UNIVERSITY OF TURKU

Latin Across Registers

A Computational Analysis of Situational Language Use Reflected in Grammar

The Norwegian Institute in Rome

September 23, 2025

Hanna-Mari Kupari, Digital language studies

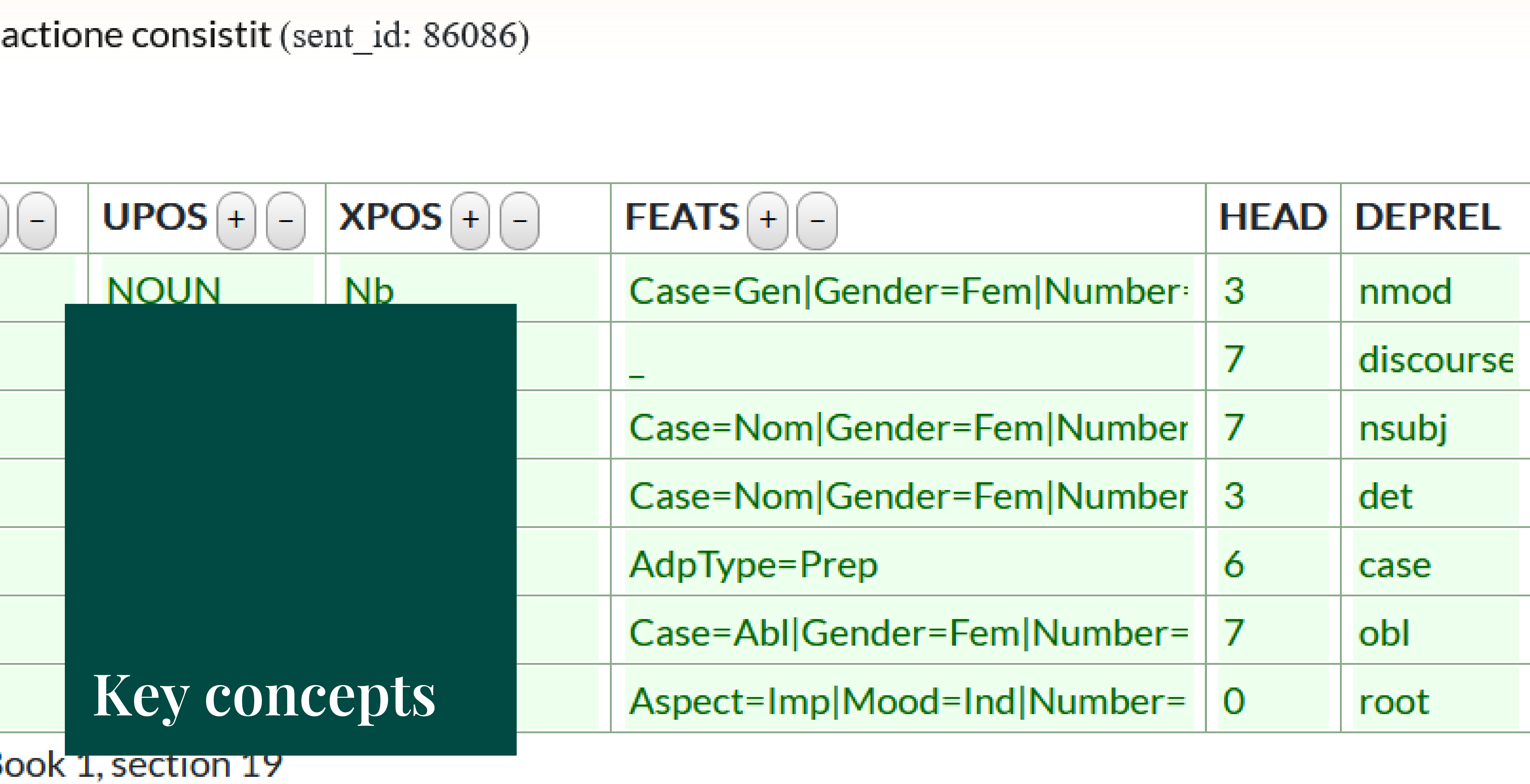
Slides as PDF

TURKUNLP.ORG

Lecture structure

1. Key concepts from Natural Language Processing (NLP) & statistics
2. What is register?
3. Examining Latin registers with computational methods
4. Close reading the results
5. Future studies

Key concepts



actione consistit (sent_id: 86086)

–	UPOS + –	XPOS + –	FEATS + –	HEAD	DEPREL
	NOUN	Nb	Case=Gen|Gender=Fem|Number=	3	nmod
			_	7	discourse
			Case=Nom|Gender=Fem|Number	7	nsubj
			Case=Nom|Gender=Fem|Number	3	det
			AdpType=Prep	6	case
			Case=Abl|Gender=Fem|Number=	7	obl
			Aspect=Imp|Mood=Ind|Number=	0	root

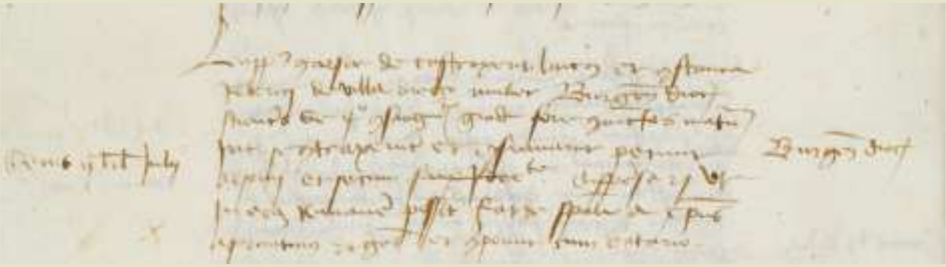
ook 1, section 19

Computational take on language

- Quantitative methods in studying language needs
- Machine readable resources
- Shared guidelines to annotate the data
- Tools to iterate the data
- Ways of calculating results
- Confirming the results with previous studies based on close reading

Ancient documents into a machine-readable language resources

- The original language use situation (Classical times, Middle Ages etc.)
- The original written document
- Centuries of textual transmission
- Modern day printed editions to improve readability
- Digital resource to increase our knowledge of Latin language use



Ancient documents into a machine-readable language resources

- The original language use situation (Classical times, Middle Ages etc.)
- The original written document
- Centuries of textual transmission
- Modern day printed editions to improve readability
- Digital resource to increase our knowledge of Latin language use

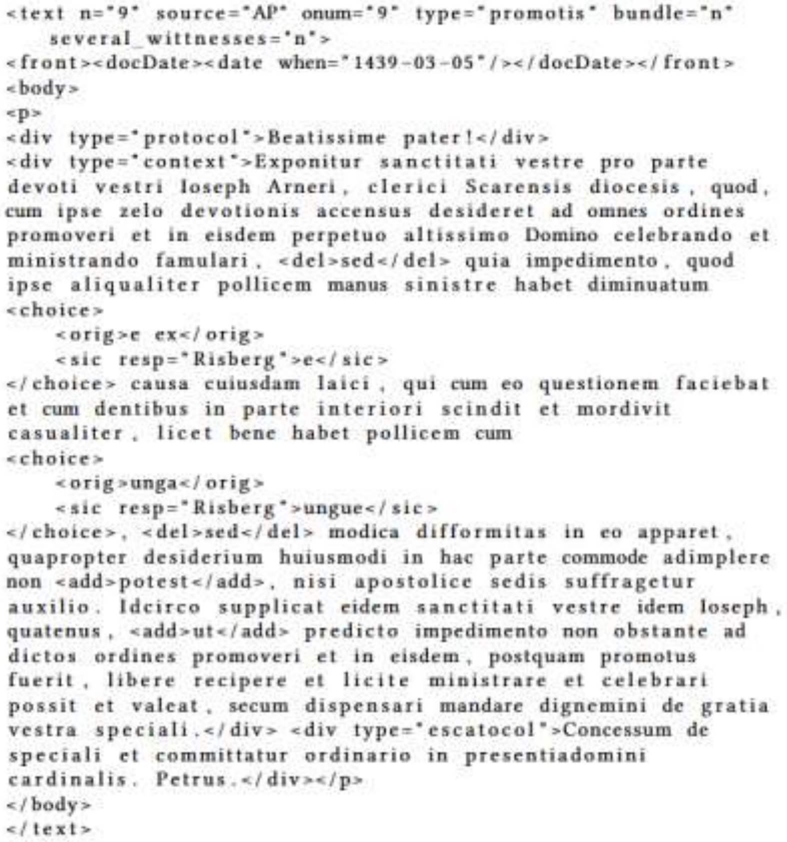
My PhD project

UNIVERSITY OF TURKU

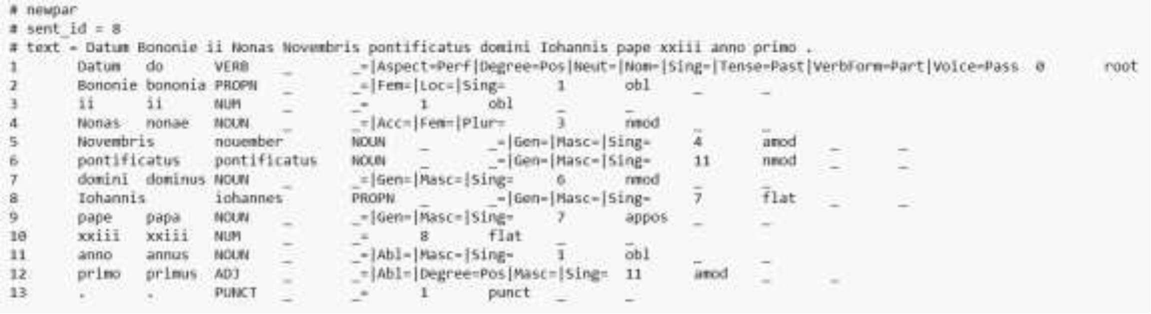
Building machine readable resources

- The text needs to have different data layers marked in a human and machine-readable way
- Common data structure is TEI-XML to tag the different parts of the text
- XML is great for text level analysis but does not work with linguistic annotation
- The Universal Dependencies framework

Two samples: TEI-XML and CoNLL-U

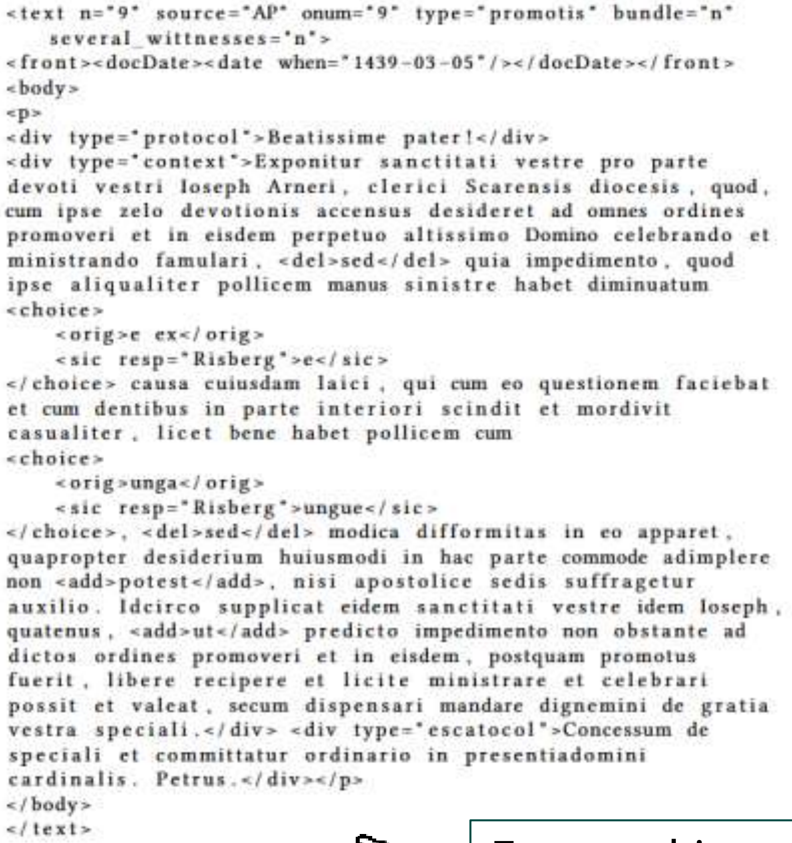


```
<text n="9" source="AP" onum="9" type="promotis" bundle="n"
   several_wittnesses="n">
<front><docDate><date when="1439-03-05"/></docDate></front>
<body>
<p>
<div type="protocol">Beatissime pater!</div>
<div type="context">Exponitur sanctitati vestre pro parte
devoti vestri Ioseph Arneri, clerici Scarensis diocesis, quod,
cum ipse zelo devotionis accensus desideret ad omnes ordines
promoveri et in eisdem perpetuo altissimo Domino celebrando et
ministrando famulari, <del>sed</del> quia impedimento, quod
ipse aliqualiter pollicem manus sinistre habet diminuatum
<choice>
    <orig>e ex</orig>
    <sic resp="Risberg">e</sic>
</choice> causa cuiusdam laici, qui cum eo questionem faciebat
et cum dentibus in parte interiori scindit et mordivit
casualiter, licet bene habet pollicem cum
<choice>
    <orig>unga</orig>
    <sic resp="Risberg">ungue</sic>
</choice>, <del>sed</del> modica difformitas in eo apparet,
quapropter desiderium huiusmodi in hac parte commode adimplere
non <add>potest</add>, nisi apostolice sedis suffragetur
auxilio. Idcirco supplicat eidem sanctitati vestre idem Ioseph,
quatenus, <add>ut</add> predicto impedimento non obstante ad
dictos ordines promoveri et in eisdem, postquam promotus
fuerit, libere recipere et licite ministrare et celebrari
possit et valeat, secum dispensari mandare dignemini de gratia
vestra speciali.</div> <div type="escatocol">Concessum de
speciali et committatur ordinario in presentiadomini
cardinalis. Petrus.</div></p>
</body>
</text>
```



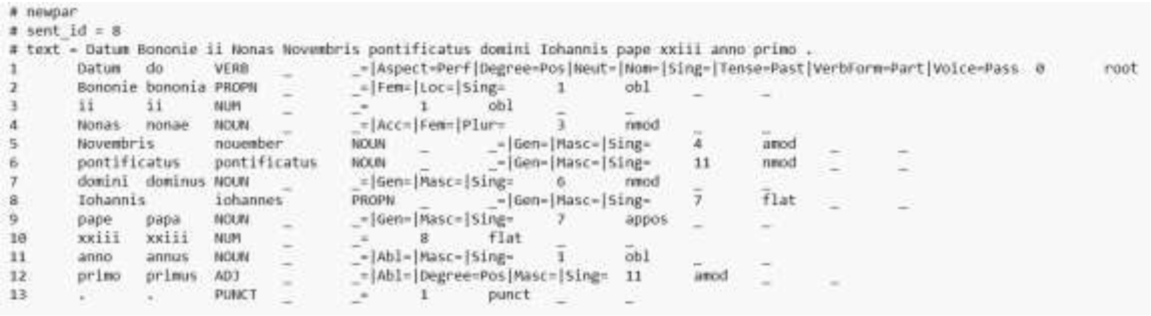
```
# newpar
# sent_id = 8
# text = Datum Bononie ii Nonas Novembris pontificatus domini Iohannis pape xxiii anno primo .
1       Datum   do      VERB    _       _=|Aspect=Perf|Degree=Pos|Neut=|Nom=|Sing=|Tense=Past|VerbForm=Part|Voice=Pass  0       root
2       Bononie bononia PROPN   _       _=|Fem=|Loc=|Sing=      1       obl     _       _
3       ii      ii      NUM     _       _=      1       obl     _       _
4       Nonas   nonae   NOUN    _       _=|Acc=|Fem=|Plur=      3       nmod    _       _
5       Novembris       nouember        NOUN    _       _=|Gen=|Masc=|Sing=     4       amod    _       _
6       pontificatus    pontificatus    NOUN    _       _=|Gen=|Masc=|Sing=     11      nmod    _       _
7       domini  dominus NOUN    _       _=|Gen=|Masc=|Sing=     6       nmod    _       _
8       Iohannis        iohannes        PROPN   _       _=|Gen=|Masc=|Sing=     7       flat    _       _
9       pape    papa    NOUN    _       _=|Gen=|Masc=|Sing=     7       appos   _       _
10      xxiii   xxiii   NUM     _       _=      8       flat    _       _
11      anno    annus   NOUN    _       _=|Abl=|Masc=|Sing=     1       obl     _       _
12      primo   primus  ADJ     _       _=|Abl=|Degree=Pos|Masc=|Sing=  11      amod    _       _
13      .       .       PUNCT   _       _=      1       punct   _       _
```

Two samples: TEI-XML and CoNLL-U



```
<text n="9" source="AP" onum="9" type="promotis" bundle="n"
    several_wittnesses="n">
<front><docDate><date when="1439-03-05"/></docDate></front>
<body>
<p>
<div type="protocol">Beatissime pater!</div>
<div type="context">Exponitur sanctitati vestre pro parte
devoti vestri Ioseph Arneri, clerici Scarensis diocesis, quod,
cum ipse zelo devotionis accensus desideret ad omnes ordines
promoveri et in eisdem perpetuo altissimo Domino celebrando et
ministrando famulari, <del>sed</del> quia impedimento, quod
ipse aliqualiter pollicem manus sinistre habet diminuatum
<choice>
    <orig>e ex</orig>
    <sic resp="Risberg">e</sic>
</choice> causa cuiusdam laici, qui cum eo questionem faciebat
et cum dentibus in parte interiori scindit et mordivit
casualiter, licet bene habet pollicem cum
<choice>
    <orig>unga</orig>
    <sic resp="Risberg">ungue</sic>
</choice>, <del>sed</del> modica difformitas in eo apparet,
quapropter desiderium huiusmodi in hac parte commode adimplere
non <add>potest</add>, nisi apostolice sedis suffragetur
auxilio. Idcirco supplicat eidem sanctitati vestre idem Ioseph,
quatenus, <add>ut</add> predicto impedimento non obstante ad
dictos ordines promoveri et in eisdem, postquam promotus
fuerit, libere recipere et licite ministrare et celebrari
possit et valeat, secum dispensari mandare dignemini de gratia
vestra speciali.</div> <div type="escatocol">Concessum de
speciali et committatur ordinario in presentiadomini
cardinalis. Petrus.</div></p>
</body>
</text>
```

For making notes about different layers of language



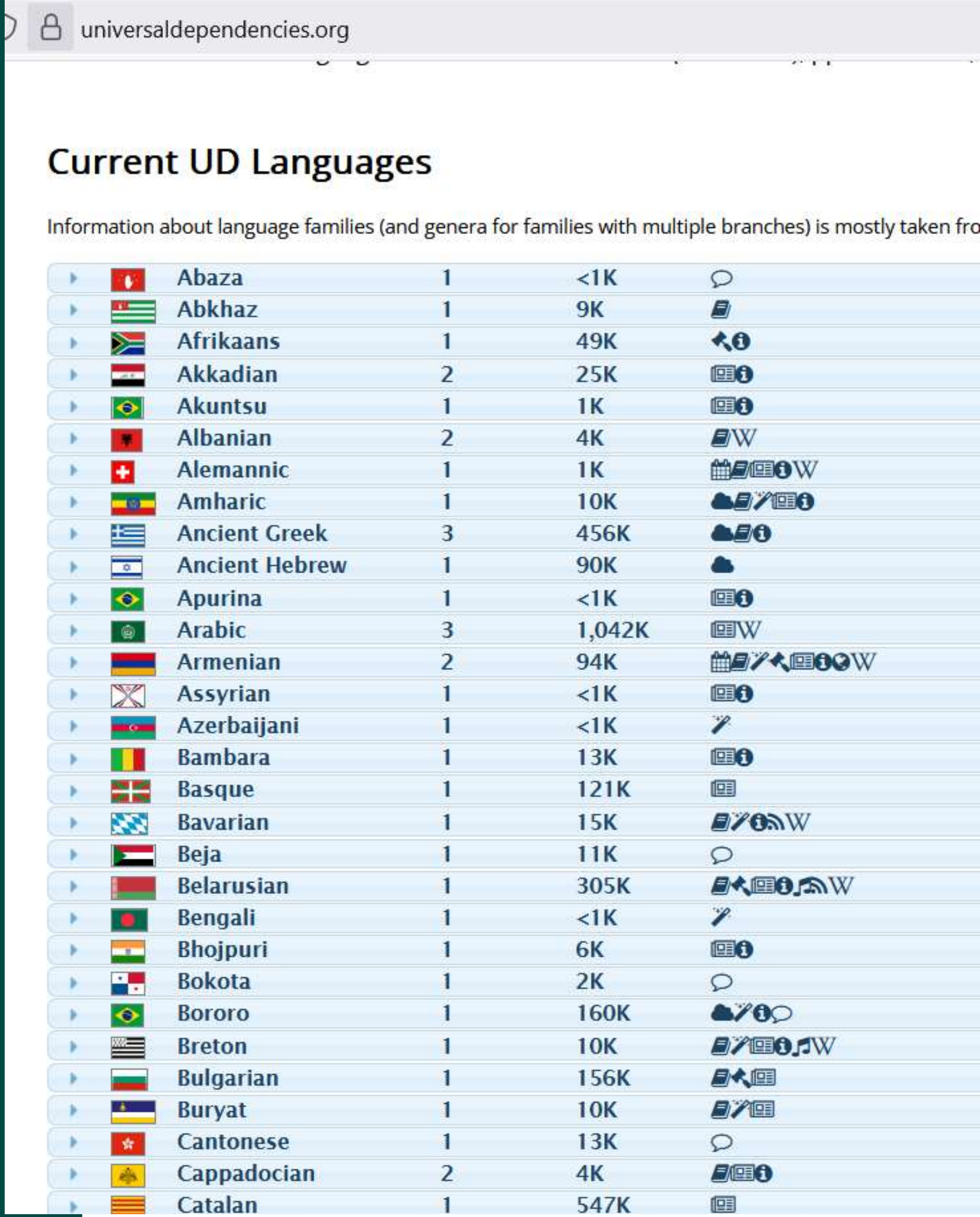
```
# newpar
# sent_id = 8
# text = Datum Bononie ii Nonas Novembris pontificatus domini Iohannis pape xxiii anno primo .
1       Datum   do      VERB    _       _=|Aspect=Perf|Degree=Pos|Neut=|Nom=|Sing=|Tense=Past|VerbForm=Part|Voice=Pass  0       root
2       Bononie bononia PROPN   _       _=|Fem=|Loc=|Sing=      1       obl     _       _
3       ii      ii      NUM     _       _=      1       obl     _       _
4       Nonas   nonae   NOUN    _       _=|Acc=|Fem=|Plur=      3       nmod    _       _
5       Novembris       nouember        NOUN    _       _=|Gen=|Masc=|Sing=     4       amod    _       _
6       pontificatus    pontificatus    NOUN    _       _=|Gen=|Masc=|Sing=     11      nmod    _       _
7       domini  dominus NOUN    _       _=|Gen=|Masc=|Sing=     6       nmod    _       _
8       Iohannis        iohannes        PROPN   _       _=|Gen=|Masc=|Sing=     7       flat    _       _
9       pape    papa    NOUN    _       _=|Gen=|Masc=|Sing=     7       appos   _       _
10      xxiii   xxiii   NUM     _       _=      8       flat    _       _
11      anno    annus   NOUN    _       _=|Abl=|Masc=|Sing=     1       obl     _       _
12      primo   primus  ADJ     _       _=|Abl=|Degree=Pos|Masc=|Sing=  11      amod    _       _
13      .       .       PUNCT   _       _=      1       punct   _       _
```

For marking text sections and notes in a hierarchical way

In-depth Universal Dependecies

- Framework for consistent annotation of grammar in human languages
- Open community effort with 600 contributors
- Over 200 treebanks
- Over 150 languages

https://universaldependencies.org/



universaldependencies.org

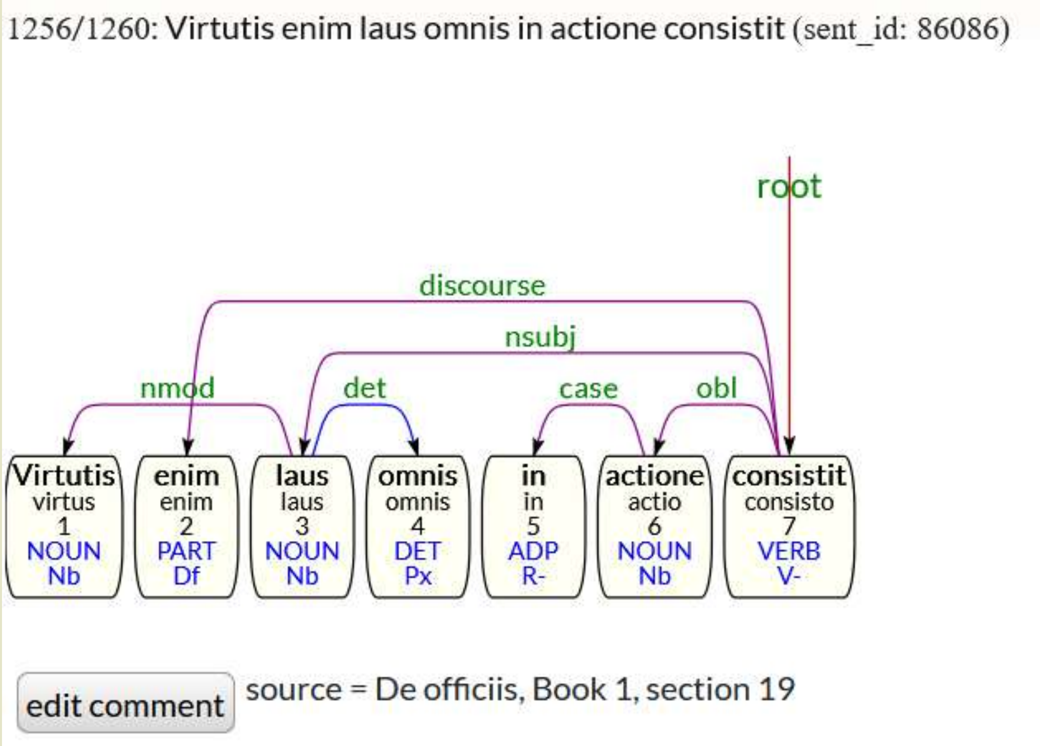
Current UD Languages

Information about language families (and genera for families with multiple branches) is mostly taken fro

▸		Abaza	1	<1K	💬
▸		Abkhaz	1	9K	📖
▸		Afrikaans	1	49K	🔨ℹ
▸		Akkadian	2	25K	📰ℹ
▸		Akuntsu	1	1K	📰ℹ
▸		Albanian	2	4K	📖W
▸		Alemannic	1	1K	📅📖📰ℹW
▸		Amharic	1	10K	☁📖🪄📰ℹ
▸		Ancient Greek	3	456K	☁📖ℹ
▸		Ancient Hebrew	1	90K	☁
▸		Apurina	1	<1K	📰ℹ
▸		Arabic	3	1,042K	📰W
▸		Armenian	2	94K	📅📖🪄🔨📰ℹ🌍W
▸		Assyrian	1	<1K	📰ℹ
▸		Azerbaijani	1	<1K	🪄
▸		Bambara	1	13K	📰ℹ
▸		Basque	1	121K	📰
▸		Bavarian	1	15K	📖🪄ℹ📶W
▸		Beja	1	11K	💬
▸		Belarusian	1	305K	📖🔨📰ℹ♫📶W
▸		Bengali	1	<1K	🪄
▸		Bhojpuri	1	6K	📰ℹ
▸		Bokota	1	2K	💬
▸		Bororo	1	160K	☁🪄ℹ💬
▸		Breton	1	10K	📖🪄📰ℹ♫W
▸		Bulgarian	1	156K	📖🔨📰
▸		Buryat	1	10K	📖🪄📰
▸		Cantonese	1	13K	💬
▸		Cappadocian	2	4K	📖📰ℹ
▸		Catalan	1	547K	📰

Dependency grammar

- Describes sentences as networks of words linked by head–dependent relations.
- E.g. “words connected like a family tree”
- "For virtue, indeed, all praise lies in action.“

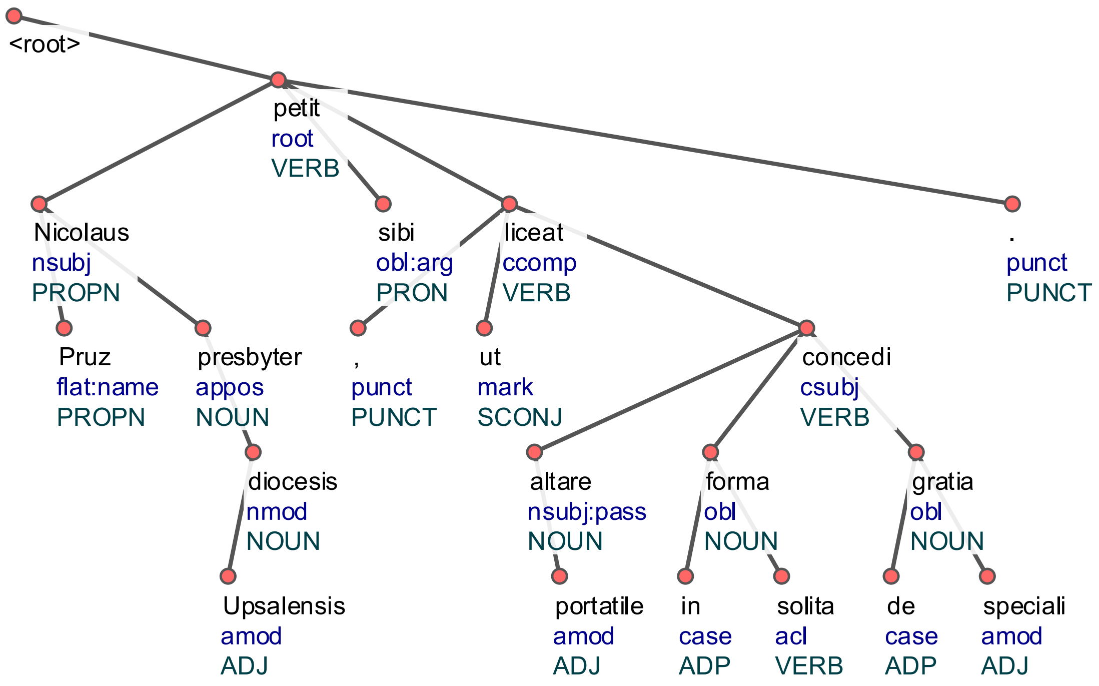


1256/1260: Virtutis enim laus omnis in actione consistit (sent_id: 86086)

source = De officiis, Book 1, section 19

https://github.com/Orange-OpenSource/conllueditor

Tree representation

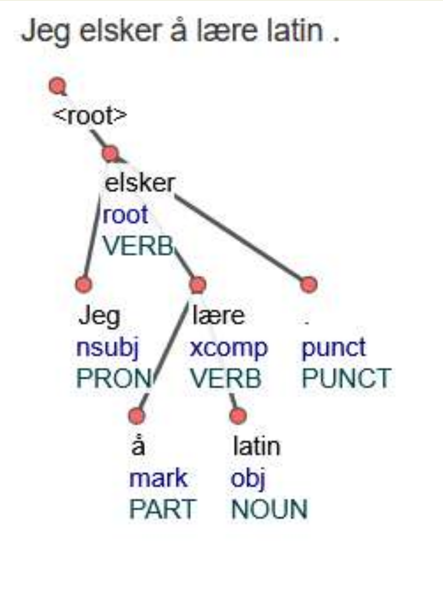


The *head* is usually the main word of a phrase (e.g. verb) and dependents give it detail

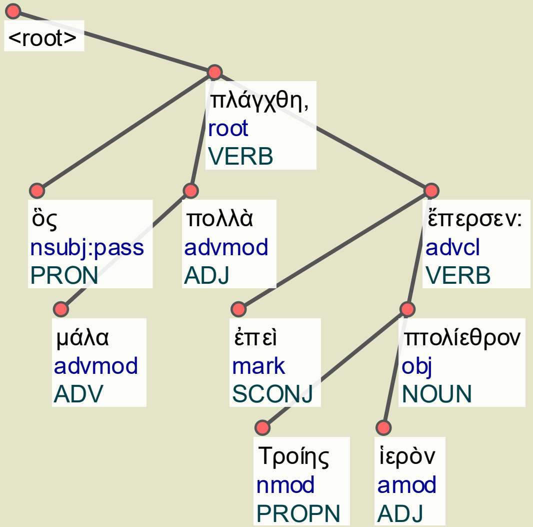
This structure is especially useful in NLP because it captures meaning and relationships between words in a way that computers can process (i.e. numbers)

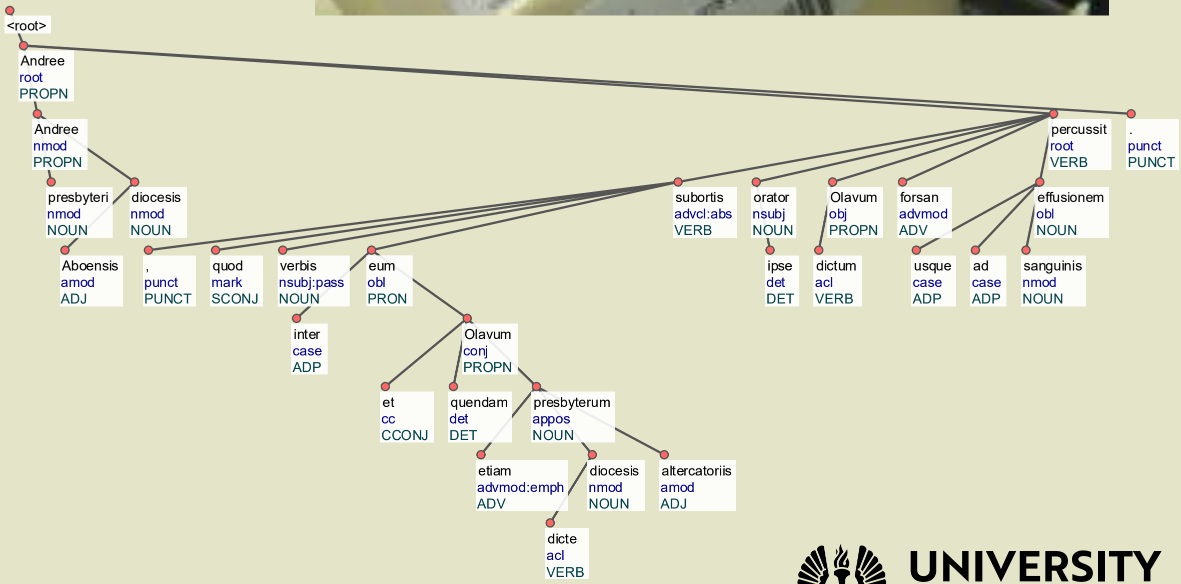
Treebanks

- With analogy from Seedbank



Picture: Wikipedia

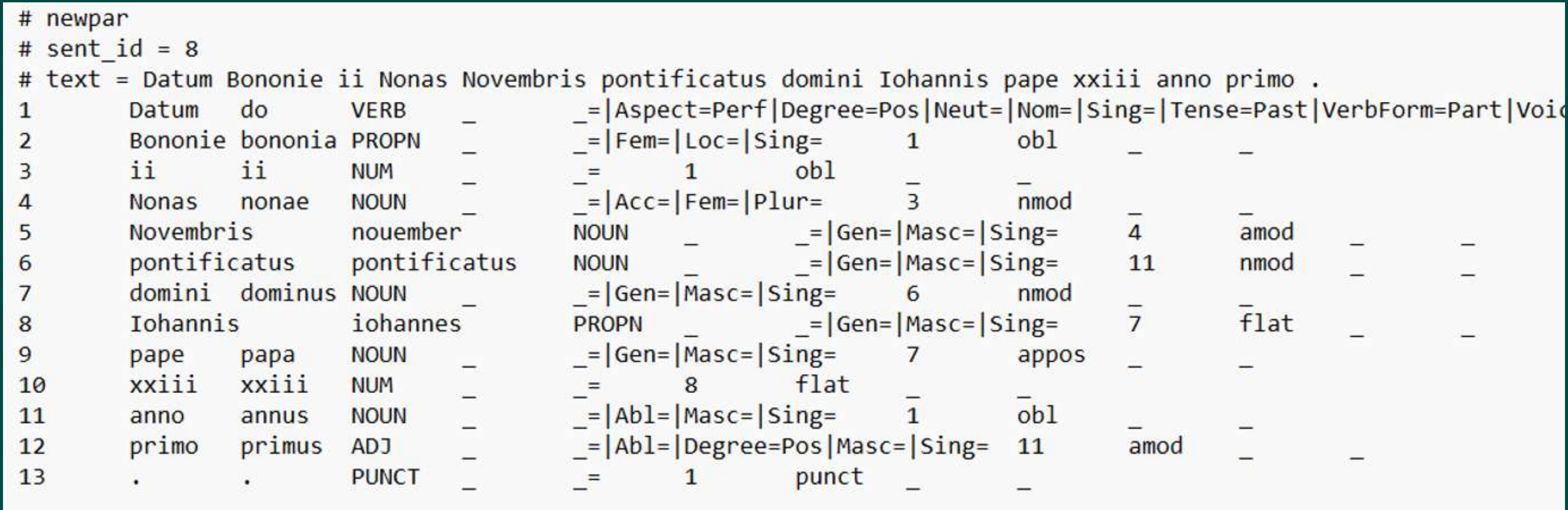




https://lindat.mff.cuni.cz/services/udpipe/

UNIVERSITY OF TURKU

CoNLL-U tables



```
# newpar
# sent_id = 8
# text = Datum Bononie ii Nonas Novembris pontificatus domini Iohannis pape xxiii anno primo .
1       Datum   do      VERB    _       _=|Aspect=Perf|Degree=Pos|Neut=|Nom=|Sing=|Tense=Past|VerbForm=Part|Voi
2       Bononie bononia PROPN   _       _=|Fem=|Loc=|Sing=      1       obl     _       _
3       ii      ii      NUM     _       _=      1       obl     _       _
4       Nonas   nonae   NOUN    _       _=|Acc=|Fem=|Plur=      3       nmod    _       _
5       Novembris       nouember        NOUN    _       _=|Gen=|Masc=|Sing=     4       amod    _       _
6       pontificatus    pontificatus    NOUN    _       _=|Gen=|Masc=|Sing=     11      nmod    _       _
7       domini  dominus NOUN    _       _=|Gen=|Masc=|Sing=     6       nmod    _       _
8       Iohannis        iohannes        PROPN   _       _=|Gen=|Masc=|Sing=     7       flat    _       _
9       pape    papa    NOUN    _       _=|Gen=|Masc=|Sing=     7       appos   _       _
10      xxiii   xxiii   NUM     _       _=      8       flat    _       _
11      anno    annus   NOUN    _       _=|Abl=|Masc=|Sing=     1       obl     _       _
12      primo   primus  ADJ     _       _=|Abl=|Degree=Pos|Masc=|Sing=  11      amod    _       _
13      .       .       PUNCT   _       _=      1       punct   _       _
```

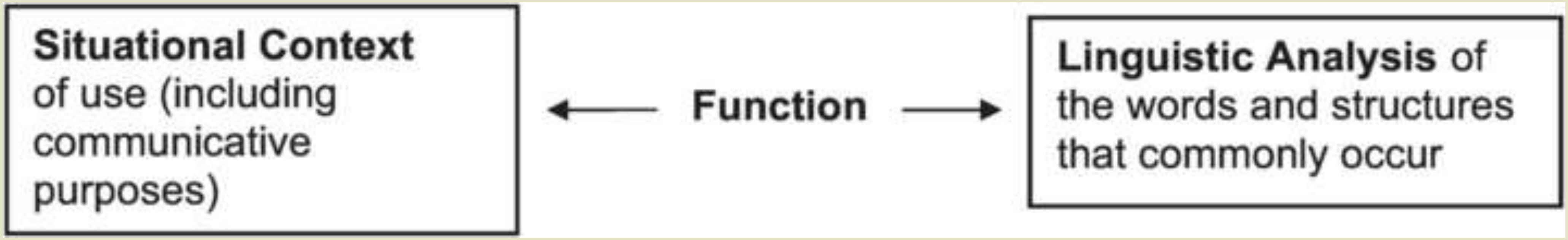
What is a register?

Context aware language use – the why?

- Grouping texts by different context traced back to Aristotle dividing rhetoric deliberative (συμβουλευτικόν), forensic (δικανικόν) and epideictic (ἐπιδεικτικόν) (*Rh* 1.3)
- Even small children intuitively mimic the style of news reading or storytelling in their role play
- But what feels intuitive is hard to define precisely—“I know it when I see it” is not enough for computational studies of language use

Terminology: register

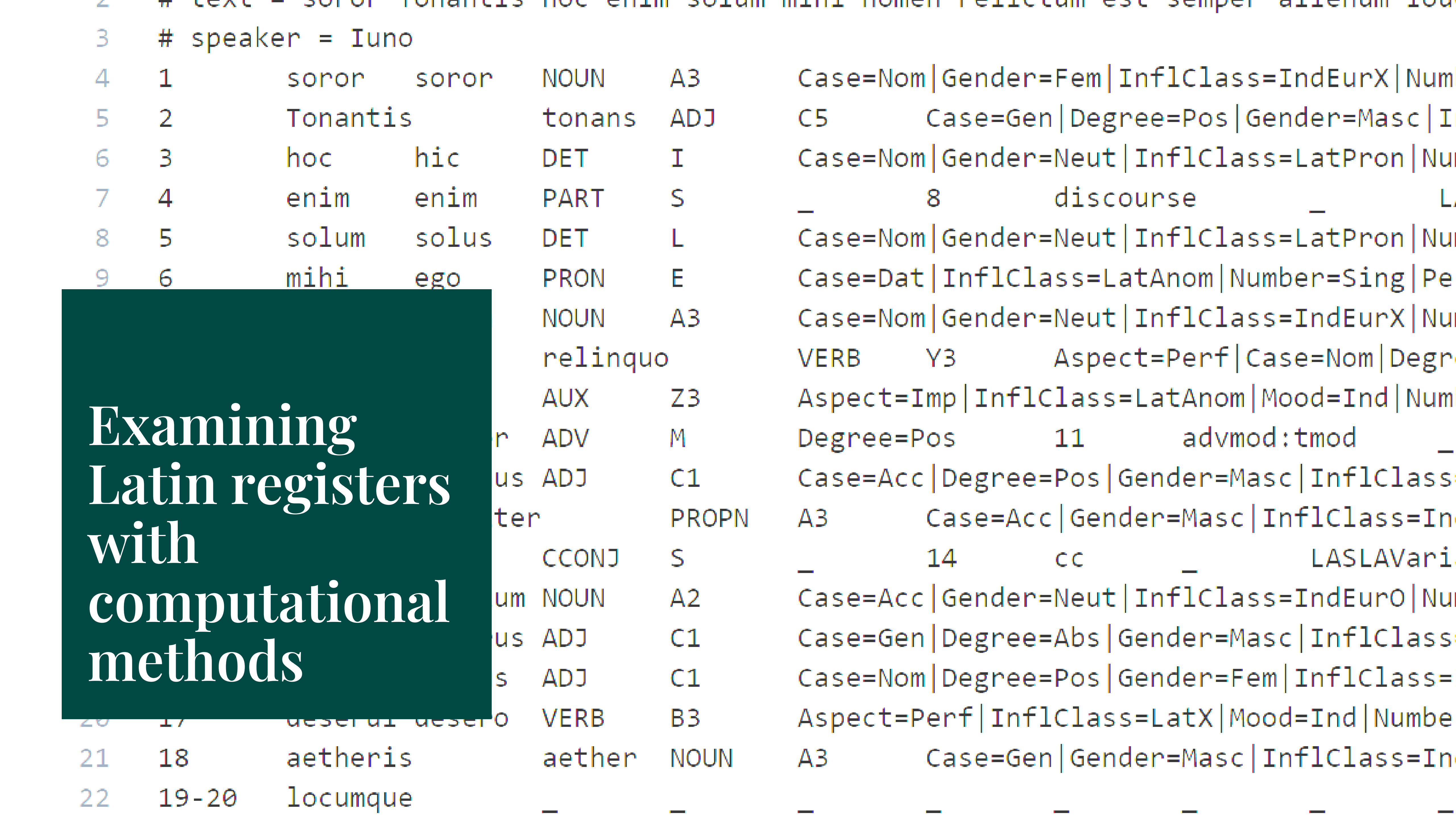
- Terminology used here adapted from Douglas Biber
- Register, as defined by Biber and Conrad (2019), is a variety of text associated with a particular situation of use:



Situational Context of use (including communicative purposes)

← **Function** →

Linguistic Analysis of the words and structures that commonly occur



```
# text = soror Tonantis hoc enim solum mihi nomen relictum est semper alienum Iou
# speaker = Iuno
1       soror   soror   NOUN    A3      Case=Nom|Gender=Fem|InflClass=IndEurX|Num
2       Tonantis        tonans  ADJ     C5      Case=Gen|Degree=Pos|Gender=Masc|I
3       hoc     hic     DET     I       Case=Nom|Gender=Neut|InflClass=LatPron|Nu
4       enim    enim    PART    S       _       8       discourse       _       L
5       solum   solus   DET     L       Case=Nom|Gender=Neut|InflClass=LatPron|Nu
6       mihi    ego     PRON    E       Case=Dat|InflClass=LatAnom|Number=Sing|Pe
                        NOUN    A3      Case=Nom|Gender=Neut|InflClass=IndEurX|Nu
                        relinquo        VERB    Y3      Aspect=Perf|Case=Nom|Degr
                        AUX     Z3      Aspect=Imp|InflClass=LatAnom|Mood=Ind|Num
                r       ADV     M       Degree=Pos      11      advmod:tmod     _
                us      ADJ     C1      Case=Acc|Degree=Pos|Gender=Masc|InflClass
                ter             PROPN   A3      Case=Acc|Gender=Masc|InflClass=In
                        CCONJ   S       _       14      cc      _       LASLAVari
                um      NOUN    A2      Case=Acc|Gender=Neut|InflClass=IndEurO|Nu
                us      ADJ     C1      Case=Gen|Degree=Abs|Gender=Masc|InflClass
                s       ADJ     C1      Case=Nom|Degree=Pos|Gender=Fem|InflClass=
17      deserui desero  VERB    B3      Aspect=Perf|InflClass=LatX|Mood=Ind|Numbe
18      aetheris        aether  NOUN    A3      Case=Gen|Gender=Masc|InflClass=In
19-20   locumque        _       _       _       _       _       _       _
```

Examining Latin registers with computational methods

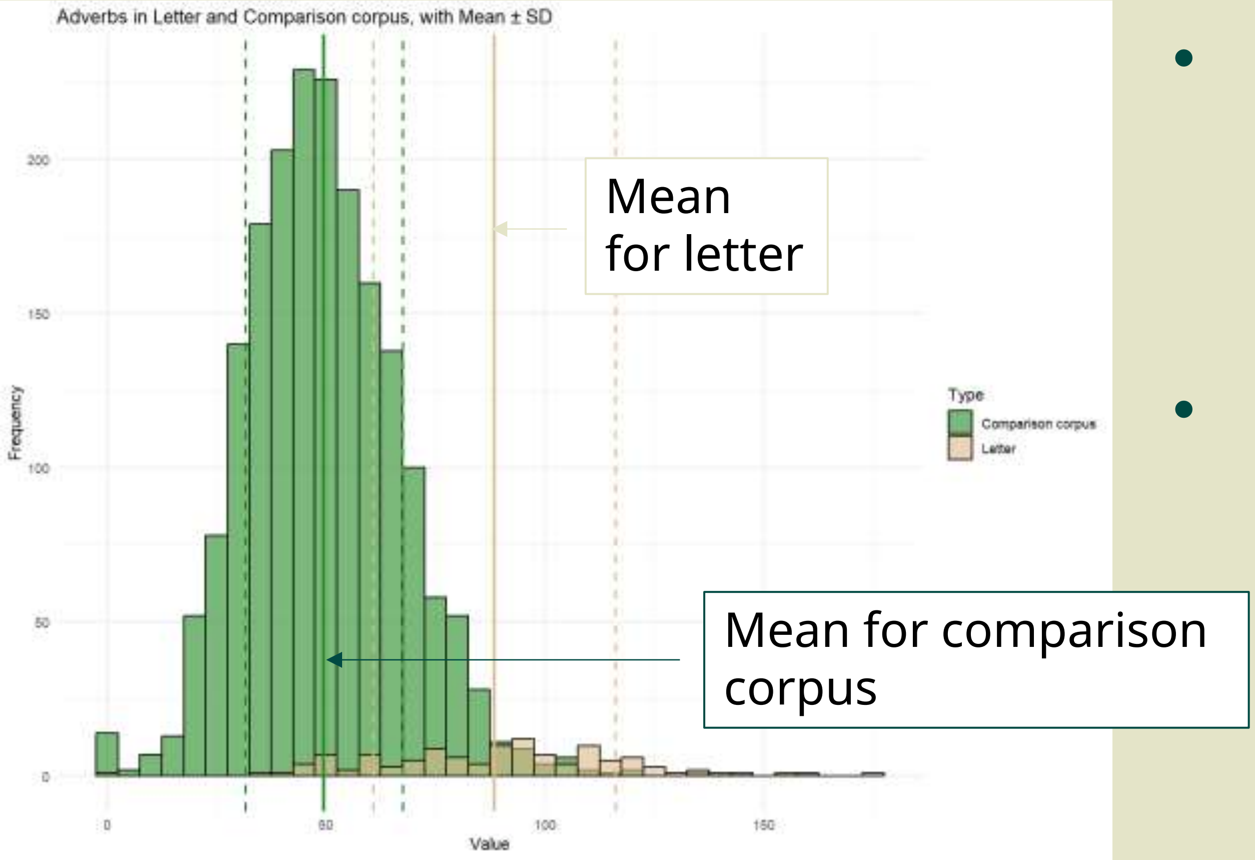
Register in Latin

- How does the statistical and computational register perspective confirm and add to previous research?
  - Situational contexts of Classical texts well (history)
  - We understand from systemic-functional linguistics the roles that language use plays (general linguistics)
  - Close reading of “school texts” the style and overall tone of writing (traditional philology)
- How all these insights relate to larger collections of texts when compared across a vast amount of material

Comparing two collections of text – how?

- Using the Key Feature Analysis, KFA, method to find distinct features
- Adapted from Egbert & Biber, 2023
  - “Key feature analysis: a simple, yet powerful method for comparing text varieties”
- https://youtu.be/tTgouKMz-eI?feature=shared
- The KFA method identifies features that are statistically over- or underrepresented in two datasets

Calculating effect size with KFA



Mean for letter

Mean for comparison corpus

- KFA applies Cohen’s *d* to detect feature differences between registers
- Cohen’s *d* measures effect size using a comparison of standard deviation and mean, formula: $Cohen's\ d = \frac{M_2 - M_1}{SD_{pooled}}$

Comparing registers step by step

1. Selecting the suitable texts for this study
2. Annotating the chosen texts
3. Exploring the effect of the division of texts (4-16 classes)
4. Chunking the texts for optimal divisions
5. Normalizing the frequencies
6. Calculating the Cohen's d effect score
7. Visualizing the results -> cross validation with close reading

The six UD treebanks – on what?

- A wide selection of texts
- A good general overview https://universaldependencies.org/la/index.html
- Working with gold standard data only
    - Issues in automatic parsing of Latin syntax
    - Validation on well known dataset – more robust results for new method
- The corpus is by no means balanced and well representative
    - Possibilities for future studies

Data annotation

- Striking a balance between class size and uniqueness of each text
- E.g. Fitting *Satyricon* into a class of
  - verse,
  - prose dialogue or
  - satire?
- The aim of register studies is to study larger collects of text – focus on single authors or works in *stylistics*

Works

Christian philosophy

- Aquinas, Forma
- Aquinas, Summa Contra Gentiles

History

- Augustus, Res Gestae
- Caesar, Gallic War
- Sallust, Bellum Catilinae
- Suetonius, Life Of Augustus
- Tacitus, Historiae
- Tacitus, Germania

Treatise

- Cicero, De Officiis
- Dante, De Vulgari Eloquentia
- Dante, Monarchia
- Dante, Questio De Aqua Et Terra
- Palladius, Opus Agriculturae

Speech

- Cicero, In Catilinam

Letter

- Cicero, Letters To Atticus 1-7
- Dante, Letters

Verse

- Dante, Eclogues
- Ovid, Metamorphoses
- Phaedrus, Fabulae
- Propertius, Elegies
- Vergil, Aeneid

Biblical

- Jerome, Vulgata

Charter

- Late Latin Charter Treebank

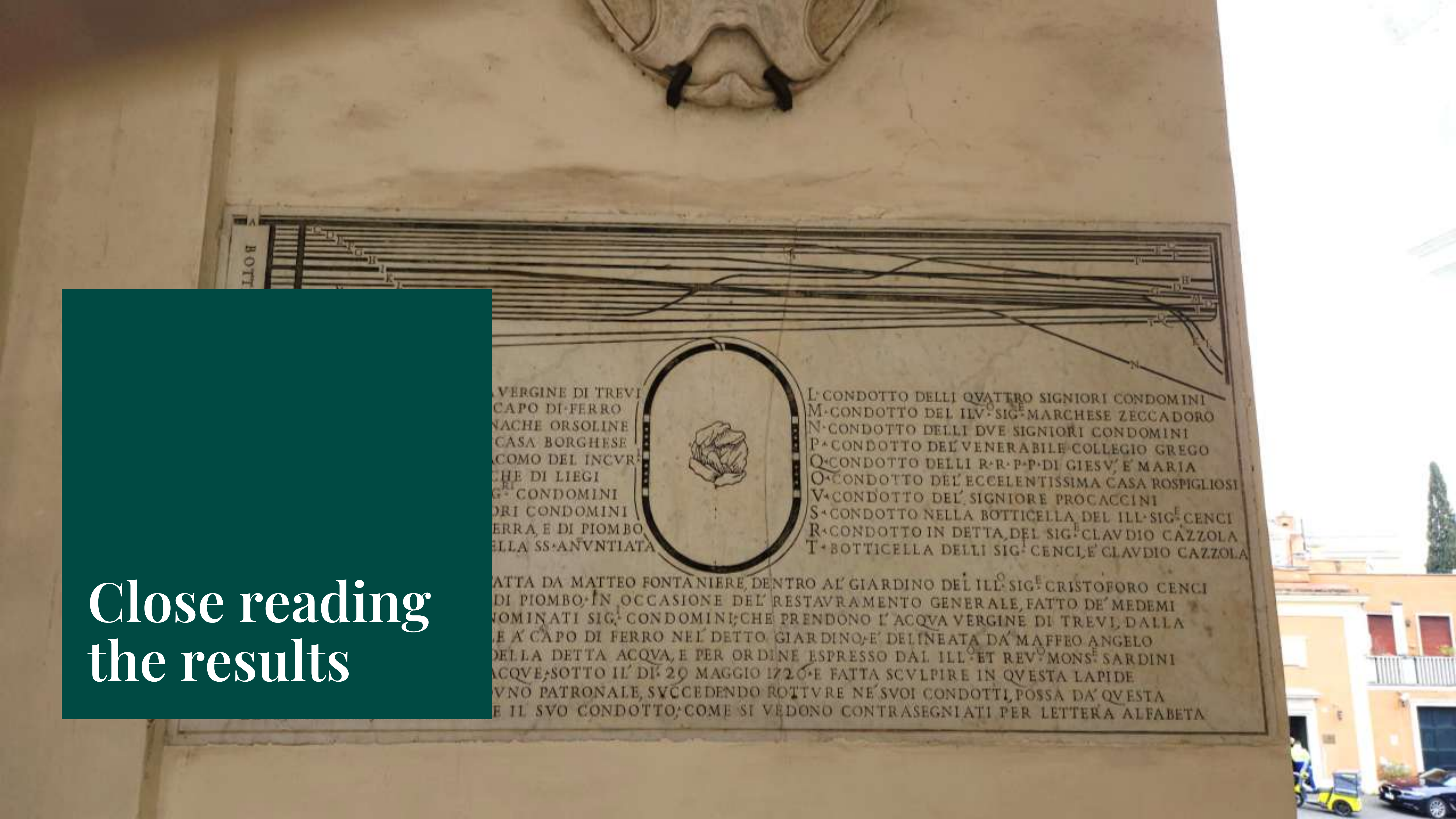
Satire

- Petronius, Satyricon

Play

- Seneca, Hercules Furens
- Seneca, Agamemnon
- Seneca, Oedipus

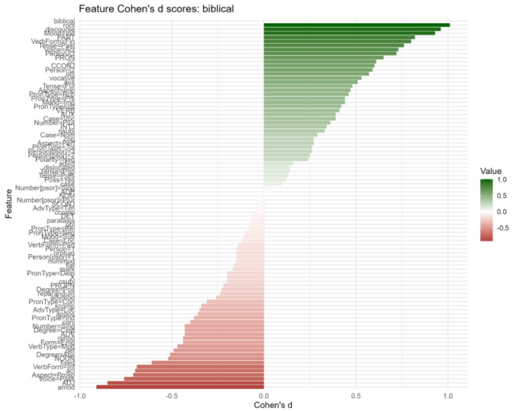
Close reading the results

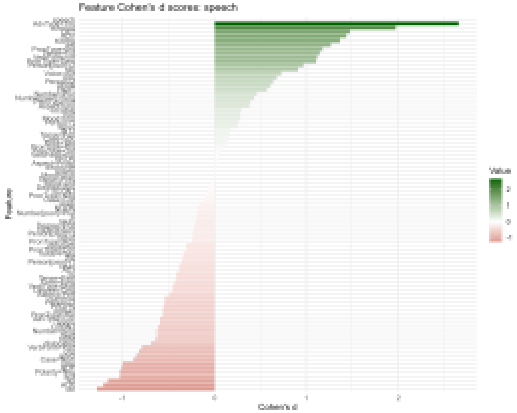


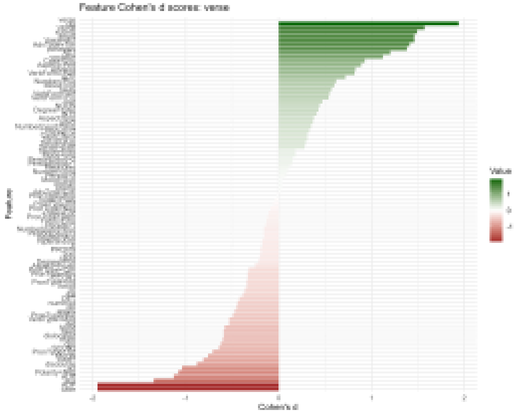
VERGINE DI TREVI
CAPO DI·FERRO
NACHE ORSOLINE
CASA BORGHESE
ACOMO DEL INCVR^{E}
CHE DI LIEGI
G^{RI} CONDOMINI
ORI CONDOMINI
ERRA, E DI PIOMBO
ELLA SS·ANVNTIATA

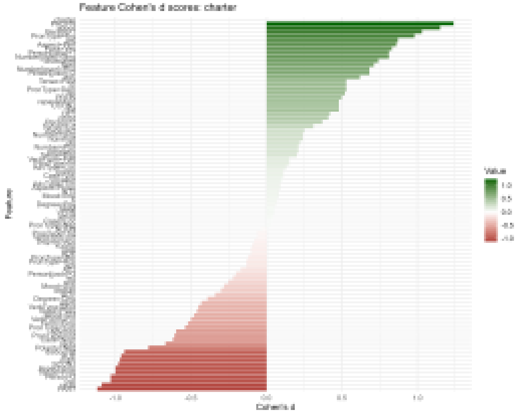
L·CONDOTTO DELLI QVATTRO SIGNIORI CONDOMINI
M·CONDOTTO DEL ILV^{O}·SIG^{E}·MARCHESE ZECCADORO
N·CONDOTTO DELLI DVE SIGNIORI CONDOMINI
P·CONDOTTO DEL VENERABILE·COLLEGIO GREGO
Q·CONDOTTO DELLI R·R·P·P·DI GIESV, E MARIA
O·CONDOTTO DEL ECCELENTISSIMA CASA ROSPIGLIOSI
V·CONDOTTO DEL SIGNIORE PROCACCINI
S·CONDOTTO NELLA BOTTICELLA DEL ILL·SIG^{E}·CENCI
R·CONDOTTO IN DETTA, DEL SIG^{E}·CLAVDIO CAZZOLA
T·BOTTICELLA DELLI SIG^{I}·CENCI, E CLAVDIO CAZZOLA

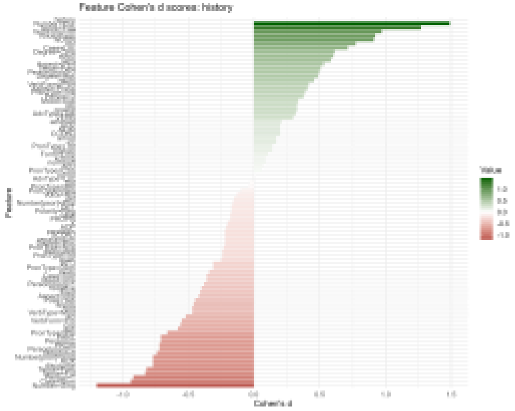
ATTA DA MATTEO FONTANIERE, DENTRO AL GIARDINO DEL ILL^{O}·SIG^{E}·CRISTOFORO CENCI
DI PIOMBO·IN OCCASIONE DEL RESTAVRAMENTO GENERALE, FATTO DE MEDEMI
NOMINATI SIG^{I}·CONDOMINI·CHE PRENDONO L ACQVA VERGINE DI TREVI, DALLA
E A CAPO DI FERRO NEL DETTO GIARDINO·E DELINEATA DA MAFFEO ANGELO
DELLA DETTA ACQVA, E PER ORDINE ESPRESSO DAL ILL^{O}·ET REV^{O}·MONS^{E}·SARDINI
ACQVE·SOTTO IL DI·29 MAGGIO 1726·E FATTA SCVLPIRE IN QVESTA LAPIDE
VNO PATRONALE, SVCCEDENDO ROTTVRE NE SVOI CONDOTTI, POSSA DA QVESTA
E IL SVO CONDOTTO·COME SI VEDONO CONTRASEGNIATI PER LETTERA ALFABETA

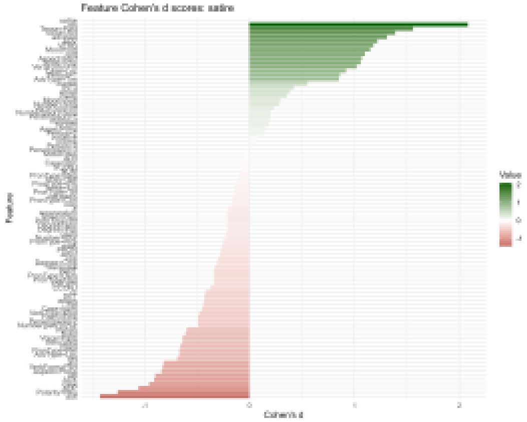


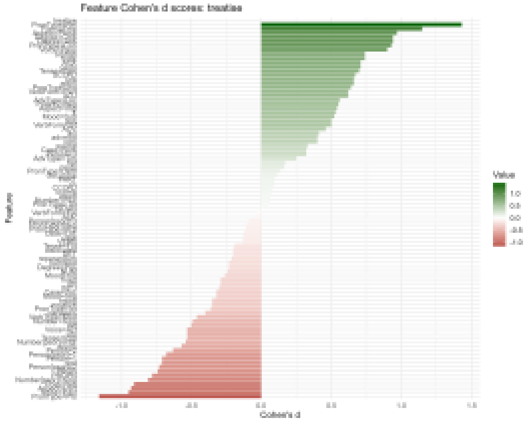


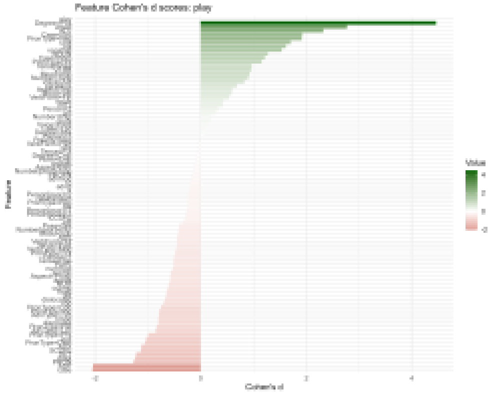


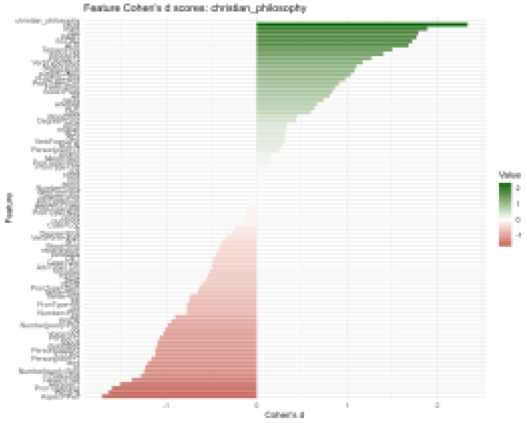


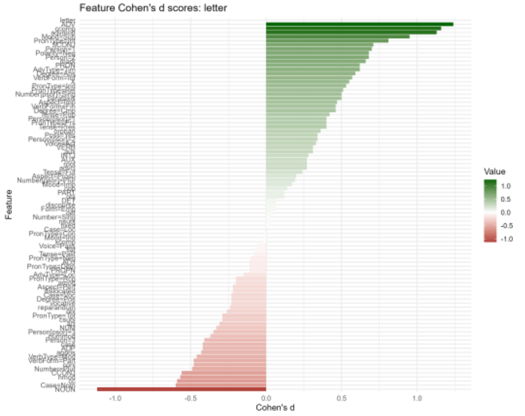






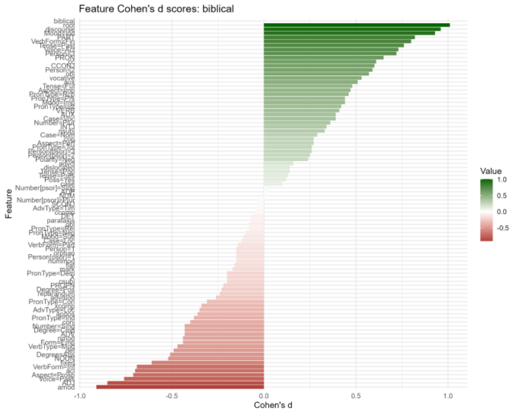


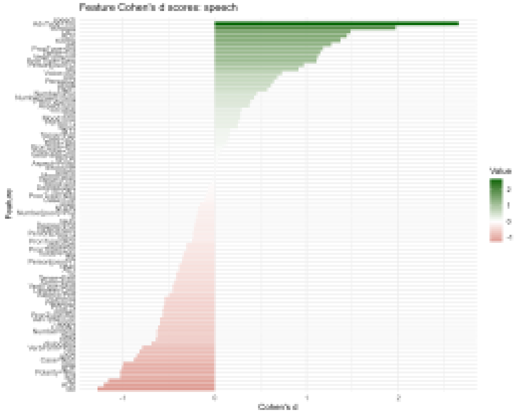


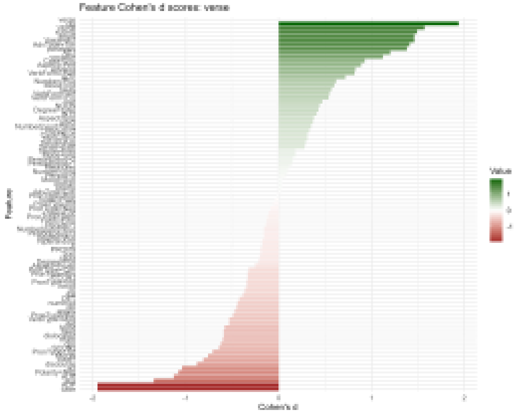


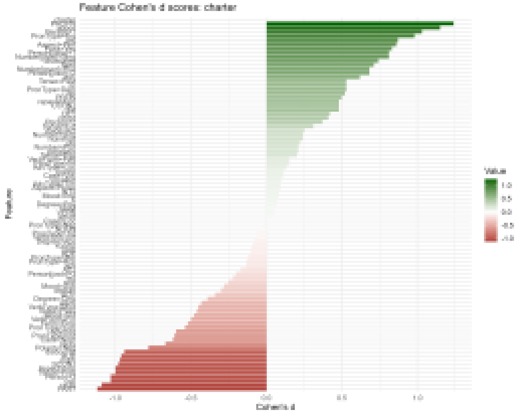
But first distant reading!

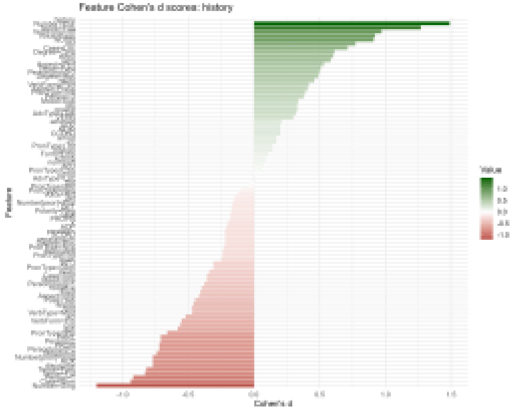
UNIVERSITY OF TURKU

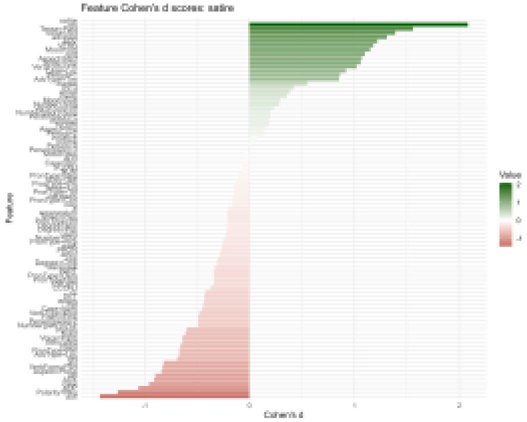

Feature Cohen's d scores: biblical
Feature
Cohen's d
-1.0
-0.5
0.0
0.5
1.0
Value
1.0
0.5
0.0
-0.5
biblical
VerbForm
PRON
vocat
Tense
Person
PronType
Case
Number[psor]
AdvType
parata
Person
PronType=Dem
Degree
VerbForm

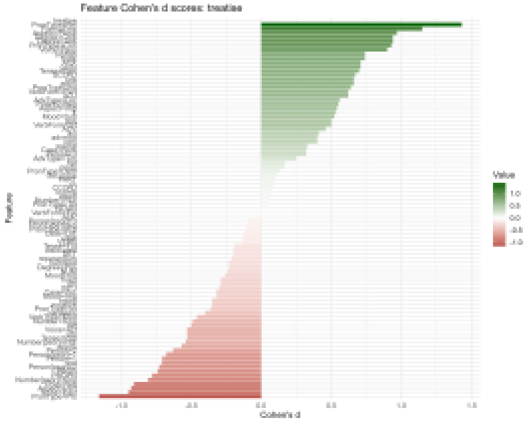

Feature Cohen's d scores: speech
Feature
Cohen's d
-1
0
1
2
Value

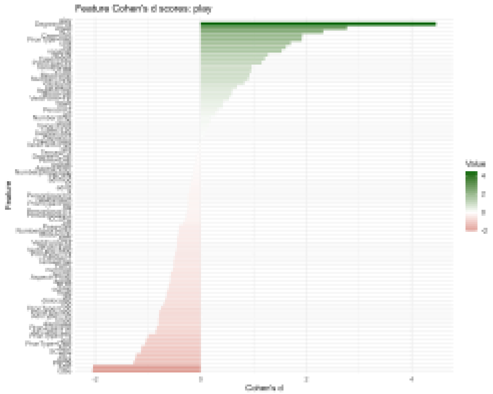

Feature Cohen's d scores: verse
Feature
Cohen's d
-2
-1
0
1
2
Value

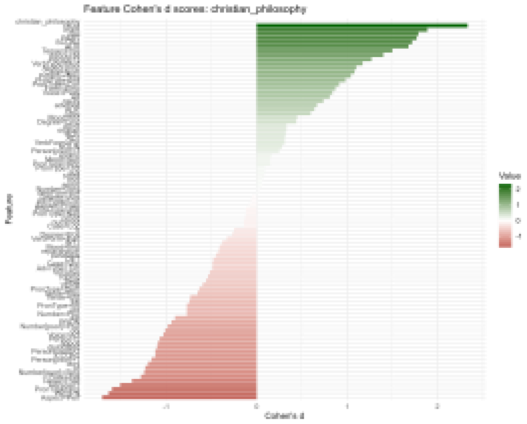

Feature Cohen's d scores: charter
Feature
Cohen's d
-1.0
-0.5
0.0
0.5
1.0
Value
1.0
0.5
0.0
-0.5
-1.0

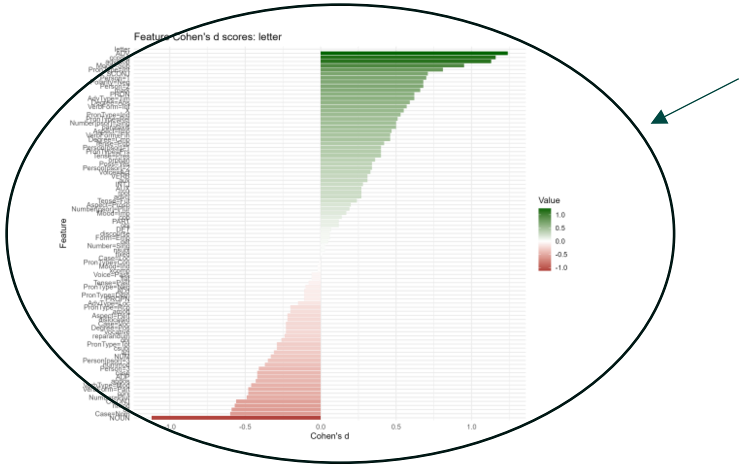

Feature Cohen's d scores: history
Feature
Cohen's d
-1.0
-0.5
0.0
0.5
1.0
1.5
Value
1.0
0.5
0.0
-0.5
-1.0


Feature Cohen's d scores: satire
Feature
Cohen's d
-1
0
1
2
Value


Feature Cohen's d scores: treatise
Feature
Cohen's d
-1.0
-0.5
0.0
0.5
1.0
1.5
Value
1.0
0.5
0.0
-0.5
-1.0

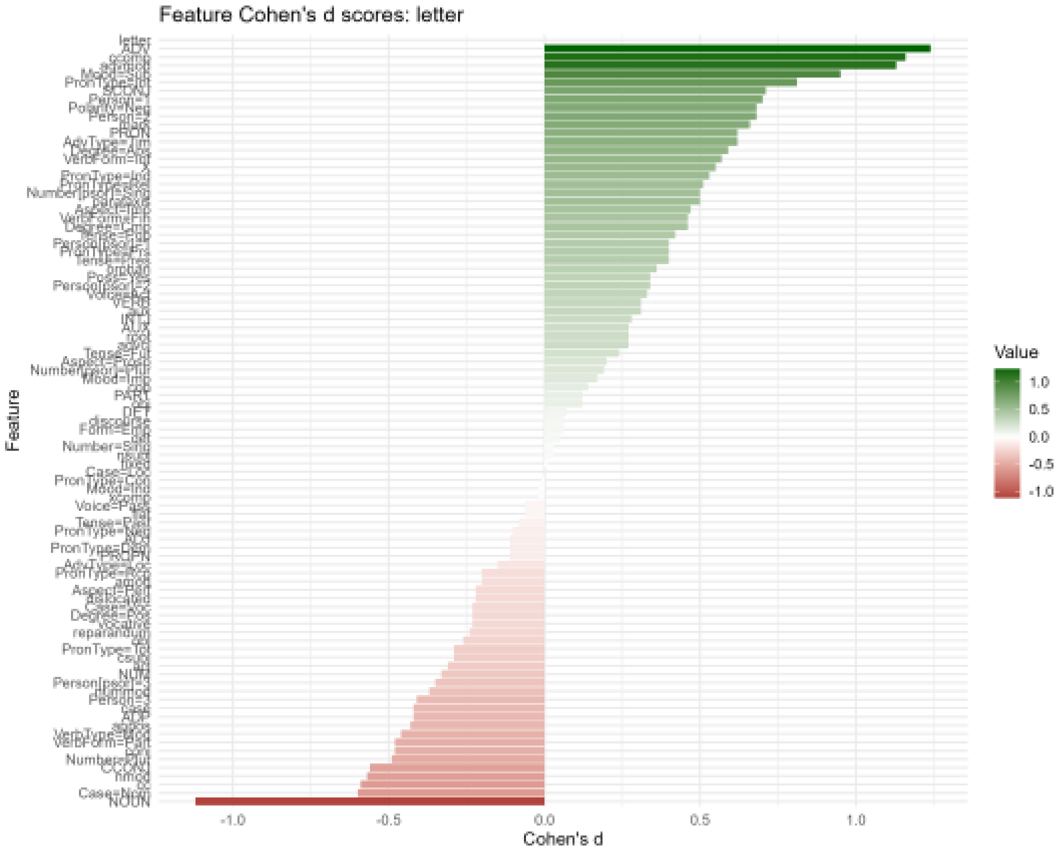

Feature Cohen's d scores: play
Feature
Cohen's d
-2
0
2
4
Value
4
2
0
-2

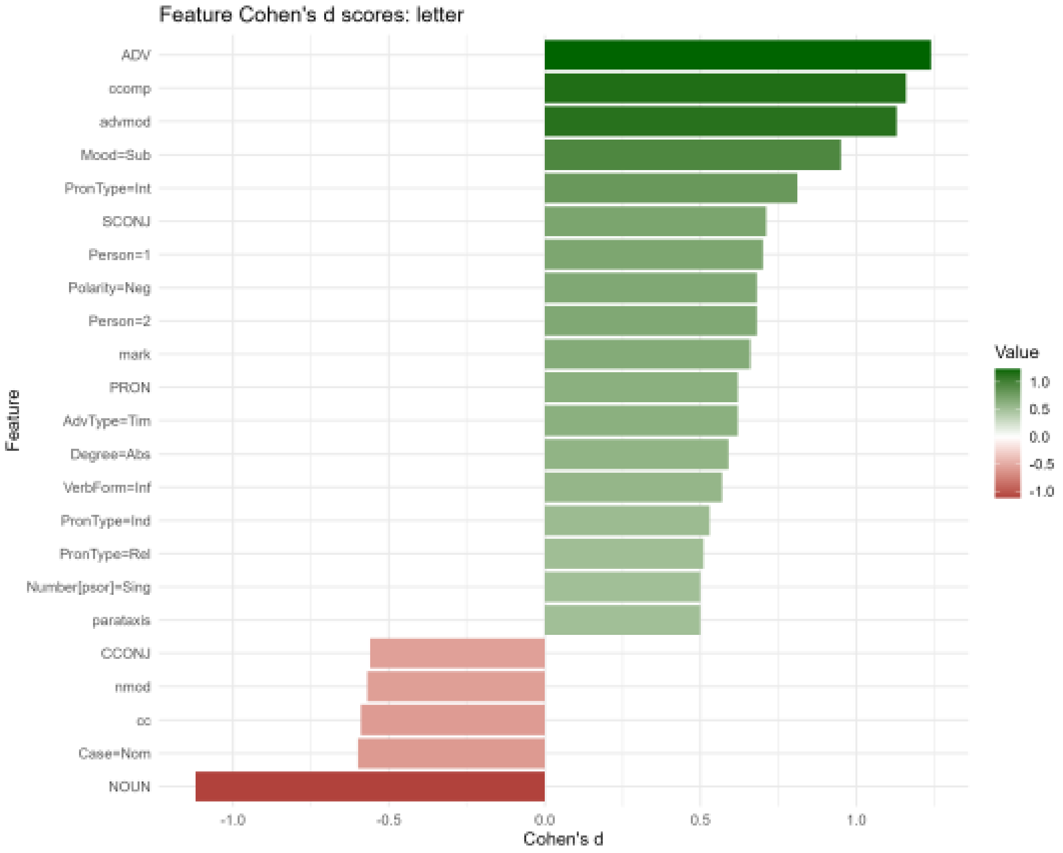

Feature Cohen's d scores: christian_philosophy
Feature
Cohen's d
-1
0
1
2
Value
2
1
0
-1


Feature Cohen's d scores: letter
Feature
Cohen's d
-1.0
-0.5
0.0
0.5
1.0
Value
1.0
0.5
0.0
-0.5
-1.0
letter
Number
AdvType
VerbForm
Person
Voice
PronType
Case=Nom
NOUN

Letter

Letter visualization with all and selected





UNIVERSITY OF TURKU

Close reading the results

- *Letter*: frequent use of **adverbs** and **first-person verbs**
    - ***dixi*** *hanc legem Publium Clodium* ***iam ante*** *servasse*
    - **I said** that Publius Clodius had **already previously** complied with this law
    - *quod* ***ego*** *non* ***credo***
    - what ***I*** don’t **believe** (Cic. *Att*. 1.16)
- Logical in **personal correspondence**, mimicking conversational **casual** speech

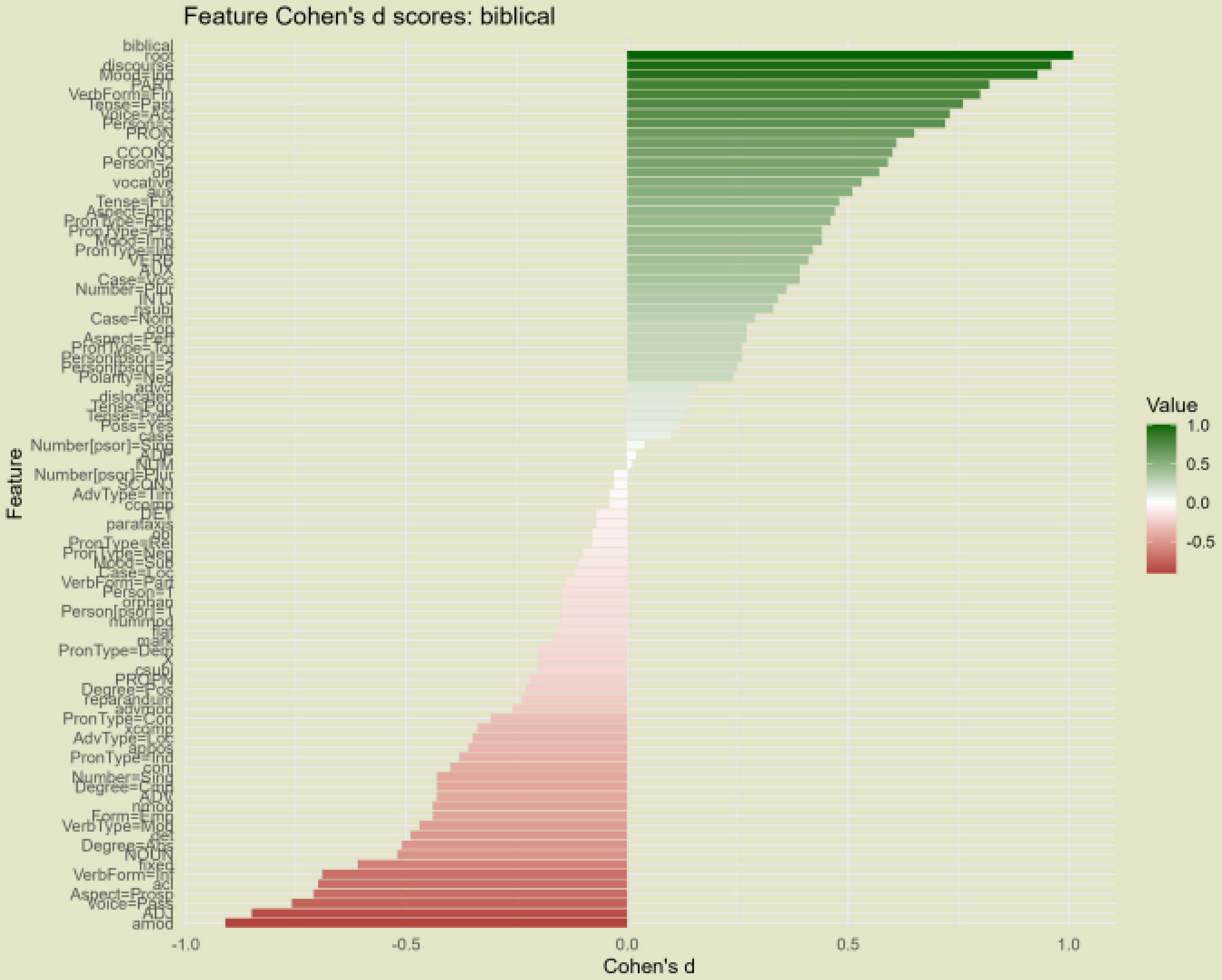
UNIVERSITY
OF TURKU

Close reading the results

- *Charter*: overrepresentation of **proper nouns** and **first-person pronouns**
    - *constat* ***me Sanitulum*** *filium quondam* ***Cicchi*** *de loco* ***Brancalo***
    - It is established that **I**, **Sanitulus**, son of the late **Cicchus** of the place **Brancale**
    - ***Gheipertum*** *clericum scribere* ***rogavi***
    - **I asked** the clerk **Gheipertus** to write (LLCT, doc. 36)

- Suitable for **legal texts** reporting the selling of personal property, precisely **identifying actors**

Results for “*Biblical*”



The Vulgate

UNIVERSITY OF TURKU

A closer look at “*Biblical*”

0.8 large positive effect

root, 1.01
discourse, 0.96
Mood=Ind, 0.93
PART, 0.82
VerbForm=Fin, 0.80

0.5 medium positive effect

Tense=Past, 0.76
Voice=Act, 0.73
Person=3, 0.72
PRON, 0.65
cc, 0.61
CCONJ, 0.60
Person=2, 0.59
obj, 0.57
vocative, 0.53
aux, 0.51

A closer look at “*Biblical*” 2

https://en.wikipedia.org/wiki/Codex_Argenteus

```
-0.5 medium negative effect
Degree=Abs, -0.51
NOUN, -0.52
fixed, -0.61
VerbForm=Inf, -0.69
acl, -0.70
Aspect=Prosp, -0.71
Voice=Pass, -0.76
-0.8 large negative effect
ADJ, -0.85
amod, -0.91
```

Discussion

- Close reading of the results aligns with established knowledge
- KFA highlights the use of adverbs in *Letter*, as is common knowledge for Cicero’s letters, pronouns are typical features of conversation (Biber and Conrad, 2019) – method verification
- In *Charters*, proper nouns are prominently emphasized, consistent with findings from previous studies (Korkiakangas, 2020)

UNIVERSITY OF TURKU

Future studies – and then?

- Expand method to include lesser-known text collections, e.g. *Corpus Corporum*
- Extending to syntactic analysis
- As Hudspeth et al. (2024) note, distinctions between Classical, Medieval, and Neo-Latin should also be considered when examining variation
- How stable are register features in texts from different time contexts?
- Other uses for the KFA method?

References

- Key feature analysis: a simple, yet powerful method for comparing text varieties. / Egbert, Jesse; Biber, Douglas. *Corpora*, Vol. 18, No. 1, 04.2023, p. 121-133.
- Marisa Hudspeth, Brendan O’Connor, and Laure Thompson. 2024. Latin Treebanks in Review: An Evaluation of Morphological Tagging Across Time. In *Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)*
- Biber, Douglas, and Susan Conrad. *Register, Genre, and Style*. Second edition, Cambridge University Press, 2019, https://doi.org/10.1017/9781108686136.

Thank you

- Emil Aaltonen foundation
- Friends of Villa Lante Society

ALFRED
KORDELIN
FOUNDATION

Happy to get in touch - LinkedIn, Bluesky and hmknie@utu.fi

UNIVERSITY
OF TURKU