



HAL
open science

Les mots du temps en latin médiéval. Enquête autour de TEMPUS

Bruno Bon

► **To cite this version:**

Bruno Bon. Les mots du temps en latin médiéval. Enquête autour de TEMPUS. Medialatinitas 2014 – VIIe Congrès international du latin médiéval, Sep 2014, Lyon, France. pp.107-124. halshs-01895103

HAL Id: halshs-01895103

<https://shs.hal.science/halshs-01895103>

Submitted on 24 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



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

Les mots du temps en latin médiéval : enquête autour de *Tempus*¹

Le temps relève des objets lexicaux difficiles à appréhender en raison du nombre très élevé de vocables concernés et d'occurrences à analyser. Les dictionnaires en proposent néanmoins une interprétation au travers de la lecture ordonnée de quelques attestations, sélectionnées en vertu de leur intérêt présumé². Pour le vocable *Tempus*, pivot *a priori* très fréquent de cette notion en latin médiéval³, un tour d'horizon des dictionnaires existants a permis de relever la récurrence des « expressions » dans la description lexicographique du temps⁴. Du latin classique au français moderne, en passant par le latin médiéval et l'ancien français, le temps est généralement appréhendé par des métaphores, « considéré comme » autre chose⁵. Il ressort de ces premières observations que, pour traiter des mots du temps en latin médiéval, il semble préférable de travailler avec les concepts (plutôt que les sens) et les collocations (*i.e.* les expressions).

A. *Tempus* dans les textes

L'observation de l'entourage lexical de *Tempus* dans les corpus textuels numérisés devrait permettre d'en faire apparaître l'évolution diachronique. Malheureusement, les rares ressources disponibles sans efforts pour le latin médiéval imposent des biais ou des limitations considérables, ces outils étant mieux adaptés à la critique d'attribution qu'à l'analyse lexicale⁶. Pour pouvoir envisager le traitement efficace d'un mot très fréquent comme *Tempus*, nous sommes donc contraints à reconstruire nous-mêmes des données directement et librement exploitables à des fins de sémantique historique.

1. Repérer les mots structurants

La première étape du travail consiste à repérer les « mots du temps », qui composent les pièces du puzzle de la pensée médiévale. Bien sûr, ce repérage peut se faire à la main, en listant tous les mots qui viennent à l'esprit, avec ou sans l'aide d'une ontologie moderne⁷. Mais les récents développements du traitement automatique des langues, sous réserve d'être adaptés au latin, devraient permettre de rechercher sans idée préconçue tous les mots « liés au temps », avec l'espoir d'autoriser ainsi la reconstruction d'une structure sémantique du temps éventuellement différente de la nôtre. C'est l'option que nous avons choisie !

Concrètement, il s'agit d'abord de rechercher les cooccurrents directs de *Tempus*, pour reconstituer son entourage lexical de premier ordre : c'est-à-dire tous les mots qui se retrouvent à proximité du pivot (ici *Tempus*) plus souvent qu'ils ne le devraient au hasard. Pour ce faire, nous avons exploité quatre corpus textuels lemmatisés⁸, en bénéficiant à la fois des résultats du programme « OMNIA » financé par l'Agence

1 Cette communication, présentée au Congrès international du Latin médiéval de Lyon en 2014, s'appuie sur les travaux préparatoires à la rédaction d'un article *TEMPUS* pour le *Novum Glossarium Mediae Latinitatis* (NGML – cf. *infra* note 2), dictionnaire international du latin médiéval (800-1200) publié par l'Union Académique Internationale, à paraître dans : Bruno Bon, « De nouvelles (co)occurrences pour le *Novum Glossarium Mediae Latinitatis* – Le traitement d'un mot fréquent : l'exemple de *Tempus* ».

2 Malheureusement, ni le *Thesaurus Linguae Latinae* (TLL) pour le latin classique, ni le NGML ou le *Mittellateinisches Wörterbuch* (MLW) pour le latin médiéval ne couvrent la lettre T : *Thesaurus Linguae Latinae*, Leipzig-Stuttgart-Berlin, 1900-2012 (derniers fascicules parus : N-Navalis et R-Rarus) ; Franz Blatt, Yves Lefèvre, Jacques Monfrin, François Dolbeau, Anita Guerreau-Jalabert (dir.), *Novum Glossarium Mediae Latinitatis*, Copenhague-Bruxelles-Genève, 1957-2011, disponible en ligne gratuitement (<http://www.glossaria.eu/ngml> – dernier fascicule paru : Plana-Plego) ; Paul Lehmann, Helmut Gneuss, Peter Stotz (dir.), *Mittellateinisches Wörterbuch bis zum ausgehenden 13. Jahrhundert*, München, 1959-2014 (dernier fascicule paru : Hospitalarius-Illibezzus).

3 La requête de *Tempus* (sous toutes ses formes fléchies) dans le *Cross Database Search Tool* de la base de données textuelles « Brepolis » (<http://clt.brepolis.net/cds/default.aspx>) fait apparaître plus de 40 000 occurrences pour la période patristique, et plus de 120 000 pour la période médiévale.

4 Dans le cadre de l'enquête lexicographique préparatoire à cette étude (cf. *supra* note 1), ont été analysés en particulier : Félix Gaffiot, *Dictionnaire latin-français*, Paris, 1934 (nouv. éd. Pierre Flobert, *Le Grand Gaffiot*, Paris, 2000) ; Charlton T. Lewis, Charles Short, *A Latin Dictionary*, Oxford, 1879 ; Aegidius Forcellini, Vincentius De Vit, *Totius Latinitatis Lexicon*, Prato, 1858-1875 (1^{re} éd. 1771) ; Peter G. W. Glare, *Oxford Latin Dictionary*, Oxford, 1968-1982 ; Charles Du Fresne Du Cange, *Glossarium Mediae et Infimae Latinitatis*, Paris-Niort, 1678-1887 (disponible en ligne gratuitement : <http://ducange.enc.sorbonne.fr>) ; Albert Blaise, *Dictionnaire latin-français des auteurs chrétiens*, Strasbourg, 1954 ; Albert Blaise, *Dictionnaire latin-français des auteurs du Moyen âge*, Turnhout, 1975 ; Jan Frederik Niermeyer, C. van de Kieft, *Mediae Latinitatis Lexicon Minus*, Leiden, 1976-2004.

5 Sur les métaphores conceptuelles, cf. *infra* note 21.

6 Pour le détail de ces observations, cf. *supra* note 1 : dans cette enquête préparatoire, sont analysés d'une part le portail libre « Logeion » (<http://logeion.uchicago.edu>) adossé à « Perseus » (<http://perseus.uchicago.edu>), d'autre part la plate-forme propriétaire « Brepolis » (*Database of Latin Dictionaries* : <http://clt.brepolis.net/dld/default.aspx> ; *Cross Database Search Tool* : cf. *supra* note 3).

7 Par exemple : Rudolf Hallig, Walther von Wartburg, *Begriffssystem als Grundlage für die Lexikographie : Versuch eines Ordnungsschemas*, Berlin, 1952.

8 L'un des principaux obstacles à l'analyse statistique du latin médiéval est posé par l'absence de lemmatisation des grands corpus

Nationale de la Recherche et du travail d'Alain Guerreau (CNRS, Paris), et de la collaboration de Krzysztof Nowak (IJP-PAN, Cracovie) dans le cadre du Soutien à la Mobilité Internationale de l'Institut des Sciences Humaines et Sociales du CNRS⁹ : des textes classiques antiques (six millions de mots), des textes de saint Augustin (six millions de mots), des textes de saint Thomas (treize millions de mots), et des textes diplomatiques médiévaux (quinze millions de mots).

A titre d'exemple, voici le relevé des premiers substantifs cooccurrents de *Tempus* dans le corpus de latin classique¹⁰. Pour chacun d'eux, le tableau donne plusieurs informations : entre autres, l'effectif total du lemme et du pivot (col. 4 et 5), la cofréquence (col. 2), et la valeur pondérée par un coefficient (col. 6 à 8). En effet, pour éliminer les cooccurrents les plus courants (comme la conjonction de coordination *Et*), un coefficient de pondération est appliqué aux résultats bruts, qui permet de faire remonter les cooccurrents dits « représentatifs » (ici classés par celui de Dice)¹¹.

PoS	Cooc	Lem	Eff_lem	Eff_pivot	Dice	PoissonS	HyperG
SUB	166	amnis	2516	7878	0,260	937,067	503,598
SUB	90	spatium	1371	7878	0,200	508,048	275,243
SUB	71	persona	705	7878	0,198	400,794	250,113
SUB	57	momentum	366	7878	0,179	321,764	228,888
SUB	88	etas	2309	7878	0,147	496,759	220,832
SUB	35	punctum	72	7878	0,123	197,574	191,806
SUB	96	nox	3276	7878	0,122	541,918	215,437
SUB	40	hiems	634	7878	0,114	225,799	124,120
SUB	54	casus	1718	7878	0,108	304,829	127,284
SUB	108	uita	4560	7878	0,099	609,658	219,306
SUB	286	locus	9478	7878	0,096	1614,465	641,297
SUB	40	memoria	1566	7878	0,084	225,799	87,430
SUB	30	necessitas	713	7878	0,083	169,349	82,010
SUB	27	estas	461	7878	0,082	152,415	83,641
SUB	30	comitium	759	7878	0,082	169,349	80,089
SUB	36	curso	1319	7878	0,081	203,219	81,574
SUB	46	mora	2221	7878	0,079	259,669	90,329
SUB	93	ratio	4926	7878	0,078	524,983	169,274
SUB	23	autumnus	148	7878	0,078	129,835	96,493
SUB	26	occasio	528	7878	0,077	146,770	76,054

Cependant, chaque coefficient permettant de pondérer la cooccurrence brute en fonction de paramètres différents, comme l'effectif ou la fréquence, ils ne donnent pas tous le même classement final. Pour s'en rendre compte, il suffit de mettre en rapport la fréquence relative et la valeur pondérée de la cooccurrence : si le coefficient de Dice (*figure 1a*) pondère relativement les mots les plus cofréquents (mais pas tous), le coefficient Hyper-Géométrique (*figure 1b*) est sensible à la cofréquence, et le coefficient de Poisson-Stirling (*figure 1c*) est visiblement fondé sur la cofréquence¹².

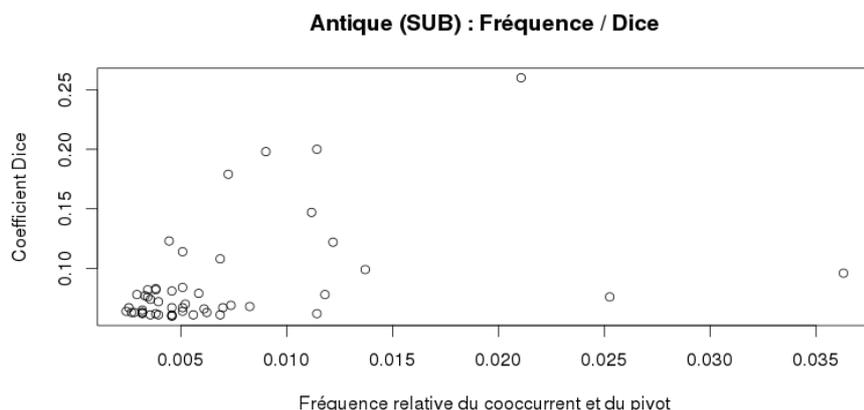
textuels. Pour nous en affranchir, nous avons utilisé le logiciel TreeTagger (<http://www.cis.uni-muenchen.de>) avec les paramètres réalisés dans le cadre du programme ANR OMNIA (*cf. infra* note 9) : toutes les analyses présentées dans cette communication traitent donc des lemmes, et non des simples formes. Sur la variation graphique et morphologique du latin médiéval, voir : Renaud Alexandre, Bruno Bon, Anita Guerreau-Jalabert, « Variations graphiques, variations morphologiques et lemmatisation du latin médiéval », dans Jacques Elfassi *et al.* (éd.), *Amicorum Societas : Mélanges offerts à François Dolbeau*, Florence, 2013, p. 3-18.

9 Projet ANR OMNIA : « Outils et méthodes numériques pour l'interrogation et l'analyse des textes médiolatins » (<http://www.glossaria.eu>). Dictionnaire polonais du latin médiéval : *Lexicon Mediae et Infimae Latinitatis Polonorum*, Instytut Języka Polskiego Polskiej Akademii Nauk (<http://www.scriptorum.pl>).

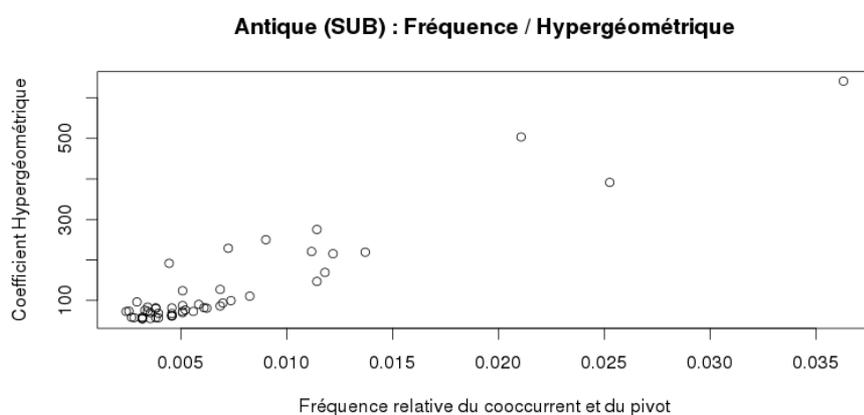
10 PoS = *Part of Speech*, nature du lemme (SUB = substantif) ; Cooc = Cooccurrence, nombre total d'apparitions conjointes du lemme et du pivot ; Lem = Lemme cooccurrent ; Eff_lem = Effectif total du lemme ; Eff_pivot = Effectif total du pivot ; Dice, PoissonS, HyperG = Dice, Poisson-Stirling, Hyper-Géométrique, algorithmes de pondération des cooccurrences (*cf. infra* note 11).

11 On parle aussi de distance de Dice : à titre d'exemple, *Amnis* (*fr.* rivière) apparaît moins souvent (166 vs. 286) avec *Tempus* que *Locus* (*fr.* lieu), mais, nettement moins fréquent que lui dans l'ensemble du corpus (2516 vs. 9478), il sort en tête du classement par le coefficient de Dice. Pour le détail des coefficients de pondération, voir la thèse et le site internet de Stefan Evert : *The Statistics of Word Cooccurrences : Word Pairs and Collocations*, Stuttgart, 2005 (<http://www.collocations.de>).

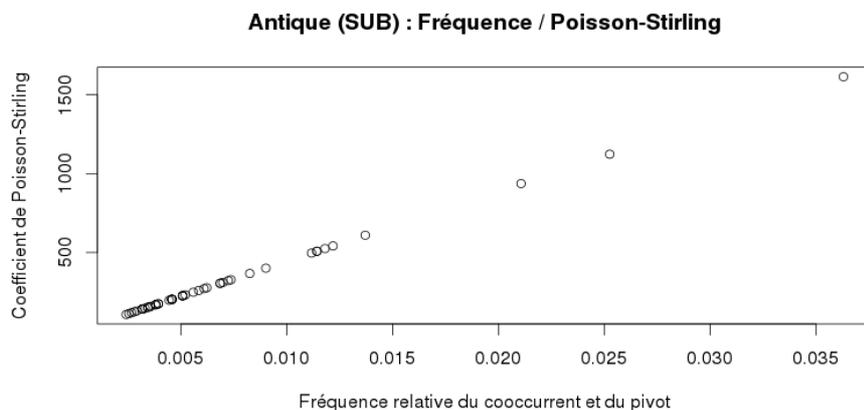
12 Sur les trois figures, chaque point représente un cooccurrent de *Tempus*, avec sa fréquence relative au pivot (rapport de la cooccurrence [Cooc] sur l'effectif total du pivot [Eff_pivot]) en abscisse et la distance pondérée en ordonnée.



<Figure 1a : Pondération des cooccurrents par le coefficient de Dice>



<Figure 1b : Pondération des cooccurrents par le coefficient Hyper-Géométrique>

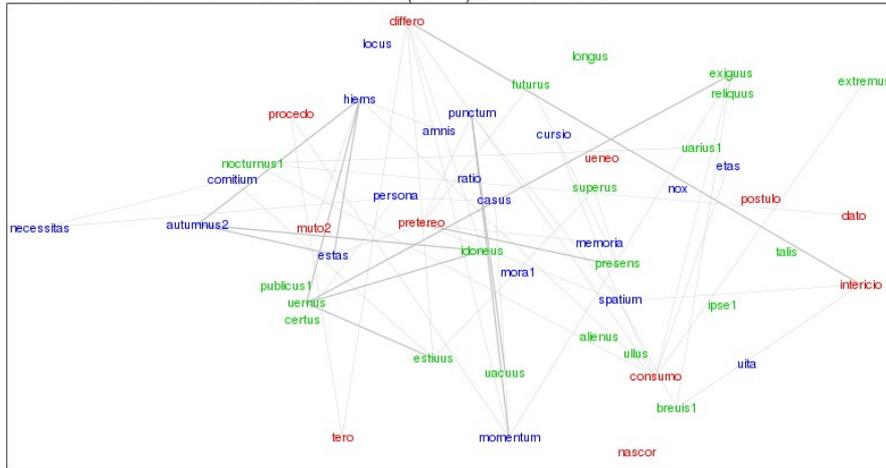


<Figure 1c : Pondération des cooccurrents par le coefficient de Poisson-Stirling>

Les cooccurrents (désormais pondérés avec le coefficient de Dice) peuvent ensuite s’afficher sur un graphe, pour faire apparaître des ensembles visuels, qui dépendent de tous les paramètres précédents (choix du coefficient, nature [Part of Speech, PoS] du cooccurrent)¹³. A titre d’exemple, dans le corpus de latin classique, les cooccurrents de *Tempus* (tous PoS confondus) ne se regroupent pas de la même façon que les seuls substantifs cooccurrents de *Tempus* : d’un côté (figure 2a), les saisons (à gauche) et la vie (à droite) ; de l’autre (figure 2b), le pouvoir (à gauche), la vie (en haut), et le temps naturel (en bas).

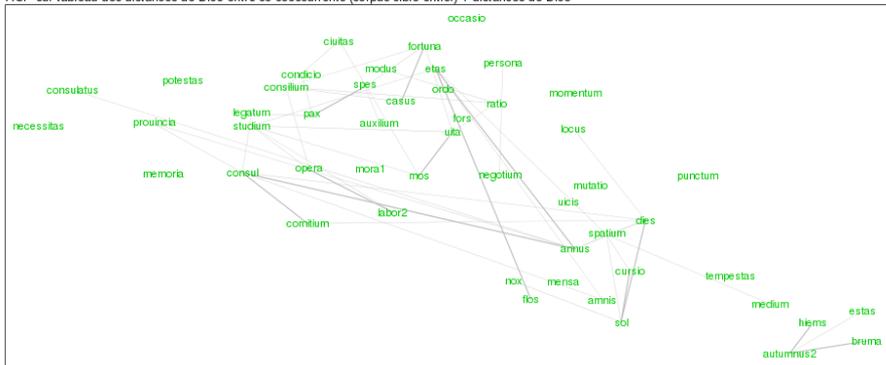
13 Sur ces graphiques, l’analyse des distances de Dice entre tous les cooccurrents du pivot permet de les rapprocher par contexte d’emploi, reflet d’une probable proximité sémantique.

corpus : ANTIQUE bornes : 0//6057529 stock cible : 6057529 lemme : tempus fréq. : 7878 fenêtre = +/-5 nb cooccurrents = 50
 ACP sur tableau des distances de Dice entre co-cooccurrents (contexte) + distances de Dice



<Figure 2a : Cooccurrents de Tempus dans le corpus Antique>

corpus : ANTIQUE bornes : 0//6057529 stock cible : 6057529 lemme : tempus fréq. : 7878 fenêtre = +/-5 nb cooccurrents = 50
 ACP sur tableau des distances de Dice entre co-cooccurrents (corpus cible entier) + distances de Dice



<Figure 2b : Cooccurrents substantifs de Tempus dans le corpus Antique>

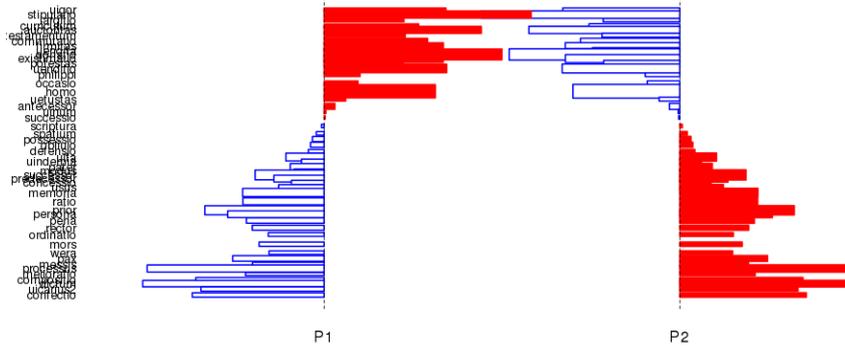
Tout cela repose sur le traitement brut de la proximité physique entre les deux mots considérés, y compris en dehors d'une même phrase, et sans contrôle *a priori* du rapport syntaxique qu'ils entretiennent. La validation des cooccurrents réels doit donc s'effectuer *a posteriori*, par un relevé d'expressions sur un échantillon aléatoire¹⁴ : dans ce tableau général des expressions, seuls trois cas sur dix-sept donnent moins de 60 % de « vraies » cooccurrences, et le plus souvent le taux est supérieur à 80 % !

Rangs	Antique	Augustin	Thomas	Cartae	Cooc. Patr.	Cooc. réels
annus	31	7	7		3081	56%
dies	22	25	9		4498	72%
etas	5	28	34		826	64%
locus	11	16	11		3982	82%
messis		11		29	341	86%
momentum	4	9	31		495	90%
ordo	49	6	42		1167	40%
pax	25		46	26	831	60%
plenitudo		3	23		957	94%
spatium	2	1	4	32	1408	78%
tribulatio		37	26		506	88%
uita	10			6	2658	56%
aduenio		10	16		940	92%
muto	7	15	12		446	80%
nascor	11	29	39		941	78%
obseruo	23	27	11	14	909	76%
pretereo	1	1	3	22	1237	84%
uenio	14	11	36	42	3262	64%

Pour donner une idée plus précise de ce dont il s'agit, nous avons procédé à un relevé des

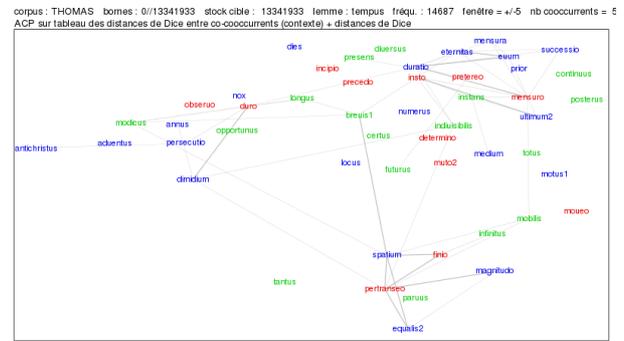
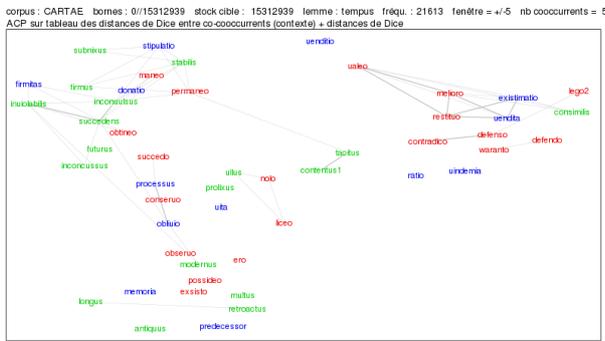
14 A titre d'exemple, *Annus* est le 31^e cooccurrent représentatif de *Tempus* dans le corpus antique, le 7^e chez saint Augustin et saint Thomas, et non relevé dans le corpus de chartes. S'il apparaît 3081 fois avec *Tempus* dans la *Patrologie Latine*, l'analyse d'un échantillon aléatoire de 50 occurrences permet de repérer 28 cooccurrences « vraies » (soit 56 %), où les deux mots entretiennent effectivement un rapport syntaxique, et 22 cooccurrences « fortuites », où la proximité n'est pas strictement signifiante.

corpus : CARTAE bomes : 0/15312939 lemme : tempus fréq. : 21613 fenêtre = +/-5 nb cooccurrents = 50
 Graphe de Bertin-Cibois de la répartition des cooccurrents par tranches chronologiques (écarts à l'indépendance)



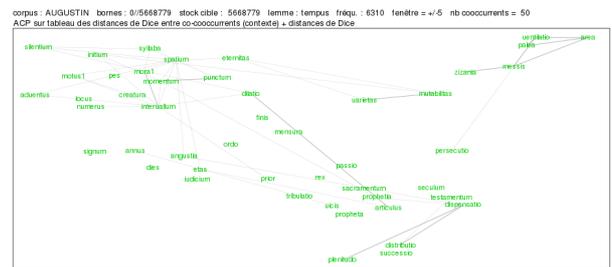
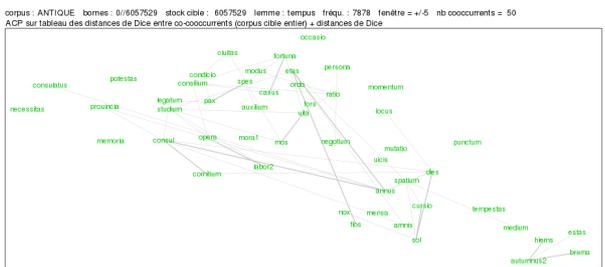
<Figure 3b : Cooccurrents substantifs de Tempus dans le corpus de Chartes>

Il est ensuite naturellement possible de comparer entre eux les cooccurrents d'un même mot dans deux corpus, et d'abord visuellement, sur deux graphes juxtaposés. Par exemple, tous PoS confondus, les cooccurrents de *Tempus* s'organisent différemment dans le corpus de Chartes (figure 4a) : pouvoir (en haut), mémoire (en bas), échange (à droite) ; et chez saint Thomas (figure 4b) : mouvement orienté (en haut, de droite à gauche : durée, début, fin), espace (en bas).



<Figures 4a et 4b : Cooccurrents de Tempus dans le corpus de Chartes et chez saint Thomas>

Un second exemple, avec les seuls cooccurrents substantifs, permet de comparer le contexte d'emploi de *Tempus* dans le corpus Antique (figure 4c) : pouvoir (à gauche), vie (en haut), temps naturel (en bas, de gauche à droite : année, saisons) ; et chez saint Augustin (figure 4d) : espace éternel (supérieur gauche), temps de la vie (au centre), fin des temps (en bas). Ces différences d'organisation, qui mériteraient une analyse approfondie, s'inscrivent au cœur de la conception médiévale (et donc chrétienne) du temps.



<Figures 4c et 4d : Cooccurrents de Tempus dans le corpus Antique et chez saint Augustin>

Mais la comparaison des cooccurrents d'un même mot dans deux corpus peut être automatisée, en utilisant les rangs de classement des cooccurrents dans chaque corpus, car il est impossible de comparer les distances pondérées d'un corpus à l'autre¹⁷. Dans ce tableau de comparaison, les listes de cooccurrents du corpus Antique et de saint Augustin sont croisées, et les cooccurrents communs apparaissent au centre¹⁸.

17 Dans chaque corpus, les cooccurrents sont affectés d'une distance pondérée au pivot qui permet de les classer ; c'est l'ordre de ce classement (i.e. le rang du cooccurrent) qui peut être comparé entre les corpus, non les distances elles-mêmes.

18 Lemme AN vs AU = cooccurrents de *Tempus* dans le corpus antique (AN) et chez saint Augustin (AU) ; rang.AN, rang.AU = rang de classement du cooccurrent (dans chaque corpus) ; cooc.AN, cooc.AU = cooccurrence brute du lemme et du pivot (dans chaque corpus) ; Dice.AN, Dice.AU = distance de Dice du cooccurrent (dans chaque corpus). A titre d'exemple, *Spatium* est le 2^e cooccurrent « représentatif » de *Tempus* dans le corpus Antique (90 occ., Dice 0.2) et le 1^{er} chez saint Augustin (136 occ., Dice 0.4).

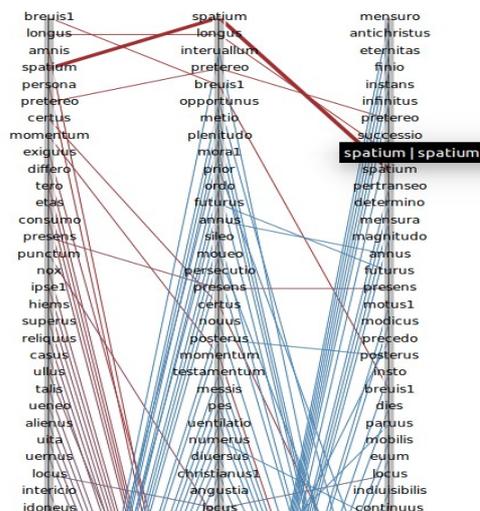
Lemme AN vs AU	rang.AN	cooc.AN	Dice.AN	rang.AU	cooc.AU	Dice.AU
casus	9	54	00.108		0	0
hiems	8	40	00.114		0	0
nox	7	96	00.122		0	0
persona	3	71	00.198		0	0
amnis	1	166	00.260		0	0
etas	5	88	00.147	28	31	00.079
<u>momentum</u>	4	57	00.179	9	37	00.131
<u>punctum</u>	6	35	00.123	17	29	00.109
<u>spatium</u>	2	90	00.200	1	136	00.401
mora	17	46	00.079	4	56	00.172
annus	31	55	00.067	7	78	00.152
ordo	49	36	00.060	6	86	00.165
interuallum		0	0	2	90	00.305
plenitudo		0	0	3	61	00.193
prior		0	0	5	66	00.165
persecutio		0	0	8	51	00.145

On peut enfin comparer les quatre listes de cooccurrents en même temps, en les regroupant manuellement dans un seul tableau, assortis de leurs rangs respectifs. Un tableau général des comparaisons permet d'évaluer l'homogénéité des cooccurrents de *Tempus* dans les différents corpus, et de relever les plus partagés d'entre eux.

SUB	Antique	Augustin	Thomas	Cartae	Total	Corpus	Moyenne	Rang
annus	31	7	7		45	3	15	4
dies	22	25	9		56	3	18,7	6
etas	5	28	34		67	3	22,3	7
locus	11	16	11		38	3	12,7	2
momentum	4	9	31		44	3	14,7	3
ordo	49	6	42		97	3	32,3	9
pax	25		46	26	97	3	32,3	10
prior		5	16	34	55	3	18,3	5
spatium	2	1	4	32	39	4	9,8	1
successio		48	3	33	84	3	28	8
Lemmes	8	9	10	4				
Proportion	80%	90%	100%	40%				

3. Visualiser les mots

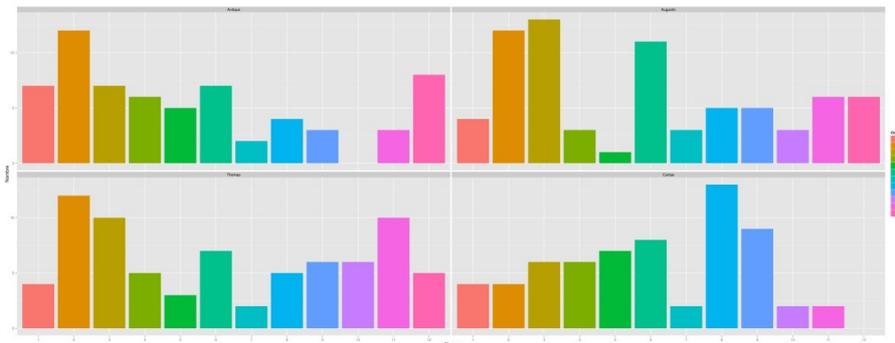
Pour un affichage graphique et une analyse visuelle de la comparaison des rangs, un outil, récemment développé à l'Accademia Europea de Bolzano¹⁹, permet de comparer automatiquement les rangs des cooccurrents d'un même mot dans plusieurs corpus : par exemple, ceux de *Tempus* dans le corpus Antique, chez saint Augustin et chez saint Thomas.



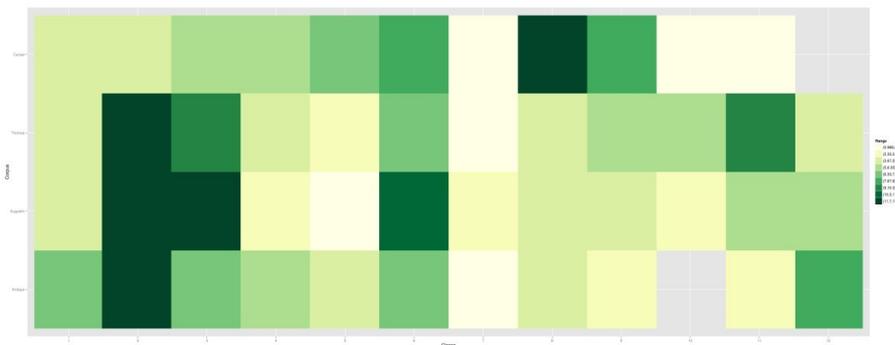
<Figure 5 : Comparaison des cooccurrents de *Tempus* (dans Antique, Augustin et Thomas)>

19 EURAC – Visualization of Linguistic Information, Structured Parallel Coordinates, Ranking Comparisons (<http://www.eurac.edu/en/research/institutes/multilingualism/Projects/LinfoVis/StructuredParallelCoordinates.html>) ; cf. Chris Culy, Verena Lyding, Henrik Dittmann, « Structured Parallel Coordinates : a visualization for analyzing structured language data », dans *Proceedings of the 3rd International Conference on Corpus Linguistics (CILC-11)*, Valencia, 2011, p. 485-493.

A partir de là, la juxtaposition de tableaux par corpus textuels, sous forme de calques, permet la comparaison deux à deux. Mais les outils de visualisation associés au logiciel de statistique R sont nettement plus efficaces²², en particulier parce qu'ils permettent une comparaison généralisée, indépendante du nombre d'ensembles de départ, sous forme de diagrammes en barres (figure 7a) ou de cartes de chaleur (figure 7b).



<Figure 7a : Classification des cooccurrents de Tempus par corpus, diagramme en barres>



<Figure 7b : Classification des cooccurrents de Tempus par corpus, carte de chaleur>

Au bout du compte, la grande difficulté de ce sujet réside dans l'utilisation de concepts métaphoriques pour définir le temps : le temps est « considéré comme » quelque chose d'autre que lui-même. En conséquence, avec un regroupement manuel des cooccurrents, on réintroduit à leur niveau le biais qu'on s'est refusé plus haut pour le pivot *Tempus*. Pour y échapper, il faudra donc à nouveau répéter l'opération pour chaque cooccurrent de *Tempus*, et passer ainsi aux cooccurrents du deuxième ordre, non seulement dans leur relevé, mais surtout dans leur classement. Une telle démultiplication de la procédure demandera nécessairement... du temps !

Bruno BON (CNRS / IRHT – Comité Du Cange)

²² R Core Team, *R : A Language and Environment for Statistical Computing*, Vienna, 2014 (<http://www.R-project.org>) ; voir en particulier : H. Wickham, *Ggplot2: elegant graphics for data analysis*, New York, 2009.