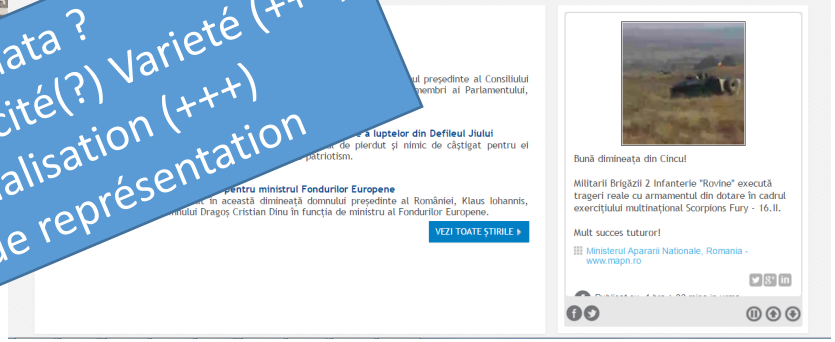
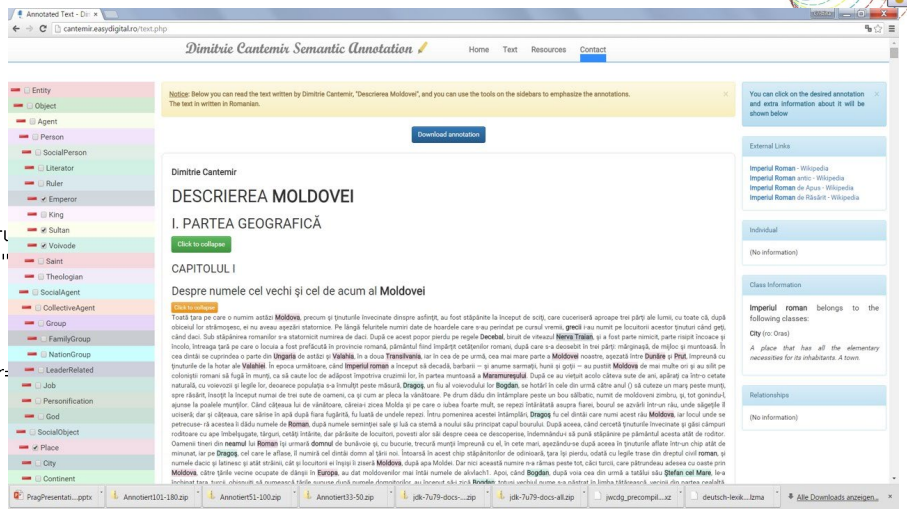
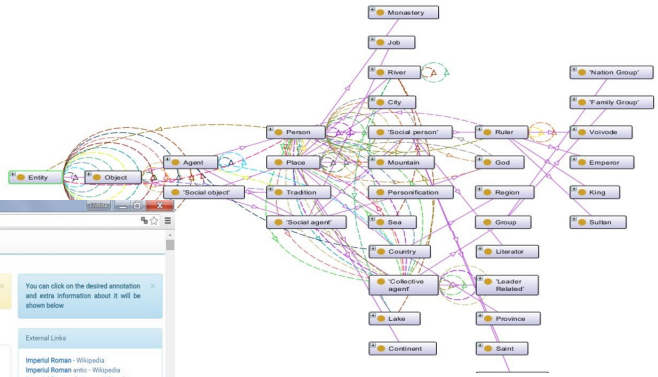


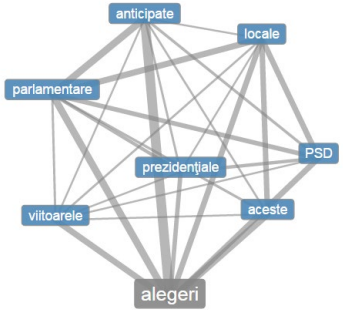
Big data ?
Volume (+) Velocité(?) Varieté (+++)
Contextualisation (+++)
Modèle de représentation

?

```
<phrase>
      <ne> <person personT="f“ first=„Dragosch“
            relationT=„son“ refT= „w123“/> </ne>
     <phraseParts constituentT="NP" mult="false">
    <tok idT="" langT="ro" styleTranscr="de">
      <pos posT="N"/>
      <string>  Dragosch</string>
     </tok></phraseParts>
</phrase>
 , ein Sohn ihres
 <phrase>
      <ne> <person personT="tl"/></ne>
     <phraseParts constituentT="NP" mult="tru
      <tok idT="" langT="de" styleTranscr="de"
      <pos posT="N"/>
      <string> Königs  </string>
     </tok>
      <tok idT=“w123" langT="ro" styleTranscr=
      <pos posT="N"/>
      <string> Bogdan</string>
     </tok> </phraseParts>
</phrase>
```







Annotation and Analysis historical texts -1-

- Heterogenous data (several historical stages of language)
- Big variations in morphology, syntax. (long time no standardization)
- Linguistic features (thus annotation-features) change over time – diachronic corpus
- Semantics and ontological space changed over time
- Not enough material for training

Annotation and Analysis historical texts -2-

- Multilinguality:
  - Across documents
  - Within one document: (Latin, Greek or other language)
- Not all necessary material is accessible and processable in digital form
- Great challenge with NEs (places, names, date)
- How to treat the vagueness-feature of the language:
  - For the moment it is ignored
  - Implication: unreliables (up to wrong) analysis .
- Multilevel annotation & discontinous phenomena
- Complex – representation -> big data ?

Workflow

Manual metadata annotation

Manual annotation image - metadata

Scan

Images

Manual image-text links

OCR + Manual correction

Manual transcription

Text

Manual layout annotation (TEI)

Manual content annotation

Automatic content annotation + Manual correction

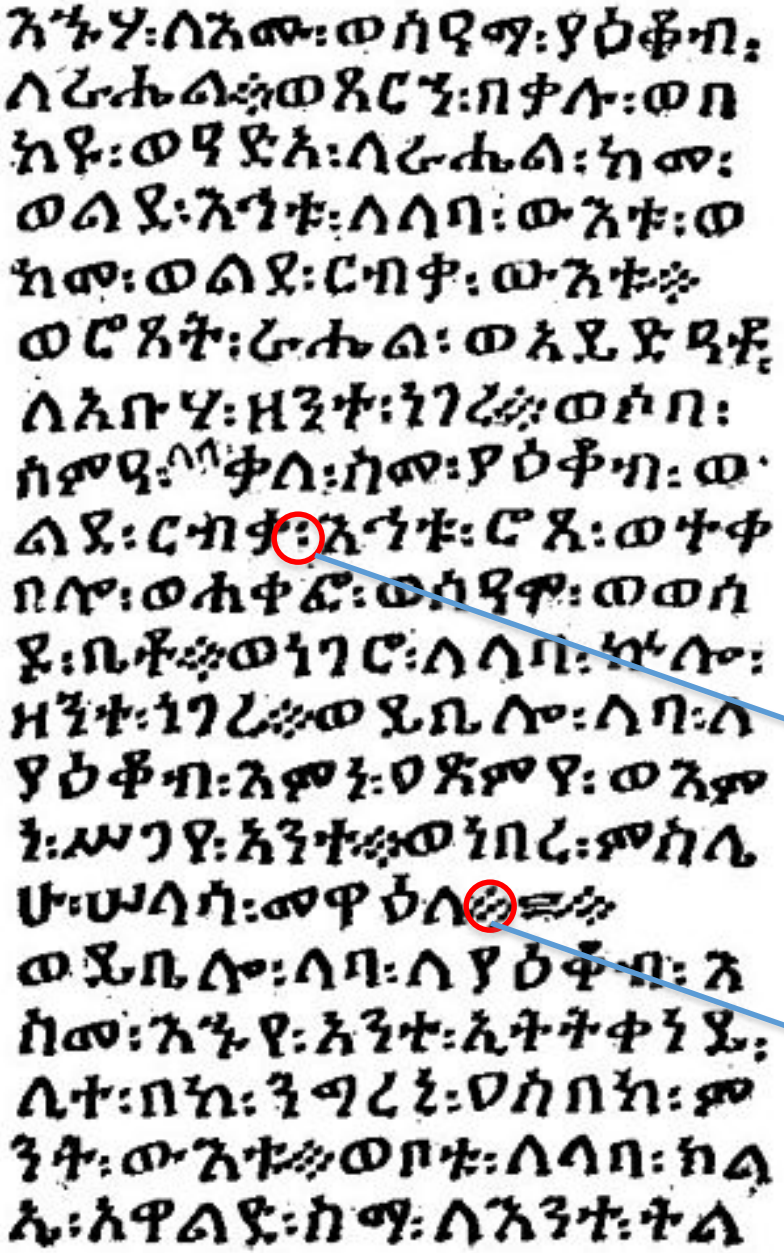
Browsing | Vizualization | Quantitative Analysis | Qualitative Interpretation ? | Search

TraCES - From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages

- ERC Advanced Grant 2014–2019

- **aim**: reliable and extensive linguistic **data** based on annotated texts for a **diachronic analysis** of Gəʿz (lexicography, morphology and style)
- **corpus**: several texts belonging to **different periods and genres** of Ethiopic literature (**text-critical editions)**

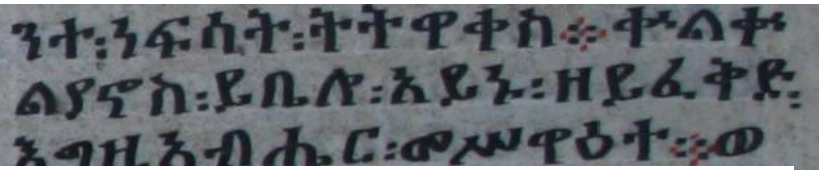
Ge‘ez Text – Example from Genesis 29.11



Word Sepparator

Section mark

The Classical Ethiopic (Gəʿəz)



Schrifttabelle

	a	u	i	ā	e	ə/ø	o
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
ḥ	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
ś	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
q	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
t	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ
ẖ	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ
n	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ
ʾ	አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ
k	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ
w	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ
c	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ
z	ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ
y	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ
d	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ
g	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ
ṭ	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ
p	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ
ṣ	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ
ḍ	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ
f	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ
p	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ
q^w	ቈ		ቊ	ቋ	ቌ	ቍ	
$ẖ^w$	ኈ		ኊ	ኋ	ኌ	ኍ	
k^w	ኰ		ኲ	ኳ	ኴ	ኵ	
g^w	ጐ		ጒ	ጓ	ጔ	ጕ	

- Each symbol represent a syllable (Fidäl)
- Wowels are represented.
- LR-Writting
- Dedicated symbols for Digits
  - For a number representation and value fora number are different
  - ፭ወ፭ Transcription 1 wa 5
    Value 6 (1+5)
- Dedicated interpunction፡ (Space) ። (. or ,) ፣ (,) ፤ (:) ፥ (;) ፧ (?) ፨ (Paragraph)
- Laryngals and Sibilants are changeblae from the phonemisc and graphemic point of view.
  ሃሌሉያ (hāleluyā)
  ሀሌሉያ ሐሌሉያ ኀሌሉያ ሓሌሉያ ኃሌሉያ
- Nonconcatenative Morphology

√qds × ʾasta1a22a3a → ʾastaqaddasa

Dillman's Lexicon - Example

አስበንያ ፡ vid. **አሶንያ** ፡

{DiL.0749}

እሳት ፡ subst., m. et fem. [[…]] *ignis*. Gen. 22,6; Gen. 22,7; Num. 21,28; Num. 31,23; Ps. 20,9; Ps. 38,4; Ps. 57,8; Ps. 77,25; Ps. 77,29; Jes. 26,11; Jes 33,14; Sir. 16,6; Sir. 27,5; **እሳት ፡ በላዒ** ፡ Sir. 30,30; Kuf. 36; Job 15,34; **በልዐቶ ፡ እሳት** ፡ Job 22,20; **እሳት ፡ ዘይነድድ** ፡ Job 31,12; Kuf. 1 **እሳት ፡ እንተ ፡ ትነድድ** ፡ Sir. 3,30; **ለእሳት ፡ ይቀድም ፡ ይጠይስ ፡ ተነ ፡ ጢሱ** ፡ Sir. 22,24; **በአምጣዙ ፡ ዕው ፡ ለእሳት ፡ የዐቢ ፡ ነዱ** ፡ Sir. 28,10; **ያነድዳ ፡ ለእሳት** ፡ Sir. 28,11; **ዘኢይጠፍእ ፡ እሳት** ፡ Sir. 28,23; Jes. 66,24; Matth. 3,10; Matth. 3,12; Matth. 7,19; Luc. 12,49; Luc. 22,55; **ነደ ፡ እሳት** ፡ Hebr. 1,7; **እሳት ንስቲት ፡ ጥቀ** ፡ Jac. 3,5; **እሳት ፡ እሳት** ፡ Hen. 14,22; ignis Dei, a coelo delabens, Gen. 19,24; Ex. 9,23; Lev. 10,2; Job 1,16; Luc. 9,54; ignis irae Dei Deut. 4,24; Deut. 32,22; Ps. 17,10; Jes. 66,16; **ገሃነም ፡ እሳት** ፡ Matth. 5,22; Matth. 18,9; Marc.9,47; **ዐዘቅተ ፡ እሳት** ፡ Apoc. 20,10; **እሳት ፡ ዘለዓለም** ፡ Matth. 18,8; Matth. 25,41; **እሳት ፡ መለኮት** ፡ Lit. 170,1. a) in Gen. positum vel st. constructo subjunctum: *igneus*, ut: **ሰረገላ ፡ ዘእሳት** ፡ 4 Reg. 2,11; **ሰይፈ ፡ እሳት** ፡ Gen. 3,24; **አእባነ ፡ እሳት ፡ ቀይሕ** ፡ Hez. 28,14; vers. nov.; Hen. 17,3-5. b) *ignis* i. e. *incendium*. **ፈቀዱ ፡ ያድኅኑ ፡ አግልኩቲሆሙ ፡ እምእክለ ፡ እሳት** ፡ Kuf. 12. c) de ignis cultu divino (Persarum) vid. Clem. f. 32; **እሙ-እቱ ፡ ጊዜ ፡ ሰገዱ ፡ ሰብአ ፡ ሰገል ፡ ለእሳት** ፡ Clem. f. 33.

እሳታዊ ፡ adj. rel. *igneus*, Phlx. 186; Jsp. p. 287; **ሳው-ዳ ፡ ዘው-እቱ ፡ ጠባይዕ ፡ እሳታዊ** ፡ Fal. f. 82; **እሳታው-ያን** ፡ *igneri* i. e. angeli Lit.170,1; Gad. Kar.

አስታር ፡ n. peregr. sc. stater (alias **ሰጥጢራ** ፡ q.v.), arab. إستار. Voc. Ae.: **አስታር ፡ ዘ ፡ ፲ወ፪ወቁት ፡ ወቦ ፡ እለ ፡ ይ ፡ መጠነ ፡ መክሊት** ። – At lectio **እስመ ፡ ኢይሁቦ ፡ አስታረ** ፡ (v. **አስታር** ፡, v. **ዐሥራተ** ፡) ሞገስ ፡ οὐ γὰρ ἐδόθη αὐτῷ παρὰ Κυρίου χάρις Sir. 37,21 non potest non esse corrupta (pro **እግዚ"** ፡ ሞገሰ ፡?).

ኢስታአስ ፡ et **ኢስታምስ** ፡ n. peregr. sc. […], *unus Deus*, Sx. Hed. 19; Mij. 30.

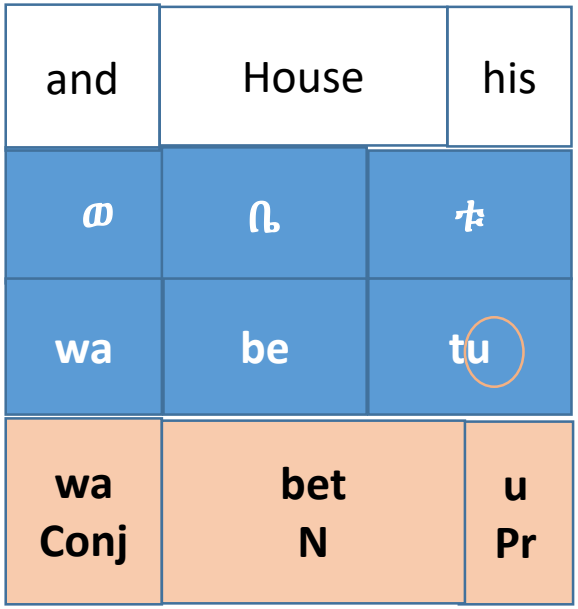
አስትዓ ፡ *munus*, *donum* vid. sub **ሰትዐ** ፡ c. 364.

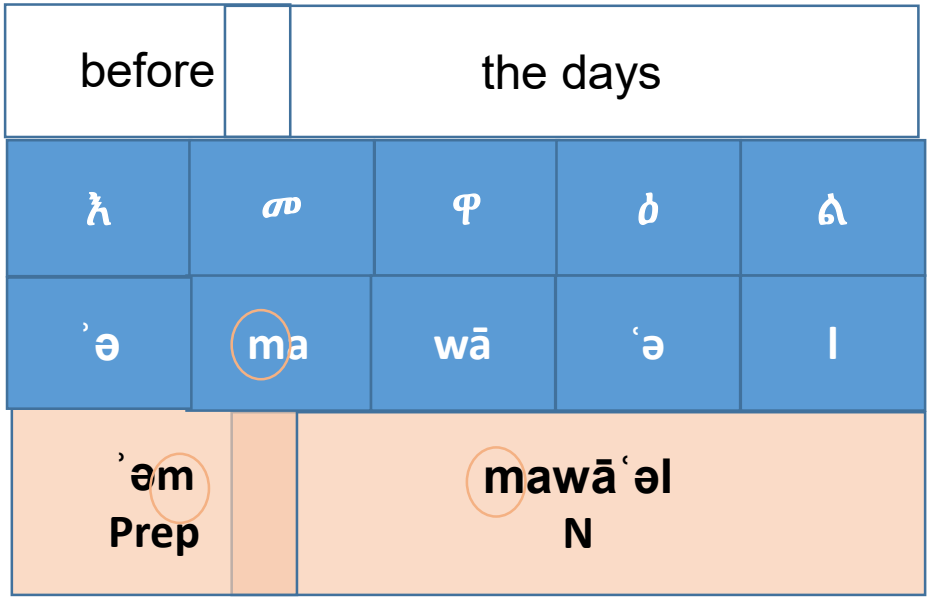
አሶንያ ፡ n. peregr. et corruptum *lunae*, Hen. 78,2; *Voc. Ae.* (in quo et forma **አስበንያ** ፡ et

Language particularities

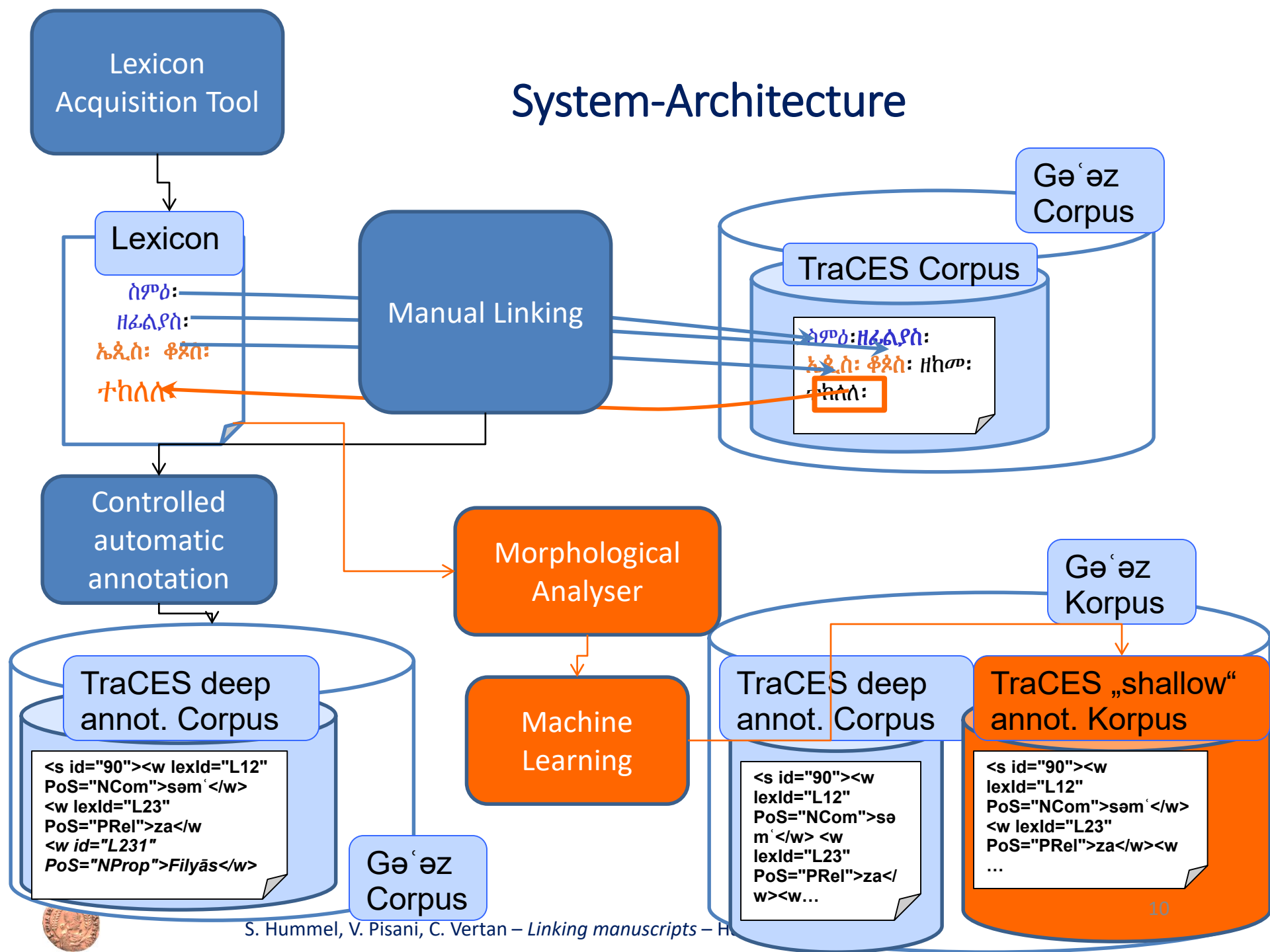
- Vowels can be themelves part-of-speech

Letter compression in Originalscript (Fidäl), but not in the transcription





System-Architecture



User requests and Challenges for the annotation

- Automatic transcription
- Synchronisation between original and transcription during the annotation
- CoOrrection of the text during the annotation (while kkeping the annotation)
- Controlled automatic Annotation:
  - Tokenising
  - Change of the text
  - Linguistic Annotation
  - Sentence Annotation
  - NE-Annotation
- Possibility of very flexible text divisions (not necessary hierarchical)
- Multilevel annotation (flexible change of annotation level)
- Approx. 30 linguistic categories (PoS)(e.g. Number following 3 categories : Nature, Pattern and Syntax)
- User-friendly GUI
- Possibility of adapting the system for other scripts and transcriptions

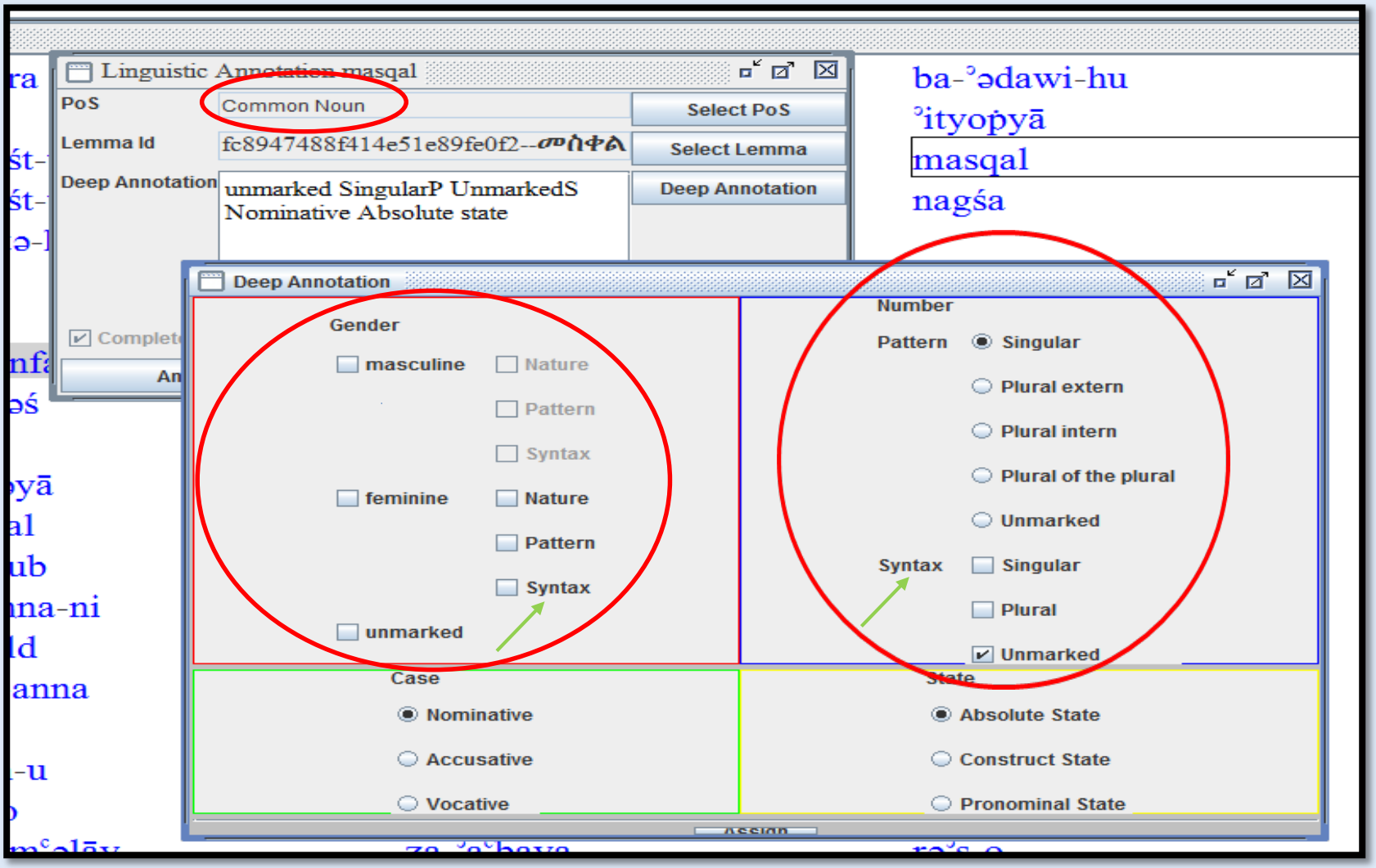
GeTa Tool

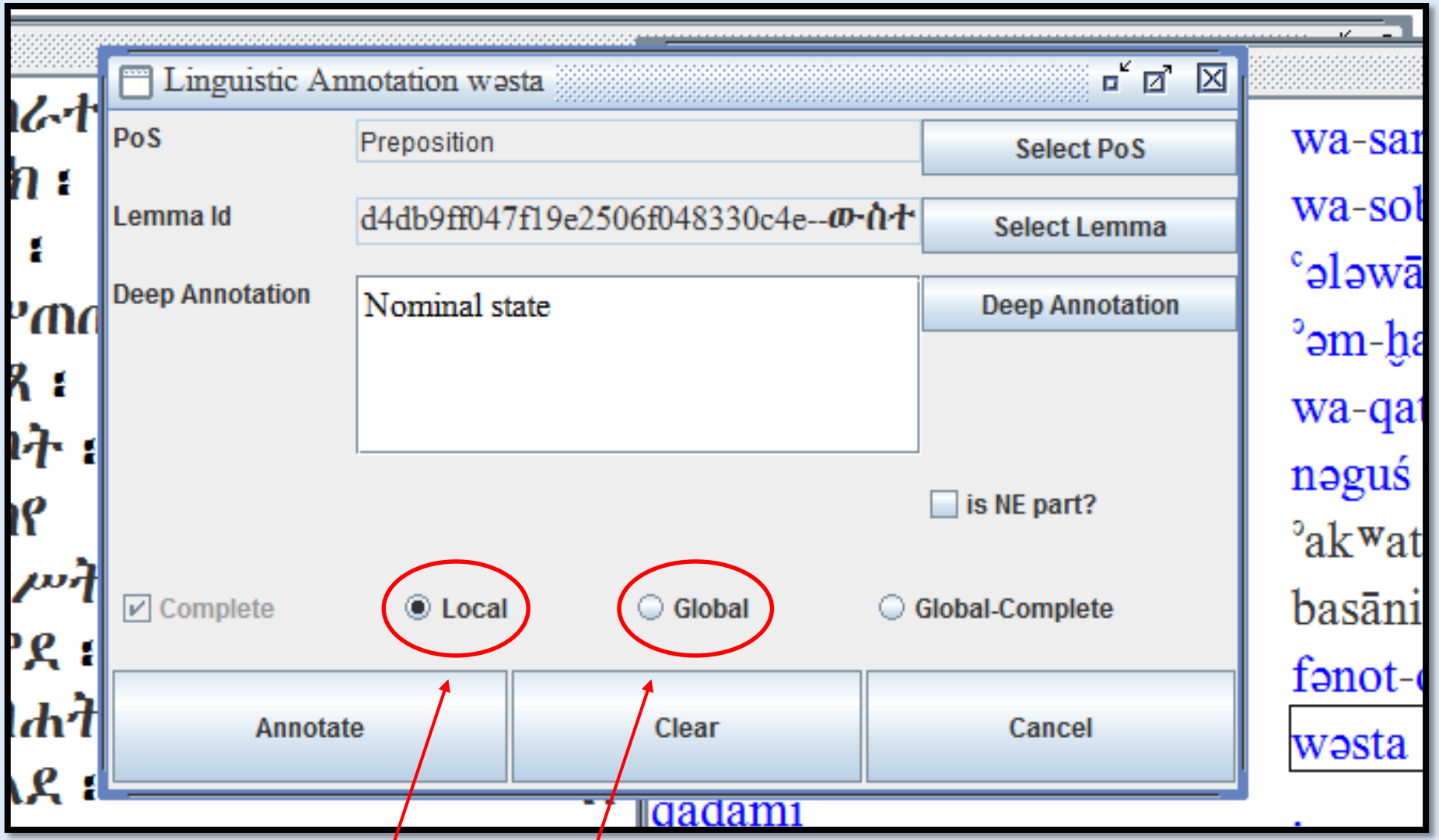
Features

- Easy to use GUI
- Automatic initial transcription (vocalized or unvocalized)
- Synchronisation between original and transcription
- Controlled changes on text while annotating possible
- Controlled semi-automatic:
  - tokenization,
  - change of the transcripts text,
  - deep linguistic annotation + link to lexicon
  - Name Entity annotation linked with the authority DB.
- Automatic „sentence“ recognition
- Visualisation of data model
- Visualisation of annotation progress
- Can read additionally Classical Ethiopic inscriptions written with South Arabic script

Software development
- Client-Application
- Open source; Java
- Data-encoding:JSON

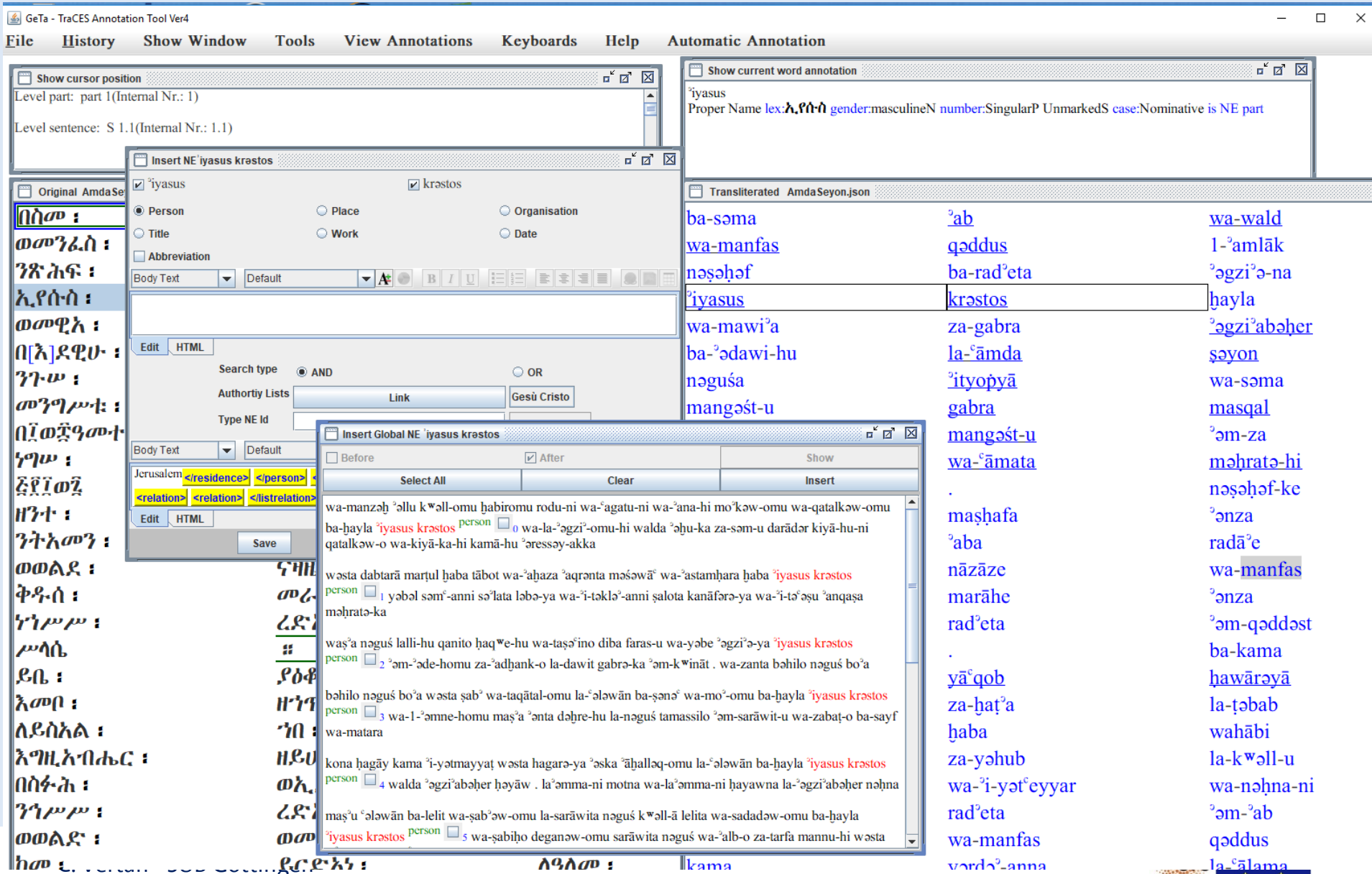
Annotation Common Noun in GeTa



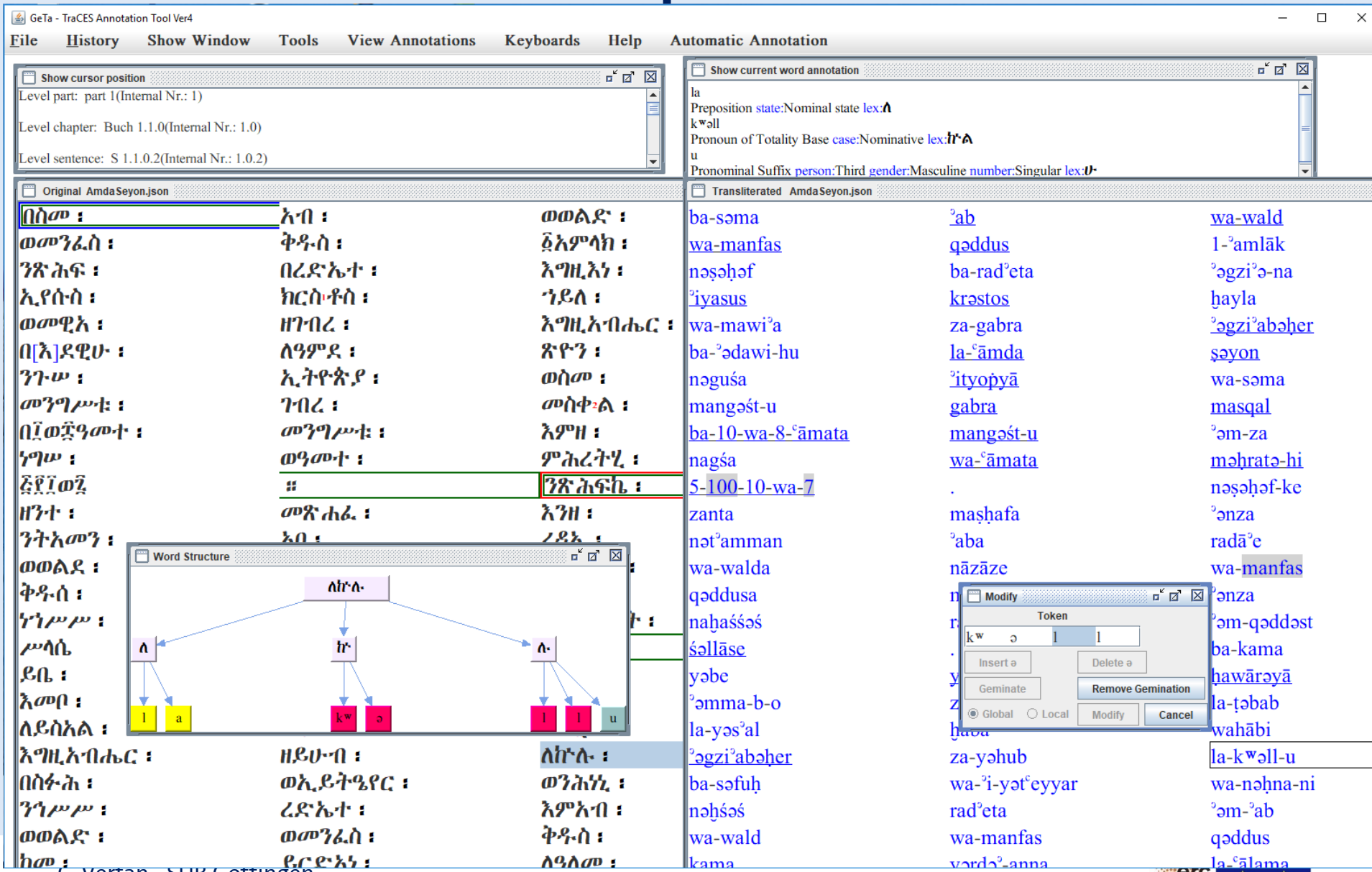


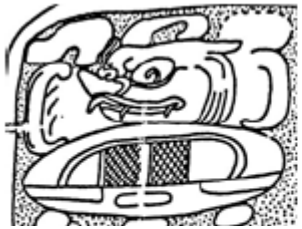
From GeTa tool

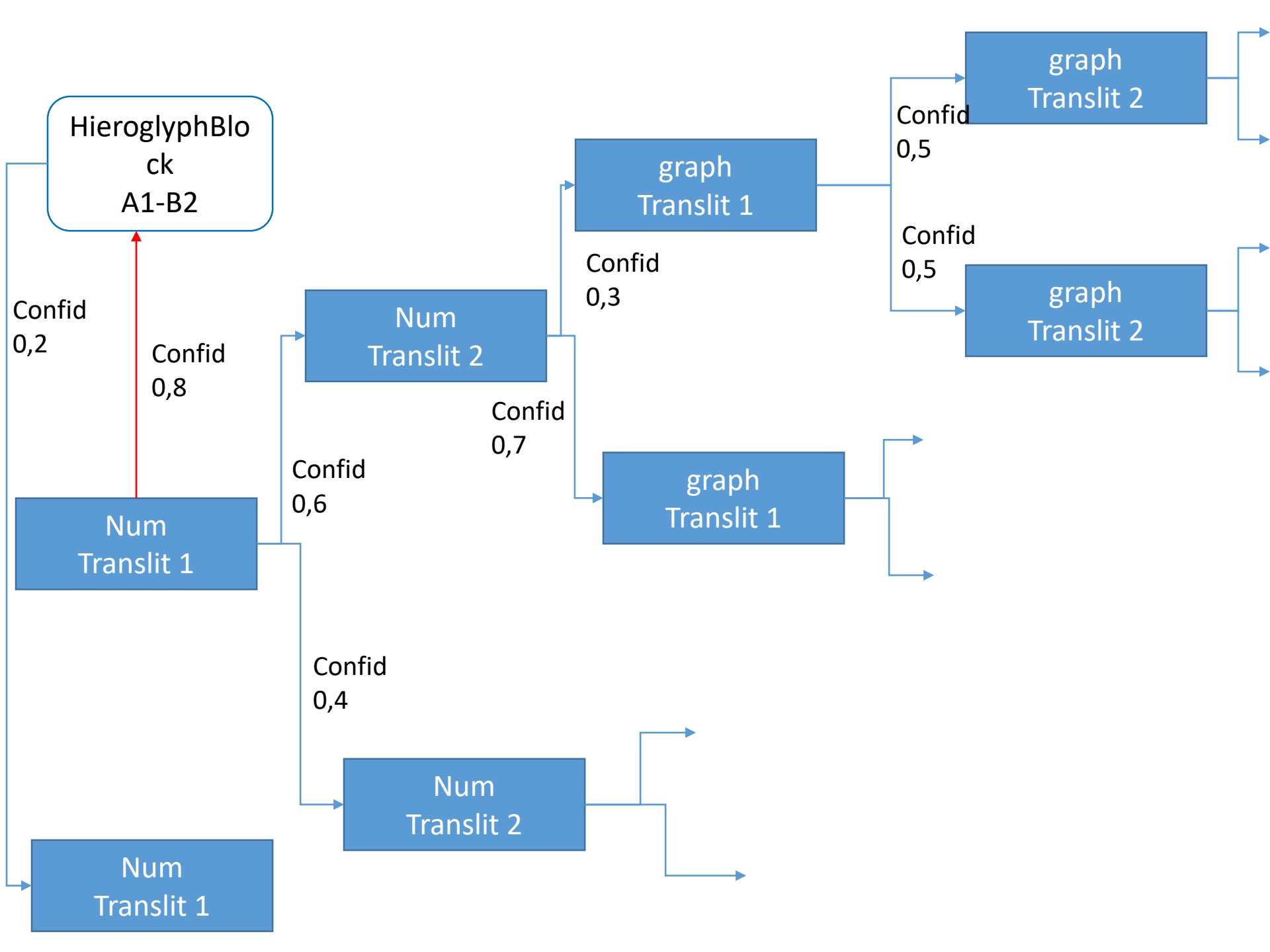
GeTa NE-Annotation



GeTa-operations



Textbeispiel						Anmerkungen
						Zeichenkatalog abfragen und muss Relation Graph → Zeichen heraussuchen
Alternative 2	124:[28°738]:548:74	–	–	–	[5:201:23].683	
Alternative 3	124:[28°738]:548:74	–	–	–		
Analyse-ebene	**graphemische Transliteration 1**					semi-automatische Einsetzung der Lautwerte entsprechend der Werte im Zeichenkatalog, Auswahl entsprechend der hinterlegten Zeichenfunktionen, sofern größer 1
Alternative 1	**tzi:ka°XOK:HAB:ma**	**9.PIK**	**SIH:ya.ja**	**SAK.ma:su**	**5:TZ'AM:na.ja**	- mit Information zur Sicherheit (Konfidenz-Angabe)
Alternative 2	**tzi:ka°XOK:HAB:ma**	-	-	-	-	
Alternative 3	**tzi:ka°XOK:HAB:ma**	-	-	-	-	
Analyse-ebene	**graphemische Transliteration 2**					Traditionelle Transliteration in der Fachwissenschaft und korrekte Lesereihenfolge.
Alternative 1	**tzi-ka XOK HAB-ma**	**9 PIK**	**SIH-ya-ja**	**SAK ma-su**	**5-TZ'AM-na-ja**	Transliterationen werden in Fettschrift angegeben. (Diese Analyseebene ist relevant für Visual Library. Export in METS/MODS ist nicht nötig. Muss per OAI PMH ausgeliefert werden können.)
Alternative 2	**TZIK XOK HAB**	–	–	–	**5-201-na-ja**	TZIK ist eine angenommene Lesung, weil wir davon ausgehen, dass 124 und 28 ein komplexes


HieroglyphBlock
A1-B2
Confid
0,2
Confid
0,8
Num
Translit 1
Num
Translit 1
Confid
0,6
Num
Translit 2
Confid
0,3
graph
Translit 1
Confid
0,5
graph
Translit 2
Confid
0,7
graph
Translit 1
Confid
0,5
graph
Translit 2
Confid
0,4
Num
Translit 2

Dimitrie Cantemir (1673 -1723)

- Prince of Moldavia (historical province) as well as „universal“ humanist (linguist, ethnographer, musicologist, historian, writer)
- As member of the Royal Academy in Berlin and at the request of this institution wrote two works :
  - Description of his own country („Descriptio Moldaviae“)
  - History of ottoman empire (History of Growth and Decay of Ottoman Empire)
- Original material written in Latin; Both originals were lost already by the end of 18th century
- Several copies were used as basis for translations into German, English, French, Russian and later in Romanian
- Sometimes the translation relies on other translation (e.g. first Romanian translation of “Descriptio Moldaviae” was done after the German version from 1774.

These translations used as reference information about the Ottoman Empire and Romanian provinces until the middle of 19th century, i.e. they give an idea about the reception about this part of the world in Western Europe.

„Nu îndrăznim să spunem ce e adevărat şi ce e fals într-o asemenea întunecime a istoriei. “

(Dimitrie Cantemir, Descrierea stării Moldaviei in vechime si azi, translation Ioan Costa 2017)

„I do not dare to decide what is the truth about this matter, given the high darkness of this story“

HerCoRe – Hermeneutic and Computer based Analysis of Reliability, Consistency and Vagueness in historical texts

- Illustrated through two main works of Dimitrie Cantemir-

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

UNIVERSITATEA DIN BUCUREȘTI
VIRTUTE ET SAPIENTIA

Funded by

VolkswagenStiftung

April 2017 –March 2020

„Mixed Methods in Humanities“

Combine hermeneutic approaches and methods from computer science for investigating reliability and consistency of original text from 18th century as well as their translations

H

Compare for the first time “original” with translations done in the 18th- 19th century

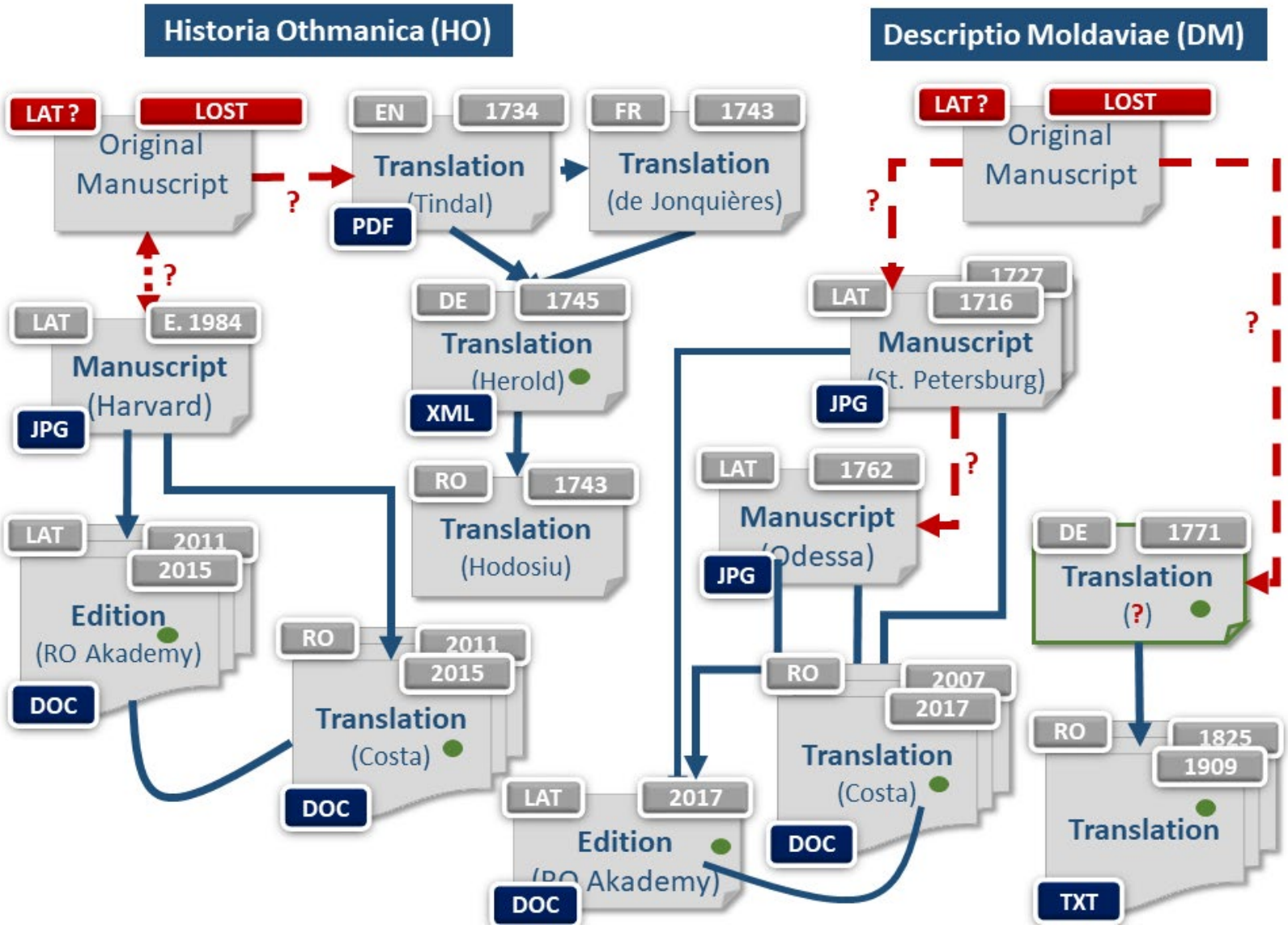
(In)Validate assumptions about source quotations in original text

CS

Demonstrate how to include vagueness and imprecision in annotations and interpretations engines

Progress work in automatic recognition of vague expressions

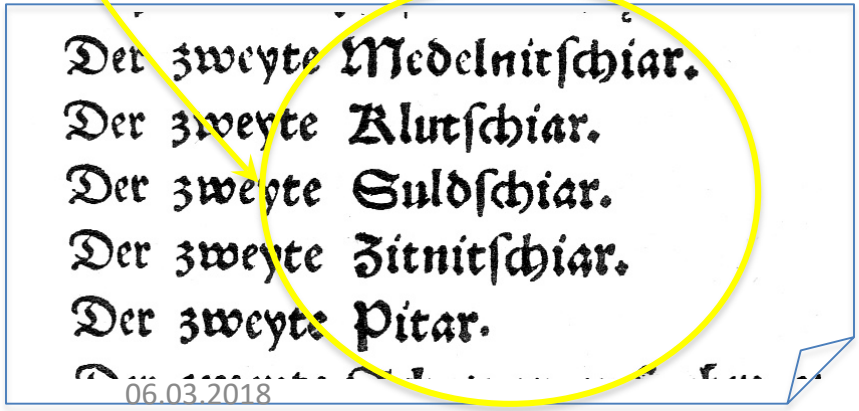
Manuscripts, editions, translations- unclear tradition

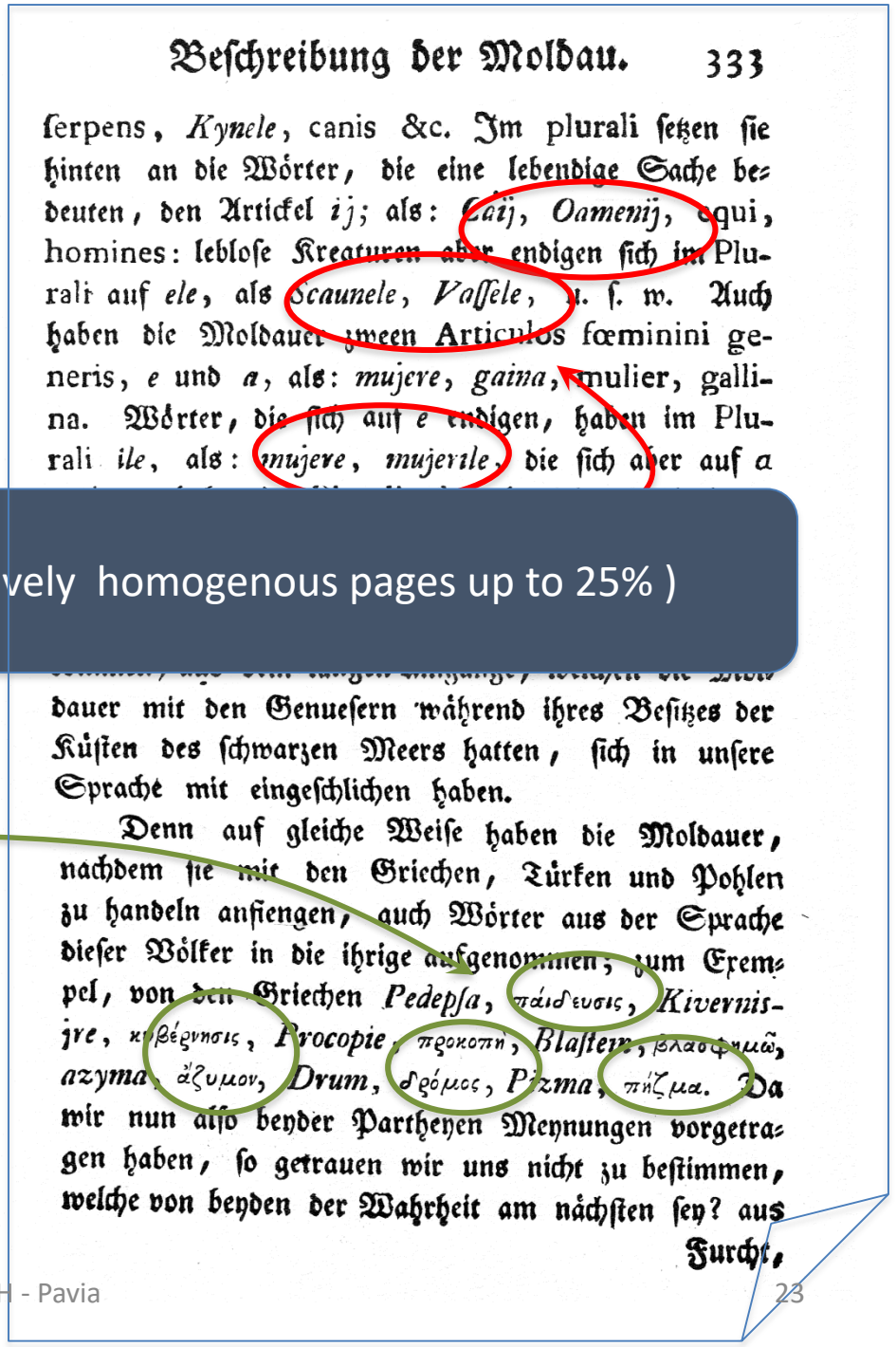


Corpus creation – challenges

- Surface form – level
  - German texts are in black-letter typeface
  - Mixed typefaces
  - Mixed scripts

Higher error rate in OCR (even on relatively homogenous pages up to 25%)


Der zweyte Medelnitschiar.
Der zweyte Klutschiar.
Der zweyte Suldschiar.
Der zweyte Zitnitschiar.
Der zweyte Pitar.


Beschreibung der Moldau. 333

ſerpens, *Kynele*, canis &c. Im plurali ſetzen ſie hinten an die Wörter, die eine lebendige Sache bedeuten, den Artickel *ij*; als: *Cäij*, *Oamenij*, equi, homines: lebloſe Kreaturen aber endigen ſich im Plurali auf *ele*, als *Scaunele*, *Vaſſele*, u. ſ. w. Auch haben die Moldauer zween Articulos fœminini generis, *e* und *a*, als: *mujere*, *gaina*, mulier, gallina. Wörter, die ſich auf *e* endigen, haben im Plurali *ile*, als: *mujere*, *mujerile*, die ſich aber auf *a*

dauer mit den Genueſern während ihres Beſitzes der Küſten des ſchwarzen Meers hatten, ſich in unſere Sprache mit eingeſchlichen haben.

Denn auf gleiche Weiſe haben die Moldauer, nachdem ſie mit den Griechen, Türken und Pohlen zu handeln anfiengen, auch Wörter aus der Sprache dieſer Völker in die ihrige aufgenommen; zum Exempel, von den Griechen *Pedepſa*, παίδευσις, *Kivernisire*, κυβέρνησις, *Procopie*, προκοπή, *Blaſtem*, βλασφημῶ, *azyma*, ἄζυμον, *Drum*, δρόμος, *Pizma*, πίζμα. Da wir nun alſo beyder Partheyen Meynungen vorgetragen haben, ſo getrauen wir uns nicht zu beſtimmen, welche von beyden der Wahrheit am nächſten ſey? aus

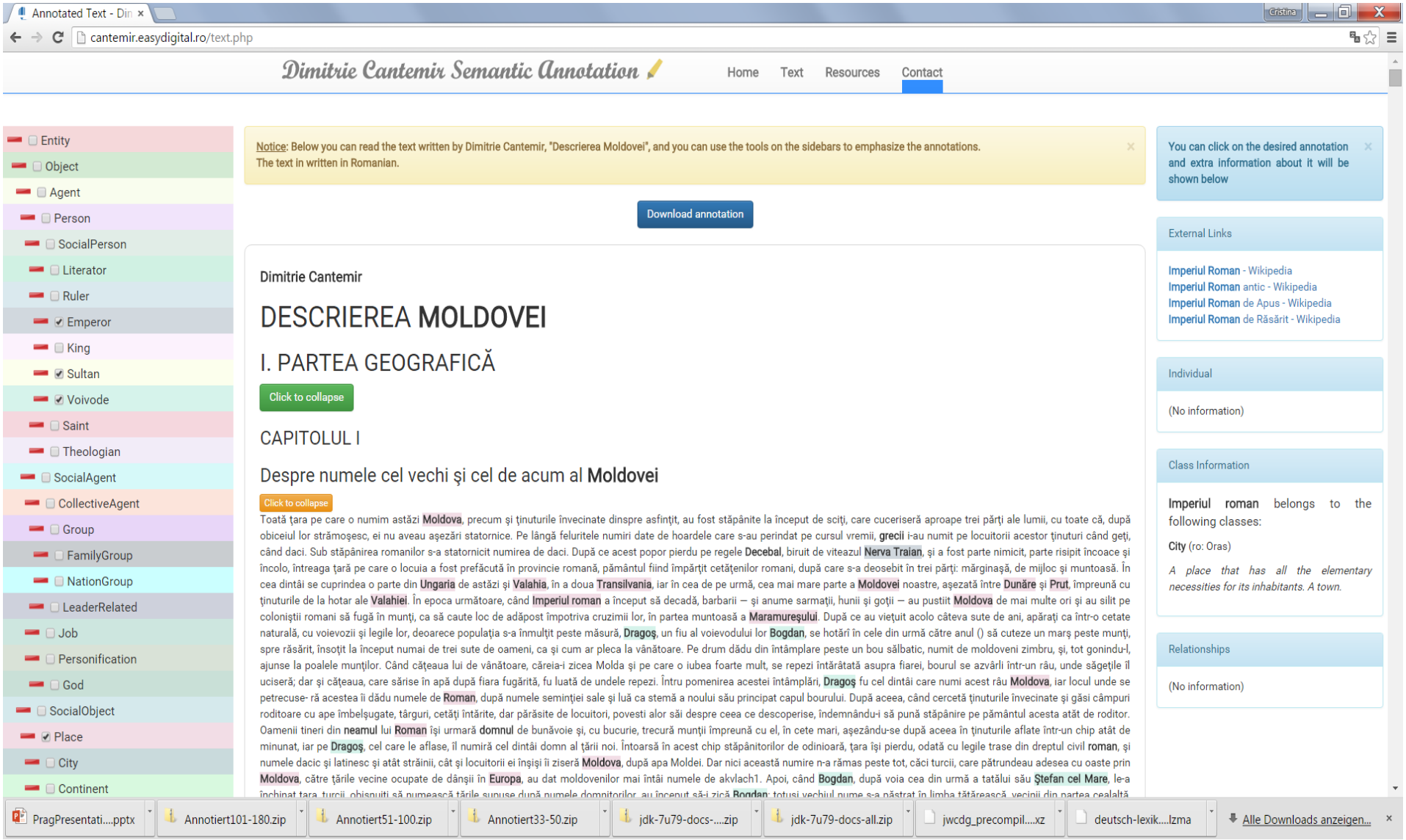
Furcht,

Domnul cel dintâi carele după năvălirea lui Batie, a agonisit iarăși strălucirea cea mai dainte a Moldovei a fost:
1. Dragoș și măcar că hronografiile noastre nu arată pentru știința neamului său, dar la noi se zice necontenit, că a fost din neamul cel vechiu al crailor Moldovinești, și a avut tată pe Bogdan fiul lui Ioan, dela carele toți Domnii obișnuesc a-și pune la iscălitură numele Ioan.
Și cuvântul acesta este mai ușor de a se adeveri și pentru aceasta, căci cu greu este de a se crede, că altul din neam mai prost, ar fi putut cu o tovărășie așa mare să meargă la vânat, carele a dat prilej la descoperirea Moldovei și ar fi putut îndemna pe ceilalți patrioți ai săi, ca să vie după dânsul.

Dimitrie Cantemir, "Descrierea Moldovei", Wikisource
(probably text of the first translation in 19th century after the German edition

Domnul cel dintâi carele după năvălirea lui Batie, a agonisit iarăși strălucirea cea mai dainte a Moldovei a fost:
1. Dragoș și măcar că hronografiile noastre nu arată pentru știința neamului său, dar la noi se zice necontenit, că a fost din neamul cel vechiu al crailor Moldovinești, și a avut tată pe Bogdan fiul lui Ioan, dela carele toți Domnii obișnuesc a-și pune la iscălitură numele Ioan. Și cuvântul acesta este mai ușor de a se adeveri și pentru aceasta, căci cu greu este de a se crede, că altul din neam mai prost, ar fi putut cu o tovărășie așa mare să meargă la vânat, carele a dat prilej la descoperirea Moldovei și ar fi putut îndemna pe ceilalți patrioți ai săi, ca să vie după dânsul.

Dragos	belongs	moldavian kings
Dragos	son_of	Bogdan
Bogdan	son_of	Johannis
Dragos	has_additional_name	Johannis
Bogdan	has_additional_name	Johannis
Drgaos	discovered	Moldau
Dragos	has_acitivity	hunting
Dragos	has_activity	development
Development	takes_place after	Batia invasion


Annotated Text - Din
cantemir.easydigital.ro/text.php
Dimitrie Cantemir Semantic Annotation
Home
Text
Resources
Contact
Entity
Object
Agent
Person
SocialPerson
Literator
Ruler
Emperor
King
Sultan
Voivode
Saint
Theologian
SocialAgent
CollectiveAgent
Group
FamilyGroup
NationGroup
LeaderRelated
Job
Personification
God
SocialObject
Place
City
Continent
Notice: Below you can read the text written by Dimitrie Cantemir, "Descrierea Moldovei", and you can use the tools on the sidebars to emphasize the annotations.
The text in written in Romanian.
Download annotation
Dimitrie Cantemir
DESCRIEREA MOLDOVEI
I. PARTEA GEOGRAFICĂ
Click to collapse
CAPITOLUL I
Despre numele cel vechi şi cel de acum al Moldovei
Click to collapse
Toată ţara pe care o numim astăzi Moldova, precum şi ţinuturile învecinate dinspre asfinţit, au fost stăpânite la început de sciţi, care cuceriseră aproape trei părţi ale lumii, cu toate că, după obiceiul lor strămoşesc, ei nu aveau aşezări statornice. Pe lângă feluritele numiri date de hoardele care s-au perindat pe cursul vremii, grecii i-au numit pe locuitorii acestor ţinuturi când geţi, când daci. Sub stăpânirea romanilor s-a statornicit numirea de daci. După ce acest popor pierdu pe regele Decebal, biruit de viteazul Nerva Traian, şi a fost parte nimicit, parte risipit încoace şi încolo, întreaga ţară pe care o locuia a fost prefăcută în provincie romană, pământul fiind împărţit cetăţenilor romani, după care s-a deosebit în trei părţi: mărginaşă, de mijloc şi muntoasă. În cea dintâi se cuprindea o parte din Ungaria de astăzi şi Valahia, în a doua Transilvania, iar în cea de pe urmă, cea mai mare parte a Moldovei noastre, aşezată între Dunăre şi Prut, împreună cu ţinuturile de la hotar ale Valahiei. În epoca următoare, când Imperiul roman a început să decadă, barbarii — şi anume sarmaţii, hunii şi goţii — au pustiit Moldova de mai multe ori şi au silit pe coloniştii romani să fugă în munţi, ca să caute loc de adăpost împotriva cruzimii lor, în partea muntoasă a Maramureşului. După ce au vieţuit acolo câteva sute de ani, apăraţi ca într-o cetate naturală, cu voievozii şi legile lor, deoarece populaţia s-a înmulţit peste măsură, Dragoş, un fiu al voievodului lor Bogdan, se hotărî în cele din urmă către anul () să cuteze un marş peste munţi, spre răsărit, însoţit la început numai de trei sute de oameni, ca şi cum ar pleca la vânătoare. Pe drum dădu din întâmplare peste un bou sălbatic, numit de moldoveni zimbru, şi, tot gonindu-l, ajunse la poalele munţilor. Când câţeaua lui de vânătoare, căreia-i zicea Molda şi pe care o iubea foarte mult, se repezi întârâtată asupra fiarei, bourul se azvârli într-un râu, unde săgeţile îl uciseră; dar şi câţeaua, care sărise în apă după fiara fugărită, fu luată de undele repezi. Întru pomenirea acestei întâmplări, Dragoş fu cel dintâi care numi acest râu Moldova, iar locul unde se petrecuse- ră acestea îi dădu numele de Roman, după numele seminţiei sale şi luă ca stemă a noului său principat capul bourului. După aceea, când cercetă ţinuturile învecinate şi găsi câmpuri roditoare cu ape îmbelşugate, târguri, cetăţi întărite, dar părăsite de locuitori, povesti alor săi despre ceea ce descoperise, îndemnându-i să pună stăpânire pe pământul acesta atât de roditor. Oamenii tineri din neamul lui Roman îşi urmară domnul de bunăvoie şi, cu bucurie, trecură munţii împreună cu el, în cete mari, aşezându-se după aceea în ţinuturile aflate într-un chip atât de minunat, iar pe Dragoş, cel care le aflase, îl numiră cel dintâi domn al ţării noi. Întoarsă în acest chip stăpânitorilor de odinioară, ţara îşi pierdu, odată cu legile trase din dreptul civil roman, şi numele dacic şi latinesc şi atât străinii, cât şi locuitorii ei înşişi îi ziseră Moldova, după apa Moldei. Dar nici această numire n-a rămas peste tot, căci turcii, care pătrundeau adesea cu oaste prin Moldova, către ţările vecine ocupate de dânşii în Europa, au dat moldovenilor mai întâi numele de akvlach1. Apoi, când Bogdan, după voia cea din urmă a tatălui său Ştefan cel Mare, le-a închinat ţara, turcii, obişnuiţi să numească ţările supuse după numele domnitorilor, au început să-i zică Bogdan; totuşi vechiul nume s-a păstrat în limba tătărească, vecinii din partea cealaltă
You can click on the desired annotation and extra information about it will be shown below
External Links
Imperiul Roman - Wikipedia
Imperiul Roman antic - Wikipedia
Imperiul Roman de Apus - Wikipedia
Imperiul Roman de Răsărit - Wikipedia
Individual
(No information)
Class Information
Imperiul roman belongs to the following classes:
City (ro: Oras)
A place that has all the elementary necessities for its inhabitants. A town.
Relationships
(No information)
PragPresentati....pptx
Annotiert101-180.zip
Annotiert51-100.zip
Annotiert33-50.zip
jdk-7u79-docs-....zip
jdk-7u79-docs-all.zip
jwcdg_precompil....xz
deutsch-lexik....lzma
Alle Downloads anzeigen...

Der erste demnach, der nach Batia Einfall (*) der Moldau ihren vorigen Glanz wieder verschafft hat, war
1. Dragosch. Obgleich unsre Jahrbücher sein Geschlechtsregister nicht angeben, so ist es doch eine beständige Sage bey uns, daß er aus dem alten königlichen moldauischen Stamme gewesen sey, und den Bogdan zum Vater gehabt habe, welcher ein Sohn des Johannis war, von welchem alle Fürsten den Namen Johannis in ihrem Titel zu führen pflegen; dieser Meinung ist desto mehr Glauben beyzumessen, weil man schwerlich glauben kan, daß einer von gemeiner Herkunft mit einem so großen Gefolge auf die Jagd (welche die Moldau zu entdecken Gelegenheit gegeben,) habe ausgehen, und seine übrigen Landsleute überreden können, ihn zu folgen.

(*) Diese Stelle bestätiget aufs neue, was wir oben schon vermuthet hatten, daß Dragosch erst nach des Tatars Bathy oder Batu Einfall, d.i. ungefähr nach 1250. aus Siebenbürgen ausgewandert ist; vielleicht aber lassen sích beede Meynungen vereinigen, wenn man zwey Auswanderungen annimmt, die eine in der lezten Hälfte des Zwölften, die andere in der ersten Hälfte des dreyzehenten Jahrhunderts **(V.)**

Dimitrie Cantemir, "Beschreibung der Moldau", Faksimiledruck der Originalausgabe von 1771, Kriterion Verlag Bukarest, 1973, pag. 111

Domnul cel dintâi carele după năvălirea lui Batie, a agonisit iarăși strălucirea cea mai dinainte a Moldovei a fost:

1. Dragoș și măcar că hronog[illegible] oastre nu arată pentru știința neamului său, dar la noi se zice necontenit, că **a fost** din neamul cel vechiu al crailor Moldovinești, și a avut tată pe Bogdan fiul lui Ioan, dela carele toți Domnii obișnuesc a-și pune la iscălitură n[illegible]

Și [illegible]

a[illegible]

d[illegible]

a[illegible]

mearga la vânat, carele a dat prilej la descoperirea Moldovei și ar fi putut îndemna pe ceilalți patrioți ai săi, ca să vie după dânsul.

was

Dragos = belongs_to Moldavian kings
Dragos = son_of Bogdan
Bogdan = son_of Johannis
Dragos = has_additional_name Johannis
Bogdan = has_additional_name Johannis

Der erste demnach, der nach Batia Einfall (*) der Moldau ihren vorigen Glanz wieder verschafft hat, war

1. Dragosch. Obgleich unsre Jahrbücher sein Geschlechtsregister nicht angeb[illegible] doch eine beständige Sage bey [illegible] aus dem alten königlichen moldau[illegible] Stamme **gewesen sey**, und den Bogdan zum Vater gehabt habe, welcher ein Sohn des Johannis war, von welchem alle Fürsten den

Dragosch ≈ belongs_to Moldavian kings
Dragosch = son_of Bogdan
Bogdan = son_of Johannis
Dragosch = has_additional_name Johannis
Bogdan = has_additional_name Johannis

die Jagd (welche die Moldau zu entdecken Gelegenheit gegeben,) habe ausgehen, und seine übrigen Landsleute überreden können, ihn zu folgen.

should have been

Der erste demnach, der nach Batia Einfall (*) der Moldau ihren vorigen Glanz wieder verschafft hat, war
1. Dragosch. **Obgleich unsre Jahrbücher sein Geschlechtsregister nicht angeben, so ist es doch eine beständige Sage bey uns, daß er aus** dem alten königlichen moldauischen Stamme **gewesen sey,** und den Bogdan zum Vater gehabt habe, welcher ein Sohn des Johannis war, von welchem alle Fürsten den Namen Johannis in ihrem Titel zu führen pflegen; **dieser Meinung ist desto mehr Glauben beyzumessen, weil man schwerlich glauben kan, daß einer von gemeiner Herkunft mit einem so** großen Gefolge auf die Jagd (welche die Moldau zu entdecken Gelegenheit gegeben,).

(*) Diese Stelle bestätiget aufs neue, was wir oben schon vermuthet hatten, daß Dragosch erst nach des Tatars Bathy oder Batu Einfall, d.i. **ungefähr** nach 1250. aus Siebenbürgen ausgewandert ist; vielleicht aber lassen sich beede Meynungen vereinigen wenn man zwey Auswanderungen annimmt, die eine in **der letzten Hälfte des Zwölften**, die andere in der **ersten Hälfte des dreyzehenten Jahrhundert** (V.)

Enriched Classical Markup

Dragosch	≈ belongs	moldavian kings
Dragosch	≈ son_of	Bogdan
Bogdan	≈ son_of	Johannis
Dragosch	≈ has_additional_name	Johannis
Bogdan	≈ has_additional_name	Johannis
Drgaosch	discovered	Moldau
Dragosch	has_acitivity	hunting
Dragosch	has_activity	development
Development	takes_place after	Batia invasion

Dragosch	has_activity	moved
Movement	takes_place	>=1150; <=1200
Movement	takes_place	>=1200; <=1250
Bathy invasion	takes_place	≈ 1250
Bathy	has_alternative_name	Batu
Bathy	is_a	Tatar

HerCoRe – Multiple-facetted fuzzy annotation

Der erste demnach, der nach Batia Einfall (*) der Moldau ihren vorigen Glanz wieder verschafft hat, war
1. Dragosch. Obgleich unsre Jahrbücher sein Geschle… …es doc… dem… gew… hab… wel… ihre… desto mehr Glauben beyzumessen, weil man schwerlich glauben kann, daß einer von gemeiner Herkunft mit einem so großen Ge… auf die Jagd (welche die Moldau zu entdeck… Gelegenheit gegeben,) habe ausgehen, und… seine ü… …ner… zu fo… …uthet hatten, daß Dragosch erst nach …athy oder Batu Einfall, D.i. ungefähr …us Siebenbürgen ausgewandert ist; …er lassen sich beide Rechnungen …venn man zwey Auswanderungen …e eine in der letzten Hälfte des …e andere in der ersten Hälfte des … Jahrhundert (B.)

Legend

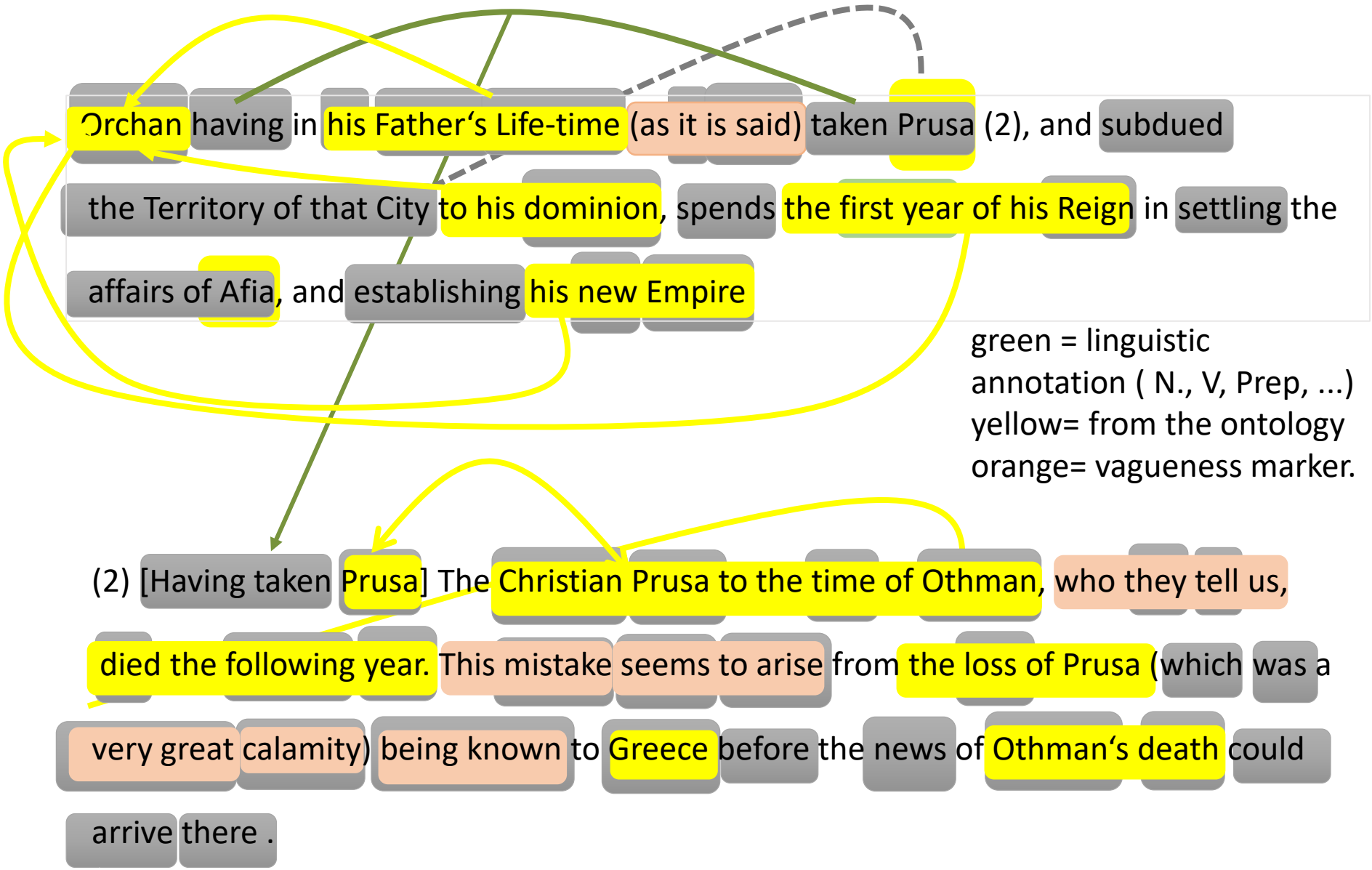
Dragosch	most_probable belongs	moldavian kings
Dragosch	discovered	Moldau
Dragosch	has_acitivity	hunting
Dragosch	has_activity	development
Development	takes_place after	Batia invasion

Cantemir

Dragosch	≈ son_of	Bogdan
Bogdan	≈ son_of	Johannis
Dragosch	≈ has_additional_name	Johannis
Bogdan	≈ has_additional_name	Johannis

Translator

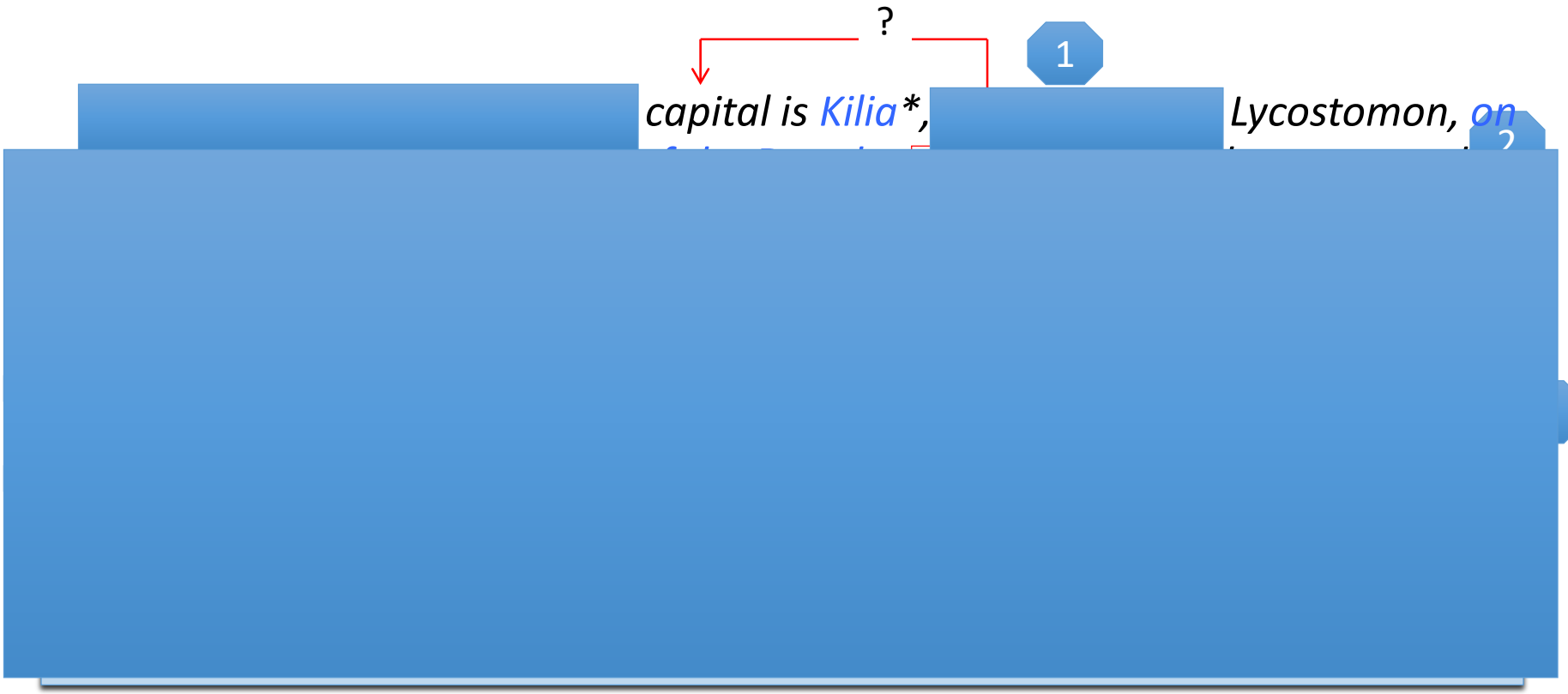
Dragosch	has_activity	moved
Movement	takes_place	> ≈ 1150; < ≈ 1200
Movement	takes_place	> ≈ 1200; < ≈ 1250
Bathy invasion	takes_place	≈ 1250
Bathy	has_alternative_name	Batu
Bathy	is_a	Tatar



History of Growth and Decay Ottoman Empire, English Translation, pag. 24

A Complicated Explicit Example: Ambiguity

(Cantemir, Descriptio Moldaviæ, p.73 transl.)



Linking NEs and Maps

Syrfia is
the abandoned name of a region in Eastern Europe, used on historical maps until 17th century, designating

- a part of Northern Dobrudja, coming from the Greek term *Σύρφοι - Syrphoi*, or
- The Cojani region from western Macedonia, today in Greece but in turkish times in the "Serfia sangiac" having the capital *Σέρβια, Servia* ;
- Sârbia, due to phonetic association.

Sources /levels of vagueness to be annotated

1. Linguistic markers for vagueness

2. Factual uncertainity

2.1 References to external written materials (publicatiosn)

2.2 References to external persons, places, names

2.3 References to events

2.4. References to other external knowledge (e.g. legends, folk beliefs)

3. Editors

3.1. Editorial marks

() pretty sure extensions

< > correction

[ ] deletion

{ } marginals /between line

3.2 „Footnotes“

4. Metadata

4.1 genre

4.2. author

4.3 translation

4.4. tradition path

Vagueness annotation is useful only if it is accompagned by inference rules and adequate ontological knowledge-base

Selected markers for linguistic vagueness

1. comparatives, inexact adjectives e.g. *“hostile”, “near”, “distant”*
2. non-intersectives e.g. „supposend“, „so-called“
3. Hedges e.g. „rather“, „more or less“
4. inexact measures *a 4 days’ journey, 10 feet”*
5. modals (attitudes) e.g. probably, hopefully
6. subjonctives v+analyse
7. lexical quotation markers
8. vague quantifiers e.g. „many“, „most“
9. complex quantifiers e.g. *"roughly half of the 20-30 thousand soldiers”*
10. numbers
11. range expressions e.g. *“The beginning of the 18. century”*
12. unclear place *„the former prince“, „Moramor“*
13. unclear person e.g. *„the former prince“*
14. unclear time e.g. *„in prehistoric times“*

Manual Annotation of Factual uncertainity

- […] He fought two Battles with Bajazet Ildirim; in the first he was Victor, and in the second he routed Him with a memorable slaughter, which seven vast piles of *Turkish* Bodies erected after the Battle, witnessed, by the Confession of *Hezarfenn* himself, the faithful *Turkish* Historian. Cantemir, pp. 47 (Annotations)

Hezarfen (Hezarfen Hüseyin Efendi) (?-1691/92), Tenkih-i Tevarih-i Mülük: is **NOT** mentioning this

- The Turkish Historians so extoll this Prince's expedition in assembling his troops, in executing his designs, and in vanquishing his enemies, that when they talk of the natural speed of the Tartars in comparison with his wonderful marches, they call the first, the creeping of a Snail.

Cantemir, pp. 48 (Anotations)

Described in Solakzade, Hoca Saadettin, Neşri

Manual Annotation of Factual uncertainity

Deutsche Übersetzung 1745

„Was seine Söhne betrifft: so wissen die christlichen Schriftsteller unter den verderbten Namen Erdogul, Issa, Kalepin, Cyricelebis und Cibelin, viele Dinge von denselben zu erzählen. Wenn man aber die Folge der GEschichte und das Zeugniß der turkischen Schriftsteller betrachtet; so siehet man offenbar, daß es bloße Erdichtungen sind. Denn diese legen **einstimmig** Bajeßid vier Söhne bei: Muståfa, der in der Schlacht mit den Tatarn um das Leben kam, Sülejman, Musa und Muhåmmed. [...] So viel ist wenigst **gewiß**, daß nicht mehr als vier Sohne Bajeßids in der ganzen Ges vorkommen, darunter aber ist **kein Erdogul**“.
(S. 79-80)

Es ist **sicher** daß **Sultan Bajeßid 4 Söhne hatte und keiner hieß Erdogul**

≠ ?

Türkische-Osmanische Quellen

- **HSE**: **Ertuğrul**, Süleyman, Mehmed, İsa Çelebi, Musa Çelebi, Mustafa Çelebi. (S. 192)
- **NES**.: **Ertuğrul**, Süleyman, Mehmed, İsa Çelebi, Musa Çelebi, Mustafa Çelebi (S. 313)
- **SOL**.: **Ertuğrul**, Süleyman, Mehmed, İsa Çelebi, Musa Çelebi, Mustafa Çelebi. (S. 70)

Treatment of (uncertain) NEs

DHO

LHO

Turkish Sources

Seråskjer·Sülejman·Pascha¤	65¤	Seraskerio·Suleiman·Baszae¤	38¤	سر عسکر سلیمان پاشا¶ Serasker· Süleyman·Paşa¤
Istifan¤	66¤	Istefan¤	38¤	¤
Matthias·(Ann.)¤	66¤	Mathiam·Corvinum¤	34¤	¤
Bajeßid·Yildirim·(Ann.)¤	66¤	Ildirìm·Baìezid¤	34¤	یلدرم بایزید¶ Yıldırım·Bayezid¤
Heßarfenn·(Ann.)¤	66¤	Hezárfenn¤	34¤	¤
Karaman·Ogli¤	67¤	Caramanougly¤	39¤	قرامان اوغلی¶ Karamanoğlu¤
Bajeßid¤	67¤	Baiezid¤	39¤	بایزید¶ Bayezid¤

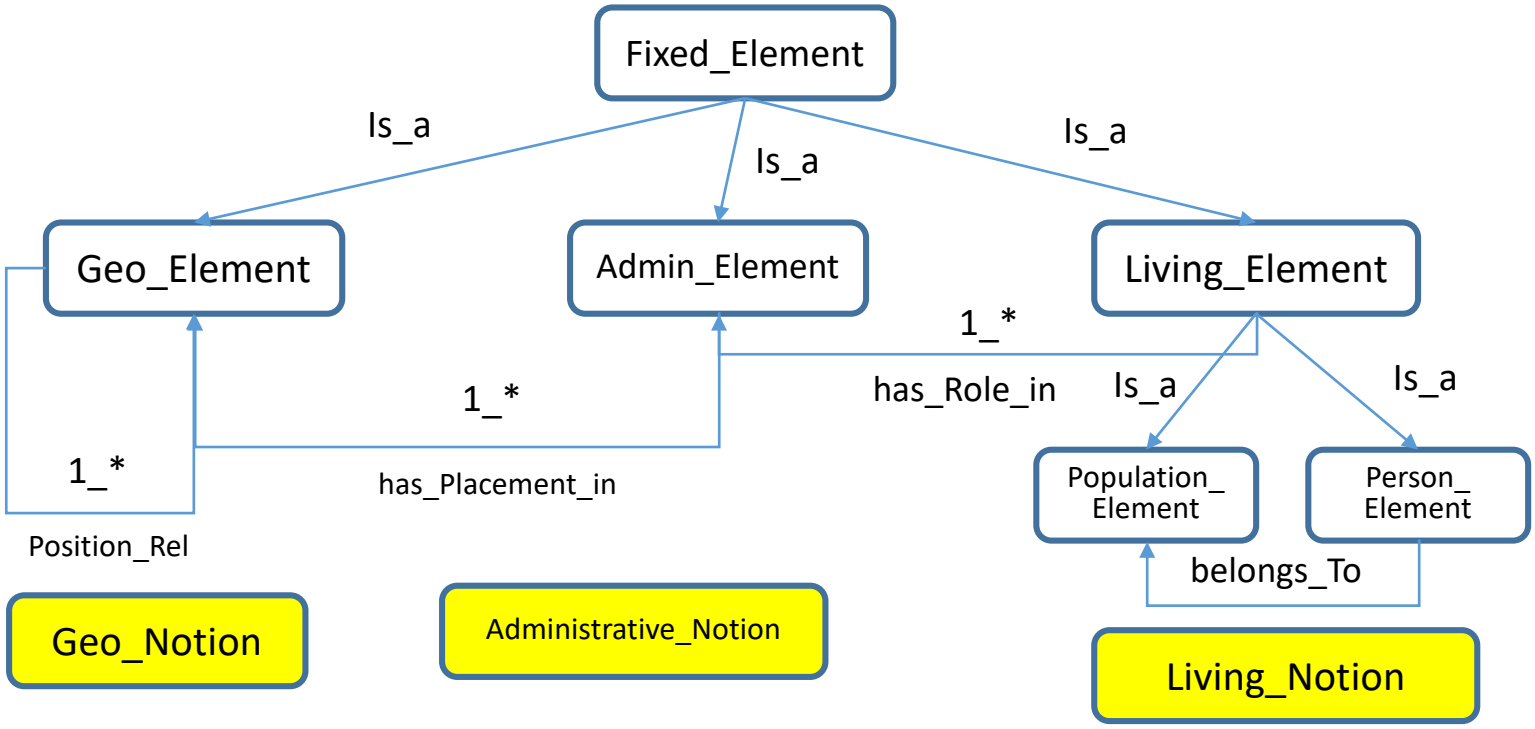
Adrianopel¤	73¤	Adrianopol¤	44¤	ادرنه·Edirne¤
Edrene·(Mewlasi)¤	73¤	Edrne·(Molasi)¤	44¤	ادرنه·Edirne¤
Edrene·(Mewlasi)·(Ann.)¤	73¤	Ederne·(Mollasì)¶ ادرنه·(مولاسی)¤	¤	ادرنه·Edirne¤
Adrianopel·(Ann.)¤	73¤	Adrianopolis¤	38¤	ادرنه·Edirne¤
Misr·(Mewlasi)·(Ann.)¤	73¤	Mýsr·(Mollasì)¶ مصر (مولاسی)¤	38¤	مصر·Mısır¤
Burusa·(Mewlasi)·(Ann.)¤	73¤	Birússa·(Mollasì)¶ بروسه·(مولاسی)¤	38¤	بروسه·Bursa¤
Haleb·(Mewlasi)·(Ann.)¤	73¤	Halèp·(Mollasì)¶ حلب·(مولاسی)¤	38¤	حلب·Haleb/·Halep¤
Dawud·Pascha¤	74¤	Daud· Pasza/· Davùd· (Ann.)¤	44,· 38· (Ann.)¤	¤

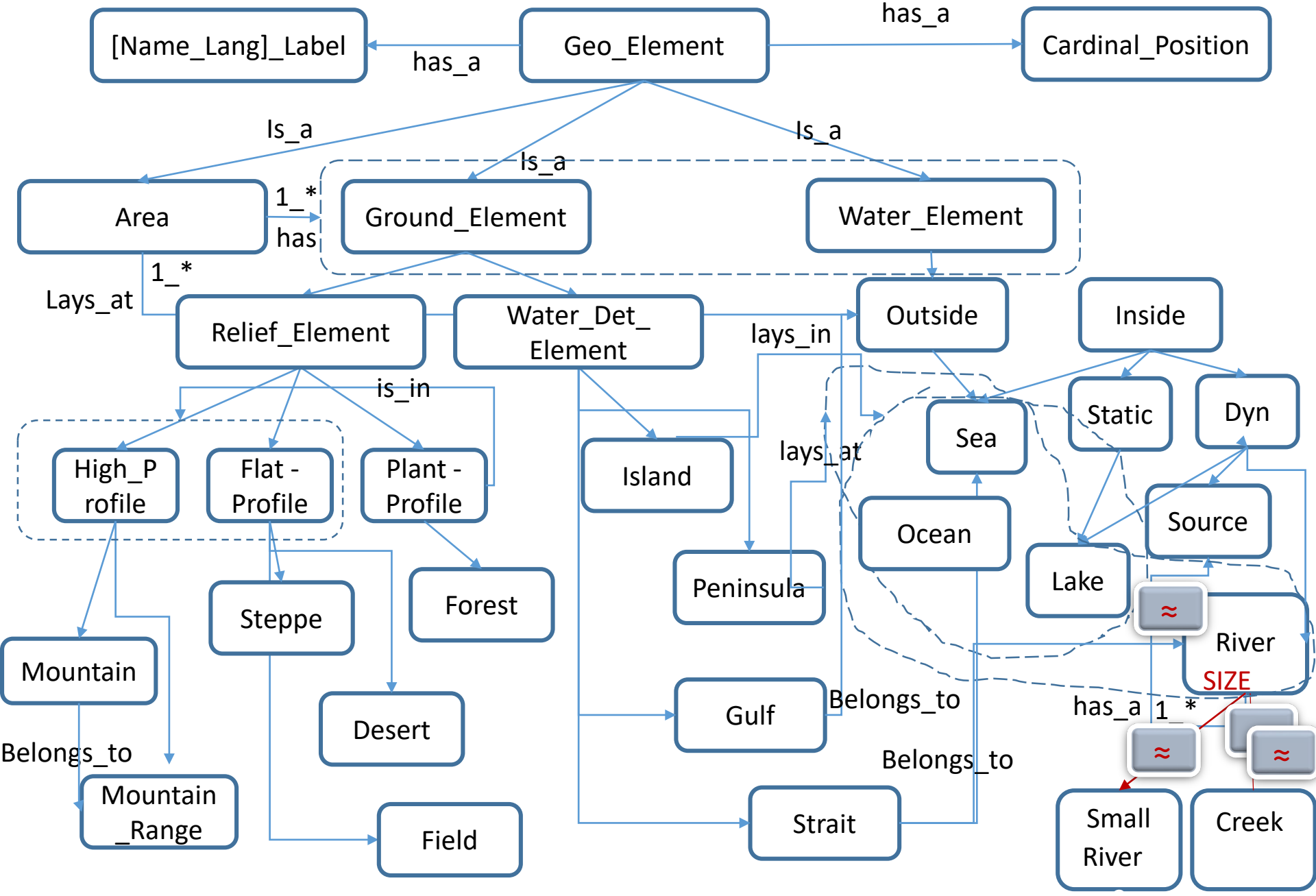
Places Database

Persons Database

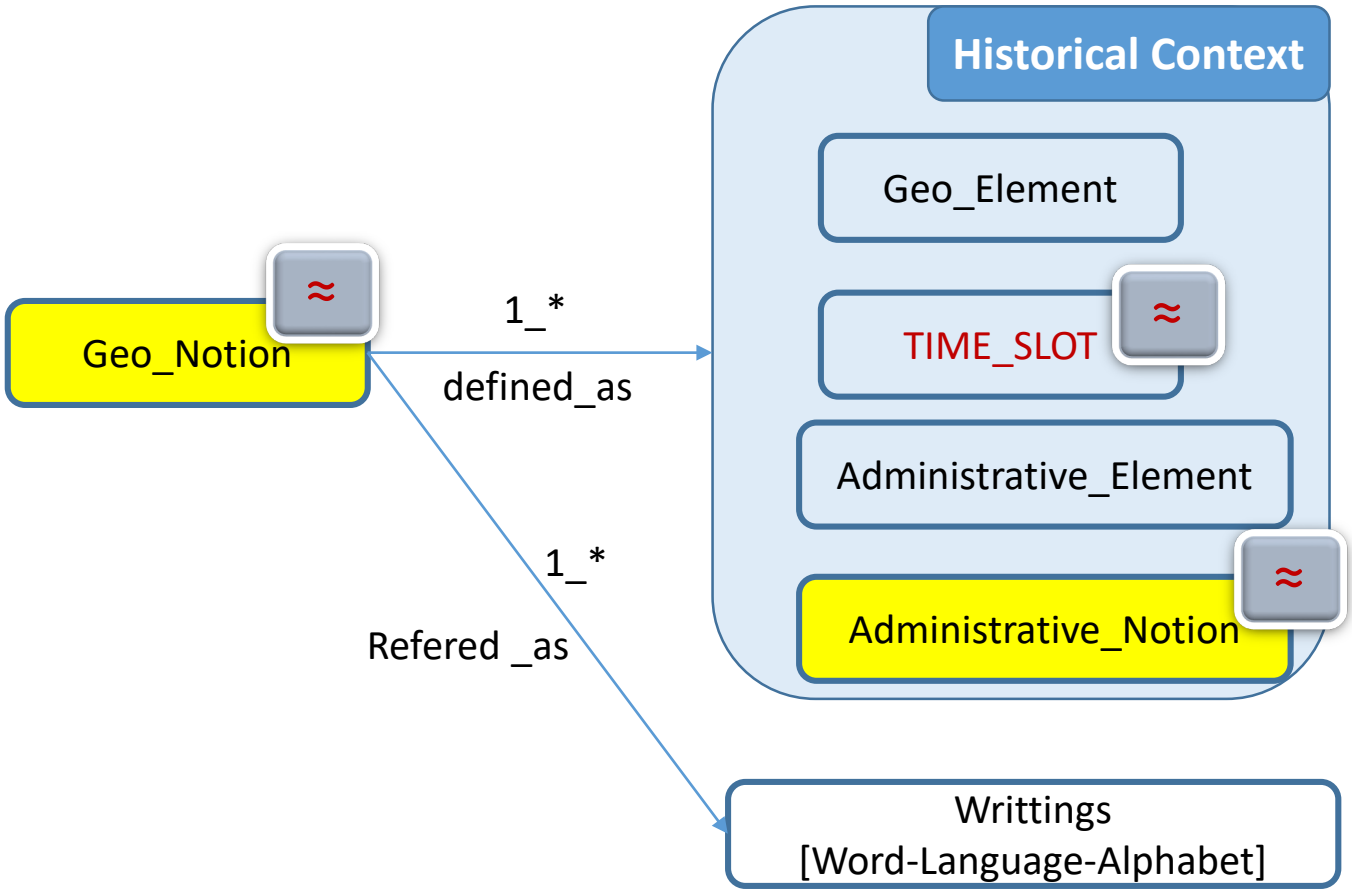
©Alptug Güney

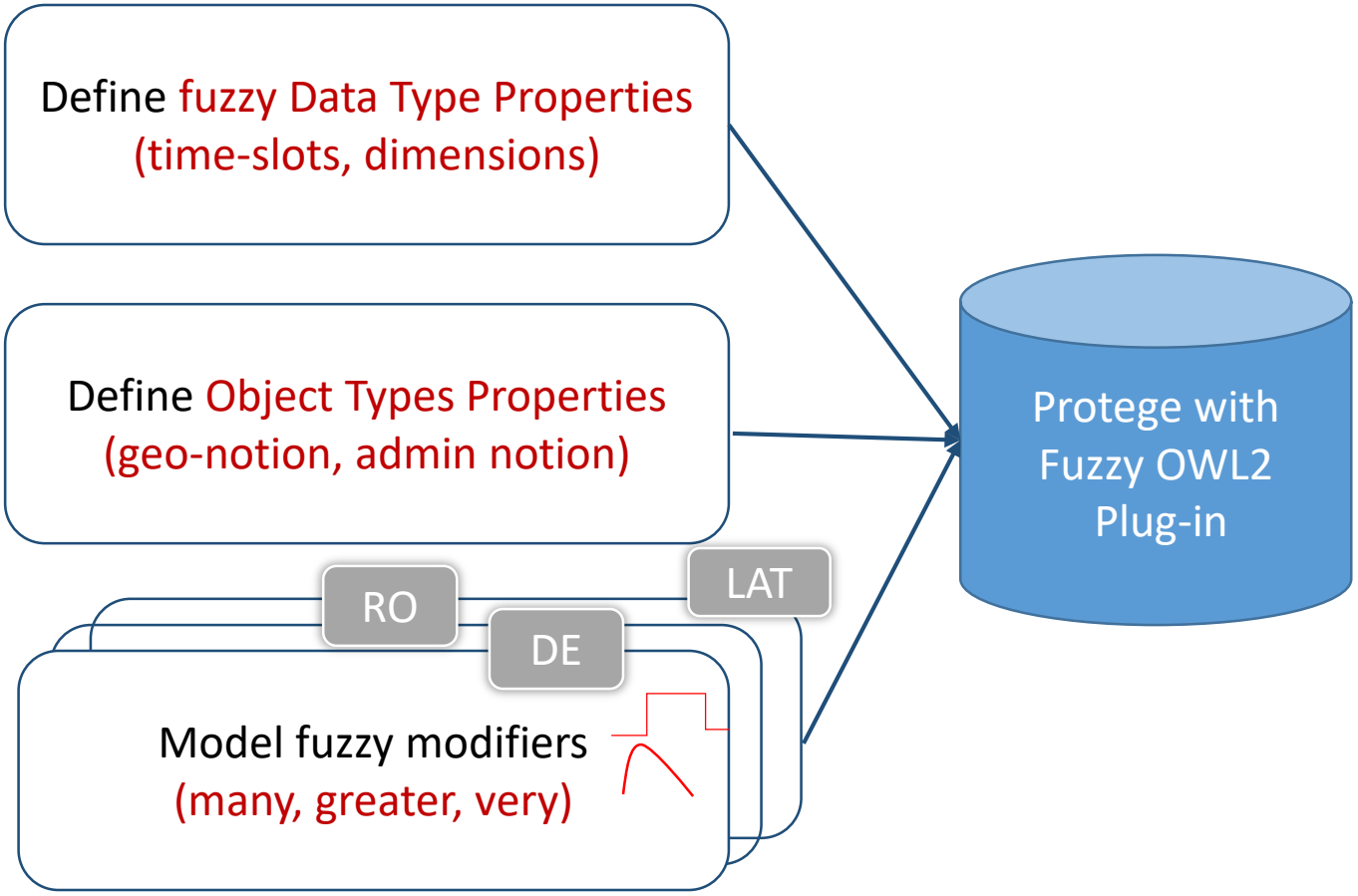
Ontological modelling -1-

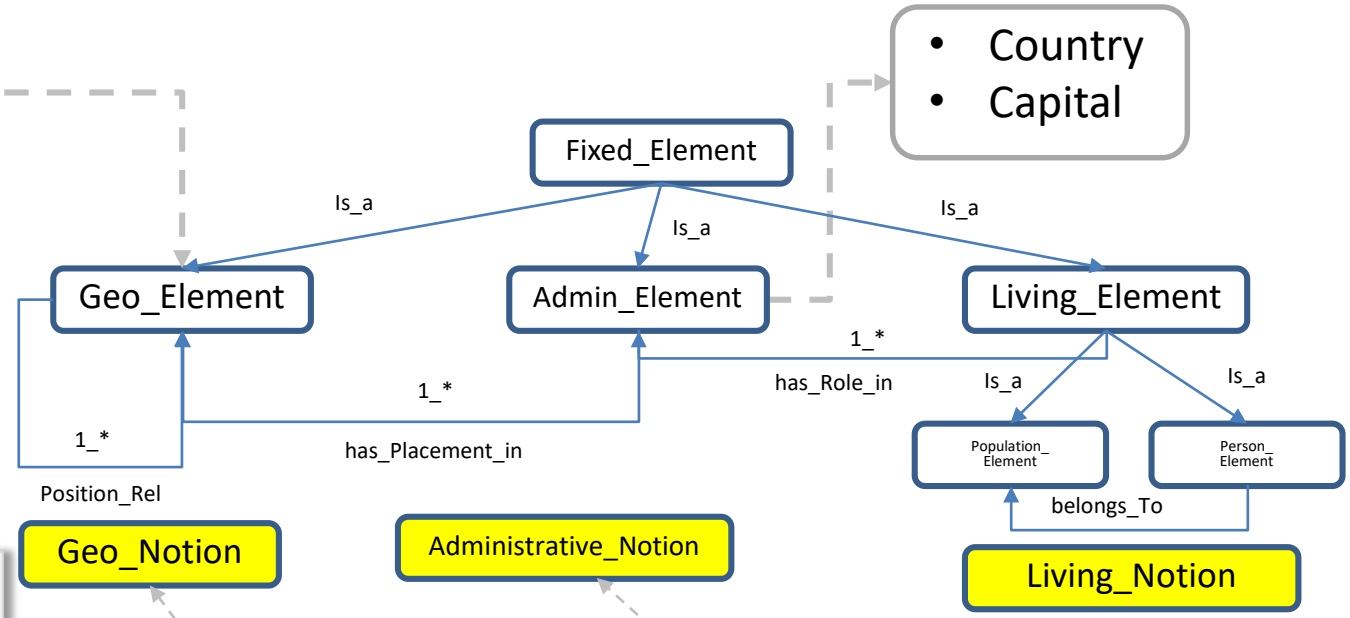



[Name_Lang]_Label
has_a
Geo_Element
has_a
Cardinal_Position
Is_a
Is_a
Is_a
Area
1_*
has
Ground_Element
Water_Element
1_*
Lays_at
Relief_Element
Water_Det_
Element
lays_in
Outside
Inside
is_in
Static
Dyn
Sea
High_P
rofile
Flat -
Profile
Plant -
Profile
Island
lays_at
Source
Ocean
Peninsula
Lake
Forest
Steppe
River
Mountain
SIZE
Belongs_to
has_a
1_*
Gulf
Belongs_to
Desert
Belongs_to
Mountain
_Range
Strait
Small
River
Creek
Field

Ontological modelling -1-




Define fuzzy Data Type Properties
(time-slots, dimensions)
Define Object Types Properties
(geo-notion, admin notion)
RO
DE
LAT
Model fuzzy modifiers
(many, greater, very)
Protege with
Fuzzy OWL2
Plug-in



- Northern Dobrudja
- Western Macedonia
- Eastern Europe

Syrfia is
the abandoned name of a region in Eastern Europe, used on historical maps until 17th century, designating

- a part of Northern Dobrudja, coming from the Greek term *Σύρφοι - Syrphoi*, or
- The Cojani region from western Macedonia, today in Greece but in turkish times in the "Serfia sangiac" having the capital *Σέρβια, Servia* ;
- Sârbia, due to phonetic association.

- Cojani Region
- Sârbia

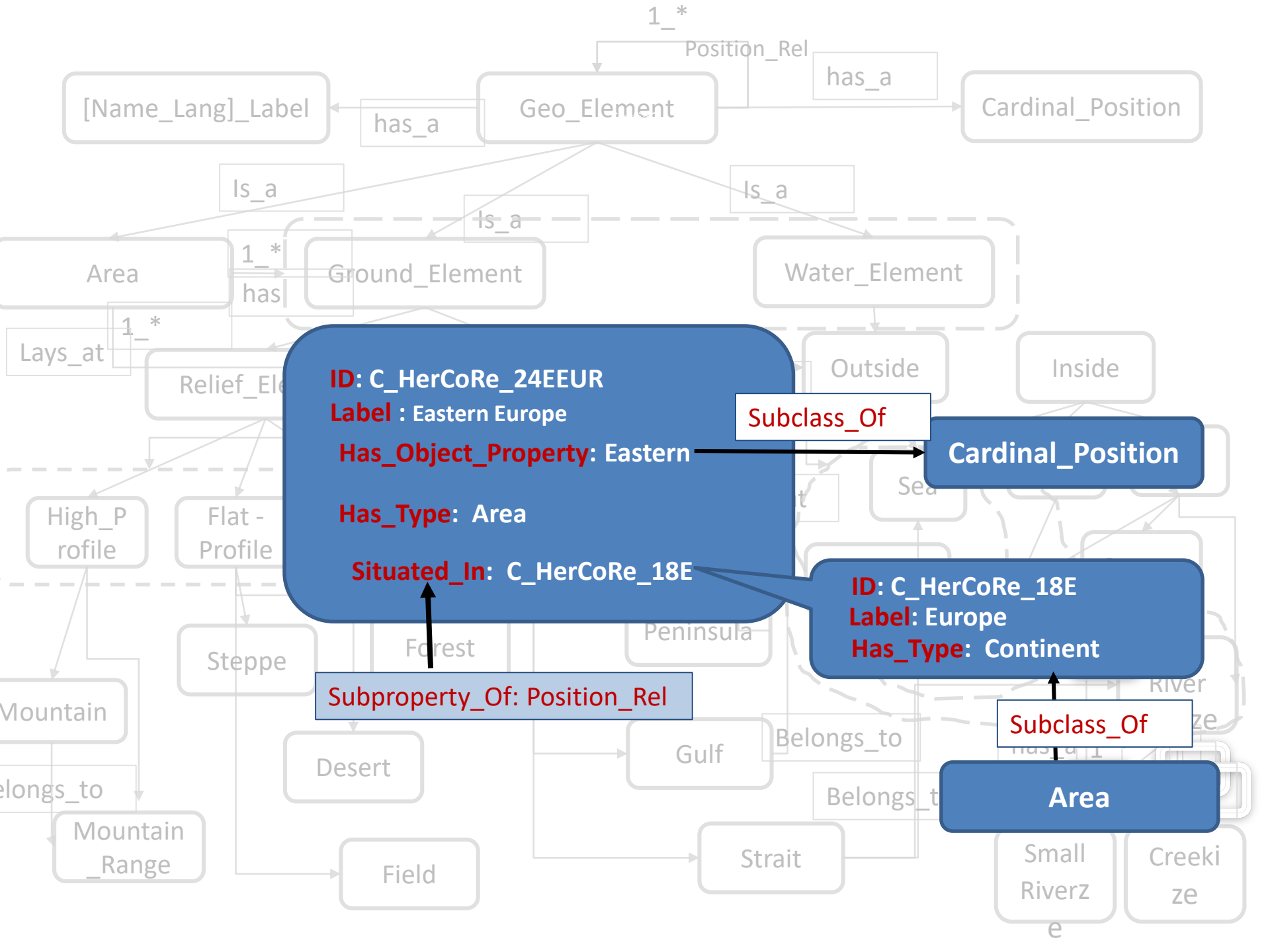
Fuzzy Concept

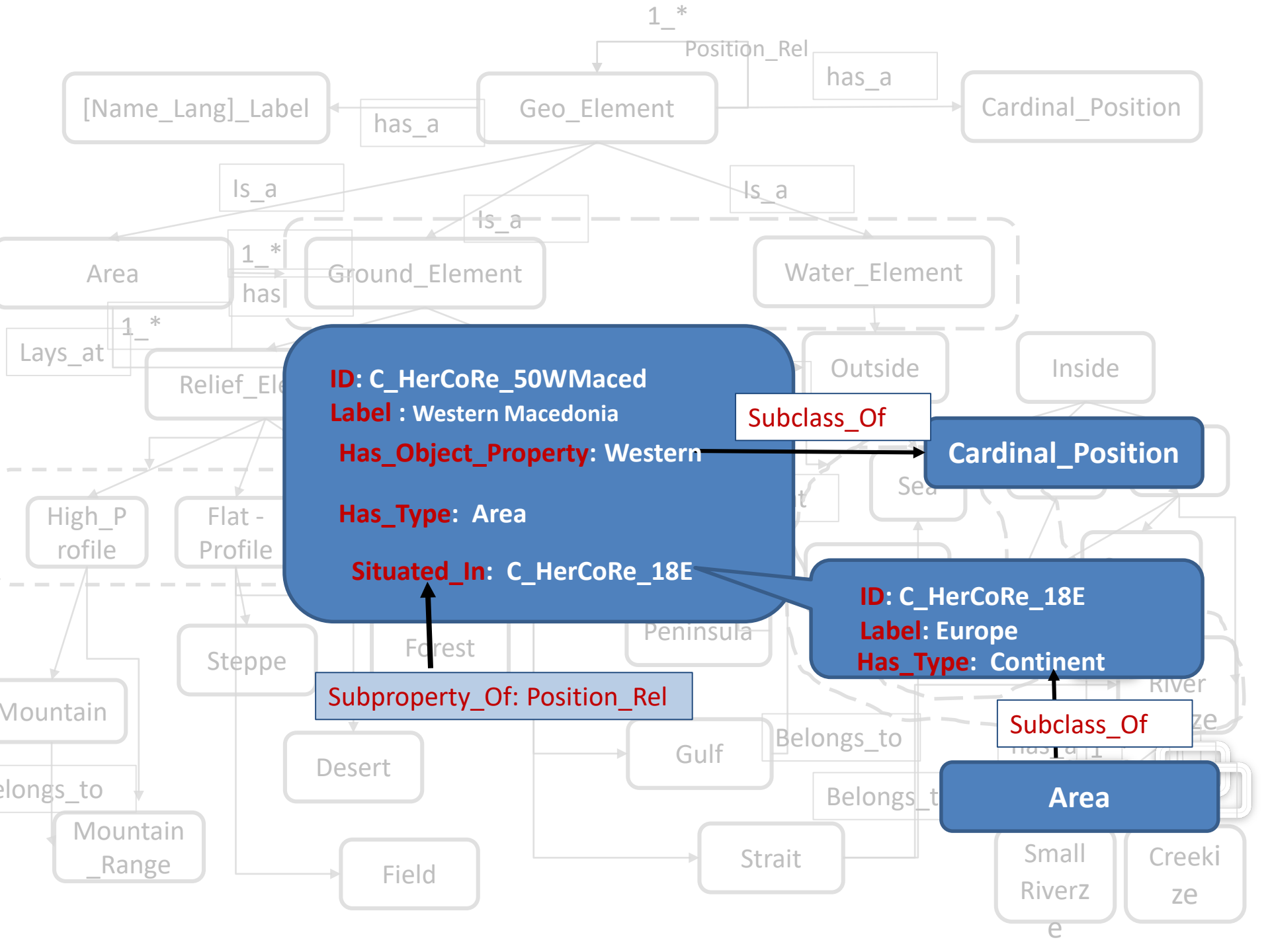
- Greece
- Serfia sangiac
- Servia

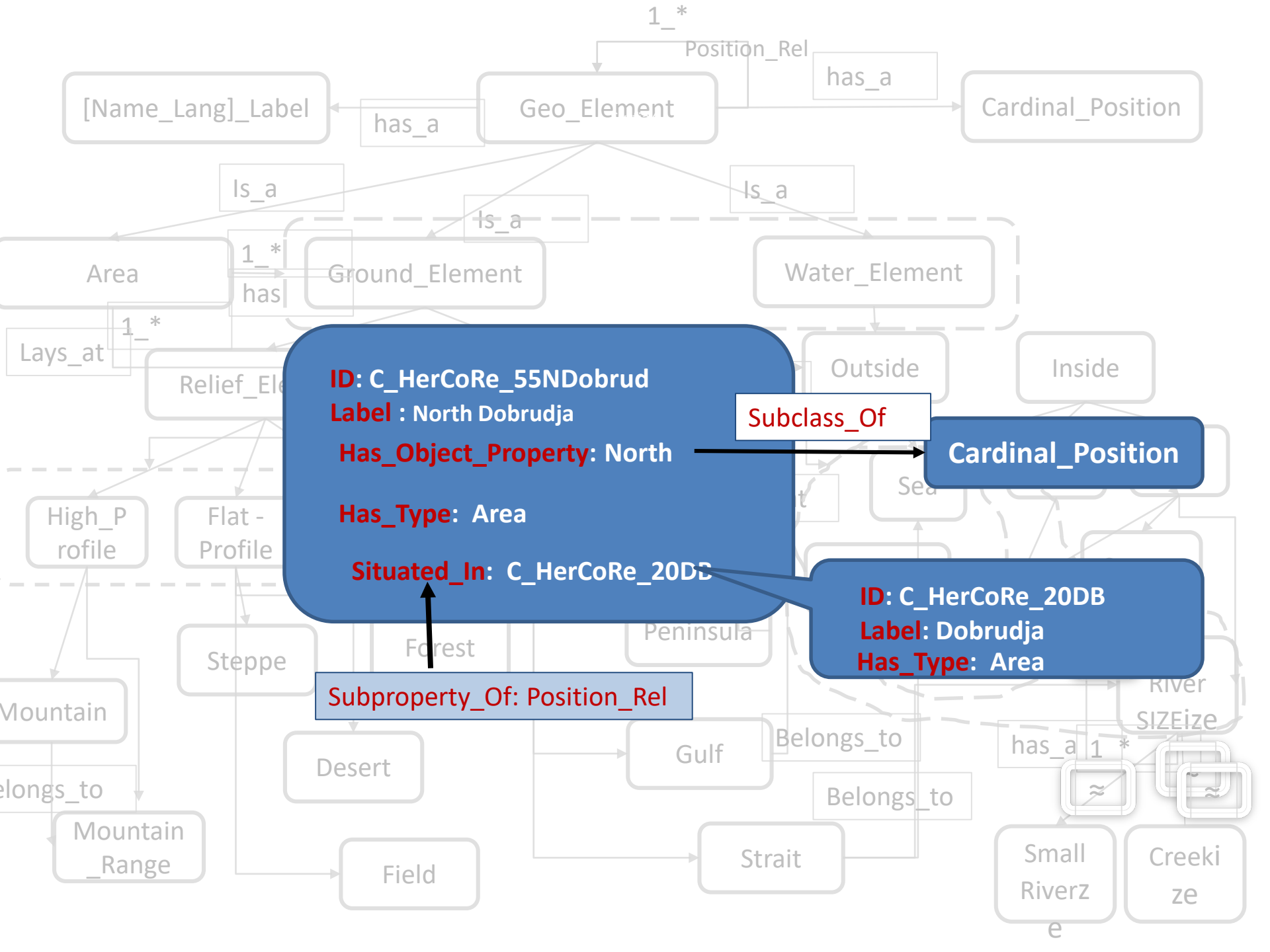
Fuzzy Concept

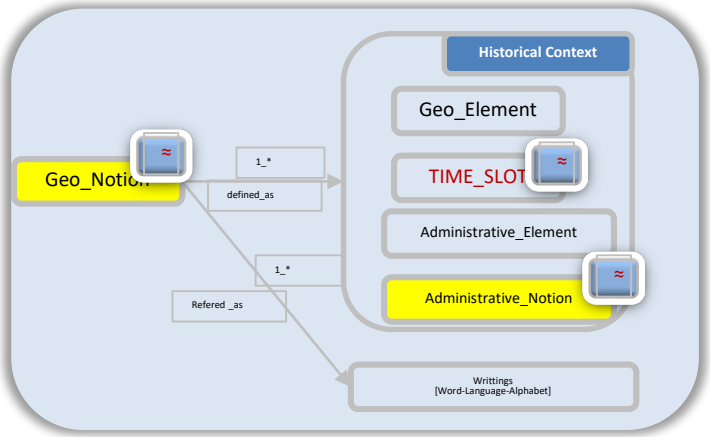
- Turkisch Times
- Greek times
- 17th century

Fuzzy Properties


1_*
Position_Rel
has_a
[Name_Lang]_Label
has_a
Geo_Element
has_a
Cardinal_Position
Is_a
Is_a
Is_a
1_*
has
Area
Ground_Element
Water_Element
1_*
Lays_at
Relief_El
Outside
Inside
ID: C_HerCoRe_24EEUR
Label : Eastern Europe
Has_Object_Property: Eastern
Subclass_Of
Cardinal_Position
Has_Type: Area
Situated_In: C_HerCoRe_18E
ID: C_HerCoRe_18E
Label: Europe
Has_Type: Continent
High_P rofile
Flat - Profile
Steppe
Forest
Peninsula
Subproperty_Of: Position_Rel
Subclass_Of
Area
Mountain
Belongs_to
Gulf
Desert
Belongs_to
Mountain _Range
Field
Strait
Small Riverz e
Creeki ze


1_*
Position_Rel
has_a
[Name_Lang]_Label
has_a
Geo_Element
has_a
Cardinal_Position
Is_a
Is_a
Is_a
1_*
has
Area
Ground_Element
Water_Element
1_*
Lays_at
Relief_El
Outside
Inside
ID: C_HerCoRe_50WMaced
Label : Western Macedonia
Has_Object_Property: Western
Subclass_Of
Cardinal_Position
Has_Type: Area
Situated_In: C_HerCoRe_18E
ID: C_HerCoRe_18E
Label: Europe
Has_Type: Continent
High_P rofile
Flat - Profile
Sea
Steppe
Forest
Peninsula
Subproperty_Of: Position_Rel
Subclass_Of
Mountain
Belongs_to
Gulf
Desert
Belongs_t
Area
elongs_to
Mountain _Range
Strait
Field
Small Riverz e
Creeki ze


1_*
Position_Rel
has_a
[Name_Lang]_Label
has_a
Geo_Element
Cardinal_Position
Is_a
Is_a
Is_a
1_*
has
Area
Ground_Element
Water_Element
1_*
Lays_at
Relief_El
ID: C_HerCoRe_55NDobrud
Label : North Dobrudja
Has_Object_Property: North
Has_Type: Area
Situated_In: C_HerCoRe_20DB
Subclass_Of
Cardinal_Position
Outside
Inside
Sea
High_P rofile
Flat - Profile
ID: C_HerCoRe_20DB
Label: Dobrudja
Has_Type: Area
Steppe
Forest
Peninsula
Subproperty_Of: Position_Rel
Mountain
River SIZEize
Belongs_to
has_a 1 *
Gulf
Desert
Belongs_to
elongs_to
Mountain _Range
Strait
Field
Small Riverz e
Creeki ze



Hist. Context 1

Geo_Element

Area

Situated_In

ID: C_HerCoRe_50MACEd
Label : Western Macedonia

≈ Time_Slot

F_TODAY

Admin_Notion

Id: C_HercoRe_17GR
Label: Greece

Hist. Context 2

Geo_Element

Area

Situated_In

ID: C_HerCoRe_50MACEd
Label : Western Macedonia

≈ Time_Slot

Turkish time

Admin_Notion

Id: C_HercoRe_71 SarfS
Label: *Serfia sangiac"*

Defined_As

Defined_As

Id: C_HercoRe_71 CojR
Label: Cojani Region:

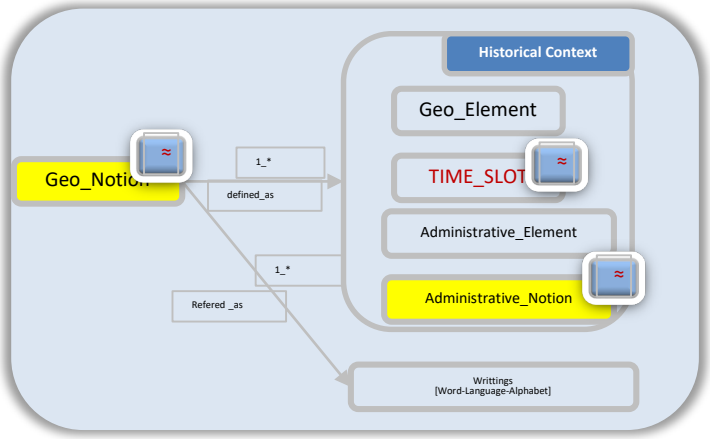
Geo_Notion

Situated_in

Admin_Notion

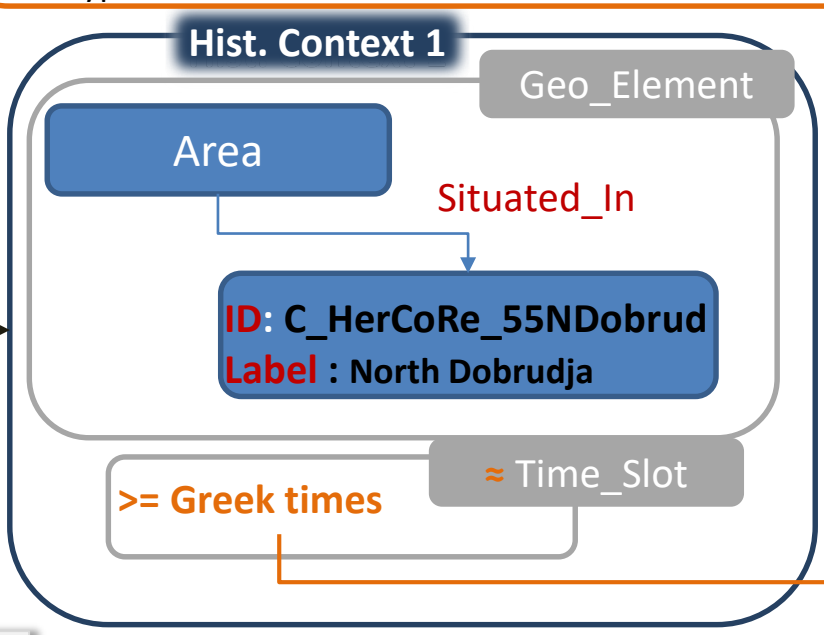
Id: C_HercoRe_74 SarfC
Label: *Σέρβια,Servia*

The Cojani region from western Macedonia, today in Greece but in Turkish times in the "Serfia sangiac" having the capital Σέρβια, Servia ;




```
<DatatypeDefinition>
    <Datatype IRI='#GreekTimes'/> <DataIntersectionOf>
    <DatatypeRestriction>

          <Datatype abbreviatedIRI='xsd:integer'/>
          <FacetRestriction facet='&xsd;minInclusive'>
                    <Literal datatypeIRI='&xsd;integer'>-750</Literal>
           </FacetRestriction>
      </DatatypeRestriction>
      <DatatypeRestrictionOf>
</DatatypeDefinition>
```



Geo_Notion

Defined_As

Id: C_HercoRe_10 DSyrf

Label: *Σύρφοι, Syrphoi*

Part of Northern Dobrudja, coming from the Greek term Σύρφοι --Syrphoi;

```
Class ( Syrfia Annotation
 (fuzzyLabel
    < fuzzyOwl2 fuzzyType =" concept " >
    < Concept type =" weightedSum " >
    < Concept type =" weighted " value ="0.33" base ="C_HercoRe_71CojR " / >
     < Concept type =" weighted " value ="0.33" base =" C_HercoRe_10DSyrf " />
     < Concept type =" weighted " value ="0.33" base =" C_HercoRe_11Srb " />
))
```

Syrfia is
the abandoned name of a region in Eastern Europe, used on historical maps until 17th century, designating

- a part of Northern Dobrudja, coming from the Greek term *Σύρφοι - Syrphoi*, or
- The Cojani region from western Macedonia, today in Greece but in turkish times in the "Serfia sangiac" having the capital *Σέρβια, Servia* ;
- Sârbia, due to phonetic association.

Source: Wikipedia

Geo_Notion

Id: C_HercoRe_70Syrfia
Label: Syrfia
Used_for: Map
<= 17 Century
≈ Time_Slot

Defined_As
Confidence 0.33

Hist. Context 1
Id: C_HercoRe_71 CojR
Label: Cojani Region:

Defined_As
Confidence 0.33

Hist. Context 2
Id: C_HercoRe_10 DSyrf
Label: *Σύρφοι, Syrphoi*

Defined_As
Confidence 0.33

Hist. Context 3
Id: C_HercoRe_11Sarb
Label:*Sarbia*

Orteliusmap 1570

Id: I_C_HerCoRe_Sy
Label: Syrfia_Ortelius
Used_for: Map_Ortelius
Time_slot: 1570

Instance_Of

Syrfia is
the abandoned name of a region in Eastern Europe, used on historical maps until 17th century, designating

- a part of Northern Dobrudja, coming from the Greek term *Σύρφοι - Syrphoi*, or
- The Cojani region from western Macedonia, today in Greece but in turkish times in the "Serfia sangiac" having the capital *Σέρβια, Servia* ;
- Sârbia, due to phonetic association.

Source: Wikipedia

Geo_Notion

Id: C_HercoRe_70Syrfia
Label: Syrfia
Used_for: Map

<= 17 Century
≈ Time_Slot

Defined_As
Confidence 0.2

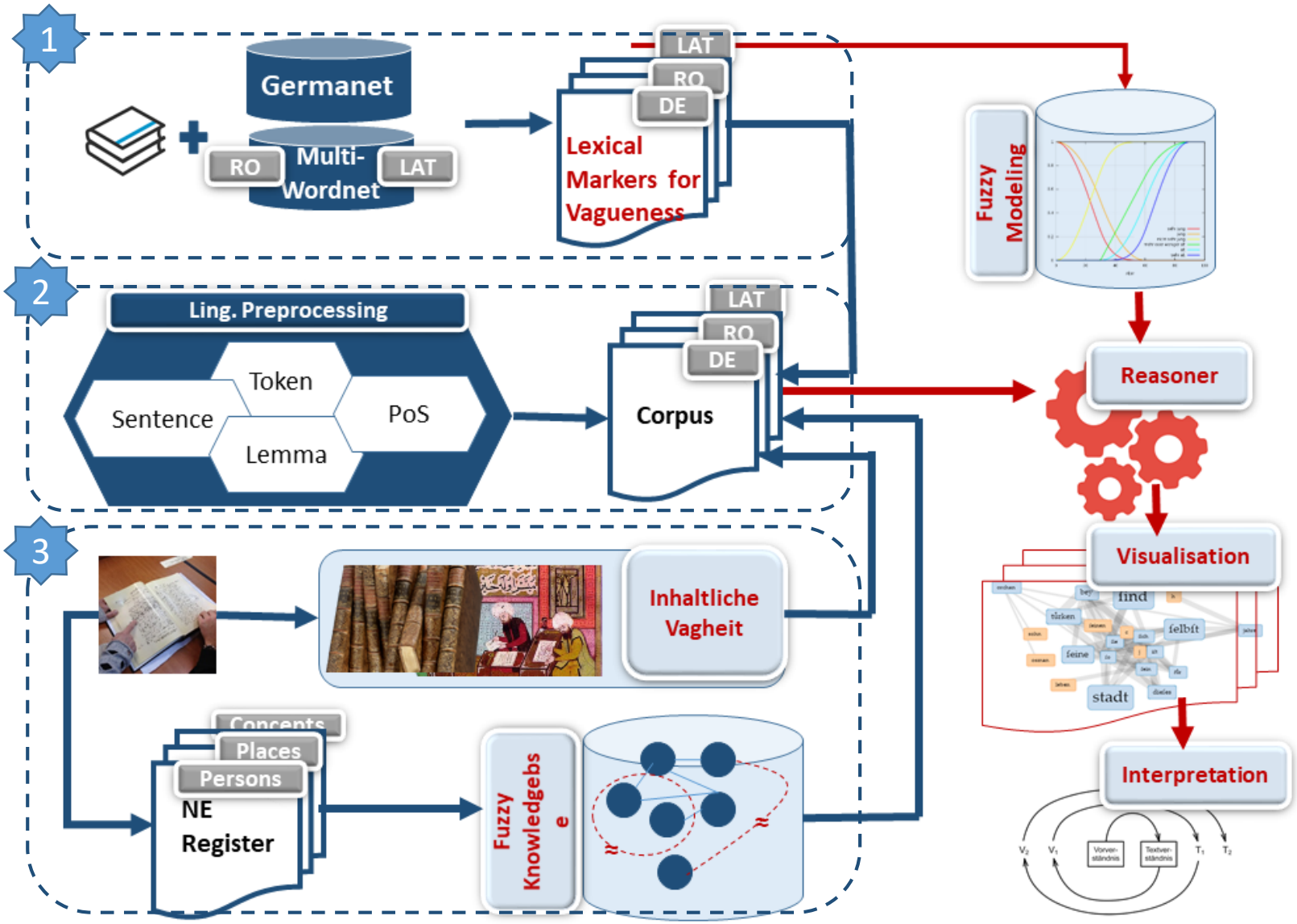
Hist. Context 1
Id: C_HercoRe_71 CojR
Label: Cojani Region:

Defined_As
Confidence 0.7

Hist. Context 2
Id: C_HercoRe_10 DSyrf
Label: *Σύρφοι, Syrphoi*

Defined_As
Confidence 0.1

Hist. Context 3
Id: C_HercoRe_11Sarb
Label: *Sarbia*


1
Germanet
RO
Multi-
Wordnet
LAT
LAT
RO
DE
Lexical
Markers for
Vagueness
Fuzzy
Modeling
2
Ling. Preprocessing
Token
Sentence
PoS
Lemma
LAT
RO
DE
Corpus
Reasoner
3
Inhaltliche
Vagheit
Visualisation
find
stärken
selbst
feine
stadt
Concepts
Places
Persons
NE
Register
Fuzzy
Knowledgebs
e
Interpretation