owards machine readable linguistic resources for Medieval Latin

from non-structured editions to automatic parsing

https://nohat.cc/f/floral-styled-letter-t-typography-free-stock-vector-564269/comrawpixel564269-201907300036.html

"Towards machine readable linguistic resources for Medieval Latin – from non-structured editions to automatic parsing"

Hanna-Mari Kupari

DH pizza lunch seminar 14th April, 2023

Aalto University

TURUN YLIOPISTO

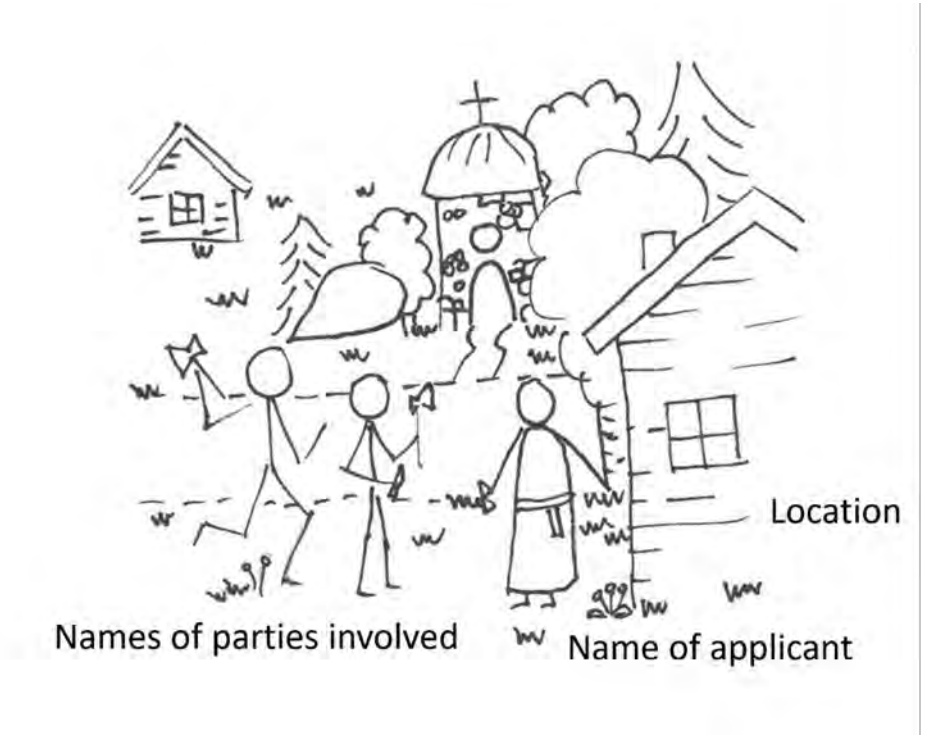
EMIL AALTOSEN SÄÄTIÖ

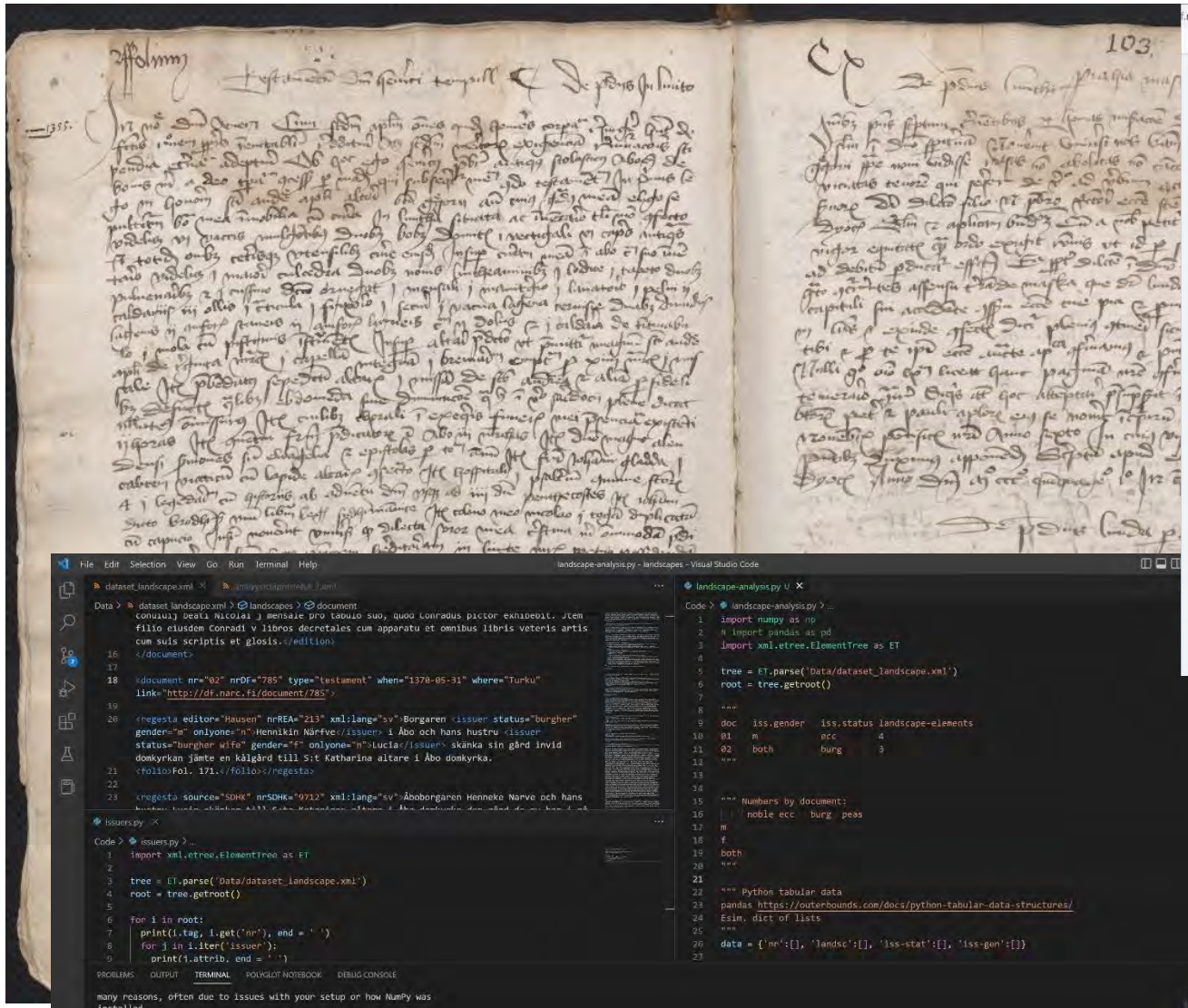
Outline of the presentation

Short background

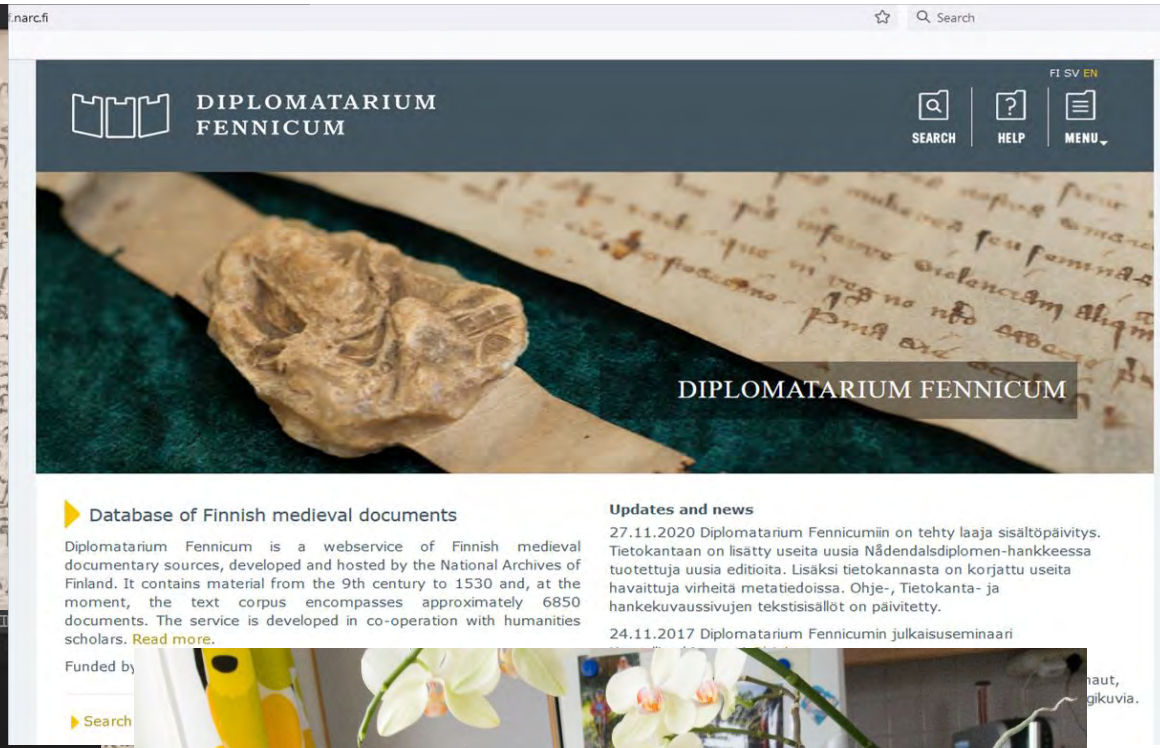
1. The data
2. The structuring
3. The automatic morpho-syntactic parsing

Discussion





landscape-analysis.py - landscapes - Visual Studio Code
dataset_landscape.xml
landscape-analysis.py
issuers.py
PROBLEMS
OUTPUT
TERMINAL
POLYGLOT NOTEBOOK
DEBUG CONSOLE
many reasons, often due to issues with your setup or how NumPy was installed.
We have compiled some common reasons and troubleshooting tips at:
https://numpy.org/devdocs/user/troubleshooting-importerror.html
Please note and check the following:
* The Python version is: Python3.9 from "C:\ProgramData\Anaconda3\python.exe"
* The NumPy version is: "1.21.5"
and make sure that they are the versions you expect.
Please carefully study the documentation linked above for further help.
Original error was: DLL load failed while importing _multiarray_umath: Määritettyä osaa ei löydy.
PS C:\Users\Hanna-Mari\Documents\Tohtorius\Kielimaisemat Lamberg\landscapes>


df.narc.fi
Search
DIPLOMATARIUM FENNICUM
FI SV EN
SEARCH
HELP
MENU
DIPLOMATARIUM FENNICUM
Database of Finnish medieval documents
Diplomatarium Fennicum is a webservice of Finnish medieval documentary sources, developed and hosted by the National Archives of Finland. It contains material from the 9th century to 1530 and, at the moment, the text corpus encompasses approximately 6850 documents. The service is developed in co-operation with humanities scholars. Read more.
Funded by
Search
Updates and news
27.11.2020 Diplomatarium Fennicumiin on tehty laaja sisältöpäivitys. Tietokantaan on lisätty useita uusia Nådendalsdiplomen-hankkeessa tuotettuja uusia editioita. Lisäksi tietokannasta on korjattu useita havaittuja virheitä metatiedoissa. Ohje-, Tietokanta- ja hankekuvaussivujen tekstisisällöt on päivitetty.
24.11.2017 Diplomatarium Fennicumin julkaisuseminaari


Conciliorum Oecumenicorum Decreta

Short bio

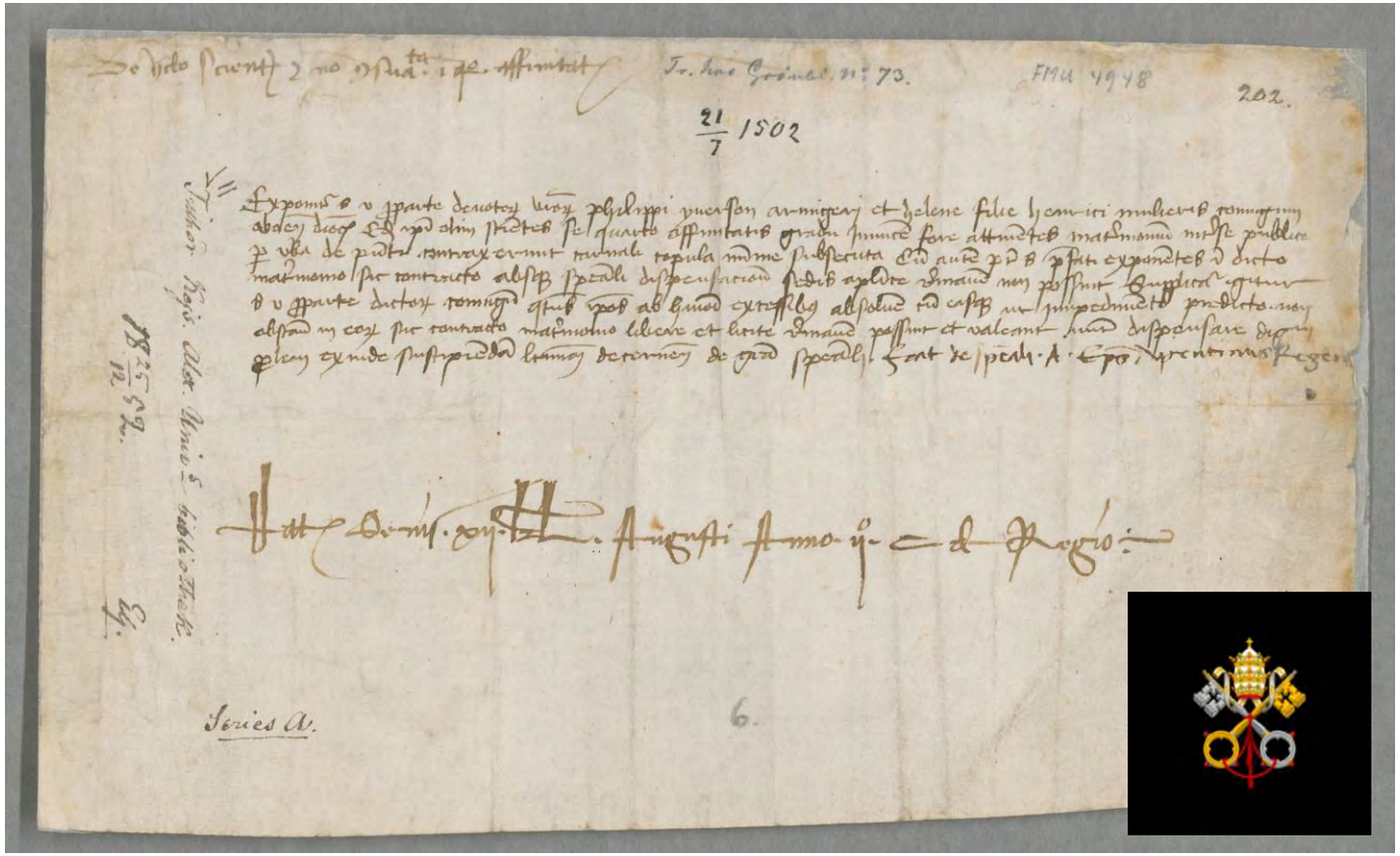
- High school Latin - 2003
- BA: Latin 2014 – late antiquity, quantitative methods and gendered language use
- MA: Latin 2019 – Medieval Latin, basic digital methods and the penitentiary documents: expressions of violence and death
- PhD studies: digital linguistics major – university of Turku

Section 1

The data

Motivation

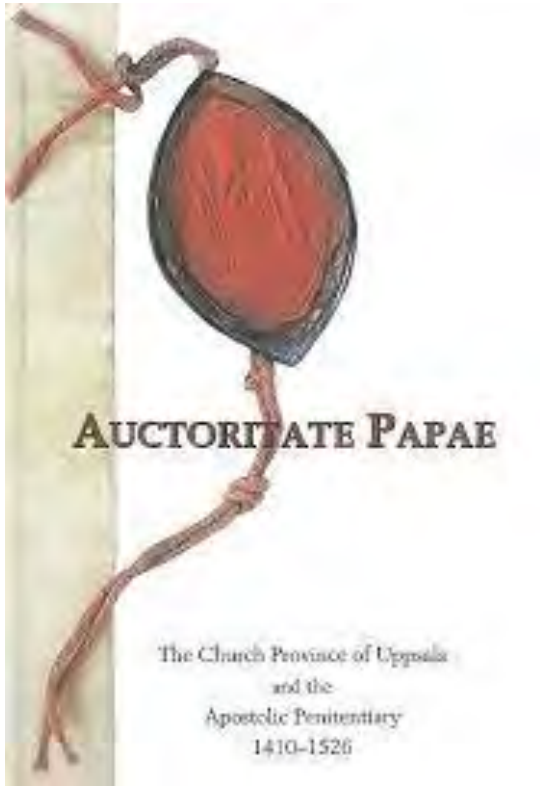
- Professor Kirsi Salonen
- Local history
- Historical linguistics
- Insightful and fun
- Novel research

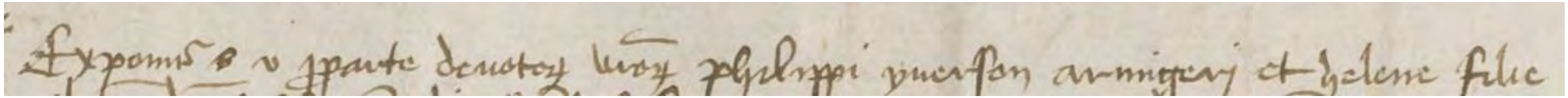


http://df.narc.fi/document/4948

Penitentiary documents

- From the Apostolic See
- Copybooks survive and edited according to modern priciples for historians using quantitative methods
- The Finnish documents ->





Exponitur sanctitati vestre pro parte devotorum vestrorum Philippi Yverson armigeri et Helene filie …..// Henrici mulieris coniugum Aboensis diocesis, quod ipsi olim scientes se quarto affinitatis gradu invicem fore attinentes matrimonium inter se publice per verba de presenti contraxerunt carnali copula minime subsecuta. /

The context of language use with the penitentiary documents

The situation – e.g. violent assault

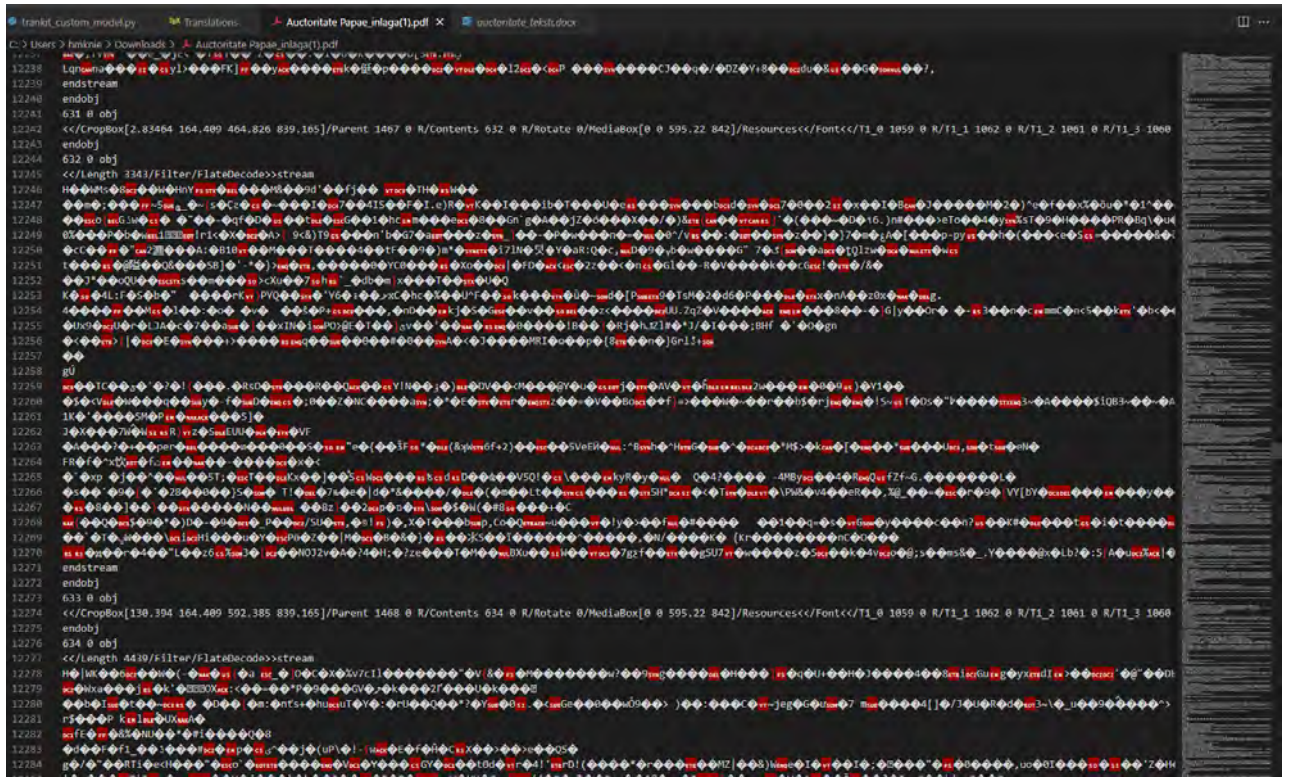
description

+
Linguistic features

Section 2

The structuring

From Data to Capta



https://www.shopsy.in

Johanna Drucker

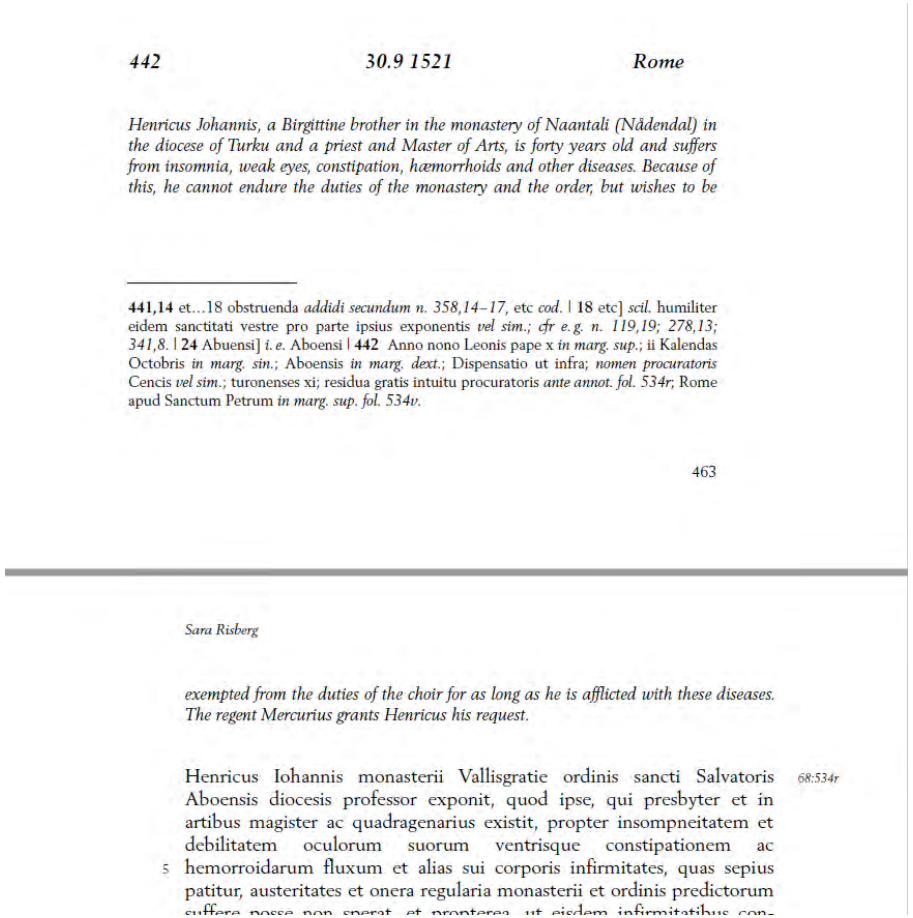
- *Humanities Approaches to Graphical Display*:
- Capta is “taken” actively while data is assumed to be a “given” able to be recorded and observed.
- From this distinction, a world of differences arises
- knowledge is constructed, taken, not simply given as a natural representation of pre-existing fact
- http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

The structuring

- What works for human consumption is really difficult for machine reading
- The end point of an editors work is where I start – pdf's
- By the means of regex and manual annotation
- Balancing between accuracy and number of categories

https://www.blumer-lehmann.com/en/buy-products/silos-and-grit-storage/wooden-silos/round-silos/300-m3-round-silo.html

TEI-XML markdown



442 30.9 1521 Rome

Henricus Johannis, a Birgittine brother in the monastery of Naantali (Nådendal) in the diocese of Turku and a priest and Master of Arts, is forty years old and suffers from insomnia, weak eyes, constipation, hæmorrhoids and other diseases. Because of this, he cannot endure the duties of the monastery and the order, but wishes to be

441,14 et…18 obstruenda *addidi secundum n. 358,14–17*, etc *cod.* | 18 etc] *scil.* humiliter eidem sanctitati vestre pro parte ipsius exponentis *vel sim.*; *cfr e.g. n. 119,19; 278,13; 341,8.* | **24** Abuensi] *i.e.* Aboensi | **442** Anno nono Leonis pape x *in marg. sup.*; ii Kalendas Octobris *in marg. sin.*; Aboensis *in marg. dext.*; Dispensatio ut infra; *nomen procuratoris* Cencis *vel sim.*; turonenses xi; residua gratis intuitu procuratoris *ante annot. fol. 534r*; Rome apud Sanctum Petrum *in marg. sup. fol. 534v.*

463

Sara Risberg

exempted from the duties of the choir for as long as he is afflicted with these diseases. The regent Mercurius grants Henricus his request.

Henricus Iohannis monasterii Vallisgratie ordinis sancti Salvatoris Aboensis diocesis professor exponit, quod ipse, qui presbyter et in artibus magister ac quadragenarius existit, propter insompneitatem et debilitatem oculorum suorum ventrisque constipationem ac hemorroidarum fluxum et alias sui corporis infirmitates, quas sepius patitur, austeritates et onera regularia monasterii et ordinis predictorum

68:534r

```
trankit_custom_model.py   Translations   auctoritate_teksti.xml ×
Data > auctoritate_teksti.xml > auctoritate > document > h2
7460  emulorum obstruenda </i>supplicatur etc., quatenus, si iudici infrascripto
7461  canonice constiterit de iusto bello et aliis premissis, ipsum oratorem
7462  nullam propter premissa inhabilitatis vel irregularitatis maculam sive
7463  notam contraxisse sed premissis non obstantibus in suis ordinibus
7464  etiam in altaris ministerio ministrare libere et licite posse nuntiari et
7465  declarari mandare dignemini prout in forma. Fiat in forma. Mercurius
7466  regens. Committatur episcopo Abuensi vel eius vicario in spiritualibus.
7467  Fiat. Mercurius.
7468  </document>
7469
7470  <document>
7471  <h2 num="442" niputus="n" several_wittnesses="n">442 <date when="1521-09-30">30.9 1521</date> Rome</h2>
7472  Henricus Iohannis monasterii Vallisgratie ordinis sancti Salvatoris
7473  Aboensis diocesis professor exponit, quod ipse, qui presbyter et in
7474  artibus magister ac quadragenarius existit, propter insompneitatem et
7475  debilitatem oculorum suorum ventrisque constipationem ac
7476  hemorroidarum fluxum et alias sui corporis infirmitates, quas sepius
7477  patitur, austeritates et onera regularia monasterii et ordinis predictorum
7478  suffere posse non sperat, et propterea, ut eisdem infirmitatibus consulere
7479  possit, cuperet ab eisdem ac oneribus chori eximi, quod sibi
7480  permitti dubitat sede apostolica desuper inconsulta. Supplicatur igitur
7481  humiliter pro parte dicti oratoris, quatenus se ab eisdem austeritatibus
7482  ac interessentia chori tam in diurnis quam nocturnis horis infirmitatibus
7483  huiusmodi durantibus, dummodo propterea divinum officium recitare
7484  non omittat, veris existentibus premissis, cuius super hoc parte licentia
7485  minime requisita eximi et sibi  indulgeri secumque desuper misericorditer
7486  dispensari non obstantibus constitutionibus et ordinationibus
7487  apostolicis ac tam provincialibus quam synodalibus necnon monasterii
7488  et ordinis predictorum statutis et consuetudinibus iuramento etc.
7489  roboratis privilegiis quoque indultis et litteris apostolicis monasterio et
7490  ordini huiusmodi sub quibusvis verborum formis et clausulis etiam
7491  derogatoriarum derogatoriis, fortioribus, efficacioribus et insolitis contra
7492  premissa forsan concessis et imposterum concedendis, etiam si de illis
7493  eorumque totis tenoribus pro illorum sufficienti derogatione specialis,
7494  specifica, expressa et individua ac de verbo ad verbum, non autem etc.
7495  habenda foret et alias, cuiuscumque tenoris existant, ceterisque contrariis
7496  quibuscumque dignemini de gratia speciali et expresso. Fiat de
7497  speciali et expresso. Mercurius regens.
7498  </document>
7499
7500  <document>
7501  <h2 num="443" niputus="n" several_wittnesses="n">443 <date when="1522-06-28">28.6 1522</date> Rome</h2>
7502  Andreas Gladh diaconus Arosiensis diocesis in artibus magister exponit,
7503  quod ipse ex magno devotionis fervore  cupit ad presbyteratus ordinem
7504  promoveri et promotus in illo ac aliis iam per eum susceptis ordinibus
7505  etiam in altaris ministerio ministrare ac beneficia ecclesiastica cum
7506  cura ac sine cura invicem tamen compatibilia sibi in futurum canonice
7507  conferenda recipere et retinere posse. Sed quia quandam maculam in
7508  oculo suo dextro, ex quo forsan nichil videt, patitur, desiderium suum
```

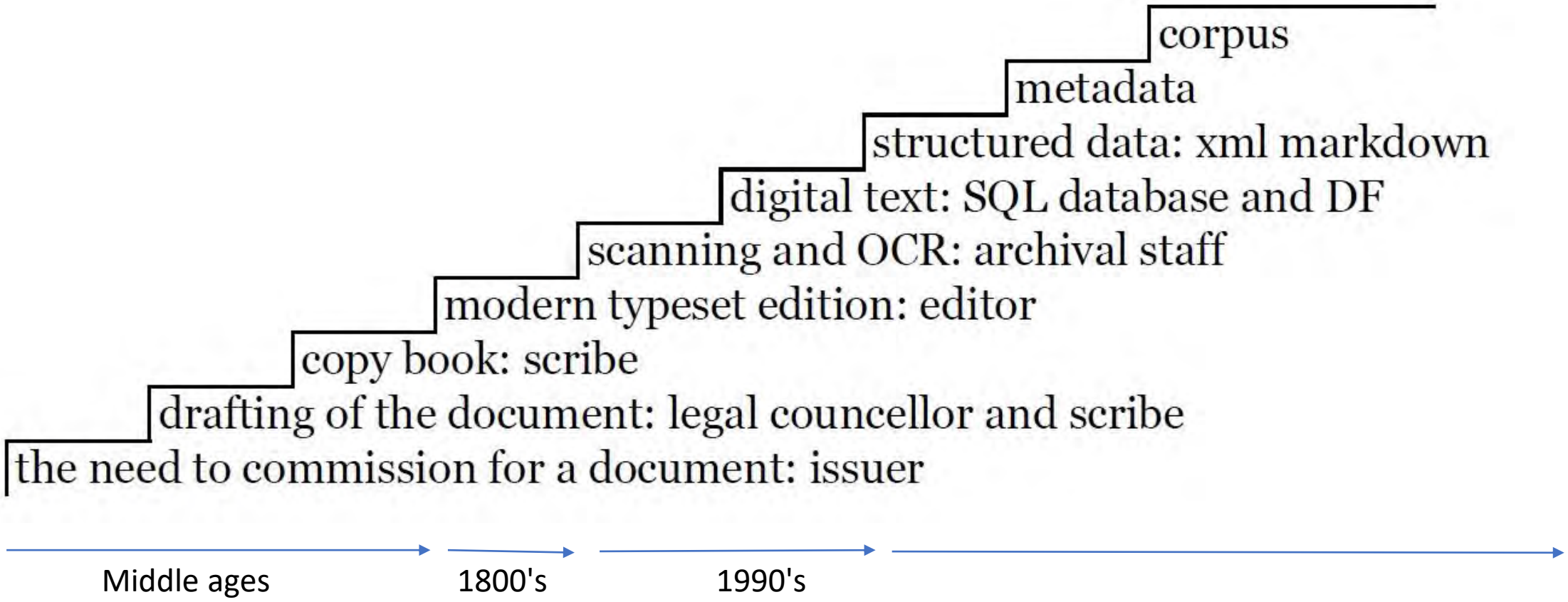
Many opportunities of xml

https://yle.fi/a/3-12159318

Linguistics landscapes: the use of landmarks in Turku copybooks

Kuvalähde: https://aamuset.fi/artikkeli/5725211

Steps of the Medieval Turku landmarks corpus



https://github.com/HannaKoo/landscapes

dataset_landscape.xml | landscapes0.xsd | landscapes1.xsd

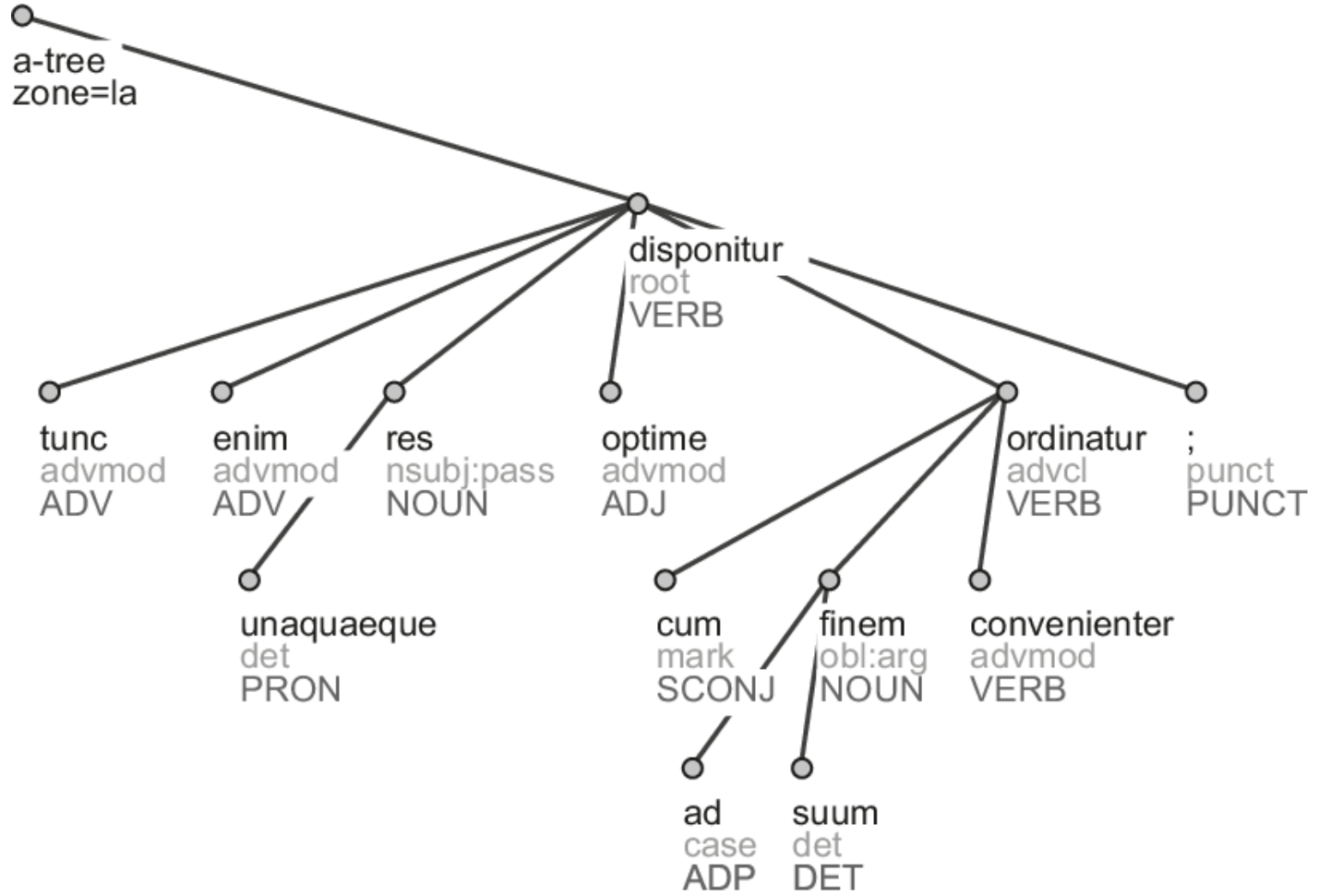
Data > dataset_landscape.xml > landscapes > document > regesta

```
meadow-land and two sheds on a hill at Kairinen. His heirs Erland, Margareta and Katarina are to
divide among themselves all his household chattels in the smaller shed at Kairinen.
115 </regesta>
116
117 <edition editor="Hausen" xml:lang="la">
118 <title>Testamentum magistri Haquini.</title>
119 <p>Altari beate patrone mee Katherine octo coclearia operis mensalis ad vnum calicem fabricandum.</
    p>
120 <p>Jtem <transaction type="lot">ffundum</transaction> prope <landscape cat="natural" type="river"
    lemma="ampnis">ampnem</landscape> cum suis <transaction type="building">edificiis</transaction>;
    jtem vnum pratellum jnsulare dictum Hartmansholm.</p>
121 <p>Jtem duo promptuaria paria jn monticulo Keeris sytuata.</p>
122 <p>Erlando, Margarete et Katherine secundum seculum heredibus meis do omnia bona mea pro
    suppellectilj domus necessaria, que in paruo promptuario Keris collocata sunt, ad inuicem
    distribuenda et nil aliud penitus nisi quod hiis litteris est insitum. Aliud facientibus occurrat
    Dei iudicium vlcionis.</p>
123 </edition>
124 </document>
125
126 <document nr="10" nrDF="3851" type="" when="1480-omkring" where="Turku" link="http://df.narc.fi/
    document/3851">
127 <regesta editor="Hausen" nrREA="677" xml:lang="sv">
```

Section 3

The automatic morpho-syntactic parsing

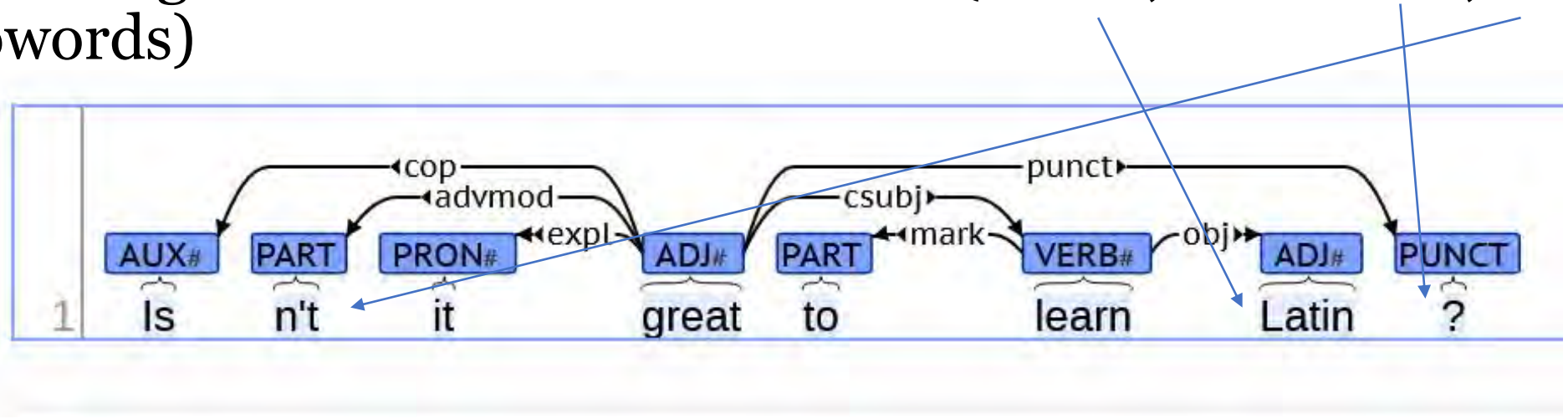
The automatic morpho-syntactic parsing



https://www.semanticscholar.org/paper/The-Project-of-the-Index-Thomisticus-Treebank-Passarotti/9c869bf2e6578a1872137f4de2ad5b87792e53c0

Automatic formo-syntactic parsing is:

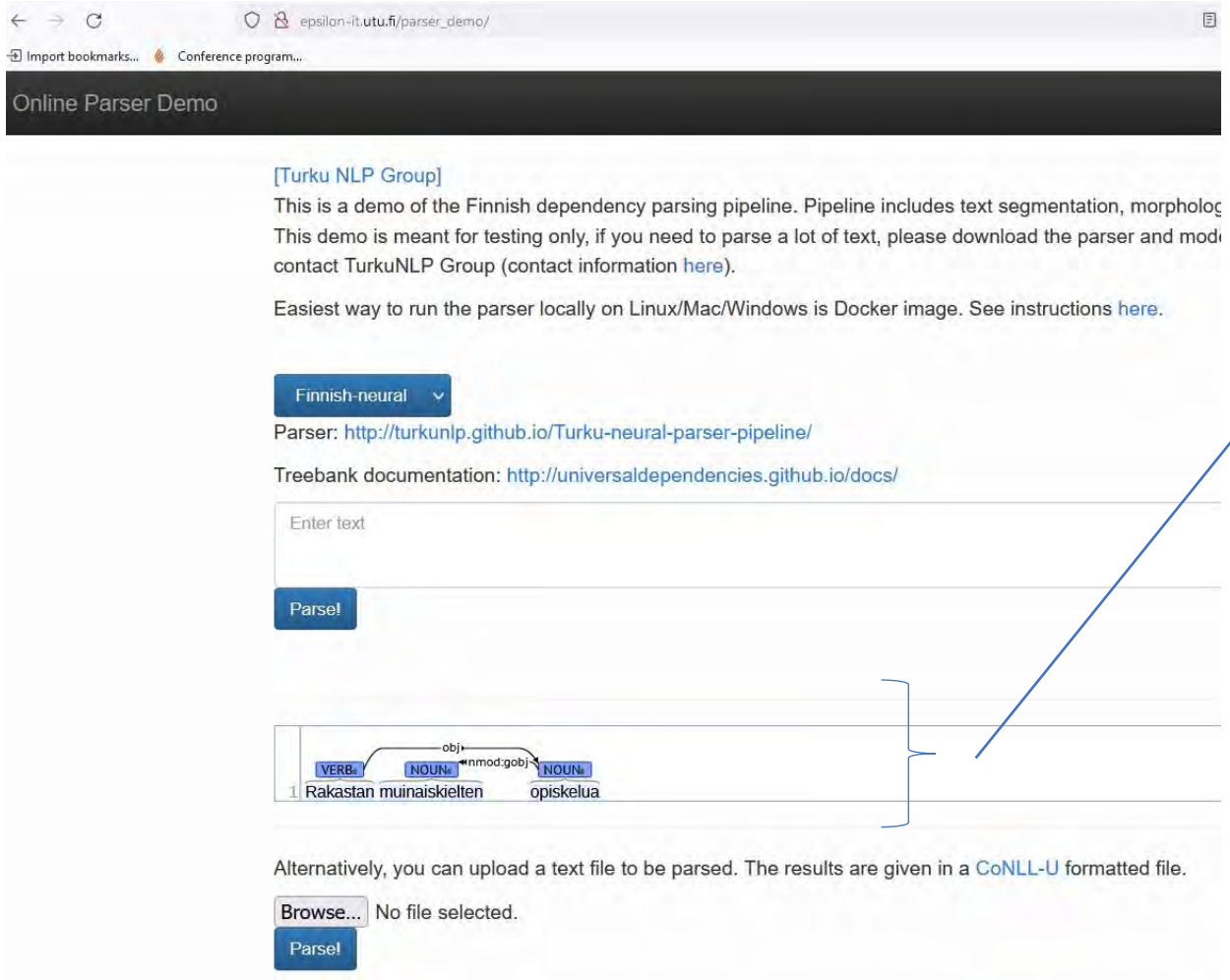
- “to analyze the input sentence in terms of grammatical constituents, identifying the parts of speech and their syntactic relations”
- producing tokens: basic blocks of text (words, characters, or subwords)

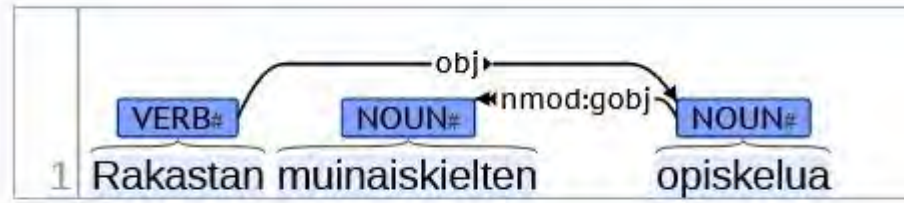


http://epsilon-it.utu.fi/parser_demo/

Warse.org & nlp.stanford.edu

Finnish: Online parser demo





More information on the parser's project page:
https://turkunlp.github.io/Turku-neural-parser-pipeline/

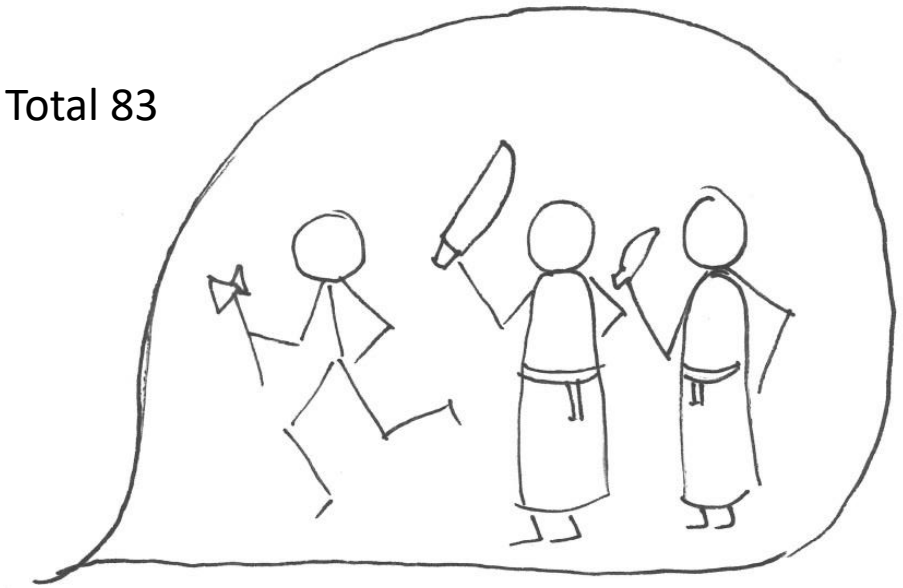
Why use parsers?

percusserit	11	119	se deffendendo aliter fugere sive evadere non valens percusserit
percusserunt	2	232	vim vi repellendo ipsum laicum adeo percusserunt
percussiendo	1	260	arrepto percussiendo in terram prostravit
percussio	1	239	quod levis fuerit percussio
percussione	23	300a	non ex percussione exponentis mors subsecuta fuerit
percussionem	1	239	prout ante percussionem predictam
percussioni	2	318	Petrus volens percussioni et periculo huiusmodi obviare
percussionibus	15	341	dictus diaconus de dictis percussionibus plene convaluerit
percussionis	1	239	ad septenam diem a die percussionis
percussis	1	239	Exinde in spatulis percussis prout ... incolumis
percussisset	2	264	quadam lancea seu cuspide percussisset
percussit	68	264	in capite ipsius percussit unico ictu
percusso	2	340	ipseque exponens cum percusso concordaverit
percussor	1	19	Quod cum percussor suus … viderent
percussorem	1	342	tam graviter vulneratum laicum percussorem
percussum	2	303	exponens dictum Laurentium sic percussum dimissit
percussus	12	340	Cum autem dictus presbyter percussus
percutere	9	360	oratorem suo gladio in capite percutere voluit
percuteret	4	160	ne ipse familiaris dictum laicum percuteret
percuteretur	1	444	quodam malleo ferreo graviter percuteretur
percuti	1	300a	percuti seu forsan interfici illum
percutiebant	1	406	pungebant et percutiebant
percutiendo	1	427	ipsum leniter percutiendo existimans
percutiens	1	360	oratorem dorsotenus percutiens
percuttere	2	145	cum deteriori parte securis percuttere voluit
percutteret	2	234	ut eum percutteret
repercussit	1	195	vim vi reppellendo cum eadem securi dictum laicum taliter repercussit
repercussiens	1	292	cum cribro repercussiens et ulterius manus eius violentas dicti … iniecit

VERBS

percusserit	11	119	se deffendendo aliter fugere sive evadere non valens percusserit
percusserunt	2	232	vim vi repellendo ipsum laicum adeo percusserunt
percussiendo	1	260	arrepto percussiendo in terram prostravit
percussio	1	239	quod levis fuerit percussio
percussione	23	300a	non ex percussione exponentis mors subsecuta fuerit
percussionem	1	239	prout ante percussionem predictam
percussioni	2	318	Petrus volens percussioni et periculo huiusmodi obviare
percussionibus	15	341	dictus diaconus de dictis percussionibus plene convaluerit
percussionis	1	239	ad septenam diem a die percussionis
percussis	1	239	Exinde in spatulis percussis prout ... incolumis
percussisset	2	264	quadam lancea seu cuspide percussisset
percussit	68	264	in capite ipsius percussit unico ictu

percutere (*percutiō*, percussī, *percussum*) 3 - Total 83



Why we need lemmas:

In Vulgar Latin, *do* becomes **dao*, by analogy with the root vowel -a-, but also by some elided third conjugation verbs like **vao*, **vare* (< *vādo, vādere*).

This table includes an archaic present subjunctive conjugation on a *du-* root that appears in the works of Plautus and Terence.

Conjugation of *dō* (first conjugation, irregular short *a* in most forms except *dās* and *dā*) [less ▲]

indicative		singular			plural		
		first	second	third	first	second	third
active	present	dō	dās	dat	damus	datis	dant, danunt
	imperfect	dabam	dabās	dabat	dabāmus	dabātis	dabant
	future	dabō	dabis	dabit	dabimus	dabitis	dabunt
	perfect	dedī	dedistī	dedit	dedimus	dedistis	dedērunt, dedēre
	pluperfect	dederam	dederās	dederat	dederāmus	dederātis	dederant
	future perfect	dederō	dederis	dederit	dederimus	dederitis	dederint
passive	present	dor	daris, dare	datur	damur	daminī	dantur
	imperfect	dabar	dabāris, dabāre	dabātur	dabāmur	dabāminī	dabantur
	future	dabor	daberis, dabere	dabitur	dabimur	dabiminī	dabuntur
	perfect	*datus* + present active indicative of *sum*					
	pluperfect	*datus* + imperfect active indicative of *sum*					
	future perfect	*datus* + future active indicative of *sum*					

subjunctive		singular			plural		
		first	second	third	first	second	third
active	present	dem, duim	dēs, duās, duīs	det, duit	dēmus	dētis	dent, duint
	imperfect	darem	darēs	daret	darēmus	darētis	darent
	perfect	dederim	dederīs	dederit	dederīmus	dederītis	dederint
	pluperfect	dedissem	dedissēs	dedisset	dedissēmus	dedissētis	dedissent
passive	present	der	dēris, dēre	dētur	dēmur	dēminī	dentur
	imperfect	darer	darēris, darēre	darētur	darēmur	darēminī	darentur
	perfect	*datus* + present active subjunctive of *sum*					
	pluperfect	*datus* + imperfect active subjunctive of *sum*					

imperative		singular			plural		
		first	second	third	first	second	third
active	present	—	dā	—	—	date	—
	future	—	datō	datō	—	datōte	dantō
passive	present	—	dare	—	—	daminī	—
	future	—	dator	dator	—	—	dantor

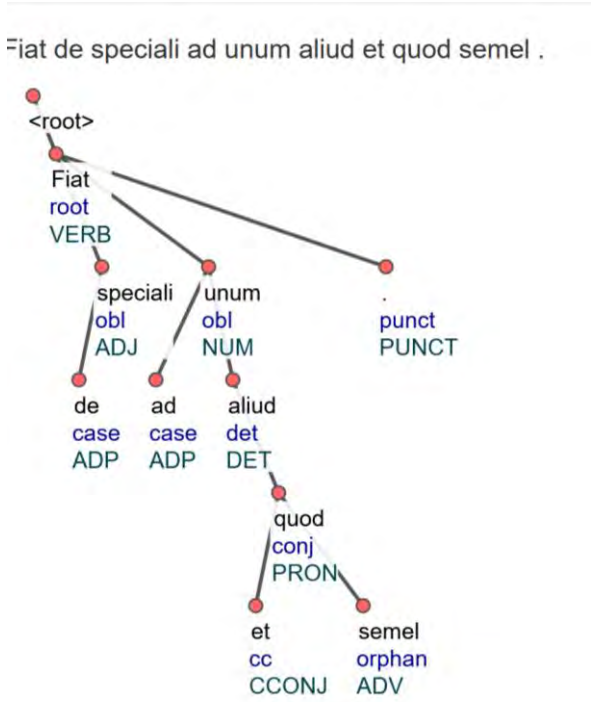
non-finite forms	active			passive		
	present	perfect	future	present	perfect	future
infinitives	dare	dedisse	datūrum esse	darī	datum esse	datum īrī
participles	dāns	—	datūrus	—	datus	dandus

verbal nouns	gerund			supine	
	genitive	dative	accusative	ablative	accusative	ablative
	dandī	dandō	dandum	dandō	datum	datū

Derived terms [edit]

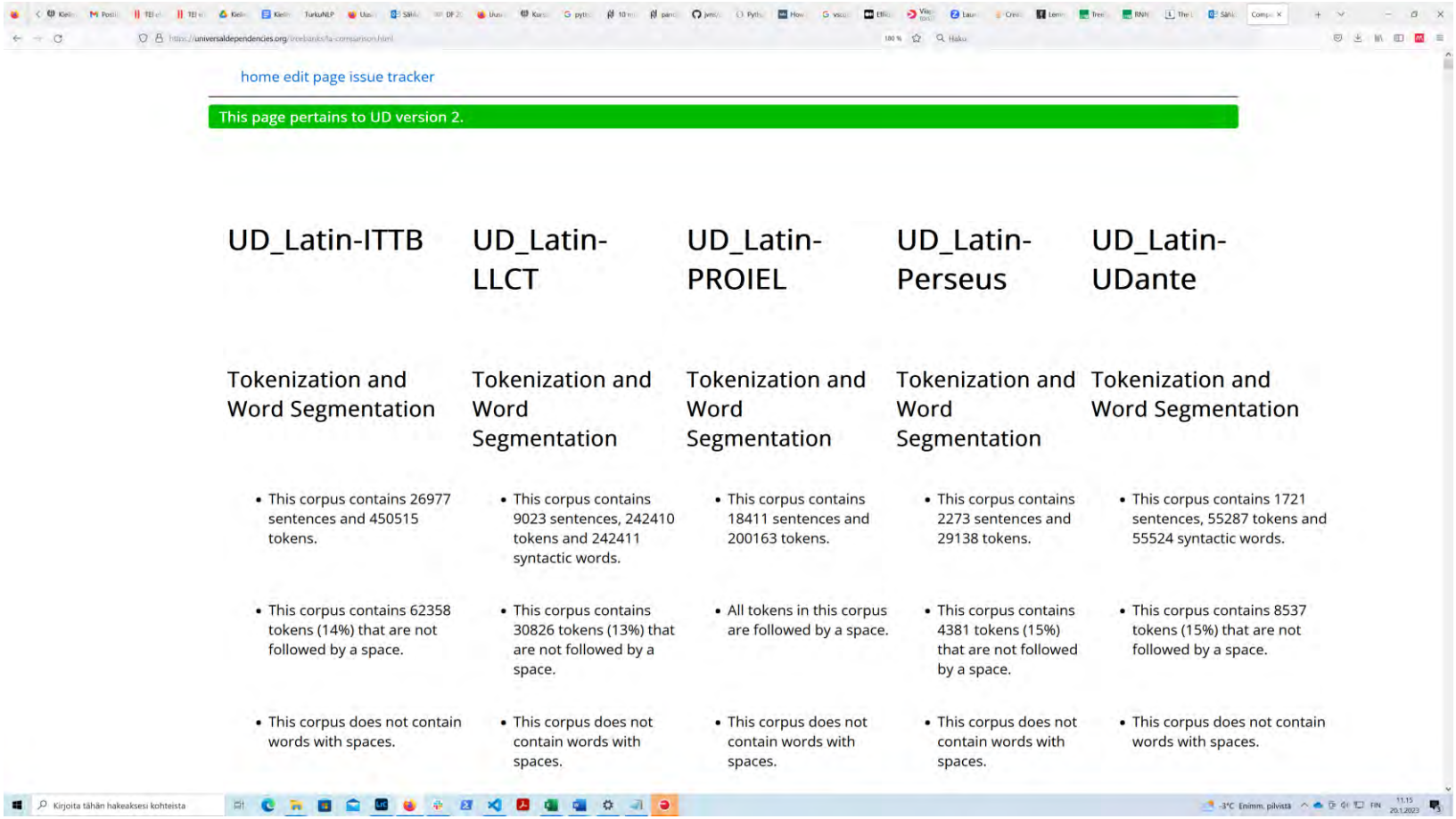
Universal Dependencies

- framework for consistent annotation of grammar (parts of speech, morphological features, and syntactic dependencies) across different human languages
- open community effort:
  - 300 contributors
  - producing nearly 200 treebanks
  - over 100 languages



https://universaldependencies.org/

UD Treebanks for Latin



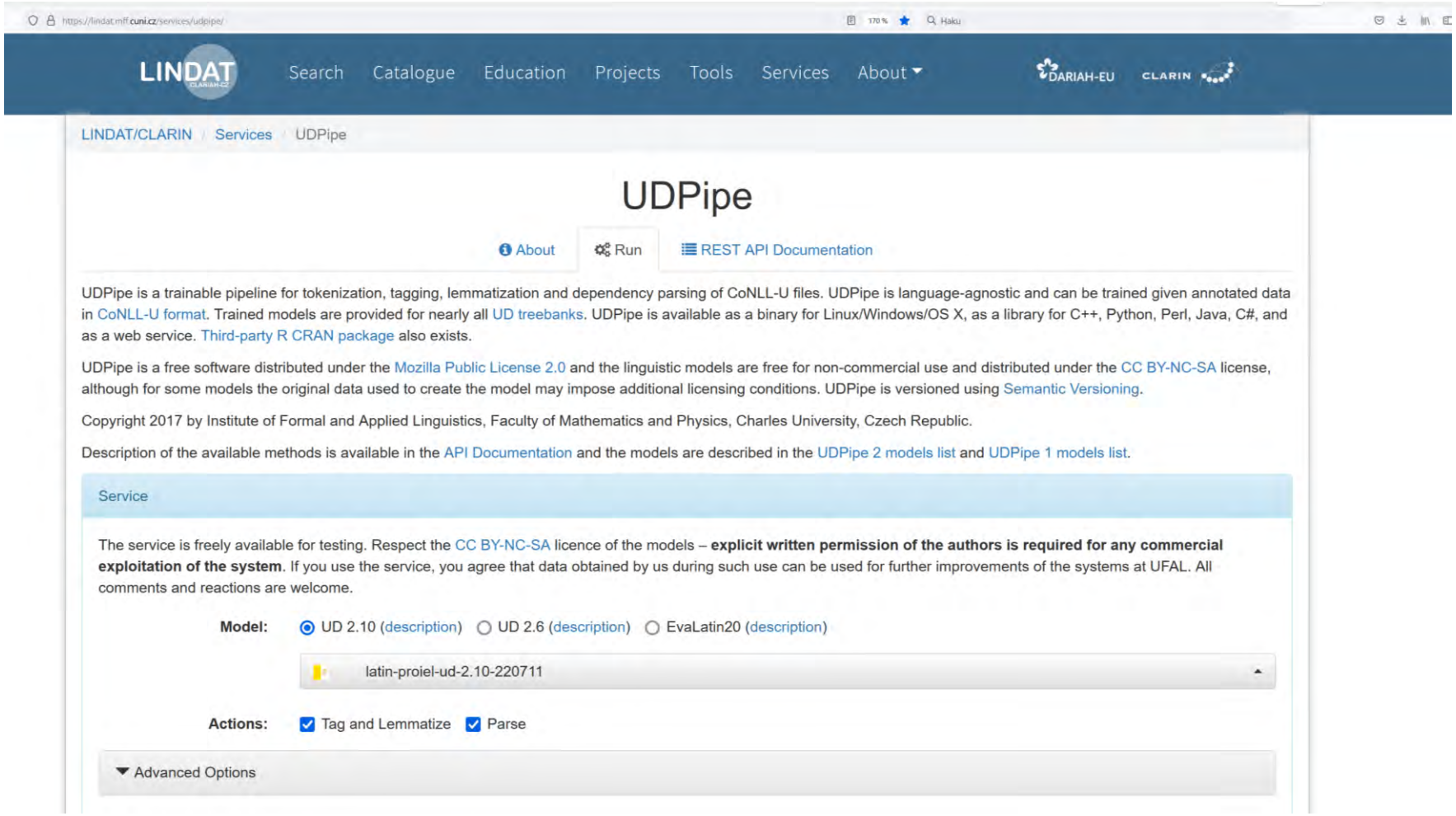
home edit page issue tracker

This page pertains to UD version 2.

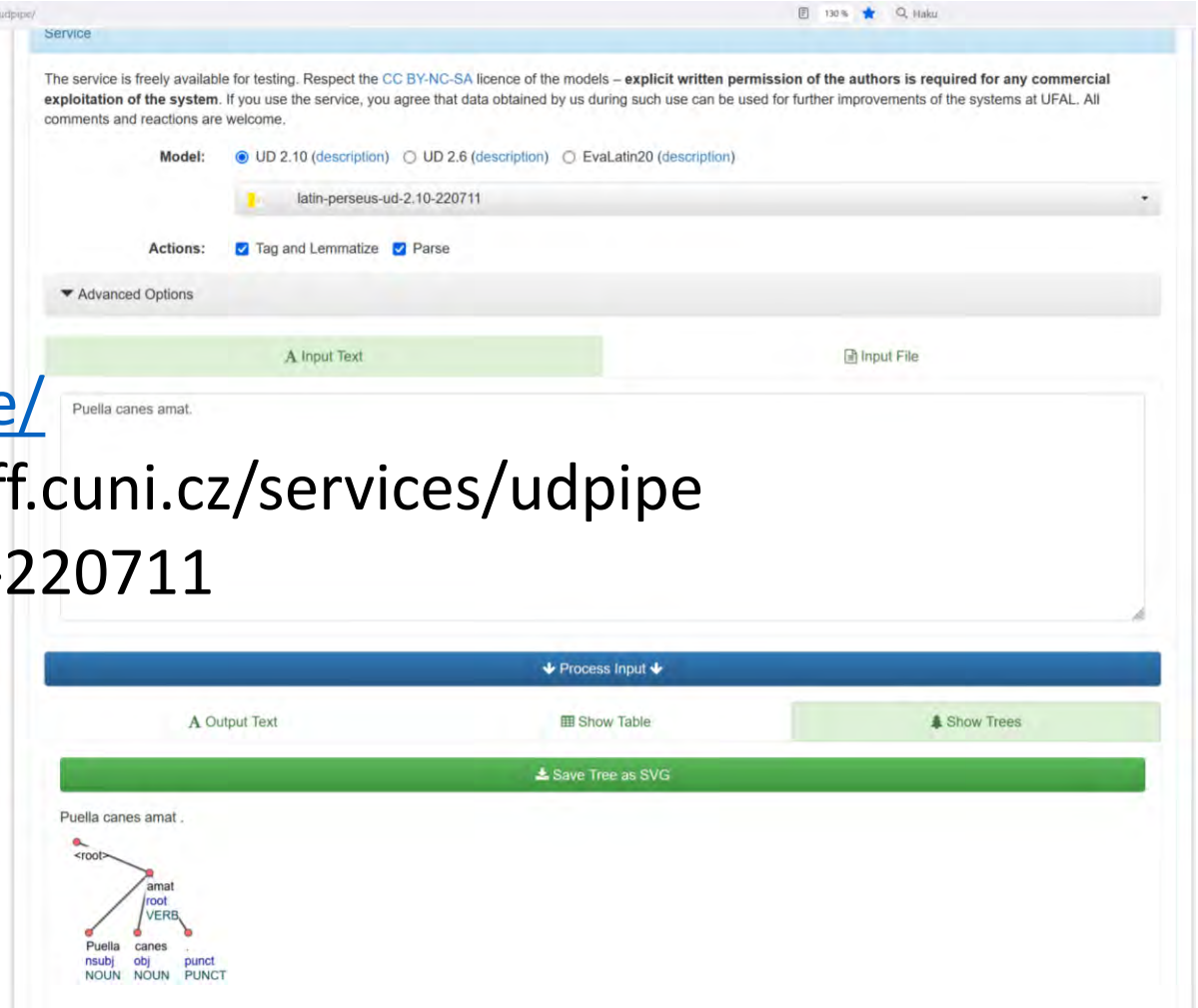
UD_Latin-ITTB	UD_Latin-LLCT	UD_Latin-PROIEL	UD_Latin-Perseus	UD_Latin-UDante
Tokenization and Word Segmentation	Tokenization and Word Segmentation	Tokenization and Word Segmentation	Tokenization and Word Segmentation	Tokenization and Word Segmentation
• This corpus contains 26977 sentences and 450515 tokens.	• This corpus contains 9023 sentences, 242410 tokens and 242411 syntactic words.	• This corpus contains 18411 sentences and 200163 tokens.	• This corpus contains 2273 sentences and 29138 tokens.	• This corpus contains 1721 sentences, 55287 tokens and 55524 syntactic words.
• This corpus contains 62358 tokens (14%) that are not followed by a space.	• This corpus contains 30826 tokens (13%) that are not followed by a space.	• All tokens in this corpus are followed by a space.	• This corpus contains 4381 tokens (15%) that are not followed by a space.	• This corpus contains 8537 tokens (15%) that are not followed by a space.
• This corpus does not contain words with spaces.	• This corpus does not contain words with spaces.	• This corpus does not contain words with spaces.	• This corpus does not contain words with spaces.	• This corpus does not contain words with spaces.

Simple example:

Puella canes amat.
The girl loves dogs.



CoNLL-U output



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

generator = UDPipe 2, https://lindat.mff.cuni.cz/services/udpipe

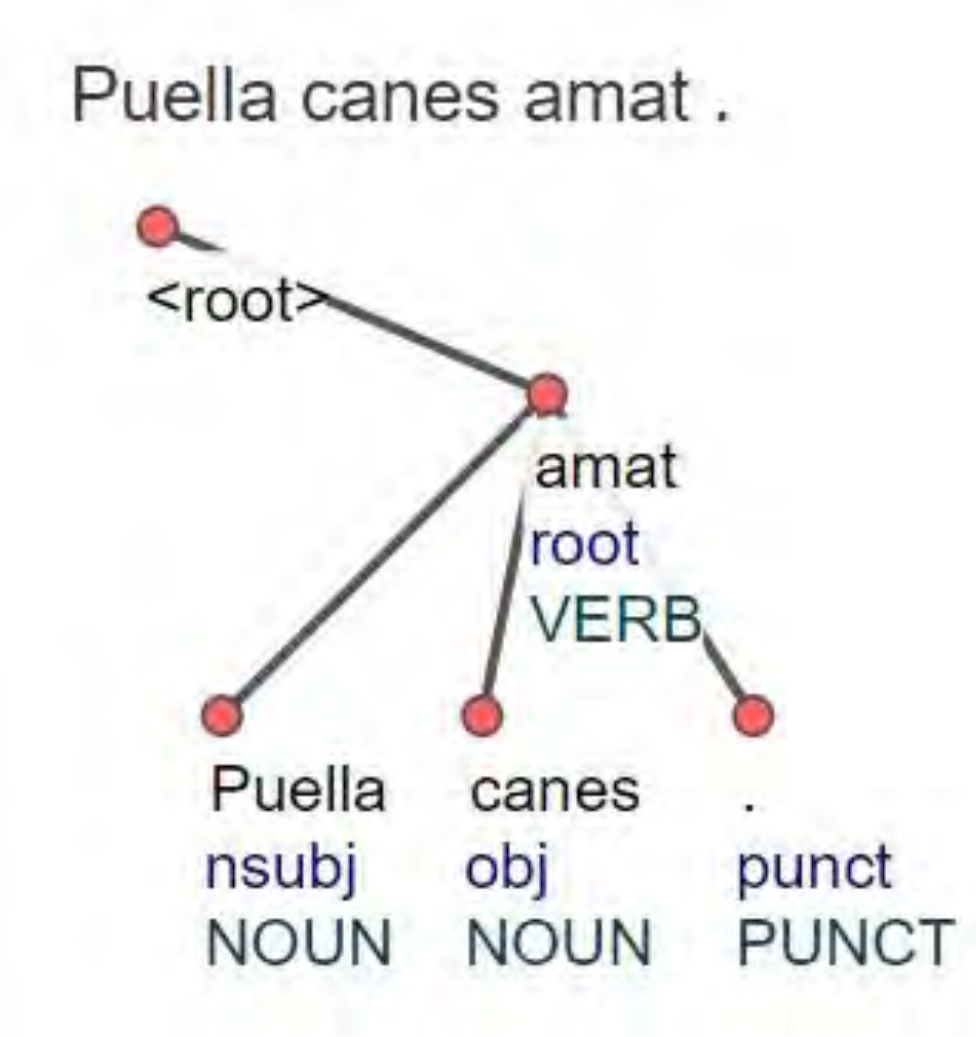
udpipe_model = latin-**perseus**-ud-2.10-220711

sent_id = 1

text = Puella canes amat.

1	Puella	puella	NOUN
2	canes	canis	NOUN
2	amat	amo	VERB
3	4	.	.	PUNCT

Tree


Puella canes amat .
<root>
amat
root
VERB
Puella
nsubj
NOUN
canes
obj
NOUN
.
punct
PUNCT

Lemmatization and morphological analysis for the Latin Dependency Treebank (Celano)

- lemmatization can be defined as consisting in the assignment of an 'ID word form' – the dictionary head word
- **dō** (*present infinitive* **dare**, *perfect active* **dedī**, *supine* **datum**); *1st conjugation* – you can **not** truncate your search at d*
- Dictionary head word **amo**

Ōdī et amō. Quārē id faciam fortasse requīris.

Nesciŏ, sed fierī sentiō et excrucior. (Catullus)

https://www.studiesaggilinguistici.it/ssl/article/view/274

Participles:

- The annotation of participles is demanding: *particeps* 'to take part in the nature of both verb and noun'
- participle **amāns -> amo** VERB ’loving’
- **amāns** m/f III -> NOUN ‘lover’
- …sed nihil difficile **amanti** puto ‘I believe nothing to be difficult for the dedicated’ Cicero (*Orator ad M. Brutum)*
- And when to consider it an adjective? - > ADJ
- Maria, femina **laundanda**, advenit. ‘Maria, a praiseworthy woman, has arrived.’

https://www.thelatinlibrary.com/101/Participles.pdf

Discussion

- Lemmatization and morphological analysis depends on digitized dictionaries (Lewis and Short, 1879)
- Printed dictionaries rely on providing definitions for humans (Preferably with holistic understanding)
- Treebanks need to be consistent
- Latin dependency treebank relies on rule based tokenization and sentence split algorithms, whose output feeds the COMBO lemmatizer PoS tagger and parser

Ensemble lemmatization with the Classical Language Toolkit (Burns)

- A combination of data-driven and rule based approaches
  - Ensemble lemmatizer
- How to handle unseen vocabulary?
- Especially for Latin: resources are extremely limited for dialect, period and genre
- There is no effective way to combine tools
- Backoff lemmatization reflects the methods of the philologically educated reader of historical languages

https://www.studiesaggilinguistici.it/ssl/article/view/273

Creating a Gold Standard

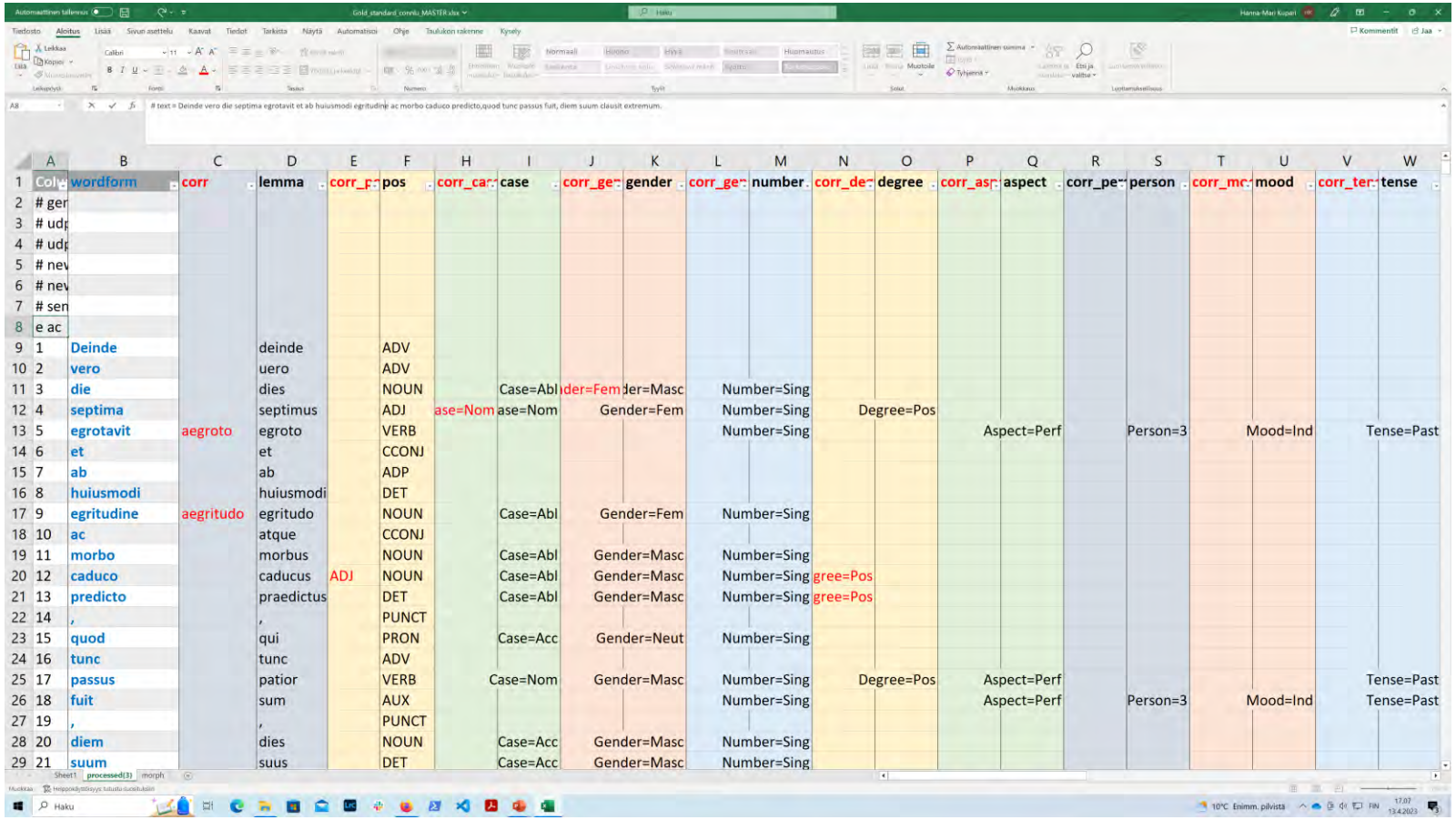
- Circa 1 000 words
- 500 chosen on chance
- 500 on deliberately picked
- Lemmas – easy
- POS – easy
- Morphology – a bit of thinking
- Syntax – guaranteed challenge

A283 # text = Cum autem dictus orator in morte dicti defuncti alias, quam ut premittitur, minime culpabilis fuerit, quia tamen bonarum mentium est culpam timere, ubi culpa minime reperitur, supplicatur etc, quatenus oratorem ipsum a quibuscumque sententiis, censuris et penis, quas propterea incurrit aut incurrere potuisset, homicidii forsan reatu et excessibus huiusmodi hac vice absolvi, demum suffragantibus sibi meritis alioque canonico non obstante super irregularitate per eum dicto modo contracta, quodque premissis non obstantibus in omnibus suis ordinibus susceptis etiam in altaris ministerio ministrare dictamque parrochialem ecclesiam et alia per eum forsan obtenta retinere sibique in futurum canonice conferenda beneficia ecclesiastica cum cura et sine cura compatibilia tamen recipere et similiter retinere libere et licite possit et valeat, cum eodem oratore auctoritate et mandato predictis misericorditer dispensari omnemque etc aboleri non obstantibus constitutionibus et ordina tionibus apostolicis ceterisque in contrarium facientibus quibuscumque mandare dignemini de gratia speciali et expresso.

	Cc	wordform	corr	lemma	corr_	pos	corr_ca	case	corr_g	gender	corr_g	numbe	corr_d	degree	corr_as	aspect	corr_pi	person	corr_m	mood	corr_te	tense	corr_ve	verbfo	corr_v	voice	Column52	Column6
274	58	mandare		mando		VERB										Aspect=Imp						Tense=Pres		VerbForm=Inf		Voice=Act		Aspect=Imp|InflClass=LatA|InflClass[nominal]=Ind|Tense=Pres|VerbForm=Inf|Voice=Act
275	59	misericorditer		misericorditer		ADV								Degree=Pos														Degree=Pos
276	60	dignemini	dignor	digneor		VERB						Number=Plur						Person=2		Mood=Ind		Tense=Pres		VerbForm=Fin		Voice=Pass		Mood=Ind|Number=Plur|Person=2|Tense=Pres|VerbForm=Fin|Voice=Pass
277	61	in		in		ADP																						AdpType=Prep
278	62	forma		forma		NOUN		Case=Abl		Gender=Fem		Number=Sing																Case=Abl|Gender=Fem|InflClass=IndEurA|Number=Sing
279	63	.		.		PUNCT																						_
280																												
281	# nev																											
282	# sen																											
283	# tex																											
284	1	Cum		cum		SCONJ																						PronType=Rel
285	2	autem		autem	PART	CCONJ																						_
286	3	dictus	dictus	dico	DET	VERB		Case=Nom		Gender=Masc		Number=Sing		Degree=Pos		Aspect=Perf						Tense=Past		VerbForm=Part		Voice=Pass		Aspect=Perf|Case=Nom|Degree=Pos|Gender=Masc|InflClass=LatX|InflClass[nominal]=IndEurO|Number=Sing|Tense=Past|
287	4	orator		orator		NOUN		Case=Nom		Gender=Masc		Number=Sing																Case=Nom|Gender=Masc|InflClass=IndEurX|Number=Sing
288	5	in		in		ADP																						AdpType=Prep
289	6	morte		mors		NOUN		Case=Abl		Gender=Fem		Number=Sing																Case=Abl|Gender=Fem|InflClass=IndEurI|Number=Sing
290	7	dicti	dictus	dictum	DET	NOUN		Case=Gen		Gender=Neut		Number=Sing																Case=Gen|Gender=Neut|InflClass=IndEurO|Number=Sing
291	8	defuncti		defungor		VERB										Aspect=Imp						Tense=Pres		VerbForm=Inf		Voice=Pass		Aspect=Imp|InflClass=LatX|InflClass[nominal]=Ind|Tense=Pres|VerbForm=Inf|Voice=Pass
292	9	alias		alias	ADV	DET		Case=Acc		Gender=Fem		Number=Plur																Case=Acc|Gender=Fem|InflClass=LatPron|Number=Plur|PronType=Con
293	10	,		,		PUNCT																						_
294	11	quam	quam	qui	SCONJ	PRON		Case=Acc		Gender=Fem		Number=Sing																Case=Acc|Gender=Fem|InflClass=LatPron|Number=Sing|PronType=Rel
295	12	ut		ut		SCONJ																						ConjType=Cmpr
296	13	premittitur	praemitto	premitto		VERB						Number=Sing				Aspect=Imp		Person=3		Mood=Ind		Tense=Pres		VerbForm=Fin		Voice=Pass		Aspect=Imp|InflClass=LatX|Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin|Voice=Pass
297	14	,		,		PUNCT																						_
298	15	minime		parum		ADV																					superlatiivi?	Degree=Abs
299	16	culpabilis		culpabilis		ADJ		Case=Nom		Gender=Masc		Number=Sing		Degree=Pos														Case=Nom|Degree=Pos|Gender=Masc|InflClass=IndEurI|Number=Sing
300	17	fuerit		sum		AUX						Number=Sing				Aspect=Perf		Person=3		Mood=Sub		Tense=Past		VerbForm=Fin				Aspect=Perf|InflClass=LatAnom|Mood=Sub|Number=Sing|Person=3|Tense=Past|VerbForm=Fin
301	18	,		,		PUNCT																						_
302	19	quia		quia		SCONJ																						PronType=Rel
303	20	tamen		tamen		ADV																						PronType=Dem
304	21	bonarum		bonus		ADJ		Case=Gen		Gender=Fem		Number=Plur																Case=Gen|Gender=Fem|InflClass=IndEurA|Number=Plur
305	22	mentium		mens		NOUN		Case=Gen		Gender=Fem		Number=Plur																Case=Gen|Gender=Fem|InflClass=IndEurI|Number=Plur
306	23	est		sum		AUX						Number=Sing				Aspect=Imp		Person=3		Mood=Ind		Tense=Pres		VerbForm=Fin				Aspect=Imp|InflClass=LatAnom|Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin
307	24	culpam		culpa		NOUN		Case=Acc		Gender=Fem		Number=Sing																Case=Acc|Gender=Fem|InflClass=IndEurA|Number=Sing
308	25	timere		timeo		VERB										Aspect=Imp						Tense=Pres		VerbForm=Inf		Voice=Act		Aspect=Imp|InflClass=LatE|InflClass[nominal]=Ind|Tense=Pres|VerbForm=Inf|Voice=Act
309	26	,		,		PUNCT																						_
310	27	ubi		ubi		SCONJ																						AdvType=Loc|PronType=Rel
311	28	culpa		culpa		NOUN		Case=Nom		Gender=Fem		Number=Sing																Case=Nom|Gender=Fem|InflClass=IndEurA|Number=Sing
312	29	minime		parum		ADV																						Degree=Abs
313	30	reperitur		reperio		VERB						Number=Sing				Aspect=Imp		Person=3		Mood=Ind		Tense=Pres		VerbForm=Fin		Voice=Pass		Aspect=Imp|InflClass=LatI|Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin|Voice=Pass
314	31	,		,		PUNCT																						_
315	32	supplicatur		supplico		VERB						Number=Sing				Aspect=Imp		Person=3		Mood=Ind		Tense=Pres		VerbForm=Fin		Voice=Pass		Aspect=Imp|InflClass=LatA|Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin|Voice=Pass
316	33	etc		etcetera		CCONJ																						_
317	34	,		,		PUNCT																						_
318	35	quatenus		quatenus		SCONJ																						Compound=Yes|ConjType=Cmpr|PronType=Rel
319	36	oratorem		orator		NOUN		Case=Acc		Gender=Masc		Number=Sing																Case=Acc|Gender=Masc|InflClass=IndEurX|Number=Sing
320	37	ipsum		ipse		DET		Case=Acc		Gender=Masc		Number=Sing																Case=Acc|Form=Emp|Gender=Masc|InflClass=LatPron|Number=Sing|PronType=Dem
321	38	a		a		ADP																						AdpType=Prep
322	39	quibuscumque	quicumque	_		DET																						_
323	40	sententiis		sententia		NOUN		Case=Abl		Gender=Fem		Number=Plur																Case=Abl|Gender=Fem|InflClass=IndEurA|Number=Plur
324	41	,		,		PUNCT																						_
325	42	censuris		censura		NOUN		Case=Abl		Gender=Fem		Number=Plur																Case=Abl|Gender=Fem|InflClass=IndEurA|Number=Plur
326	43	et		et		CCONJ																						_
327	44	penis	poena	pena		NOUN		Case=Abl		Gender=Fem		Number=Plur																Case=Abl|Gender=Fem|InflClass=IndEurA|Number=Plur
328	45	,		,		PUNCT																						_
329	46	quae		qui		PRON		Case=Acc		Gender=Fem		Number=Plur																Case=Acc|Gender=Fem|InflClass=LatPron|Number=Plur|PronType=Rel

UDante and UD pipe

- 134/1 251 lemma -> 10 % needed to be corrected
- 63/1 251 POS -> 5 %



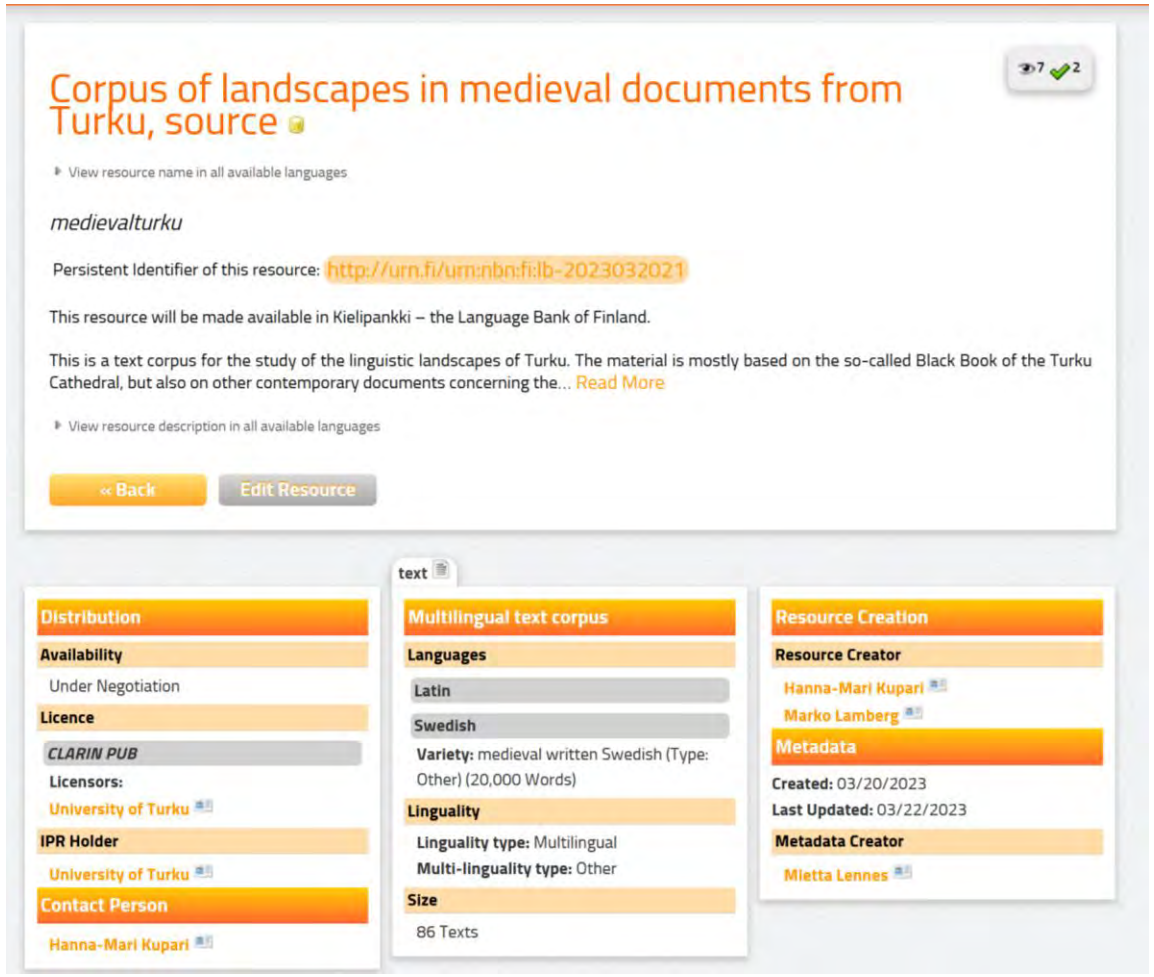
	A	B	C	D	E	F	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Colu	wordform	corr	lemma	corr_p	pos	corr_ca	case	corr_ge	gender	corr_ge	number	corr_de	degree	corr_as	aspect	corr_pe	person	corr_mo	mood	corr_ter	tense
2	# ger																					
3	# udp																					
4	# udp																					
5	# nev																					
6	# nev																					
7	# sen																					
8	e ac																					
9	1	Deinde		deinde		ADV																
10	2	vero		uero		ADV																
11	3	die		dies		NOUN		Case=Abl	der=Fem	der=Masc		Number=Sing										
12	4	septima		septimus		ADJ	ase=Nom	ase=Nom		Gender=Fem		Number=Sing		Degree=Pos								
13	5	egrotavit	aegroto	egroto		VERB						Number=Sing				Aspect=Perf		Person=3		Mood=Ind		Tense=Past
14	6	et		et		CCONJ																
15	7	ab		ab		ADP																
16	8	huiusmodi		huiusmodi		DET																
17	9	egritudine	aegritudo	egritudo		NOUN		Case=Abl		Gender=Fem		Number=Sing										
18	10	ac		atque		CCONJ																
19	11	morbo		morbus		NOUN		Case=Abl		Gender=Masc		Number=Sing										
20	12	caduco		caducus	ADJ	NOUN		Case=Abl		Gender=Masc		Number=Sing	gree=Pos									
21	13	predicto		praedictus		DET		Case=Abl		Gender=Masc		Number=Sing	gree=Pos									
22	14	,		,		PUNCT																
23	15	quod		qui		PRON		Case=Acc		Gender=Neut		Number=Sing										
24	16	tunc		tunc		ADV																
25	17	passus		patior		VERB		Case=Nom		Gender=Masc		Number=Sing		Degree=Pos		Aspect=Perf						Tense=Past
26	18	fuit		sum		AUX						Number=Sing				Aspect=Perf		Person=3		Mood=Ind		Tense=Past
27	19	,		,		PUNCT																
28	20	diem		dies		NOUN		Case=Acc		Gender=Masc		Number=Sing										
29	21	suum		suus		DET		Case=Acc		Gender=Masc		Number=Sing										

Key take home messages:

- We want to ensure as much consistency as possible *id est*:
- human use allows for a number of irregularities that impinge on computational resources derived from them
- There are no guidelines for any of the UD treebanks in Latin – a new global effort – UD4HL

Future outlooks

- MDA Multi-Dimensional Analysis -> registers
- Social history and everyday life in medieval Turku
- Linguistics landscaped data on Kielipankki and opportunities for new research topics



Thanks

- hmknie@utu.fi

- https://github.com/HannaKoo
- Twitter: @kuparimari
- Instagram: hannasuperturkulainen

- Kalmistopiiri