

Vocabularium Bruxellense

Towards Quantitative Analysis of Medieval Lexicography



Krzysztof Nowak (IJP PAN, eFontes)

Iwona Krawczyk (IJP PAN)

Renaud Alexandre (IRHT CNRS)



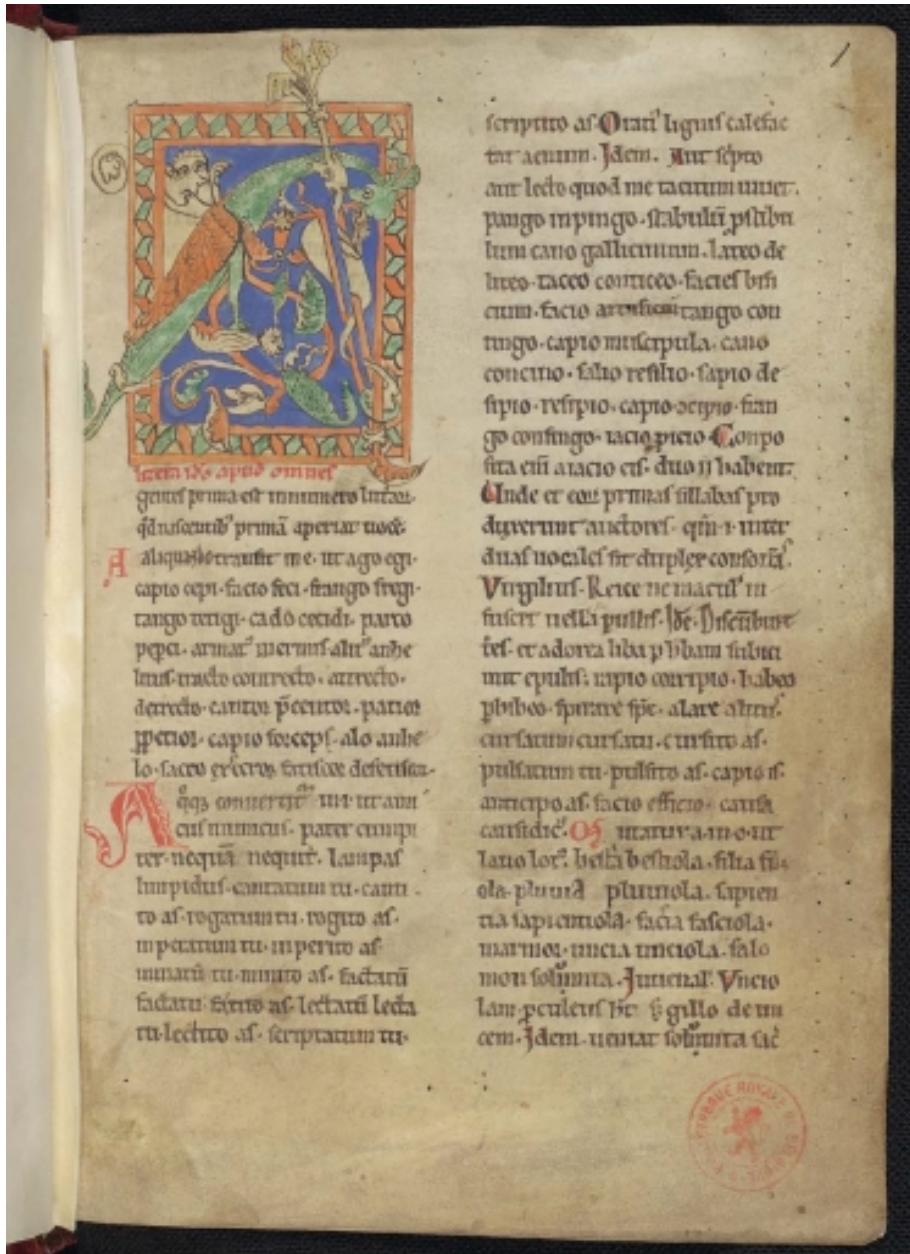
TEI Conference and Members' Meeting 2022
September 12 - 16, 2022
Newcastle, UK

Outline

- *Vocabularium Bruxellense*
 - Annotation
 - Digital edition
 - Analytical Framework
 - Results
 - Conclusions



Ms. Bruxelles, B.R. II 1049



ad. tri
lunii
al. sedi
Ind. tu
sqdoli
i. ddu
E. Tull
accu
ssi. na
culans

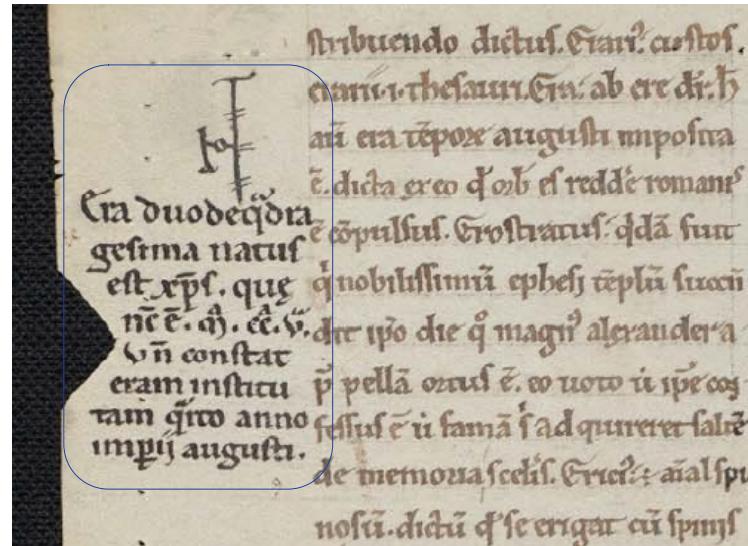
rigid. rigida. dum.
littera pnde
interdu nichil i. q. alicubi
posita nec uocali n. glonias
est. Vocali non i. quia. se
quuntur. glonias n. i. q. q. pcedit. Ideo
g. nichil q. n. glonias n. vocali. Iadet
a ges. digamoni uocat qn u. sibi ista
luis uocalib; ringit. q. duplex i. sic.
que duplex gamam it. s. align tribit

Place and Time

p. 49v

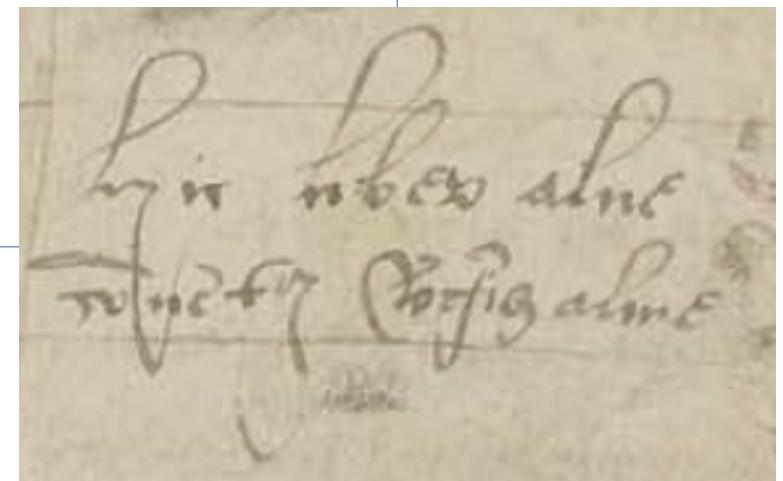
Nota

Era duodequadra
gesima natus
est Christus. Que
nunc est M^a CC^a V^a
unde constat
eram institu
tam quarto anno
imperii Augusti



Aulne Abbey, Belgium

p. 143v



Towards the Digital Edition



ANR Omnia

Outils et méthodes numériques pour
l'interrogation et l'analyse des textes médiolatins
(2009-2013)



ANR Velum

Visualisation, exploration et liaison de ressources
innovantes pour le latin médiéval (2018-2022)



Vocabularium
Bruxellense

Edition électronique d'un glossaire médiolatin
(H.J. Reijnders, 2018-2019)

<https://glossaria.eu/projets/>

Macrostructure of the Vocabularium

Venim illi q. i. ea illi
Cedone: sc̄ qdā popli q̄ pa-
 rent suos sepelire impiu-
 dum h̄c̄ carnes m̄ntas epuli-
 et, ossa; caput autem cincta po-
 cula facit. Et dī a splendore
 acris. Vñ ip̄ ac dipteron fūb̄-
 rū alienū dī pecunia fēicitia.
 Estor, pœlla. Estus, calor l' estas.
 Estimū, estimatio. Escō, pascō.
 Escor, pascor. H̄i esculetū i. esca
 dedit. Escul̄, arbor qdā glan-
 disera. Essedū, uehuelū, baster-
 na. Eshoponū, uestis m̄tior i.
 canaria, q̄ amarciano m̄tula
Eappellat.
Ethir q̄. magl. Enī dī q̄ian-
 ti. Ethnos q̄. gen. Ethnic̄. gen
 cil. Ethna, ē mons qdā h̄c̄ die
 igne i. sulphure dict. H̄i eth-
 ne, aū, p̄ducta pen. Eturia.

fol. 50r

fol. 36r

in modo aqle inuenit. Ethica, pa-
 sio ē. sp̄es qdā febris. Etheria loc̄
 in q̄ sydā sc̄. Dī aū h̄ etheria h̄
 echere. H̄ etheri h̄ns. Vñ in h̄j-
 no eant̄. Quē tñ pontis ḡthera
 colunt adorant p̄. Salibi. Octoū
 nitida unū in ethera. Ethimo-
 logia q̄. libloqū. Ethimoni-
 ubū, logo, oīo. Ethimologia
 cause redditio. Ú idō te diligo q̄
 ate diligor. Etiam si una iunc-
 tio cōposita, er i. si. ú ibi. E-
 tias oportuerit me mou tec-
 n̄ te negabo.

Eua hebraice m̄p̄r uita.
 V̄ calamitas. Eulat̄, ē indi, re-
 gio, dicta ab eulath filio iectan-
 aq̄ possessa ē p̄ diluumi. Accen-
 tat̄ aū h̄ nom̄ in fine. sic cesta
 qm̄ p̄a sc̄ i indeclinabila. Cor-
 ripuit̄ i pen. sic ect̄ p̄a q̄ h̄uit̄

fol. 139v

Veneo, uendor; Vatis, catheus ferri
 vbi p̄qua in loco. nunc. Vagurrit de-
 pit, uagit, vagab̄ h̄ebice, rhythm̄, mis-
 sus, h̄uus, h̄umid̄, ita distin-
 gunt̄. Vd̄ extia sit i. infundit. Uuid̄
 exudat humorē, q̄ ad mediū usq̄ p̄-
 uenit. humid̄ aū qui tot̄ oīo p̄ma-
 duit. In uidez, i. udit̄, i. udo, das;

Venetia est quedā y talie pumcia.
 cuī metropol̄, aqleia. Verona, q̄
 dā ciuitas y talie. Venusū, oppi-
 du und̄ fuit orat̄. Vectones sc̄ celtib̄i
 gens quedā ex gall̄ mixta z h̄beris.

Carico, c̄insequente
 hostē fugiōnum
 frequēt̄
 m̄ponit̄
 dictioib̄
 cōposita, hiat̄ phibendi ú redigo,
 redarguo, redibeo. Subētit̄ i q̄ se
 quen̄ sillaba, ab. s. i alia consonante
 incip̄, ú aspero, asclo, aslo. **D**ā, aliquā
 pon̄, p̄, ú mediant̄, mezzant̄. **D**aliquā
 ruit̄ in. e, ú accedo, accelō, accimulo,
 accendo, accit̄, accip̄, accino, accent̄, acci-
 dia, qd̄, qd̄. **D**ā mutat̄ in. f, ú afficio,
 affini, afflo, affus, affore. Aliquā i. G, ú
 aggruo, aggo, aggello, doceo, dogma. Al-

si, spondeo, m̄ordeo, mōr̄si, ardō si.
 rassū, ordior, si. Aliqñ i mudas, es-
 tū cudo, si. scindo, scissū, cedo, ce si, fo-
 dio, fissis, sindo, fissus. **D**ā, dūc̄, dūc̄, i. c.
 ú attamino, attingo, attineo, atten-
 do, attraho, attonis, attonit̄, incen-
 do, incenit̄, dā, dautie.

Dalmacia ē qdā gr̄e p̄uincia.
 Dachi, siue dache, s. qdā
 seytbie popli, a datia eorū
 p̄uincia dicti, q̄ a gothis originem
 ducit̄. Danai, sc̄ ḡci. Danis, sc̄ qdā
 ḡmanie popli, q̄ regio danomarchia
 nūcupat̄. Unū i. ip̄i q̄ danomarchia
 nichilomin̄ appellant̄. Daniste, se-

Microstructure

Decutere excutere, depellere. **Peredia** edendi auditas. **Trapetum** ⁴⁶¹¹: concha olearia qua oleum exprimitur.

Nubere operire. Inde **nubere** dicitur mulier quando traditur uito quia amborum ibi capita solent nubi, id est “operiri”. Inde etiam **nubes** dicitur aeris densitas a uento congregabati. **Nubes** in Exodo quod dicitur : *tenebrosa illuminans noctem* intelligendus (?) est Egyptiis tenbrosam fuisse, Israelitis uero luminosam. **Nubere** et **enubere** ita distinguntur: **nubit** ciuis ciui, **enubit** extraneo. **Nuptare** . sepe nubere. Nuptiari est nuptias agere.

Scilla ⁴²³⁹ dicitur a greco scillo ⁴²⁴⁰, id est "spolio". **Scylla**: Forci filia. Scylle due fuerunt: filia Nisi ⁴²⁴¹ una et filia Forci ⁴²⁴² altera. Filia Nisi dum patrem suum Minoi tradidisset, uersa est iuxta fabulas in alaudam, pater uero eius in <smirilium> ⁴²⁴³ ut eam persequeretur. **Scyllam** uero, Forci filiam, in nympham mutatam iuxta Siciliam fabulantur. Fingitur uero uitgo pubetenus, reliqua pars cincta dicitur esse lupis et canibus. Que dum facie uitginali blanditur inprouidis, repente arripit incautos suisque canibus et lupis proicit deuorandos. In ueritate autem **Scylla** non marina belua, sed maritima aliqua femina fuit non succincta canibus et lupis, sed rapax ut canes et lupi et superuenientibus inhospitalis. Unde et **Scyllas** uocant saxa in mari latentia. **Scyllam** iuxta Siciliam uocant accolae saxum mari imminens simile a longe prospectantibus celebrate forme. Unde et monstrosam speciem quasi hominis illi fabule dederunt succinctam canibus quia ex collisione fluctuum latrare uidetur.

Layout and entry structure

```
<cb type="foliotation_manuscrit" n="1ra"/>
<!-- manuscript paragraphs -->
<div>
    <entryFree/>
    <entryFree/>
</div>
```

```
<div>
    <entryFree>
        <form>
            <orth type="lemma">Botricus</orth>;
        </form>
        <sense>
            <def>vas vinarium.</def>
        </sense>
    </entryFree>
</div>
```

Forms



```
<form>
    <orth type="lemma">C</orth>
</form>
```

```
<def>
    <oRef>C</oRef> aliquando transit in ...
</def>
```

```
<oRef>C</oRef> aliquando transit in <seg>g</seg>, ut
<mentioned>efficio</mentioned><pc>/</pc><mentioned>effigies</
mentioned>, <mentioned>centum</mentioned><pc>/</pc>
<mentioned>quadrингента</mentioned>,
<mentioned>ango</mentioned> pro <mentioned>anco</mentioned>
aliquando in <seg>q</seg>, ut <mentioned>stercus</mentioned>
<pc>/</pc> <mentioned>sterquilinium</mentioned>,
<mentioned>acumen</mentioned> <pc>/</pc>
<mentioned>aquila</mentioned>...
```

Sense relations



```
<form>
    <orth type="lemma">Carenum</orth>
</form>
<sense>
    <def>
        cui contraria est
        <mentioned type="contr">sappa</mentioned>
    </def>
</sense>
```

```
<form>
    <orth type="lemma">Belue </orth>
</form>
<sense>
    <def>
        proprie dicuntur marina animalia,
        <mentioned type="dif">Bestie</mentioned>
        que dentibus aut unguibus seviunt
        <mentioned type="dif">Fere</mentioned> ...
    </def>
</sense>
```

Quotations

```
<cit>
  <bibl>
    <author norm="Horatius" xml:id="A_auth_1">
      Oratius
    </author>
    <title norm="Epodon"/>
    <biblScope norm="2:2:169"/>
    <quote>lignis calefactat aenum </quote>.
  </bibl>
</cit>
<cit>
  <bibl>
    <author sameAs="#A_auth_1">
      Idem
    </author>
    <title norm="Saturaे"/>
    <biblScope norm="1:6:122-123"/>
    <quote>Aut scipto aut lecto quod me tacitum iuvet.</quote>
  </bibl>
</cit>
```

Sources, notes and responsibility

```
<entryFree>
    <form><orth type="lemma">Abominari</orth></form>
    <note n="6" place="above" type="attested" resp="#HRijnders">
        <bibl>
            <author>Papias</author>
        </bibl>
    </note>
</entryFree>
```

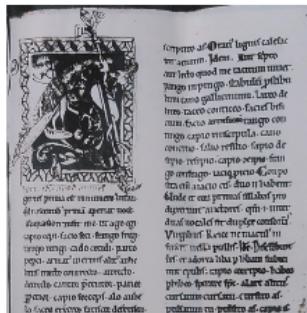
```
<bibl>
    <author>Papias</author>
    <biblScope>p. 23</biblScope>
    <ref target="https://gallica.bnf.fr/ark:/12148/bpt6k594399/f24.item"/>
</bibl>
```

The Digital Edition



Vocabularium Bruxellense

Transcription annotée du manuscrit Bruxelles, Bibliothèque royale de Belgique II 1049



Transcription et recherche des sources initiale : Hatty F. Reijnders.

Édition annotée : Hatty Reijnders (sainte informatique), Iwona Krawczyk (convention en XML-TEI, correction des données et recherche des sources), Elenia Ghittighelli (ODD), Renaud Alexandre (responsable de la publication), Krzysztof Nowak (conseil scientifique et technique).

Édition annotée sous licence Creative Commons BY-NC-SA. Consulter la [notice explicative](#).



<https://ngml.scriptores.pl/>

Renaud Alexandre
Bruno Bon



Iwona Krawczyk
Krzysztof Nowak



```
<form>
  <orth type="lemma">Absentale nomen</orth>
</form>
<note n="4" place="above">
  <bibl>
    <choice>
      <abbr>J. Engels, C. H. Kneepken:</abbr>
      <expan>Sig.</expan>
    </choice> p. 57</bibl> s.v. <ref>ab:
  alicuius rei absentia imponitur, ut 'silentium' (abest', 'calvitium' "capillorum privatio", 'tenel-
```

cappa/cappula, pennula/penna, turca/turcula, caliga/caligula, caconiugationem composita vadunt, ut 'do das/reddo -is, gero ge

Absolutum nomen ³: dicitur quod per se intelligitur nec

Absoluta verba dicuntur que nullis casibus agent, licet figura vivit asinum, id est "more asini".

How it works?



1. Read in the TEI XML file and parse the XML tree

```
xml = os.path.join('.../.../xml/Vocabularium-Bruxellense.xml')
nss = {'tei':'http://www.tei-c.org/ns/1.0'}
tree = ET.parse(xml).getroot()
```

```
entries_tree = [ entry for entry in tree.findall("./{http://www.tei-c.org/ns/1.0}entryFree") ]
forms_tree = [ form for form in tree.findall("./{http://www.tei-c.org/ns/1.0}form") ]
lemmas_tree = [ orth for orth in tree.findall("./{http://www.tei-c.org/ns/1.0}orth[@type='lemma']") ]
```

2. Process elements and tokenize strings

```
lemmas = [ ''.join(lem.itertext()).strip() for lem in lemmas_tree]
lemmas_tokens = [len(lem.split()) for lem in lemmas]
lemmas_first = [lem[0].upper() for lem in lemmas]
```

How it works?



3. Count and tabulate

```
: vocbrux = pd.DataFrame({"lemma":lemmas,"letter":lemmas_first,"lemma_tokens":  
vocbrux["folium"] = [ entry.xpath('.//preceding::tei:pb[1]', namespaces=nss)  
#exclude els  
excl = ['tei:note']  
#entry strings  
lemmas_ents = [ lem.xpath('.//ar  
#with lemma  
lemmas_ents_list = [ clean_xml(e  
#without lemma  
lemmas_ents_withoutlemma_list =  
  
#all entries  
vocbrux.count()  
#entries per letter  
entries_count = vocbrux.groupby("letter").size().reset_index(na  
entries_count["perc"] = (entries_count["count"]/entries_count["  
entries_count["folia"] = vocbrux.groupby(["letter"])["folium"].  
entries_count["perc_cum"] = entries_count["perc"].cumsum()  
entries_count["folia_cum"] = entries_count["folia"].cumsum()  
entries_count.sort_values("folia", ascending=False)
```

	letter	count	perc	folia	perc_cum	folia_cum
0	A	978	12.607967	67	12.607967	67
14	P	791	10.197241	53	77.994070	443
2	C	910	11.731339	51	28.670878	140
17	S	758	9.771819	50	92.161918	522
4	E	489	6.303983	36	40.105711	205
11	M	437	5.633621	34	61.815135	356

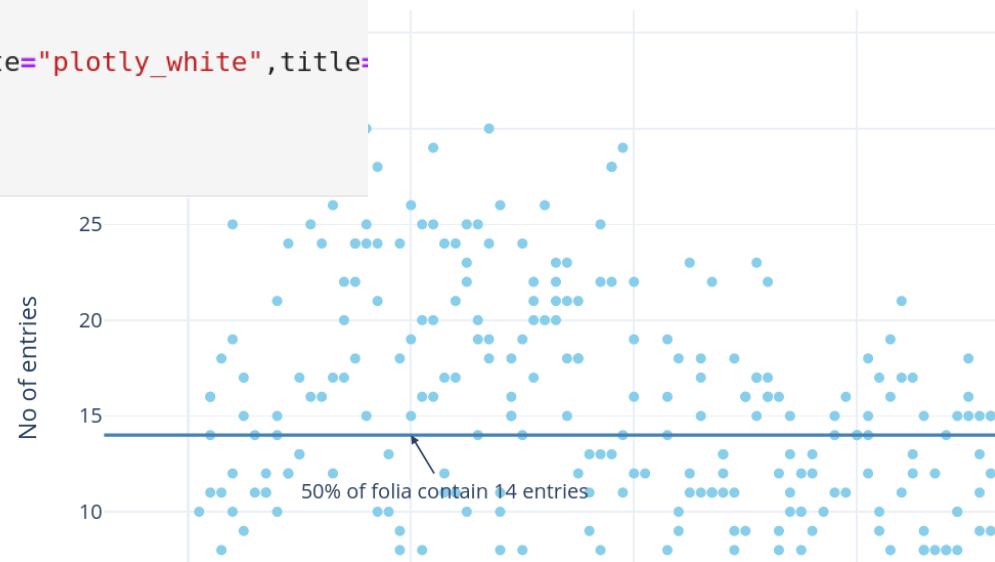
How it works?



4. Visualize

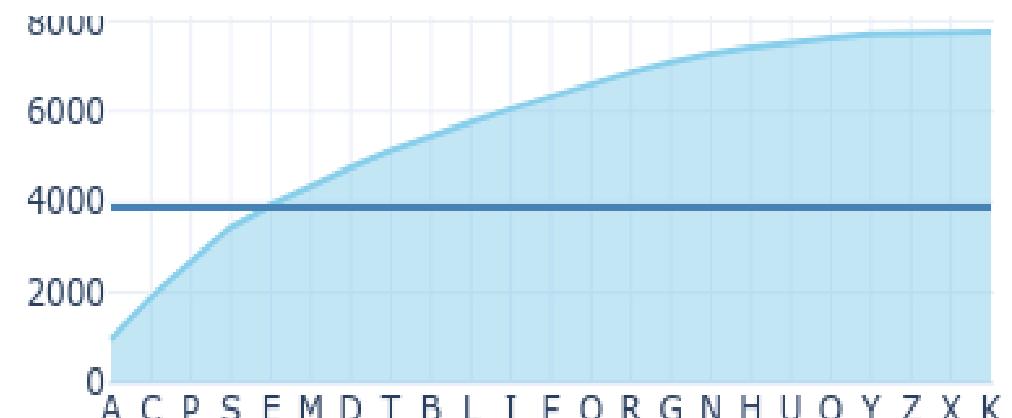
```
#list of lemmas by folium
fig = go.Figure(data=[go.Scatter(
    x=df.fol_int,
    y=df["lem_per_fol"], mode="markers", marker_color="skyblue")]
)
fig.add_hline(y=df["lem_per_fol"].median(), line_color="steelblue")
fig.add_annotation(x=20, y=df["lem_per_fol"].median(),
                    text="50% of folia contain " + str(int(df["lem_per_fol"].r
arrowhead=2))

fig.update_layout(width=1200, height=600, template="plotly_white", title=
fig.update_xaxes(title="Folium")
fig.update_yaxes(title="No of entries")
fig.show()
```



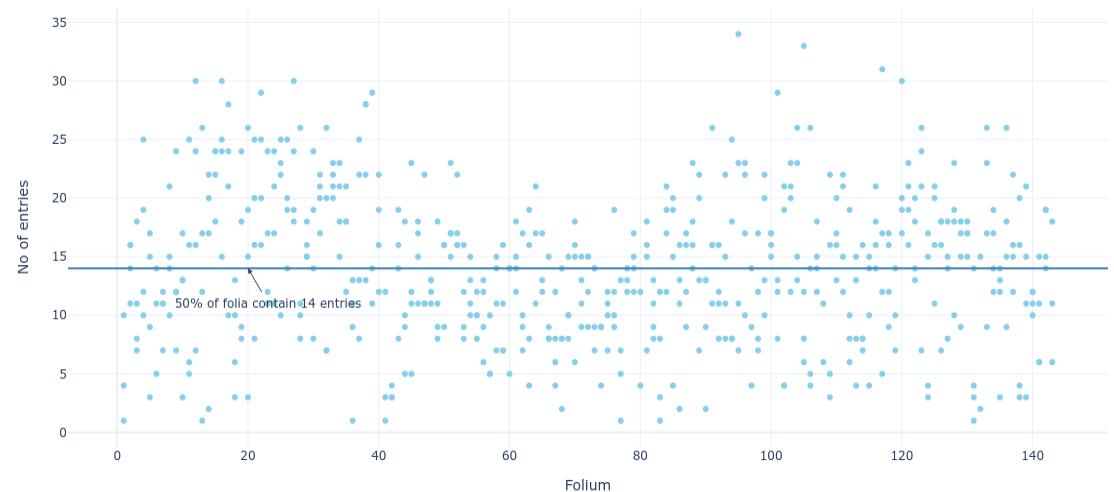
Macrostructure: Entry Distribution

Letter	No of entries	No of pages
A	978	67
C	910	51
P	791	53
S	758	50
F	489	36



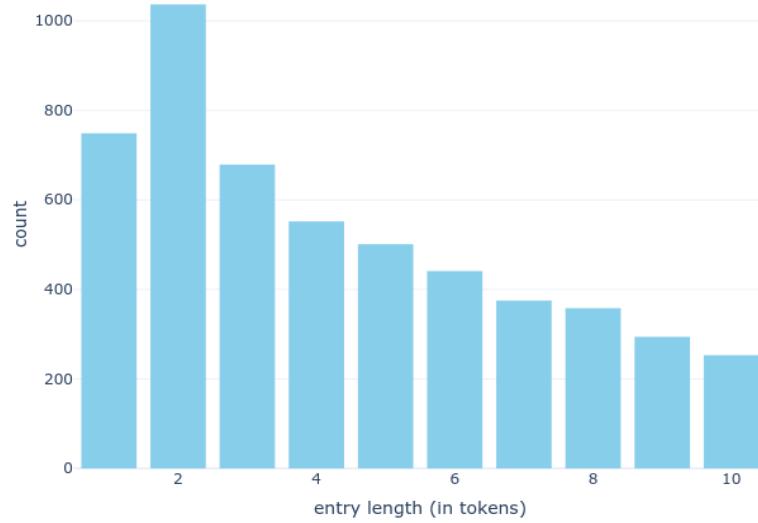
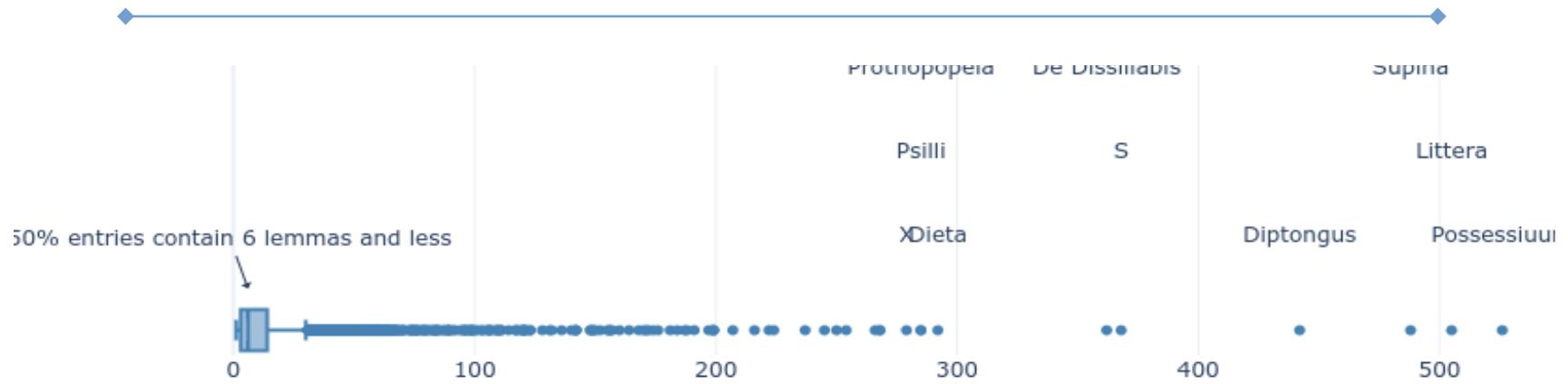
Letter	No of entries	No of pages
Q	92	11
Y	79	6
Z	24	3
X	17	3
K	10	2

Number of entries (by the first letter)

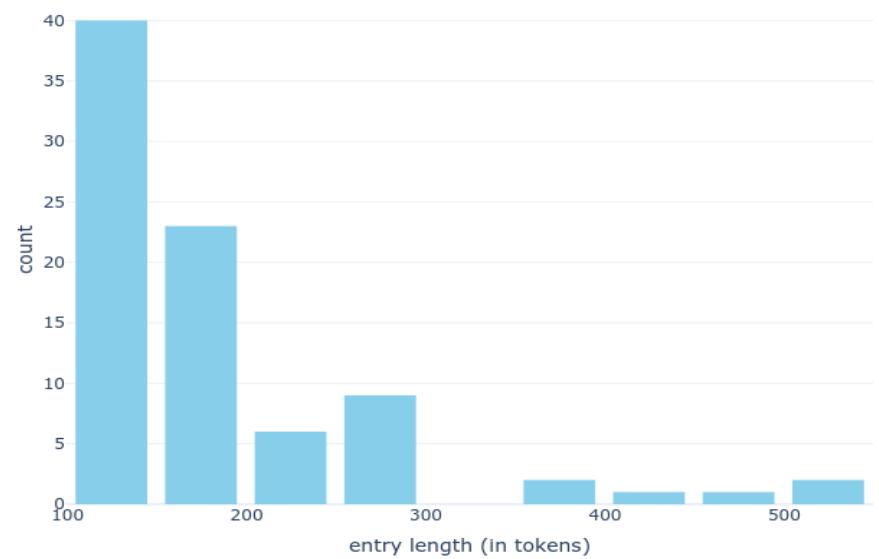


Number of entries per folium
(2 clusters visible)

Macro-structure: Entry Length



Short entries (up to 10 tokens)



Long entries (more than 100 tokens)

Micro-structure: Source quotations

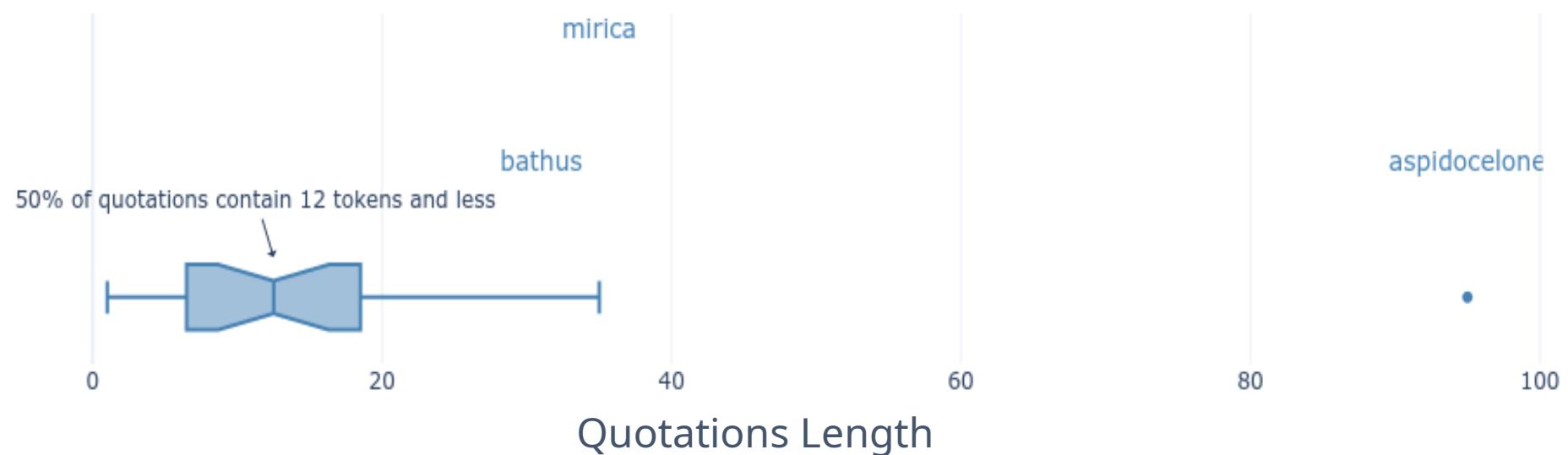
Number of Quotations	Number of Tokens
0	6958
1	589
2	125
3	48
4	18
5	7
6	5
7	3
8	1
9	2

Super et **supra** hoc differunt. **Super** est quod eminet, **supra** quod substratum aliquid habet. **Super** aliquando prepositio est, ut Uirgilius : *Multa super Priamo rogitans super Hectore multa*, aliquando aduerbiū, ut idem Uirgilius :

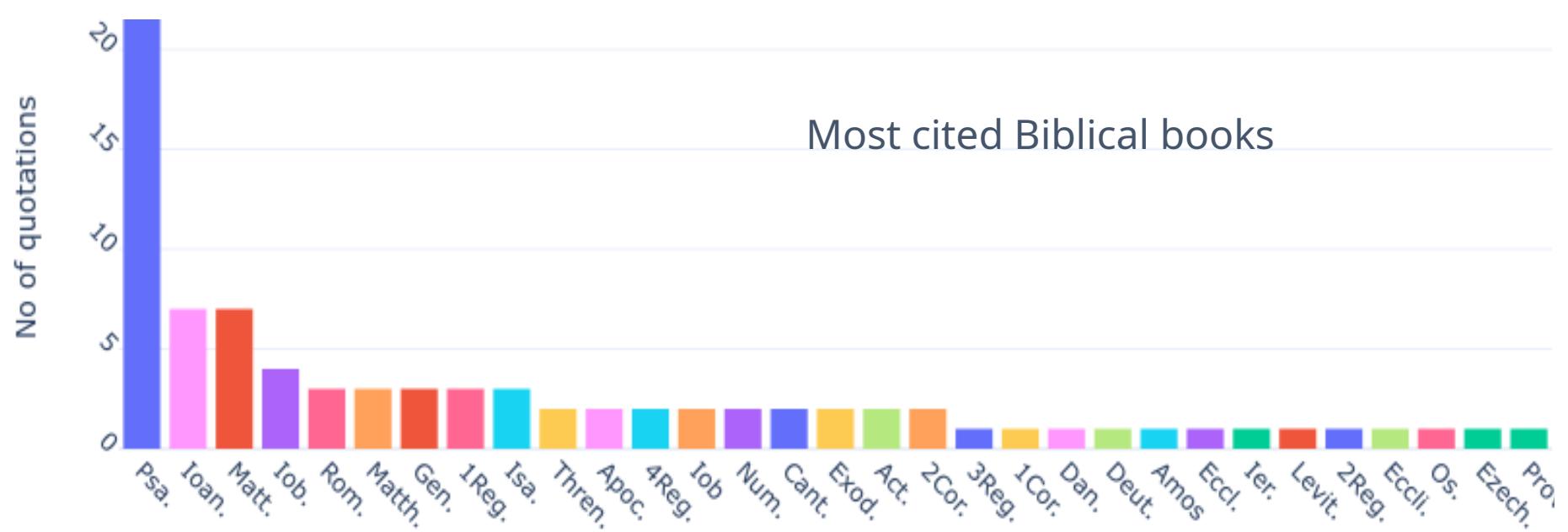
Satis una superque
Uidimus excidia

. **Super** et **subter** contraria sunt. Utraque tamen tam accusatio quam ablatiuo iungitur, licet quibusdam uideantur figurate ablatiuo iungi quomodo 'post' et 'ante' et 'circum'. Uirgilius : *Ille admirans uenerabile donum / fatalis uirgo longo post tempore uisum*. Idem : *Omnia que multo ante memor prouisa reponas*. Idem : *Anna ui*

Number of Quotations



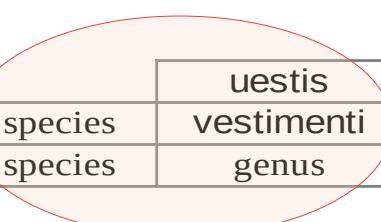
Micro-structure: Source quotations



Conclusions

- TEI XML data + ad-hoc scripts for text analysis
- digital edition as a research tool
- improving reproducibility of research
- Future work
 - extending research on *Vocabularium's* sources
 - automatic alignment and source detection

VocBrux	gauina	uestis	super	humeros	purpuram	habens
Pap	gavina	species	super	humeros	purpuram	
LibGloss	gauina	species	super	humeros	purpuram	



Conclusions

Towards framework of comparative medieval lexicography research

- data availability
- degree of interoperability

iwona.krawczyk@ijp.pan.pl

krzysztof.nowak@ijp.pan.pl

renaud.alexandre@irht.cnrs.fr

