

# Eulalia – lexicon: work in progress

**Guido Milanese**

**`guido.milanese@unicatt.it`**

UCSC – Brescia / Milano

18th of January 2021

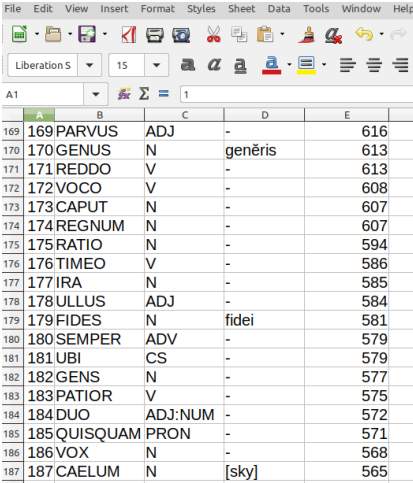
# The Core Vocabulary

---

L. Delatte *et alii* *Dictionnaire fréquentiel et index inverse de la langue latine*, LASLA, Liège 1981 (*corpus*: Catullus, Cicero, Horace, Juvenal, Ovid, Persius, Propertius, Curtius Rufus, Sallustius, Seneca, Tacitus, Livius, Virgil, Vitruvius).

Cauquil, G. & Guillaumin, J. *Vocabulaire de base du latin (alphabétique, fréquentiel, étymologique)* ARELAB, Besançon 1984

# The CSV file



	A	B	C	D	E
169	169	PARVUS	ADJ	-	616
170	170	GENUS	N	genĕris	613
171	171	REDDO	V	-	613
172	172	VOCO	V	-	608
173	173	CAPUT	N	-	607
174	174	REGNUM	N	-	607
175	175	RATIO	N	-	594
176	176	TIMEO	V	-	586
177	177	IRA	N	-	585
178	178	ULLUS	ADJ	-	584
179	179	FIDES	N	fidei	581
180	180	SEMPER	ADV	-	579
181	181	UBI	CS	-	579
182	182	GENS	N	-	577
183	183	PATIOR	V	-	575
184	184	DUO	ADJ:NUM	-	572
185	185	QUISQUAM	PRON	-	571
186	186	VOX	N	-	568
187	187	CAELUM	N	[sky]	565

Figure 1: CSV file

# The CSV file

```
169,PARVUS,ADJ,-,616
170,GENUS,N,genĕris,613
171,REDDO,V,-,613|
172,VOCO,V,-,608
173,CAPUT,N,-,607
174,REGNUM,N,-,607
175,RATIO,N,-,594
176,TIMEO,V,-,586
177,IRA,N,-,585
178,ULLUS,ADJ,-,584
179,FIDES,N,fidei,581
180,SEMPER,ADV,-,579
181,UBI,CS,-,579
182,GENS,N,-,577
183,PATIOR,V,-,575
184,DUO,ADJ:NUM,-,572
185,QUISQUAM,PRON,-,571
186,VOX,N,-,568
187,CAELUM,N,[sky],565
```

Figure 2: CSV file

# The CSV file

---

169	PARVUS	ADJ	-	616
170	GENUS	N	geněris	613
171	REDDO	V	-	613
172	VOCO	V	-	608
173	CAPUT	N	-	607
174	REGNUM	N	-	607
175	RATIO	N	-	594
176	TIMEO	V	-	586
177	IRA	N	-	585

Figure 3: CSV file

# The VP programme



Figure 4: Main menu

# The VP programme: choice

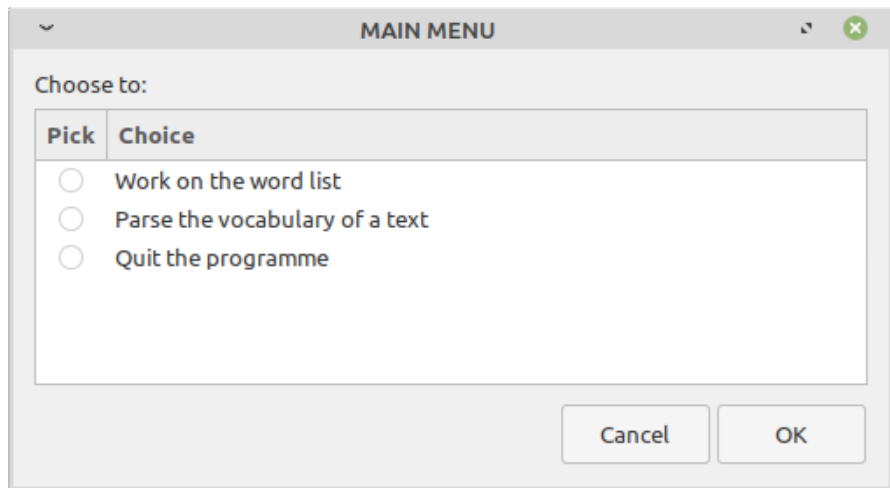


Figure 5: Choice menu

# The VP programme: word list

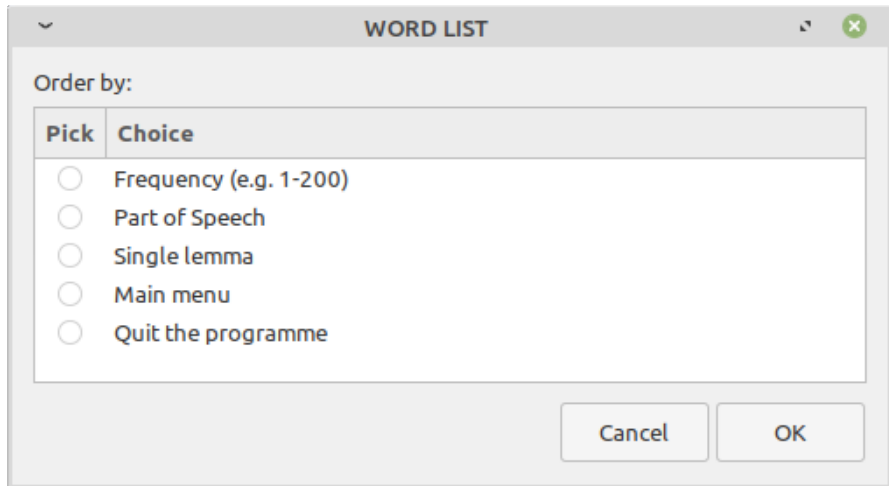


Figure 6: Word list



# The VP programme: word list

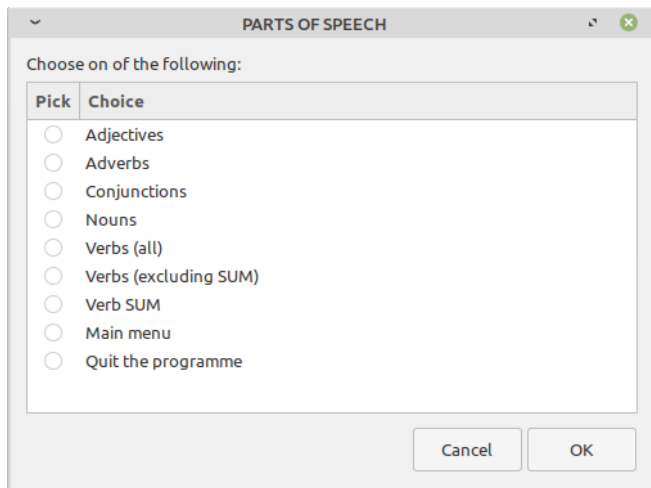


Figure 7: Word list: Parts of Speech

# The VP programme: word list

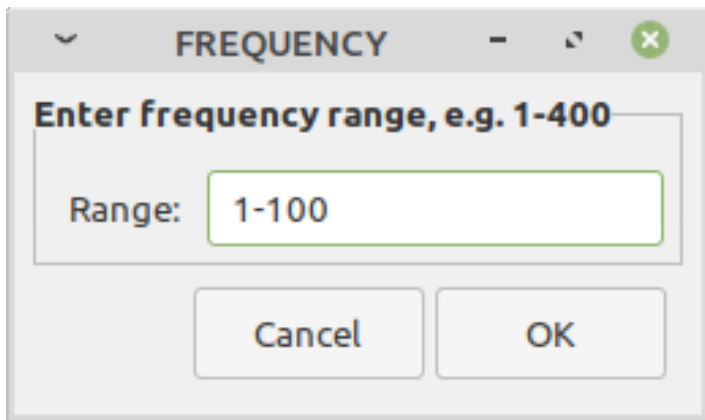
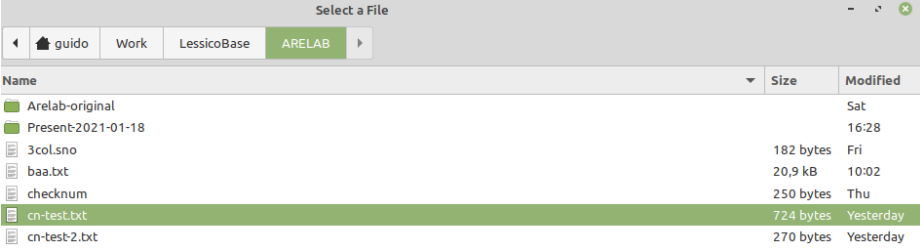


Figure 8: Word list: Range

# The VP programme: analyze a text



Select a File

guido Work LessicoBase ARELAB

Name	Size	Modified
Arelab-original		Sat
Present-2021-01-18		16:28
3col.sno	182 bytes	Fri
baa.txt	20,9 kB	10:02
checknum	250 bytes	Thu
cn-test.txt	724 bytes	Yesterday
cn-test-2.txt	270 bytes	Yesterday

Figure 9: Select a file

# Tree tagger in the background

```
guido@vivarium:~/Work/LessicoBase/ARELAB$ echo "Roma est in Italia" | tree-tagger-latin
reading parameters ...
tagging ...
Roma      N:voc   Roma
est       ESSE:IND   sum
in        PREP    in
Italia    N:abl   Italia
          finished.
```

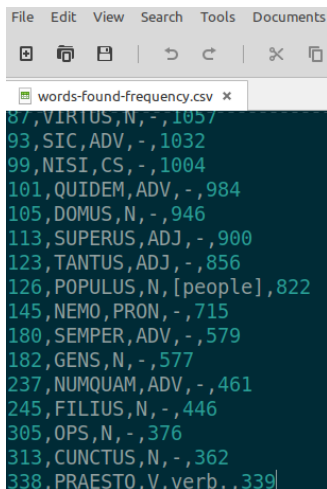
Figure 10: Tree Tagger CLI

# Analyze a text: results



Figure 11: Results message

# Output files (1)



```
File Edit View Search Tools Documents
+ | | | | | | |
words-found-frequency.csv x
87, VIRTUS, N, -, 1057
93, SIC, ADV, -, 1032
99, NISI, CS, -, 1004
101, QUIDEM, ADV, -, 984
105, DOMUS, N, -, 946
113, SUPERUS, ADJ, -, 900
123, TANTUS, ADJ, -, 856
126, POPULUS, N, [people], 822
145, NEMO, PRON, -, 715
180, SEMPER, ADV, -, 579
182, GENS, N, -, 577
237, NUMQUAM, ADV, -, 461
245, FILIUS, N, -, 446
305, OPS, N, -, 376
313, CUNCTUS, N, -, 362
338, PRAESTO, V, verb., 339
```

Figure 12: Output files: lemmas found in Core Lexicon

## Output files (2)

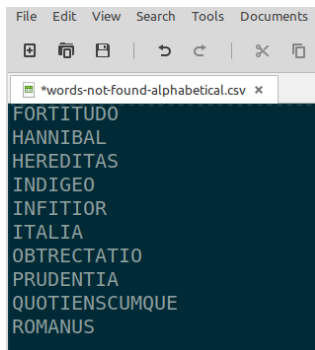


Figure 13: Output files: lemmas not found in Core Lexicon

*Please notice that proper nouns are not listed in the Core Lexicon*

hopefully

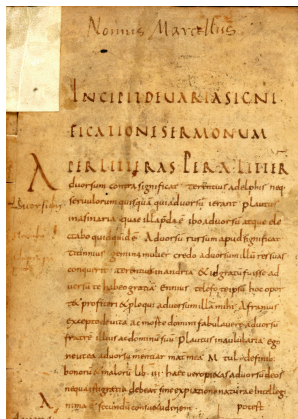


Figure 14: a good example