

# *Dictionary with Code*

## Towards Quantitative Analysis of Medieval Lexicography

Renaud **Alexandre**  
Bruno **Bon**

Iwona **Krawczyk**  
Krzysztof **Nowak**

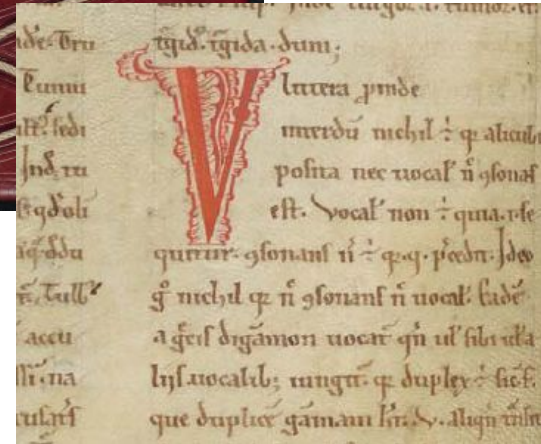
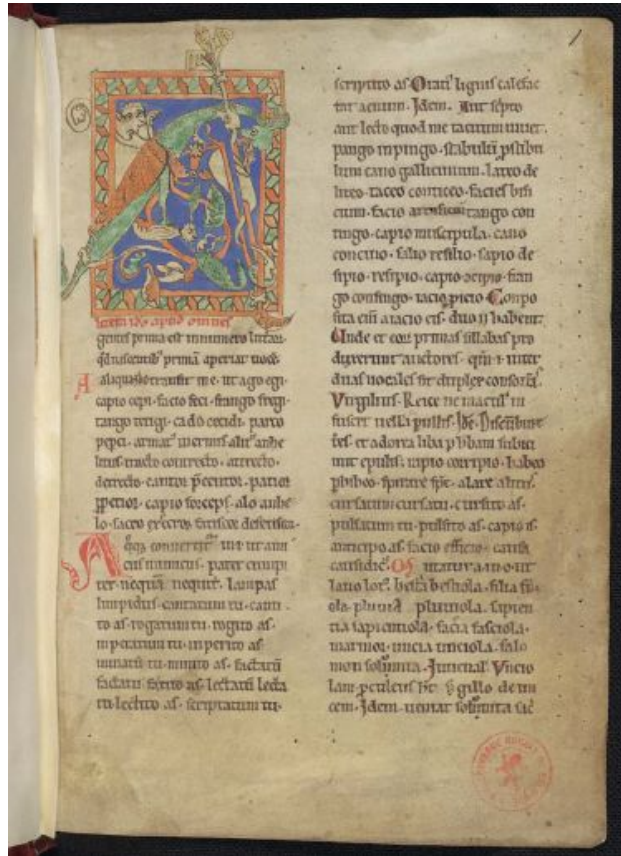


# Outline

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



# Vocabularium Bruxellense: an invisible dictionary

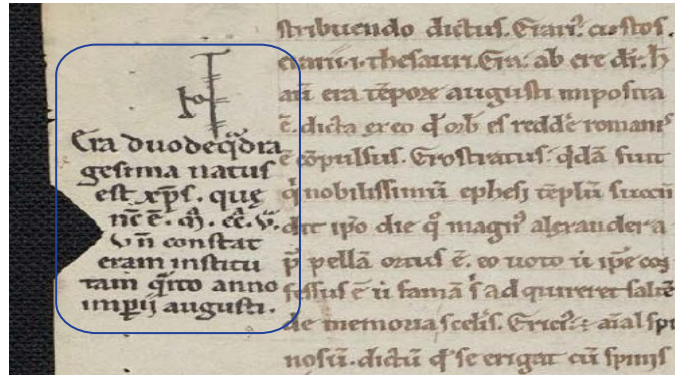


Ms. Bruxelles, B.R. II 1049

# Place and Time

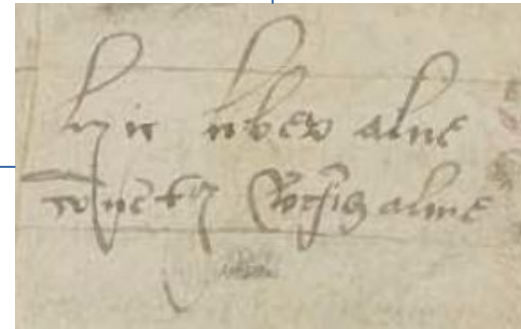
p. 49v

Nota  
Era duodequadra  
gesima natus  
est Christus. Que  
nunc est M<sup>a</sup> CC<sup>a</sup> V<sup>a</sup>  
unde constat  
eram institu  
tam quarto anno  
imperii Augusti



Aulne Abbey, Belgium

p. 143v



# Towards the digital edition

---

H.F. Reijnders



IRHT

Institut de recherche  
et d'histoire des textes

Iwona **Krawczyk**  
Krzysztof **Nowak**



Renaud **Alexandre**  
Bruno **Bon**



<https://ngml.scriptor.es.pl/>



# Macrostructure

Veniunt illi q. i. ca. m. i.  
**E**ssedone: se qda popli q pa  
reunt suos sepelire impu  
ducte h. ay. carnes murtas q. ule  
ede. ossaq. capm. aut. emela po  
cda factur. Et dr a splendore  
aer. vñ q. p. ac. diphougon sebi  
Et alonā dr. parana feratū a.  
Estor. pella. Estus. caloi. l. estas.  
Estumi. estumario. Estco. pasco.  
Estor. pascoz. In estilar. i. estil  
dadr. Estul. arbor qda glau  
dsera. Estedū. ueludū. balsa  
na. Estophonū. uelst mūoz. i.  
caustia. q. a mariano mūala  
**E**uā hebraice mīpē ura.  
V. calamita. Estylar. ē indiq. re  
gro. dīda ab euilard filio rethan  
aq. possessa ē p. deluūm. decen  
tatē aū h. nom. in fine. sic. cēsla  
qm. pa. se. i. m. decmalūta. Cor  
ripit. i. pen. sic. cēta. pa. q. hūc

fol. 50r

oe. Venec. uendoz. Vatik. cath. em. ferri  
Vbi p. qua. in low. nunc. Sagurrit. de  
pit. uagur. Vagedab. hebraice. v. r. h. m. i. g. m. i. s.  
**U**dul. z. huuyd. z. humid. ra. distin  
gunt. Vd. extra. fit. i. infundit. Vup  
grudat. humore. z. q. ad. mediū. usq. p.  
uenit. humid. aū. qui. tot. oīo. p. ma  
durr. In. udoz. z. udtal. z. udo. das.  
**V**enecia. est. quēda. y. talie. p. uencia.  
V. cur. mē. pol. z. aq. lera. Verona. z. q.  
da. ciuitas. y. talie. V. enū. i. y. o. p. i.  
du. und. furr. orat. V. edonel. h. et. i. b. i.  
gen. quēda. ex. gall. m. i. r. a. z. h. y. ber. i.

fol. 139v

**O**uestro. i. insequente  
hoste fugio  
frequē  
mipont  
dicitur  
dopost. a. h. u. r. p. h. i. b. e. n. d. i. u. r. e. d. i. g. o.  
red. u. q. u. o. r. e. d. i. b. e. o. S. u. b. t. i. n. e. i. q. u. i. s. e.  
quē. s. i. l. l. a. b. s. r. a. l. i. a. c. o. n. s. o. n. a. n. t. e.  
m. e. a. p. u. i. p. r. o. n. i. c. h. o. a. l. t. o. O. q. a. l. i. q. u. i.  
p. o. i. t. p. x. u. i. m. e. d. i. t. i. n. e. z. e. n. t. i. B. a. l. i. q. u. i.  
f. u. i. t. i. n. e. u. i. a. c. e. d. o. a. c. c. e. l. o. a. c. c. e. m. u. l. o.  
a. c. c. e. d. o. a. c. c. i. p. e. a. c. c. e. r. o. a. c. c. e. n. t. o. a. c. c. e. r.  
t. u. a. q. d. s. e. p. t. O. q. m. u. c. a. t. i. n. f. i. u. a. f. f. i. c. i. o.  
a. f. f. i. m. a. f. f. i. u. s. a. f. f. o. r. e. A. l. i. q. u. i. l. o. u. i.  
a. q. u. i. s. a. q. u. i. s. a. q. u. i. s. t. o. d. o. c. e. o. d. o. g. m. a. A. l. i.  
s. i. s. p. o. n. d. e. s. s. i. m. o. d. e. o. m. o. s. t. r. a. a. r. d. o. s. i.  
r. a. s. s. i. o. d. i. o. s. i. s. A. l. i. q. u. i. t. u. i. d. u. a. l. e. s. s.  
u. c. u. d. o. s. i. s. s. e. c. u. d. o. s. e. s. s. i. s. C. e. d. o. c. e. s. i. s. s. o.  
d. i. o. s. o. s. i. s. s. i. n. d. o. s. i. s. s. i. s. O. q. u. i. o. u. e. r. t. i. t.  
u. a. r. t. a. n. i. m. o. a. r. t. i. n. g. o. a. r. t. i. n. c. o. a. r. t. e. n.  
d. o. a. r. t. a. b. o. a. r. t. o. n. i. s. a. c. c. o. n. i. t. t. u. e. n. t.  
d. o. u. e. n. t. i. a. s. d. d. d. a. u. e. r. e. l.  
**D**almatia ē qda gric p. uencia.  
Dachi siue dache. s. q. d. a. m.  
s. e. y. t. h. e. p. o. p. l. i. a. d. a. t. a. c. o. y.  
p. u. e. n. t. a. d. i. c. t. i. q. a. g. o. t. h. i. s. o. r. i. g. i. n. e. m.  
d. u. c. i. t. Danu. s. e. g. r. i. Dan. s. e. q. d. a.  
q. u. a. m. e. p. o. p. l. i. i. q. u. i. r. e. g. i. o. d. a. n. o. m. a. r. c. h. a.  
n. u. c. u. p. a. t. C. i. a. i. p. i. q. d. a. n. o. m. a. r. c. h. i.  
n. e. b. i. l. o. m. i. n. i. a. p. e. l. l. a. n. t. Dan. s. e. s. e.  
Dan. m. i. p. e. u. o. t. e. u. i.

fol. 36r

# Microstructure

---



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

**Scilla** <sup>4239</sup> dicitur a greco scillo <sup>4240</sup>, idest "spolio". **Scylla**: Forci filia. **Scylle** due fuerunt: filia Nisi <sup>4241</sup> una et filia Forci <sup>4242</sup> altera. Filia Nisi dum patrem suum Minoi tradidisset, uersa est iuxta fabulas in alaudam, pater uero eius in <smirilium> <sup>4243</sup> ut eam persequeretur. **Scyllam** uero, Forci filiam, in nympham mutatam iuxta Siciliam fabulantur. Fingitur uero uirgo pubetenus, reliqua pars cincta dicitur esse lupis et canibus. Que dum facie uirginali 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 accole 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/>  
  <entryFree/>  
  <entryFree/>  
</div>
```



# Forms

---

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

canonic form

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

form reference

```
<oRef>C</oRef> aliquando transit in  
<seg>g</seg>, ut  
<mentioned>efficio</mentioned>  
<pc>/</pc>  
<mentioned>effigies</mentioned>
```

levels of analysis

[C sometimes becomes g, as in *efficio* / *effigies*]

# Sense relations

---

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

antonym

```
<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>
```

near-synonym

# 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="Saturae"/>
    <biblScope norm="1:6:122-123"/>
    <quote>Aut scripto aut lecto quod me tacitum iuuet.</quote>
  </bibl>
</cit>
```

explicit

hidden

# 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>
```

source

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

source  
link

# Making the dictionary visible (after 8 centuries)

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

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

**Absolutum nomen**<sup>3</sup>: dicitur quod per se intelligitur nec

**Absoluta verba** dicuntur que nullis casibus egent, licet figu  
vivit asinum, idest "more asini".



## Vocabularium Bruxellense

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



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

Édition numérique : Hatus Reijnders (système informatique), Iwona Krawczyk (conversion en XML-TEI, c  
sources), Elena Ghiringhelli (ODD), Renatid Alexandre (responsable de la publication), Krzysztof Nowak (



<https://ngml.scriptor.es.pl/>



# Dictionary-with-code (1)

---

## 1. Read in the 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 ]
```

# Dictionary-with-code (2)

## 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('..ancestor::tei'  
#with lemma  
lemmas_ents_list = [ clean_xml(ent[0].xpath  
#without lemma  
lemmas_ents_withoutlemma_list = [ clean_xml
```

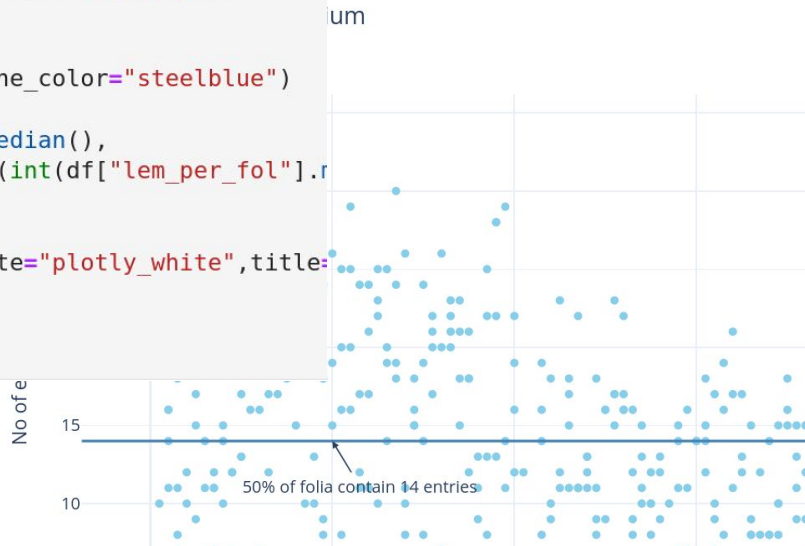
```
#all entries  
vocbrux.count()  
#entries per letter  
entries_count = vocbrux.groupby("letter").size().reset_index(na  
entries_count["perc"] = (entries_count["count"]/entries_count["  
entries_count["folia"] = vocbrux.groupby(["letter"])[ "folium"].  
entries_count["perc_cum"] = entries_count["perc"].cumsum()  
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

# Dictionary-with-code: 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"].median())),
    arrowhead=2)
fig.update_layout(width=1200,height=600,template="plotly_white",title="Folium")
fig.update_xaxes(title="Folium")
fig.update_yaxes(title="No of entries")
fig.show()
```



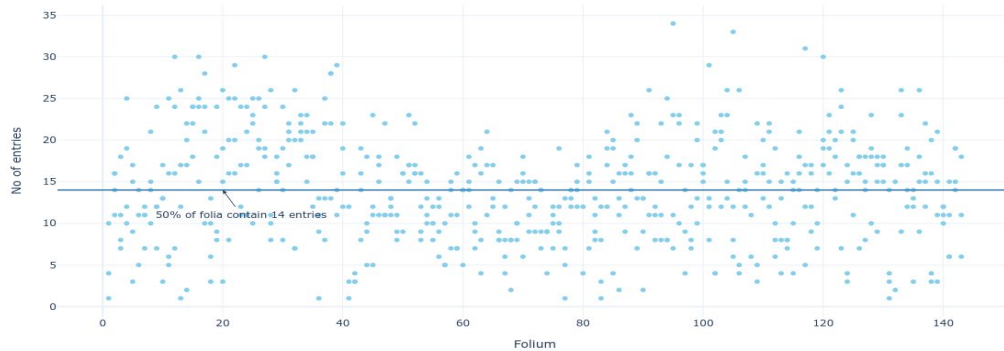
# Macrostructure: Entry Distribution

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



Cumulative sum of the number of entries

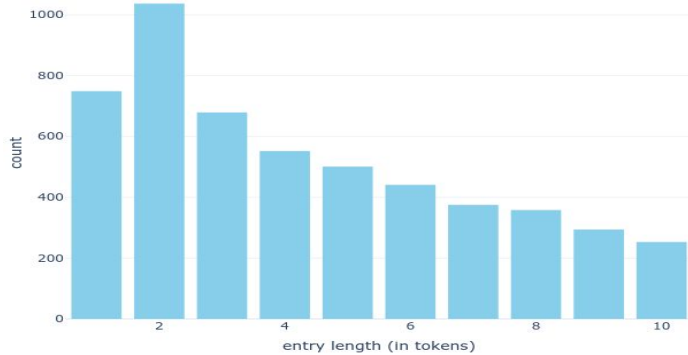
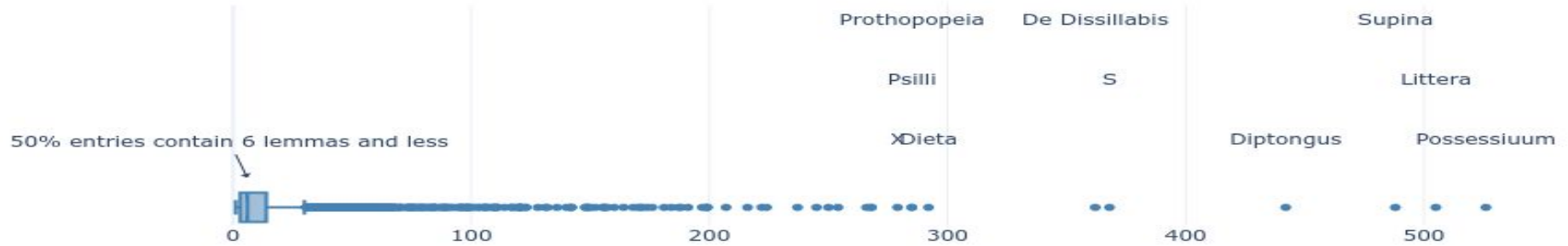
Letter	No of entries	No of pages
Q	92	11
Y	79	6
Z	24	3
X	17	3
K	10	2



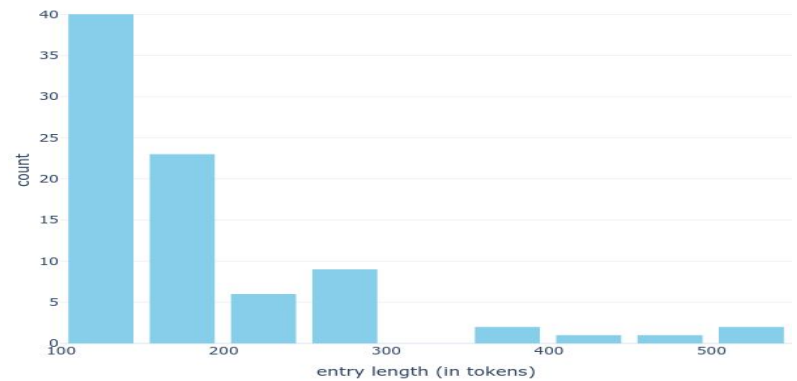
Number of entries per folium  
(2 clusters)

Number of entries (by the first letter)

# Macro-structure: Entry Length



Short entries (up to 10 tokens)



Long entries (more than 100 tokens)



# Micro-structure: Source quotations

No of quotations	No of entries
0	6958
1	589
2	125
3	48
4	18
5	7
6	5
7	3
8	1
9	2
10	1

Number of Quotations

**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 aduerbium, ut idem Uirgilius :

*Satis una superque*

*Uidimus excidia*

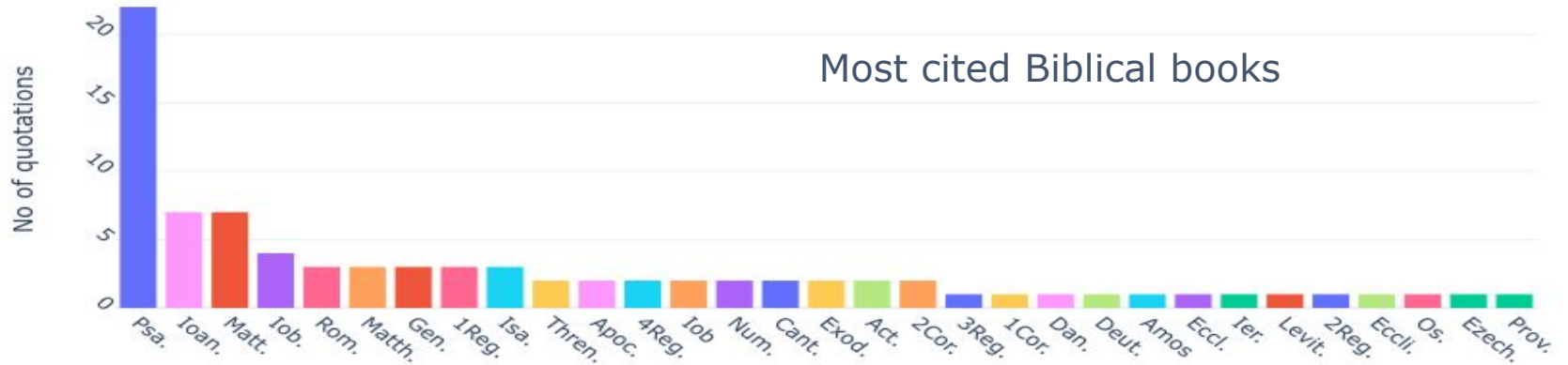
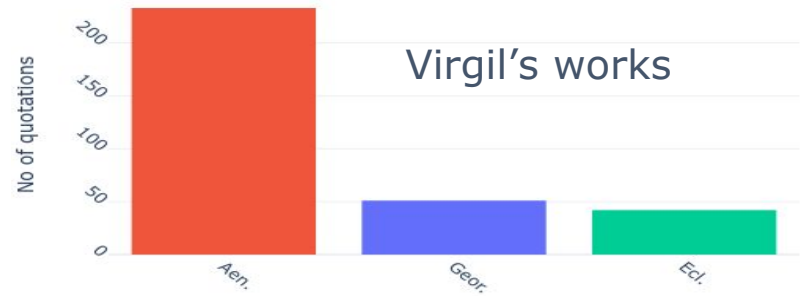
. **Super** et **subter** contraria sunt. Utraque tamen tam accusatiuo quam ablatiui iungitur, licet quibusdam uideantur figurate ablatiui iungi quomodo 'post' et 'ante' et 'circum'. Uirgilius : *Ille admirans uenerabile donum / fatalis uirgo longo post tempore uisum*. Idem : *Omnia que multo ante memor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



aspidocelone



# Micro-structure: Source quotations





# Perspectives: Investigating Text Reuse (2)

**Gauina** uestis super humeros purpuram habens

<b>VocBrux</b>	<b>gauina</b>		uestis	super	humeros		purpuram	habens
<b>Pap</b>	<b>gavina</b>	speci es	vestimenti	super	humeros	habens	purpuram	
<b>LibGloss</b>	<b>gauina</b>	speci es	genus	super	humeros	habens	purpuram	

**Gorgones** fingunt poete fuisse meretrices crinitas serpentibus que aspicientes uertebant in lapides habentes unum oculum quo inuicem utebantur. Fuerunt uero tres sorores: Stenno, Euriale, Medusa, Forcinis filie, unius pulcritudinis quasi unius oculi. Que spectatores suos ita in suo decore stupefaciebant, ut eos quodammodo uertere uiderentur in lapides

Pap: Quas **fingunt poetae meretrices fuisse crinitas serpentibus quae se aspicientes conuertebant in lapides habentes unum oculum quo inuicem utebantur. Fuerunt uero tres sorores id est Stenno, Euriale et Medusa, unius pulchritudinis Forcilis filiae quasi unius oculi, quae ita spectatores stupere faciebant, ut uertere eos putarentur in lapides**

Hug: **Gorgodes** insule sunt Oceani in fine Yspanie, quas incolerunt filie Phorci, unde et Gorgones dicte sunt. Vel gorgon interpretatur terror, unde **tres filie Phorci**, scilicet **Stenio, Euriale et Medusa**, que interpretantur debilitas, lata profunditas et obliuio – dicte sunt Gorgones, idest terrores, quia terror erant uidentibus eas, quia ille tres **unum habebant oculum quo inuicem utebantur**, et **quicumque eas uidebat uertebatur in lapidem**. Quod nichil aliud fuit nisi quia ille tres sorores **unius** erant **pulchritudinis**, **quas cum uiderent adolescentes stupore torpebant**, unde fictum est **quod si quis eas uidisset uertebatur in lapidem**

LibGloss: **Gorgones** — **finguntque poete fuisse meretrices crinitas serpentibus, quae aspicientes conustebant in lapides, habentes+ unum oculum quem inuicem utebantur. Fuerunt autem tres sorores unius pulchritudinis, quasi unius oculi, quae ita expectatores suos stupescere faciebant ut uertere eos putarentur in lapides**

# Conclusions

---

## **Future work: text reuse**

- automatic alignment
- source detection

## **Dictionary-with-code model**

- dictionary edition as a research tool
- improving reproducibility

## **Towards comparative analysis of medieval lexicography**

- data availability
- sticking to standards

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

<https://scriptores.pl/efontes>

<https://glossaria.eu>



**@FontesLatini**



# References

---

- Weijers, O. (1989). Lexicography in the Middle Ages, *Viator*, 20: 139-53.
- Merrilees, B. (1996). The Shape of the Medieval Dictionary Entry. *Digital Studies/le Champ Numérique*, (4).
- Berger S. (1879). De glossariis et compendiis exegeticis quibusdam medii aevi: sive de libris Ansileubi, Papiiae, Hugutionis, Guill. Britonis, de catholicon mammotrecto aliis, Paris
- Weijers, O. (1996). Notice sur le "Vocabularium Bruxellense" (ms. Bruxelles, B. R. II 1049), *ALMA*, 54: 233-238.
- Saenger, Silke-Margarete (1977). *Vocabularium Bruxellense: A Study Of Its Sources With Particular Reference To Paulus's Epitome Of Festus.*
- Cinato F. Le 'Goth Ansileubus', les *Glossae Salomonis* et les glossaires wisigothiques: Mise aupoint sur les attributions et les sources glossographiques du *Liber glossarum*, *Dossiers d'HEL* **8** (2015): 37–56.

# Conclusions



## ➤ **Future work**

- extending research on *Vocabularium's* sources
- automatic alignment and source detection

## ➤ **TEI XML data + ad-hoc scripts for text analysis**

- digital edition as a research tool
- improving reproducibility of research