, ou de verset et de livres, etc.

Requête: `[[src="*MI.4.*"]][src="*Sg.*"]| [src="*Sg.*"]][src="*MI.4.*"] within 100 expand to seg`

Clés de tri : #1 Aucun #2 Aucun #3 Aucun #4 Aucun Tri

src. ref	Contexte gauche	Pivot	Contexte droit
[Vg.Sg.7.26], Sermo 45, ch. 9	figura substantiae Patris et quidam perpetuus minimeque fucatus	candor vitae aeternae Quam mihi decorus es Do...	de corde terrae resurgis Quam formosus in stola tua demum Rex gloriae
[Vg.Sg.5.6], Sermo 70, ch. 6	bonos et malos Neque enim qui dicti sunt	Sol iustitiae non ortus est nobis lucem illius quan...	Ergo candor huius illi apud iustos fragrantia etiam usque ad iniquos diffunditur
[Vg.Sg.3.7][Vg.Mt.13.43], Sermo 72, ch. 9	sic inter filios tenebrarum in tetra mortis caligine	fulgebunt iusti et videbunt eo utique clarius quo ex...	- ii inquam ibunt de tenebris in tenebras densiores ut qui in
[Vg.Mt.4.2][Vg.Sg.5.6], Sermo 72, ch. 9	respiraverunt - nec enim quaesierunt inspirantis diei lumen	et Sol iustitiae non ortus est eis	- ii inquam ibunt de tenebris in tenebras densiores ut qui in

Illustration 27. Recherche des cooccurrences de MI 4, 2 avec un verset de la Sagesse.

Nec nox mortis praevalebit adversus redivivum hunc diem ; magis autem in tenebris lucet , et tenebrae eum non comprehenderunt [+][9]. In tantum non reor nec vita decedente cedere lumen vitae [+][9] ut nemini congruentius quam sic mortuo assignandam censeam vocem illam : *Et nox illuminatio mea in deliciis meis* [+][9]. Quidni clarius videat , nube , vel potius faece corporis evolutus ? Erit sine dubio vinculis solutus corporeis inter mortuos liber [+][9], et inter caecos videns . Nam quemadmodum olim , omni oculo caligante per universam Aegyptum , solus in mediis tenebris clare videbat populus videns Deum , id est populus Israel , dicente Scriptura , quia *ubicumque Israel erat , lux erat* [+][9], sic inter filios tenebrarum [+][9], in tetra mortis caligine **fulgebunt iusti** [+][9] **et videbunt , eo utique clarius quo exuti corporum umbris . Nam ei ii qui ante non respiraverunt - nec enim quaesierunt inspirantis diei lumen , et Sol iustitiae non ortus est eis** [+][9] Vg.Mt.4.2 Vg.Sg.5.6[9] - ii , inquam , ibunt de tenebris in tenebras densiores , ut qui in tenebris sunt tenebrescant adhuc , et qui vident videant magis [+][9].

Illustration 28. Retour au texte correspondant en SCT 72, 10.

Ces visualisations rendent aussi possible de distinguer entre les types de citations. La requête ci-dessous ne donne que les citations modifiées de MI 4, 2. Très prochainement, les citations typées Vieilles Latines seront également isolables.

2At talis visio non est vitae praesentis, sed in novissimis reservatur, his dumtaxat qui dicere possunt : *Scimus quia, cum apparuerit, similes ei erimus, quia videbimus eum sicuti est* [+]⁹¹. Et nunc quidem apparet quibus vult, sed sicuti vult, non *sicuti est* [+]⁹¹. Non sapiens, non sanctus, non propheta videre illum *sicuti est* [+]⁹¹ potest aut potuit in corpore hoc mortali ; poterit autem in immortalis, qui dignus habebitur. Itaque videtur et hic, sed sicut videtur ipsi, et non *sicuti est* [+]⁹¹. Nam neque hoc *luminare magnum* [+]⁹¹, solem loquor istum, quem quotidie vides, vidisti tamen aliquando *sicuti est* [+]⁹¹, sed tantum sicut illuminat, verbi gratia, aërem, montem, parietem. Quod ne ipsum quidem aliquatenus posses, si non aliqua ex parte ipsum lumen corporis tui [+]⁹¹, pro sui ingentia serenitate et perspicuitate, caelesti luminis simile esset. Non denique alterum membrum corporis capax est luminis, ob multam utique dissimilitudinem. Sed nec ipse oculus, cum turbatus fuerit, luminis propinquabit, nimirum ob amissam similitudinem. Qui ergo turbatus nullatenus serenam solem videt propter dissimilitudinem, serenus aliquatenus videt propter nonnullam similitudinem. Profecto si pari prorsus puritate vigeret, videret omnino inoffensa acie eum *sicuti est* [+]⁹¹, propter omnimodam similitudinem. Ita et *Solem iustitiae* [+]⁹¹ illum, qui *illuminat omnem hominem venientem in hunc mundum* [+]⁹¹, videre in hoc mundo, sicut illuminat : *illuminatus potes*, tanquam iam in aliquo similis ; *sicuti est* [+]⁹¹, omnino non potes, tanquam nondum perfecte similis. Propterea dicit : *Accedite ad eum, et illuminamini, et facies vestras non confundentur* [+]⁹¹. Ita sane, sed si quantum satis est illuminamur, ut *revelata facie speculantes*

default 3 / 11

SCT[ana="MI.4."inexactQuotation.""] expand to seg

Requête : [ana="MI.4."inexactQuotation.""] expand to seg

Clés de tri : #1 Aucun #2 Aucun #3 Aucun #4 Aucun Tri

src, ref	Contexte gauche	Pivot	Contexte droit
[Vg.MI.4.2], Sermo 31, ch. 2	eum sicuti est propter omnimodam similitudinem Ita et	Solem iustitiae	illum qui illuminat omnem hominem venientem in hunc mundum videre in hoc
[Vg.MI.4.2], Sermo 70, ch. 6	lucem illius quandoque viderunt Viderunt autem quotquot audierunt	Vobis qui timetis Deum oriatur sol iustitiae	Ergo candor huius lili apud iustos fragrantia etiam usque ad iniquos diffunditur
[Vg.MI.4.2][Vg.Sg.5.6], Sermo 72, ch. 9	respiraverunt – nec enim quaesierunt inspirantis diei lumen	et Sol iustitiae non ortus est eis	– Il inquam ibunt de tenebris in tenebras densiores ut qui in
[Vg.MI.4.2], Sermo 75, ch. 10	quot sectae tot noctes Frustra per has noctes	iustitiae solem	et lumen quaeritis veritatis id est Sponsum quia nulla societas luci ad

Illustration 29. Citations modifiées de MI 4, 2 dans les Sermons sur le Cantique.

En complément de ces visualisations ponctuelles, le projet développe aussi des présentations linéaires du texte :

- La consultation, par entrée biblique, des lieux des œuvres où figure cette entrée, avec toutes les métadonnées, et par lien hypertexte un accès au texte de Bernard ;
- Inversement, la consultation, au fil du texte latin, des citations bibliques, avec indication de la référence scripturaire chiffrée, un lien vers le texte correspondant de la Vulgate et éventuellement d'autres typages ;
- La consultation, via des formulaires modulables d'interrogation, d'une reconstitution de la Bible de Bernard, élaborée à partir de toutes les occurrences scripturaires de ses œuvres, avec leurs variantes, classées dans l'ordre des livres de la Vulgate.

VI. – CONCLUSION

Nous retrouvons chez Bernard, comme chez tous les Pères, une illustration de la fameuse idée de Grégoire le Grand, selon laquelle les Écritures croissent avec ceux qui les lisent⁴⁰ : Bernard déploie en ses écrits le développement d'une interprétation qui intègre de façon originale des siècles de lectures antérieures, qui se nourrit des exigences propres à son contexte particulier. Mais dans ce langage scripturairement façonné, l'assimilation est plus intime encore : si la « Bible de Bernard » reste fondamentalement la Bible au sens courant, un texte distinct, universel, elle en est aussi cette appropriation unique, tissée dans une œuvre autre, qui n'est pas une simple sélection de versets mais une œuvre biblique nouvelle.

⁴⁰ *Homélies sur Ézéchiel.*, éd. C. Morel, t. I, SC 327, Paris 1986, p. 244-245.

Pour en rendre compte, il faut nous tenir constamment dans une mise en abyme : lire Bernard au prisme de ses usages bibliques, c'est faire l'exégèse d'un texte qui s'est fait lui-même exégèse pour se donner comme Parole assimilée et rendue transmissible. Bernard n'écrit pas pour qu'on commente ses œuvres, mais pour qu'on accède à travers elles à leur source : oublier cette perspective, vouloir défaire analytiquement l'écheveau serait manquer ce qui précisément constitue l'originalité de l'œuvre.

Il nous semble que la conception d'un outil numérique qui articule références canoniques et retour au texte peut s'inscrire dans la continuité de cette démarche d'enrichissement perpétuel : en rendant possible des parcours de lecture diversifiés, des visualisations de rapprochements sans occulter les contextes, il laisse vivre le texte tout en donnant des clefs pour l'analyser et le décrire.