

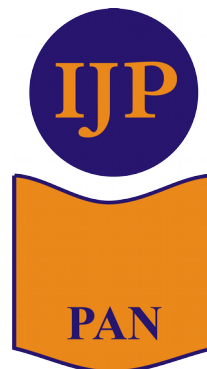
International Medieval Congress
Leeds, 4-7 July 2016

Cena Cibus et cetera

Automatic Extraction of Food-Related Terminology from the *Patrologia Latina*

Krzysztof Nowak

Department of Medieval Latin
*Lexicon Mediae et Infimae
Latinitatis Polonorum*



Institute of the Polish Language
Polish Academy of Sciences
Kraków, PL



Goals

- Query Expansion
- Word Clustering
- Semantic Roles Labelling etc.








NLP Tasks

- Lexical Acquisition
- Thesaurus Generation
- Word Similarity

Sort by relevance ▾ Prod

Show comparison Slideshow

Search result for "writing table"

 LINNMON / FINNVARD Table £75	 LINNMON / FINNVARD Table £70	 LINNMON / FINNVARD Table £70	 FINNVARD Trestle with shelf £25 Width: 46 cm Depth: 70 cm Min. height: 71 cm More options <input type="checkbox"/> Compare Add to basket Save to list	FI Tre £2
				

Outline

- toolbox
 - methods
 - corpus
- food-related words in the *Patrologia Latina*
 - clustering similar terms
 - term frequency
- CIBUS
 - similar terms
 - collocations
- distribution of CIBUS
 - time
 - texts
- CIBUS vs. PANIS
 - texts
 - collocations
- constructing semantic frame

Taming Data

- corpus linguistics
 - concordances
 - collocations
- context and semantics
 - Firth (1957): „You shall know a word by the company it keeps”
- distributional hypothesis
 - Baroni (2010): semantic similarity
 - ← overlap among linguistic contexts



Image Source: *Initiale* database (IRHT CNRS)

Déluge (Besançon, BM,
ms. 0148, 15 c.)

The Corpus: *Patrologia Latina*

Image Source: Wikimedia



Jacques-Paul Migne
(1800-1875)

About the corpus...

- > 102M tokens (incl. interpunction)
- 5238 texts
- electronic version: OMNIA Project (IRHT CNRS + ENC)
- PoS tagged and lemmatized
- time coverage: 2-13 century
 - ca. 32% earlier than 8th century

... that is not a corpus (strictly speaking)

- representative
 - missing some genres of medieval writing
- ~~balanced~~
 - huge subcorpora of one author
 - St. Augustinus: > 140 texts (5% of the total)
 - St. Jerome: > 100 texts

Top similar terms in the PL

	word	freq		word	freq		word	freq
1.	<u>cibus</u>	8	11.	subsidium	4	21.	coquo	2
2.	<u>alimentum</u>	7	12.	comedo	3	22.	elinguo?	2
3.	alimonia	7	13.	copia	3	23.	esurio	2
4.	daps	6	14.	delicia	3	24.	holus	2
5.	refectio	6	15.	edulium	3	25.	legumen	2
6.	<u>epule</u>	5	16.	esus	3	26.	nutrimentum	2
7.	<u>esca</u>	5	17.	<u>pabulum</u>	3	27.	pascua	2
8.	uictus	5	18.	panis	3	28.	<u>pastus</u>	2
9.	alo	4	19.	potus	3	29.	reficio	2
10.	potus	4	20.	saturitas	3	30.	ruminor	2

... to *Patrologia*'s neighbourhoods

alimentarius, alimentum, ambrosia, cibaria, cibus, coliphia, **crudus**, culina, **depasco**, edendum, epulae, esca, **esculentus**, **gulosus**, inedia, macies, pabulum, **pasco**, pastus2, penus, popina, pransus, pulmentum, pulpamentum, **tumidus**, **vescor**, visceratio

—————▶ alimonia,cibus,subsidiu ...
—————▶ manutius,acinaces,exhalatus,picotus ...
—————▶ esca,alimentum,epule,refectio...
—————▶ delicia,edulium,potus2,sapor...

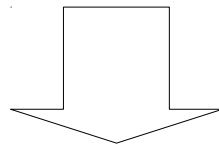
27 LS entries

x

15 similar terms in PL

=

279 unique similar terms



30 top terms

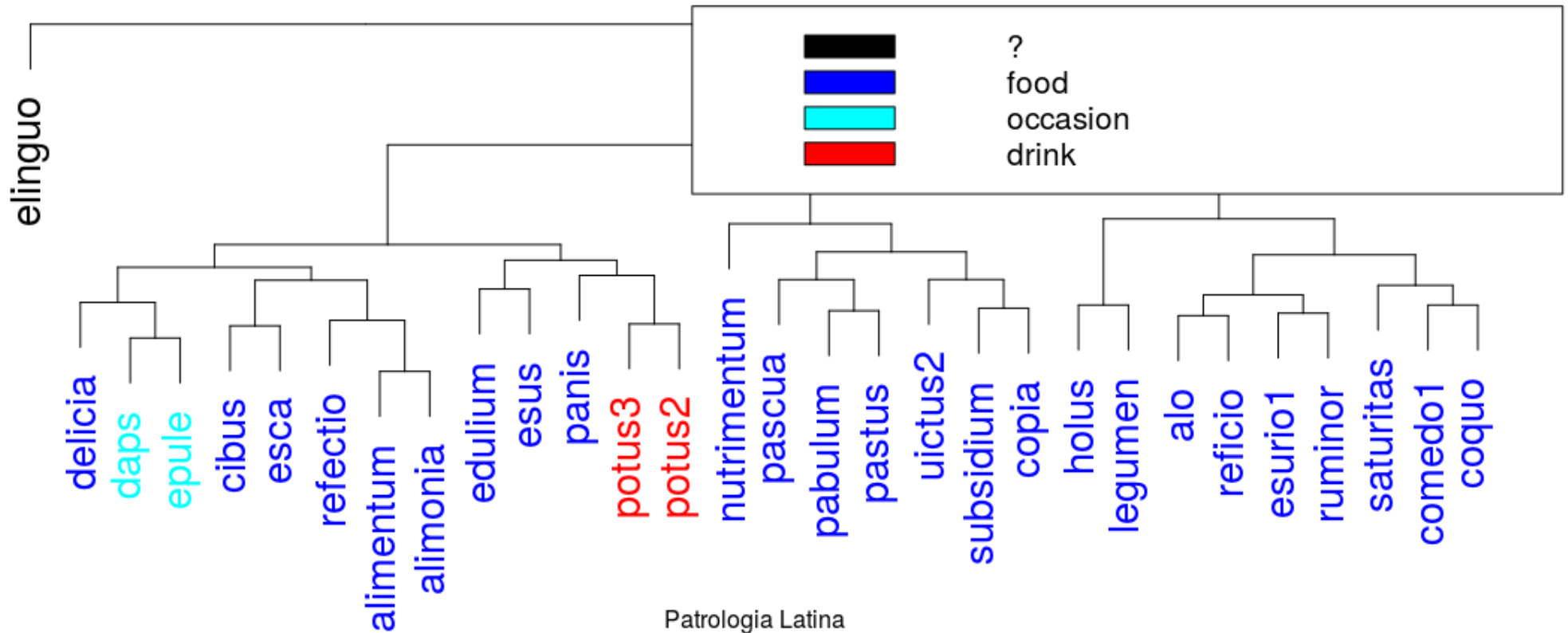
(shared by at least 2 LS entries)

Top similar terms in the PL

	word	freq		word	freq		word	freq
1.	<u>cibus</u>	8	11.	subsidium	4	21.	coquo	2
2.	<u>alimentum</u>	7	12.	comedo	3	22.	elinguo?	2
3.	alimonia	7	13.	copia	3	23.	esurio	2
4.	daps	6	14.	delicia	3	24.	holus	2
5.	refectio	6	15.	edulium	3	25.	legumen	2
6.	<u>epule</u>	5	16.	esus	3	26.	nutrimentum	2
7.	<u>esca</u>	5	17.	<u>pabulum</u>	3	27.	pascua	2
8.	uictus	5	18.	panis	3	28.	<u>pastus</u>	2
9.	alo	4	19.	potus	3	29.	reficio	2
10.	potus	4	20.	saturitas	3	30.	ruminor	2

Top similar terms: clustering

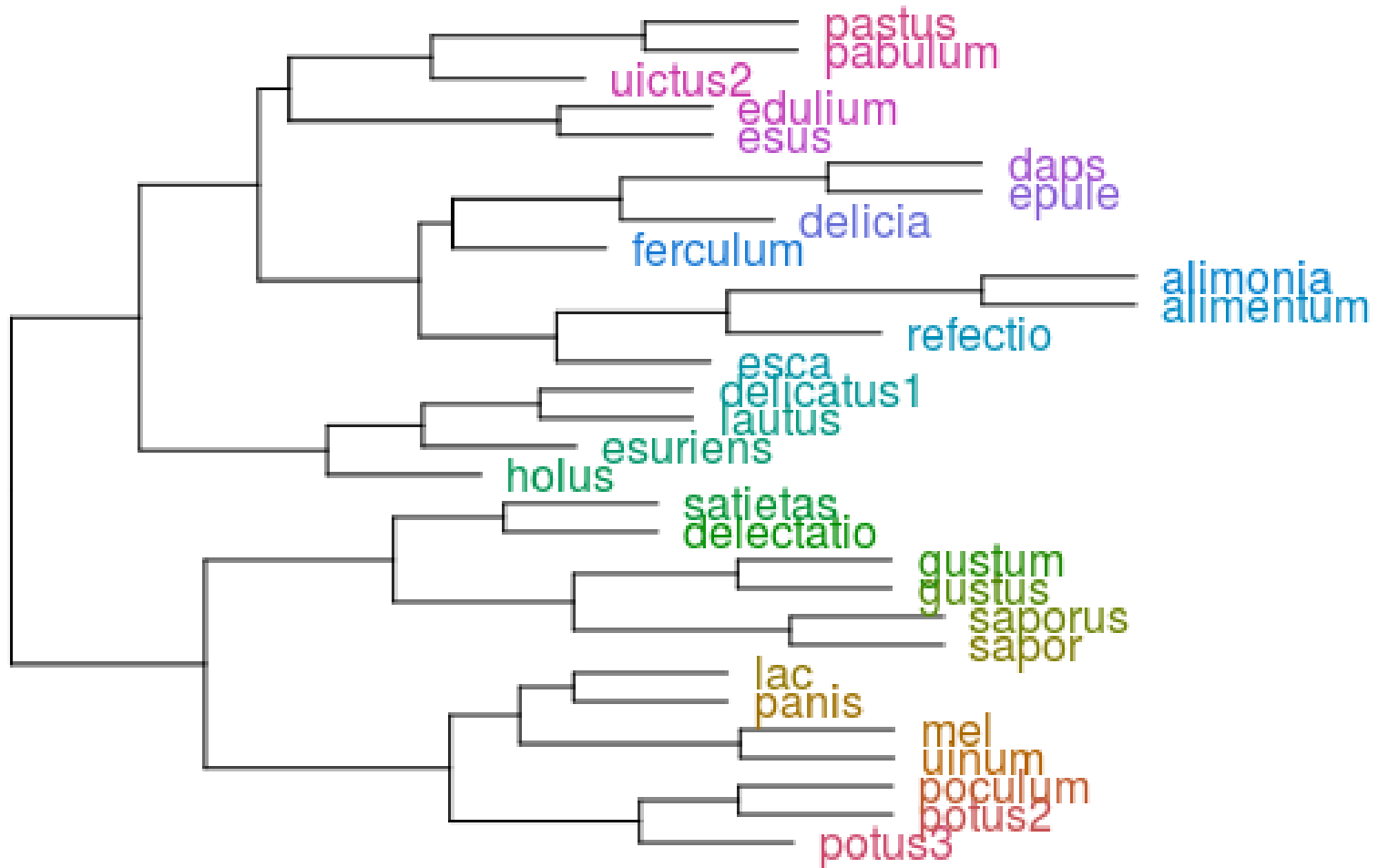
Sense Clusters of 30 Most Frequent Food Words



Similar terms: frequencies

	lemma	count		lemma	count		lemma	count
1	panis	29670	11	uictus	3001	21	esus	1053
2	cibus	19443	12	alo	2631	22	nutrimentum	1041
3	comedo	16281	13	alimentum	2548	23	ruminor	1000
4	copia	6648	14	pascua	2411	24	alimonia	978
5	reficio	6506	15	subsidium	2333	25	saturitas	940
6	delicia	5819	16	coquo	2088	26	pastus	852
7	esca	5691	17	epule	2079	27	daps	727
8	esurio	5315	18	potus	1427	28	legumen	507
9	potus	4493	19	pabulum	1427	29	edulium	409
10	refectio	3521	20	holus	1270	30	elinguo	37

CIBUS: similar terms (2)



CIBUS: collocations vs. similar terms

potus	<u>comedo</u>	utor
<u>potus</u>	uesco	manduco
sumo	dulcis	<u>refectio</u>
solidus	<u>panis</u>	<u>alo</u>
abstineo	pasco	delicatus
<u>reficio</u>	lautus	communis
solidum	<u>esurio</u>	pauper
abstinentia	uenter	paro
capio	corporalis	lac
spiritalis	uilis	preparo

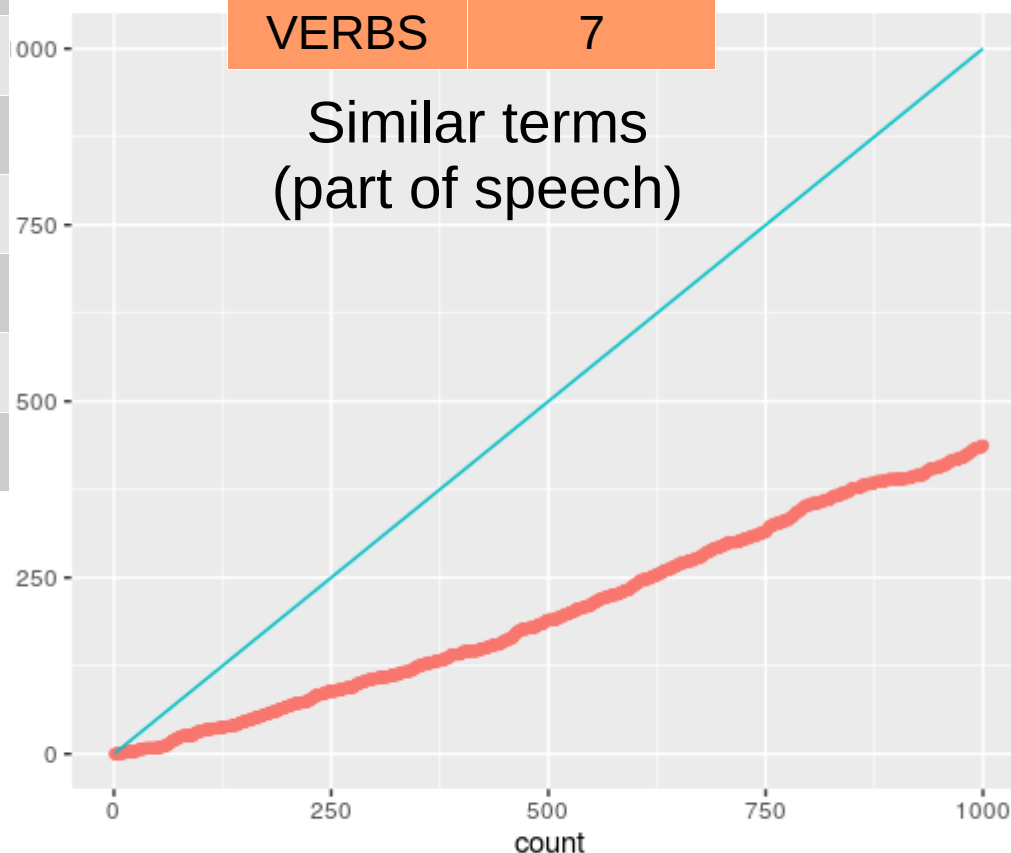
30 strongest collocates of CIBUS

POS	N	POS	N
NOUNS	42	ADV	2
VERBS	34	CON	2
ADJ	19	PRE	1

Collocates (part of speech)

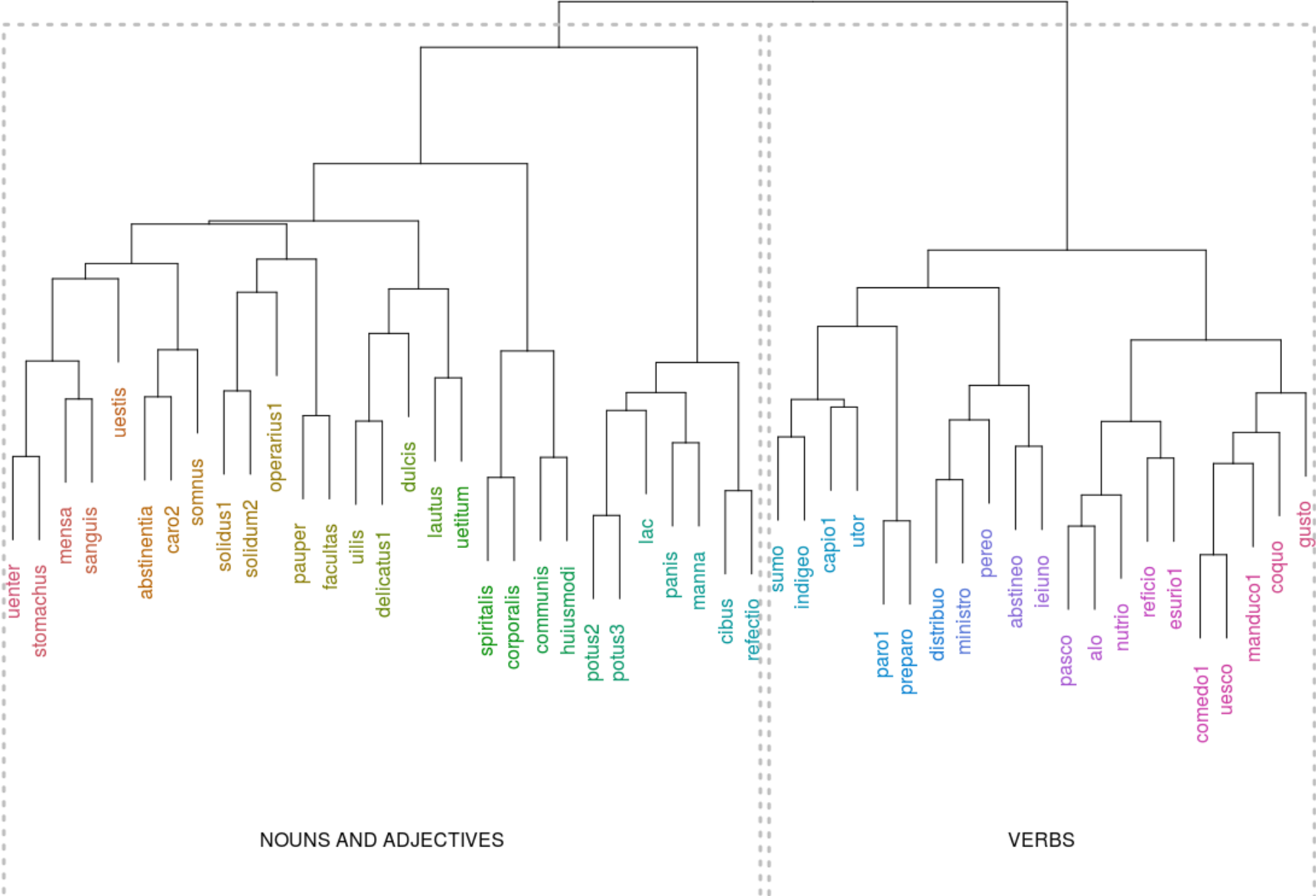
POS	N
NOUNS	80
ADJ	13
VERBS	7

Similar terms
(part of speech)

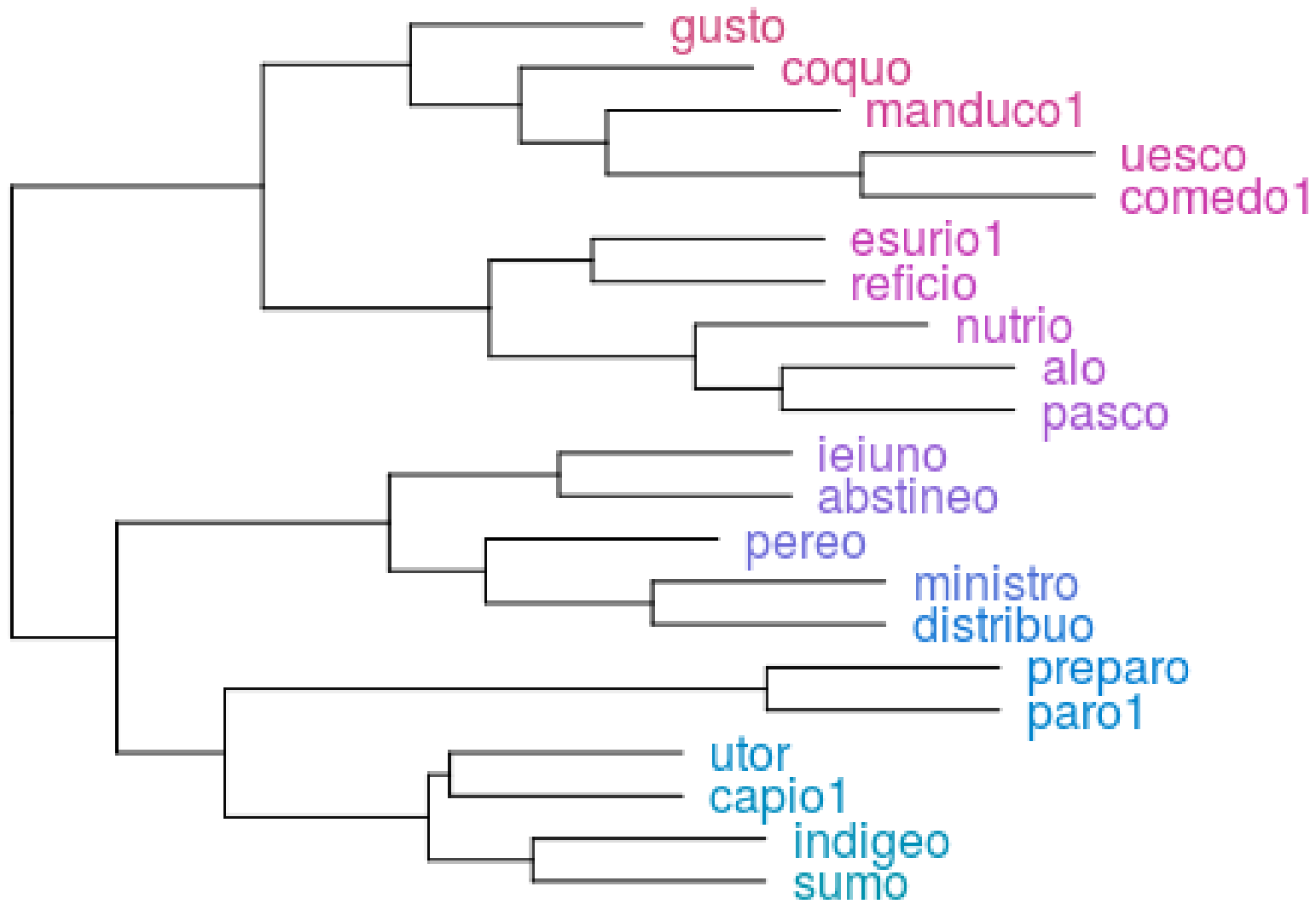


Overlap between collocates (red)
and similar terms

CIBUS: collocations



CIBUS collocations: verbs



CIBUS collocations: difficult cases

INDIGEO

wide range of applications

- expositione, interpretatione
- auxilio, adiutorio, ope
- penitentia, purgatione, correctione...
- poculis, cibo, pane, victu, lacte
- nullo, medico, spoliis
- sapientia, misericordia

VS.

ESURIO

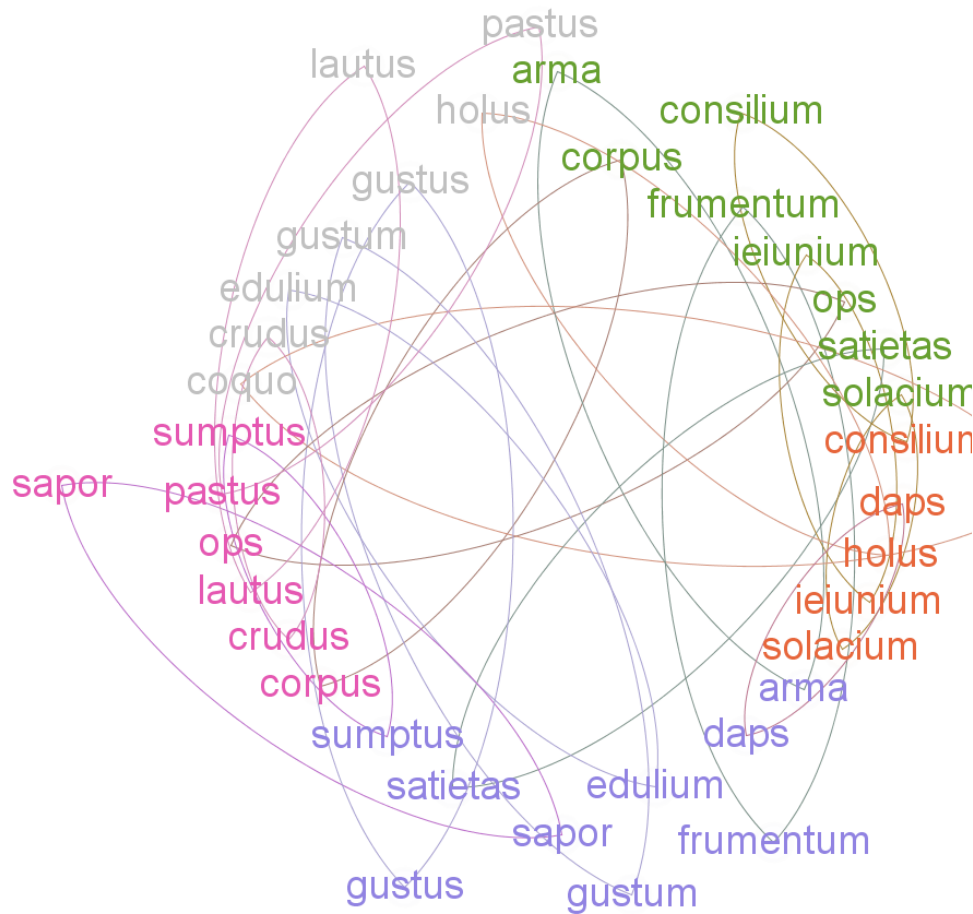
limited application

- *esurientes et sitientes*
- *frange esurienti panem*
- *beati qui esuritis*

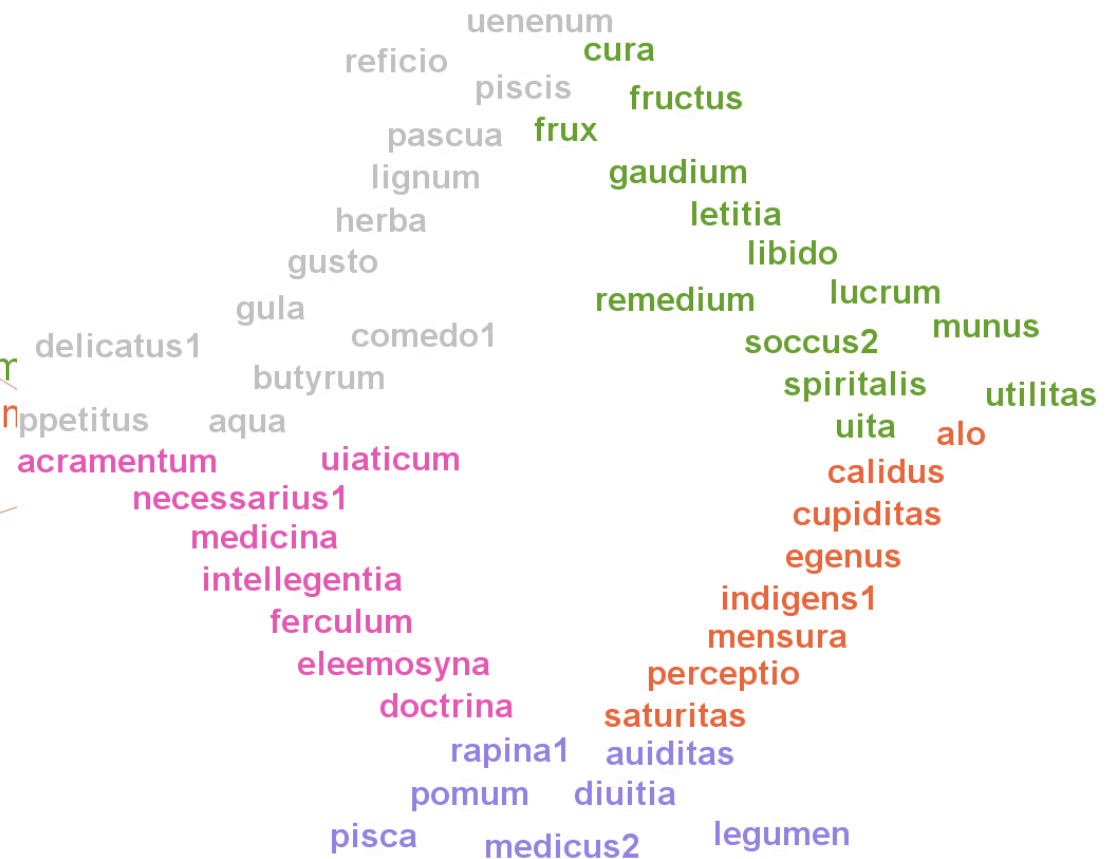
PEREO 137 occurrences

- *Operamini non cibum qui perit, sed qui permanet* (J 6:27) [Do not work for the food that perishes, but for the food that endures to eternal life]
- *Carnalis cibus perit, spiritualis uero permanet* (Beda, In Evangelium S. Joannis)

CIBUS similar terms: time

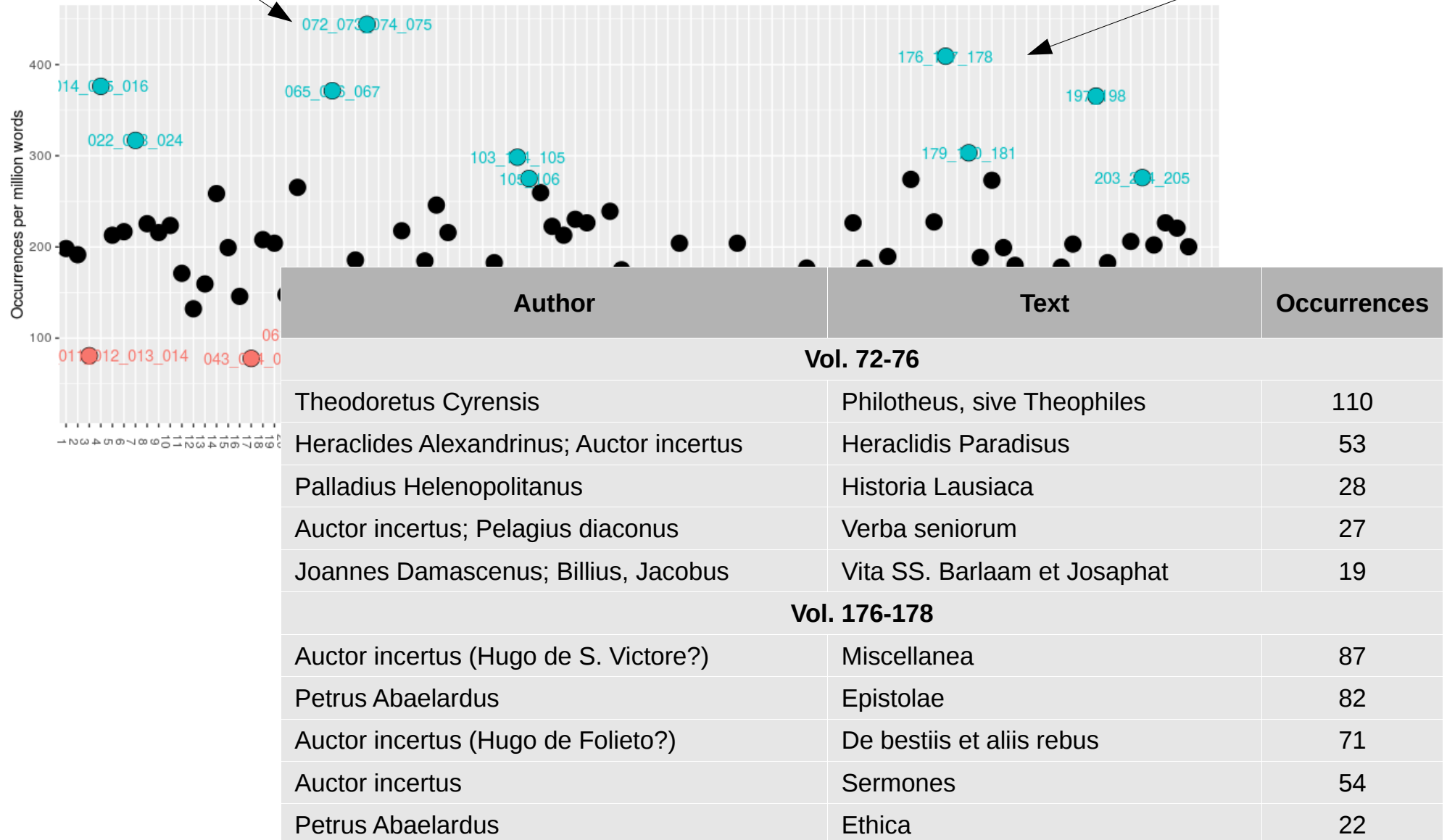


Similar terms attested in 2 periods



Similar terms unique to 1 period

CIBUS frequency: time



PANIS vs. CIBUS: texts

Author	Work
Ratramnus Corbeiensis	De corpore et sanguine Domini
Auctor incertus	Liber poenitentialis
Haltgarius Cameracensis	Liber poenitentialis
Cummianus Hibernus	De mensura poenitentiarum
Petrus Cellensis	De panibus
Lanfrancus Cantuariensis	De corpore et sanguine Domini
Auctor incertus	Expositio canonis missae
Durandus Trobarnensis	Liber de corpore et sanguine C.
Hildeburtus Cenomanensis	De sacra eucharista
Algerus Leodiensis	De sacramentis
Odo Cameracensis	Expositio in canonem missae
Adalhardus Corbeiensis	Statuta pro monachis suis
Hildeburtus Cenomanensis	De expositione missae
Innocentius III	Mysteria evangelicae legis et sacramenti eucharistiae
Auctor incertus	Canones edit
Humbertus Silvae Candidae	Adversus Graecorum calumnias
Auctor incertus	Expositio super missam
Auctor incertus	Sermones
Auctor incertus	In feria II Paschatis
Abedoc; Ethelwolfus	Canones Hibernenses

Author	Work
Wandalbertus Prumiensis	Vita S. Goaris
Ambrosius Mediolanensis	De Elia et jejuniu
Guillelmus abbas	De natura corporis et animae
Lupus de Olmeto	Regula monachorum...
Lactantius	Liber de opificio Dei
Hugo de Folieto?	De medicina animae
Basilus Caesareae	Regula S. Basilii
Hugo de S. Victore	De institutione novitiorum
Ambrosius Mediolanensis	De Noe et Arca
Jonas Bobiensis	Vita Burgundofarae
Auctor incertus	Regula ad virgines
Marbodus Redonensis	Carmina varia
Augustinus Hipponensis	De beata vita
Leandrus Hispalensis	Regula
Hildegardis	Subtilitates diversarum...
Alanus de Insulis	Liber poenitentialis
Cyprianus Carthaginensis	De lapsis
Hieronymus Stridonensis?	De expositione Psalmorum
Hildegardis	Solutiones quaestionum...
Patricius Hiberniae	Confessio

PANIS vs. CIBUS: collocations

uinum	solus
manduco	hordeaceus
quinque	lacrima
aqua	calix
comedo	subcinericius
frango	azymus
uiuus	saturo
propositio	lapis
quotidianus	piscis
fractio	species

panis et uinum
panis et calix
panis et aqua



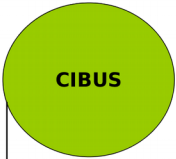
quinque panes et duo
pisces (Matth 14:19)

potus	reficio
potus	spiritalis
sumo	uilis
abstineo	facultas
solidus	uenter
abstinentia	utor
solidum	corporalis
capio	delicatus
lautus	somnus
dulcis	uetitum

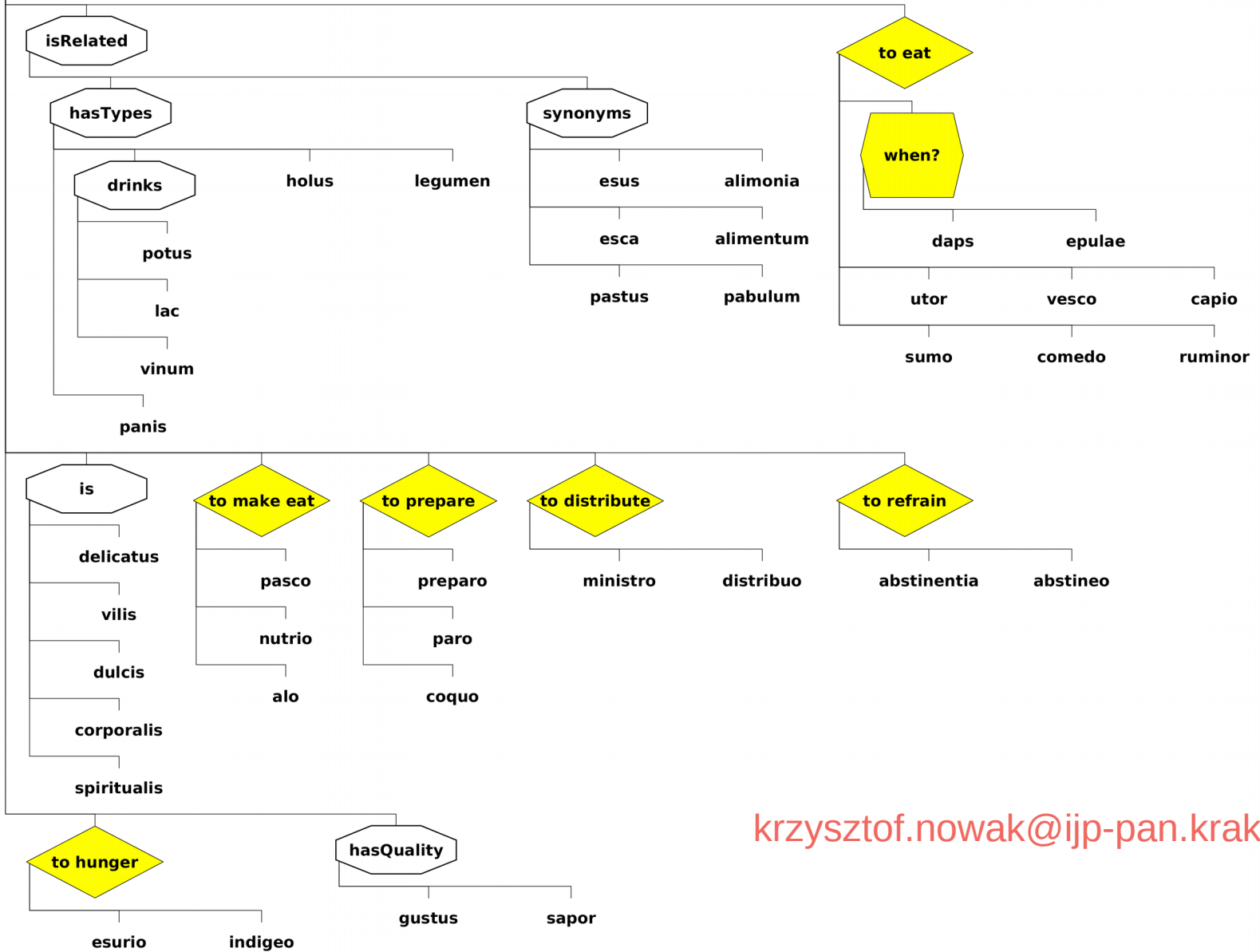
frange esurienti panem
 (Is 58:7)

fuerunt mihi lacrimae meae panis die
 ac nocte (Ps 41:4)

panem nostrum *cotidianum* da nobis cotidie
 (L 11:3)



Conclusions



krzysztof.nowak@ijp-pan.krakow.pl