

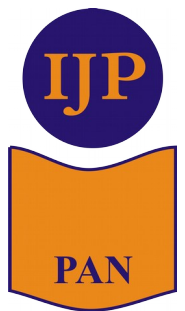
Les représentations du livre aux époques carolingienne et ottonienne
15-17 octobre 2015 - Paris

Autour de *LIBER* Étude (e-)lexicographique



Bruno Bon
Institut de Recherche et
d'Histoire des Textes, CNRS
Paris, France

Krzysztof Nowak
Institut de la Langue
Polonaise (IJP-PAN)
Cracovie, Pologne



I. Etat des lieux lexicographique

II. *Liber* dans les corpus textuels

1. Distribution du mot dans le corpus
2. Analyse manuelle des concordances
3. Analyse automatique des cooccurrences

III. Conclusion



I. Etat des lieux lexicographique : les dictionnaires

- **Le latin classique**

- *Gaffiot* : **I** écorce **II** écrit composé de plusieurs feuilles **1** ouvrage **2** division d'un ouvrage, recueil **3** toute espèce d'écrit

- **Le latin patristique**

- *Blaise Patristique* : **1** volume, rouleau **2** registre **3** l'Écriture sainte **4** toutes sortes d'écrits

- **Le latin médiéval**

- *Blaise Médiéval* : **1** l'Écriture **2** antiphonaire, responsorial, rituel, bréviaire **3** livres de magie noire **4** charte, pouillé



I. Etat des lieux lexicographique : les glossaires

- **Isidorus Hispalensis - Etymologiarum sive Originum liber VI 13 :**

*Codex multorum **librorum** est, **liber** unius voluminis. Et dictus codex per translationem a codicibus arborum seu vitium, quasi caudex, quod ex se multitudinem **librorum** quasi ramorum contineat. Volumen **liber** est a volvendo dictus, sicut apud Hebraeos volumina Legis, volumina Prophetarum. **Liber** est interior tunica corticis, que ligno coheret... Unde et **liber** dicitur in quo scribimus, quia ante usum carte vel membranarum de **libris** arborum volumina fiebant, id est conpaginabantur. Unde et scriptores a **libris** arborum librarios vocaverunt.*



II. *Liber* dans les corpus textuels

- Le corpus
 - Migne, *Patrologia Latina*
 - Volumes 97 (Carolus Magnus) à 139 (Abbo Floriacensis)
 - 1221 textes lemmatisés
 - <http://glossaria.eu/treetagger>
 - 20 437 560 « tokens »
 - 12 354 occurrences de *Liber*_SUB

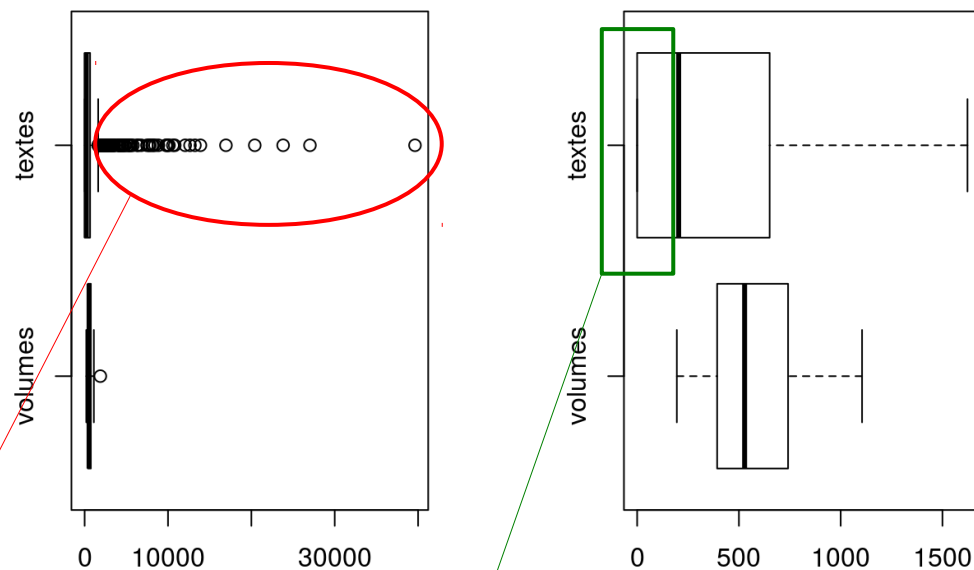


II. 1. Distribution du mot dans le corpus

Nombre moyen d'occurrences du lemme *Liber*_SUB

- par volume : 596
- par texte : 755

Dispersion de fréquence PPM



quelques textes contiennent la plupart des occurrences de *Liber*

près de la moitié des textes (522, ≈ 43 %) sans occurrence de *Liber*



II. 1. Distribution du mot dans le corpus : les maxima par texte

Auteur	Titre	Par million de mots
Auctores varii	<i>Testimonia veterum et recentiorum</i>	39627
Hericus Antissiodorensis	<i>Collectanea</i>	27027
Hucbaldus S. Amandi	<i>Epitaphium Milonis</i>	23810
Erkembaldus Argentinensis	<i>Poematia</i>	20408
Walafrius Strabo	<i>Prophetia Baruch</i>	16949
Joannes Trithemius	<i>Vita Smaragdi</i>	13889
Notkerus Balbulus	<i>De interpretibus divinarum Scripturarum</i>	13230
Auctores varii	<i>Elogia</i>	12638
Nicolaus de Lira	<i>Prologus de commendatione Sacrae Scripturae</i>	12059
Joannes Trithemius	<i>Vita D. Haymonis</i>	10811
Angilbertus Corbeiensis	<i>Versus in librum S. Augustini ad Ludovicum</i>	10582



II. 1. Distribution du mot dans le corpus : les minima par texte

Auteur	Titre	Par million de mots
Ludovicus Pius	<i>Diplomata ecclesiastica</i>	15
Walafrius Strabo	<i>Evangelium secundum Lucam</i>	19
Basilii Caesareae	<i>Regula S. Basilii</i>	32
Joannes Venetus	<i>Chronicon Venetum et Grandense</i>	33
Adrianus II	<i>Epistola et decreta</i>	36
Hrotsuita Gandersheimensis	<i>Comoediae</i>	42
Joannes VIII	<i>Epistolae et decreta</i>	44
Gerbertus Auriliacensis	<i>De geometria</i>	47
Joannes XIII	<i>Epistolae et decreta</i>	50
Grimaldus Sangallensis	<i>Praefationes antiquae</i>	58



II. 1. Distribution du mot dans le corpus : les titres

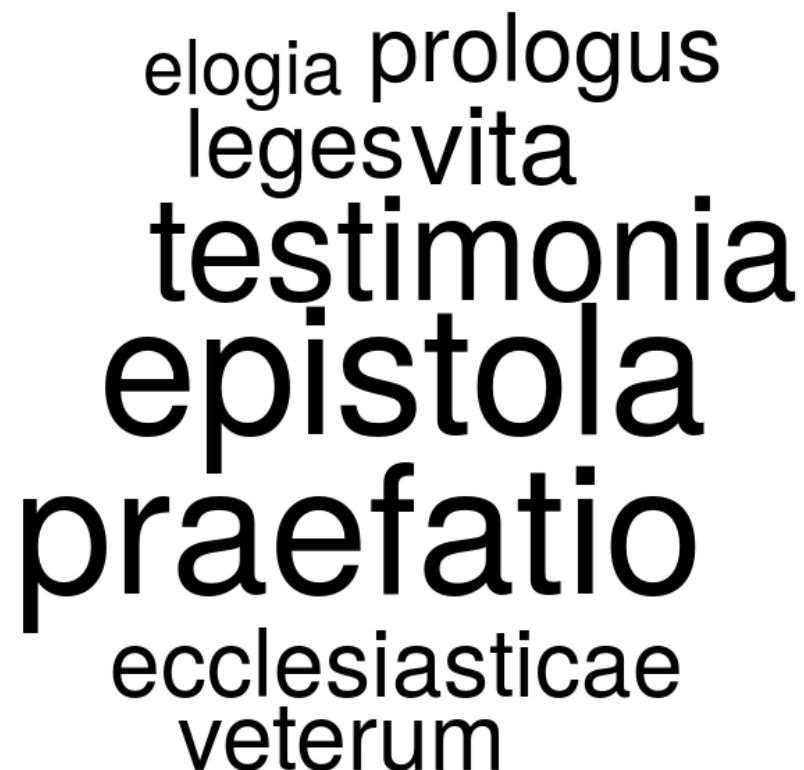
Les mots dans les titres des œuvres avec la plus grande fréquence de *Liber*

avec les « stop words »

sans les « stop words »



A word cloud showing the distribution of words in titles with 'Liber' and stop words. The word 'de' is the largest and most prominent. Other words include 'epistola', 'praefatio', 'prologus', 'versus', 'veterum', 'et', 'leges', 'ecclesiasticae', 'in', 'liber', 'vita', 'elogia', and 'testimonia'.



A word cloud showing the distribution of words in titles without 'Liber' and stop words. The words 'epistola' and 'praefatio' are the largest. Other words include 'elogia', 'prologus', 'legesvita', and 'ecclesiasticae'.



II. 1. Distribution du mot dans le corpus : les maxima par auteur

Auteur	Hits*	Titres	Nombre de mots dans les textes	Par million de mots
Nicolaus de Lira	40	Prologus de commendatione Sacrae Scripturae, Prologus de intentione auctoris, Prologus in moralitates Bibliorum	5538	7223
Isaacus Lingonensis	71	Canones	14675	4838
Joannes Trithemius	68	Vita Smaragdi, Vita altera Rabani Mauri, Vita D. Haymonis	15757	4316
Rudolfus scholasticus	35	Vita Rabani Mauri	10995	3183
Antonius Hispalensis	46	Disquisitio	15506	2967
Aeneas Parisiensis	98	Liber adversus Graecos	35060	2795
Carolus Magnus et al.	171	Capitularia	69079	2475
Adrianus I	61	Epistola	25828	2362
Gotteschalculus Orbacensis	24	Confessio, Confessio prolixior, Epistola, Fragmenta omnia	10334	2322
Conventus Parisiensis	66	Acta	34033	1939

* Auteurs avec au moins 20 occurrences du mot *liber*



II. 2. Analyse des concordances

- **La méthode**

- John Sinclair, *Reading Concordances : An Introduction*, 2003 :

- "the information that a corpus can give you about the usage of a word or phrase is often just a confirmation of what a competent speaker of the language **already knows**"
 - "when we are studying the common words that are used thousands of times a day by everyone, then it is a strange experience to see those words as - in part - **strangers**"



II. 2. Analyse des concordances : le cœur

- **Les formes**

- Singulier : 70%
- Pluriel : 23%

<i>Forme</i>	<i>dont</i>		<i>%</i>
Libro		78	39
Liber		43	21
	QLF	14	7
	SUB	29	14
Librum		21	10
Libris		19	9
Libros		17	8
Libri		16	8
	Sing.	12	6
	Plur.	4	2
Librorum		6	3



II. 2. Analyse des concordances : la colligation

- **Les fonctions**

- dans 50 % des cas :

- $PRE > Liber$
 - $Liber > SUB$

- dans 25 % des cas :

- $Liber > PRE$
 - $SUB > Liber$

PoS	Total N-	Total N+
ADV	21	9
CON	58	27
NAM	47	18
NUM	18	43
PON	44	76
PRE	101	51
PRO	14	17
QLF	63	47
SENT	40	36
SUB	48	109
VBE	48	57



II. 2. Analyse des concordances : la préférence sémantique

- **Les expressions**

- très grande préférence de *Liber* pour le génitif
- grande préférence de *Liber* pour les numéraux (et les adjectifs)

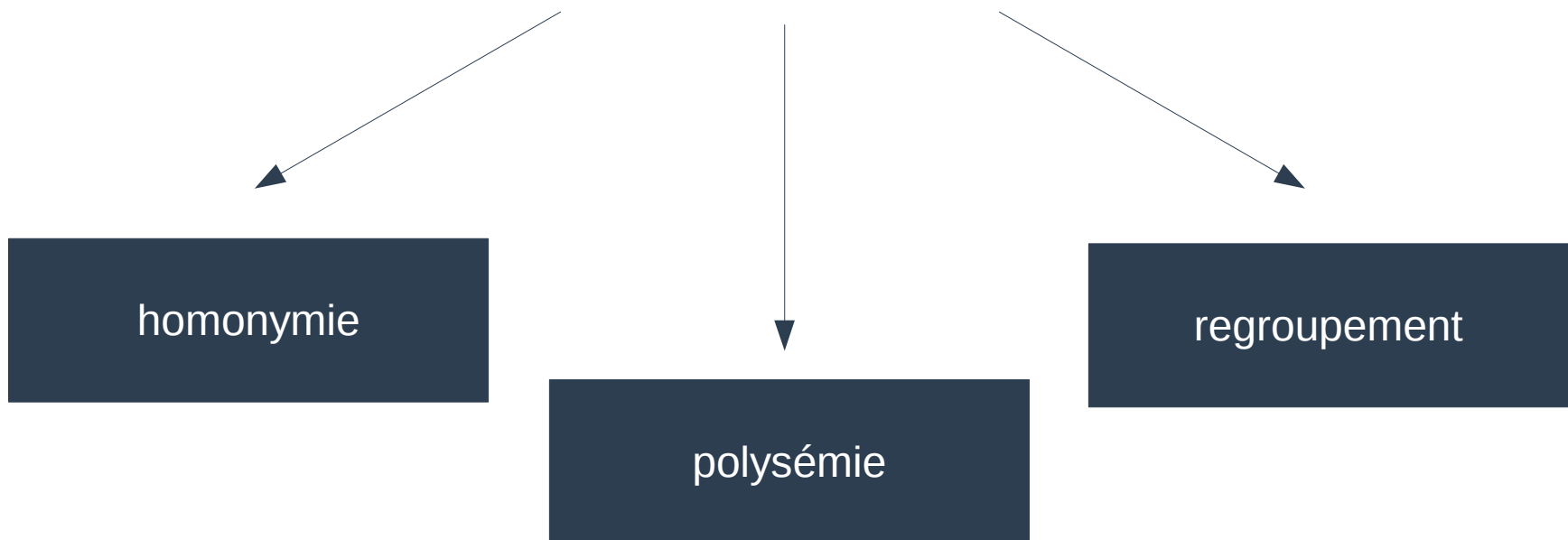
<i>Expression</i>		%
Liber + Génitif	68	34
Liber + NUM	44	22
Liber + QLF	38	19
VBE + (in / ex) libro / libris	34	17
Liber de...	26	13
NAM + (in) libro	25	12
VBE + librum / libros	21	10
SUB + libri / librorum	17	8
Liber + VBE	9	4
Liber + PRE	4	2
Liber + PRO	4	2
Liber + Apposition	3	1
Dans une énumération	2	1



II. 3. Analyse des cooccurrences

- **John Firth (1957) : « *You shall know a word by the company it keeps* »**

mot-5 mot-4 mot-3 mot-2 mot-1 **liber** mot+1 mot+2 mot+3 mot+4 mot+5



II. 3. Cooccurrences et sens du mot : l'homonymie (1)

Liber = 'livre' ou 'libre' ?

secundus, tertius, incipit, primus, quartus,
generationis, seruus, explicit, decimus, quintus,
mortuos, caput, unus, appellatur, sextus, actuum,
septimus, ecclesiasticus, liber, psalmorum, prefatio,
paralipomenon, narrat, numerorum, octauus, i, iste,
exodi, septimo, gratis, apostolorum, continet,
apocryphus, homo, prologus, sapientie, nonus,
quicumque, inuolutus, regum, iesu, incipiunt,
leuiticus, uicesimus, inscribitur, geneseos,
deuteronomii, capitula, scriptus, demonstrat



II. 3. Cooccurrences et sens du mot : l'homonymie (2)

Liber = 'livre' ou 'libre' ?

secundus, tertius, incipit, primus, quartus,
generationis, seruus, explicit, decimus, quintus,
mortuos, caput, unus, appellatur, sextus, actuum,
septimus, ecclesiasticus, liber, psalmorum, prefatio,
paralipomenon, narrat, numerorum, octauus, i, iste,
exodi, septimo, gratis, apostolorum, continet,
apocryphus, homo, prologus, sapientie, nonus,
quicumque, inuolutus, regum, iesu, incipiunt,
leuiticus, uicesimus, inscribitur, geneseos,
deuteronomii, capitula, scriptus, demonstrat



II. 3. Cooccurrences et sens du mot : l'homonymie (3)

Liber = 'livre' ou 'libre' ?

mortuos → *inter mortuos liber*

appellatur → *liber qui appellatur*

gratis → *dimittatur/egredietur liber gratis*

quicumque → *quicumque liber homo*

inuolutus → *recessit celum/ecclesia sicut liber inuolutus*

iesu → *liber generationis Iesu Christi*



II. 3. Cooccurrences et sens du mot : l'homonymie (4)

Liber = 'livre' ou 'libre' ?

secundus, tertius, incipit, primus, quartus,
generationis, seruus, explicit, decimus, quintus,
mortuos, caput, unus, appellatur, sextus, actuum,
septimus, ecclesiasticus, liber, psalmorum, prefatio,
paralipomenon, narrat, numerorum, octauus, i, iste,
exodi, septimo, gratis, apostolorum, continet,
apocryphus, homo, prologus, sapientie, nonus,
quicumque, inuolutus, regum, iesu, incipiunt,
leuiticus, uicesimus, inscribitur, geneseos,
deuteronomii, capitula, scriptus, demonstrat



II. 3. Cooccurrences et sens du mot : l'homonymie (5)

Liber = 'livre' ou 'libre' ?

Sur 57 occurrences, seulement 3 au sens de 'libre'

(114_WalStr_ViSMa10)

... **liber** sed spiritus **iste** In Domini uirtute manet

(113_WalStr_LibPsa3)

ille **liber**, **iste** seruus

(120_PasRad_ExInMa2)

Inter mortuos **liber**. Sed **iste** ...



II. 3. Cooccurrences et sens du mot : méthodologie (1)

Les cooccurrents de *LIBER* 'livre' (par fréquence brute)

, in . - et de qui sum : idem hic scribo -- ut ex is rex
sanctus dico primus lego ad ; secundus beatus item
a unus caput suus non sicut tertius omnis ille diuinus
ipse lex numerus capitulum uita quinque habeo si
deus quia sed quartus quattuor que



II. 3. Cooccurrences et sens du mot : méthodologie (2)

Les 15 premiers cooccurrents de *LIBER* 'livre' (par fréquence brute)

1 (1)	,	6 (17)	de	11 (13)	hic
2 (6)	in	7 (5)	qui	12 (119)	scribo
3 (2)	.	8 (4)	sum	13 (14)	--
4 (7)	-	9 (9)	:	14 (12)	ut
5 (3)	et	10 (56)	idem	15 (34)	ex

* Entre parenthèses : la position du « mot » dans le corpus entier



II. 3. Cooccurrences et sens du mot : méthodologie (3)

L'usage des coefficients

- **Objectif** : filtrer les effets qui ne relèvent pas seulement du hasard
- **Moyens** : différents coefficients soulignent différents types de relation entre les mots, et privilégient certains effets
- **Choix** : cette étude utilise le coefficient de Dice

$$D = \frac{\text{fréquence}_{\text{mot A et mot B}}}{\text{fréquence}_{\text{mot A}} + \text{fréquence}_{\text{mot B}}}$$



II. 3. Cooccurrences et sens du mot : méthodologie (4)

Augustinus

$$D_{Augustinus} = \frac{\text{fréquence}_{Augustinus \text{ avec } liber}}{\text{fréquence}_{Augustinus} + \text{fréquence}_{liber}} = \frac{666}{3290 + 12354} \approx 0,0426$$

10^e dans le corpus entier 1^{er} à la puissance d'association

in

$$D_{in} = \frac{\text{fréquence}_{in \text{ avec } liber}}{\text{fréquence}_{in} + \text{fréquence}_{liber}} = \frac{5543}{463796 + 12354} \approx 0,0116$$

1^{er} dans le corpus entier 10^e à la puissance d'association



II. 3. Cooccurrences et sens du mot : la polysémie (1)

Les 50 premiers cooccurrents de *LIBER*

Augustinus Regum eodem item de legitur
primo quinque scriptum in secundo
Paralipomenon secundus beatus Augustini
idem Numerorum primus Moraliu quatuor
incipit legis Iob sanctus Iudicum Gregorius
caput tertio Trinitate scripsit Sapientie
quem i tertius Ciuitate ex quarto beati uite
huius diuinis legimus Iesu Moysi capite
decimo Euangelii : sancti eiusdem



II. 3. Cooccurrences et sens du mot : la polysémie (2)

1. référence textuelle (localisation)

in libro (-is) legimus (legitur, scripsit, scriptum est)

ex libro (-is)

2. identification d'une partie d'œuvre

in libro primo (tertio, quarto, decimo, eodem)

liber secundus (I, tertius)

caput libri, liber incipit

3. identification d'une œuvre (titre et auteur)

liber Augustini (eiusdem)

in libro (sanctus, beatus) Augustinus (Gregorius, idem)

liber Regum (Paralipomenon, Numerorum, Iudicum, Iob, Sapientie, Moysi, Euangelii)

diuini libri, quattuor libri (= Evang.), quinque libri (= Pent.)

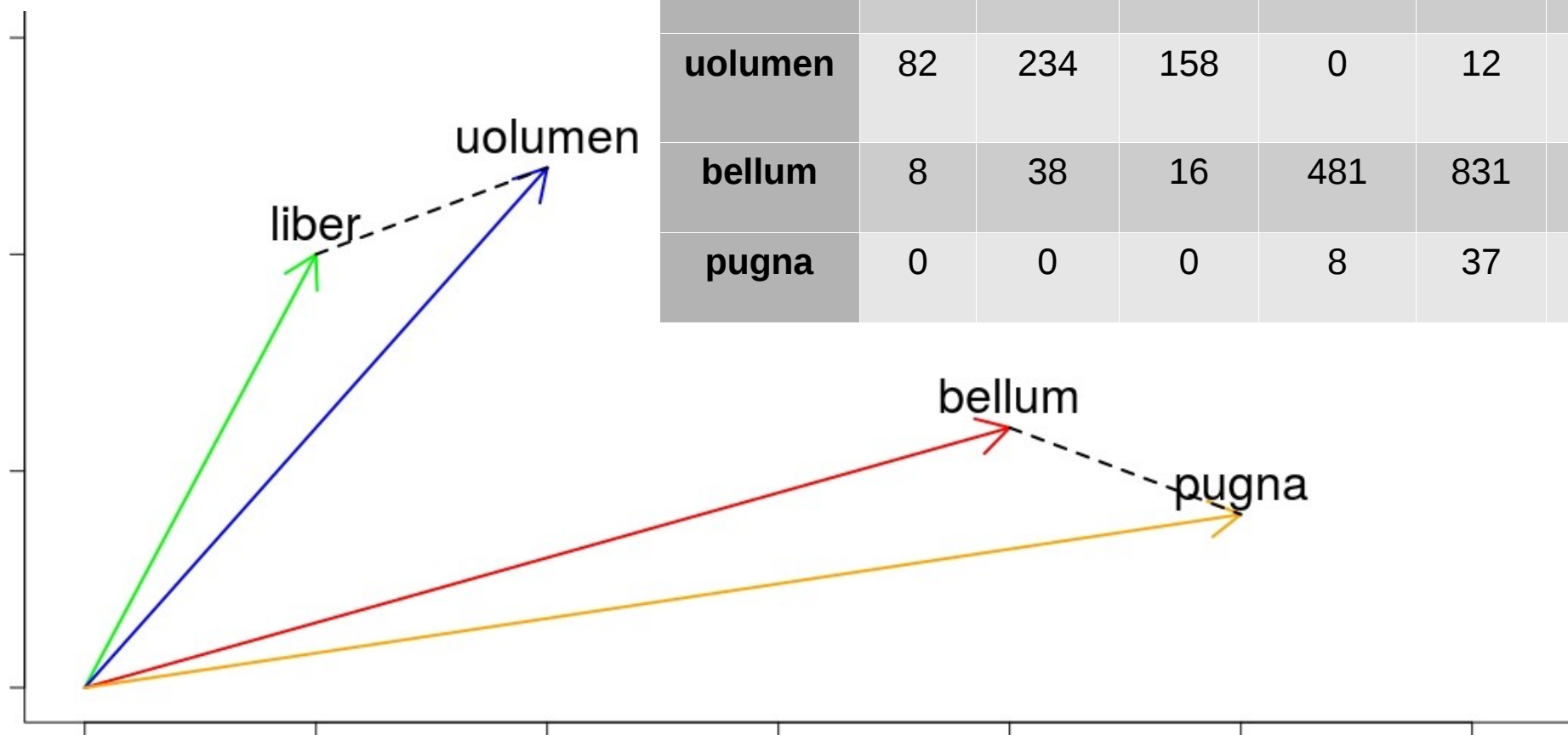
liber Moralium | de (Ciuitate, Trinitate)

4. métaph. 'liber vite' (< Apoc.)

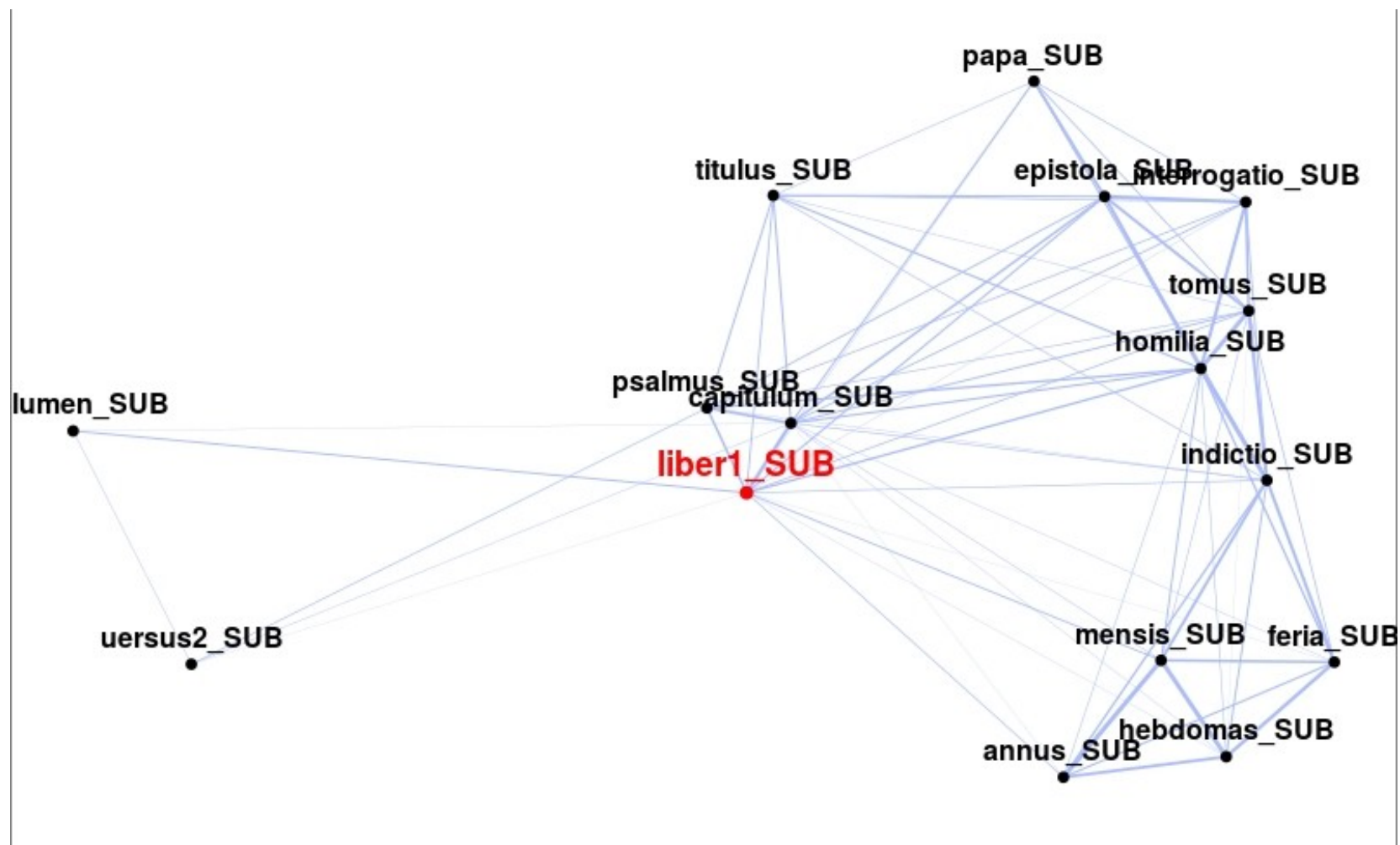


II. 3. Cooccurrences et sens du mot : la sémantique distributionnelle (1)

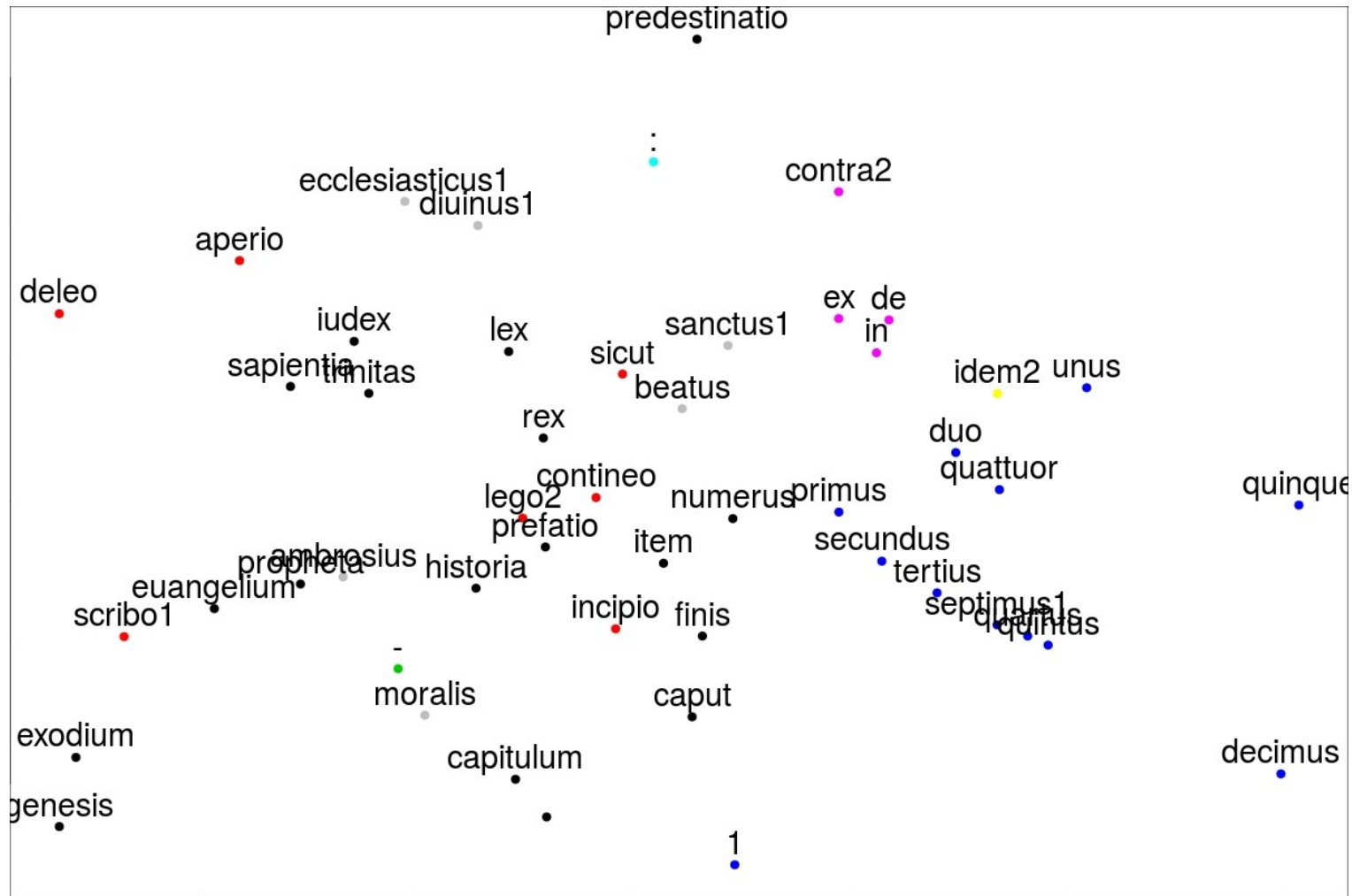
	edo	sacer	lego	ciuilis	gero	infero
liber	55	73	422	0	9	0
uolumen	82	234	158	0	12	0
bellum	8	38	16	481	831	331
pugna	0	0	0	8	37	17



II. 3. Cooccurrences et sens du mot : la sémantique distributionnelle (2)



II. 3. Cooccurrences et sens du mot : la sémantique distributionnelle (3)



III. Conclusion

- **Résultats**

- Validation des observations de la lexicographie traditionnelle
- Relevé de modèles d'usage souvent négligés par les lexicographes

- **Perspectives**

- Approfondir la dimension géographique, et surtout diachronique
- Questions nouvelles



Les représentations du livre aux époques carolingienne et ottonienne
15-17 octobre 2015 - Paris

• **Bruno Bon**

- Institut de Recherche et d'Histoire des Textes (CNRS), Paris, France
- *Novum Glossarium Mediae Latinitatis* (glossaria.eu)
- bruno.bon@irht.cnrs.fr



IRHT

• **Krzysztof Nowak**

- Institut de la Langue Polonaise (IJP-PAN), Cracovie, Pologne
- *Lexicon Mediae et Infimae Latinitatis Polonorum* (scriptores.pl)
- krzysztofn@ijp-pan.krakow.pl

