



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

DEPARTMENT INFORMATIK

NATÜRLICHSPRACHLICHE SYSTEME (NATS) AB.  
„Computerphilologie“-Arbeitsstelle

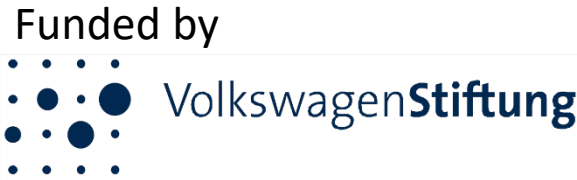
# Digitale Methoden zur Analyse von Zuverlässigkeit, Konsistenz und Vagheit in historischen Texten

**Cristina Vertan**

cristina.vertan@uni-hamburg.de

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

- Illustrated through two main works of Dimitrie Cantemir-



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

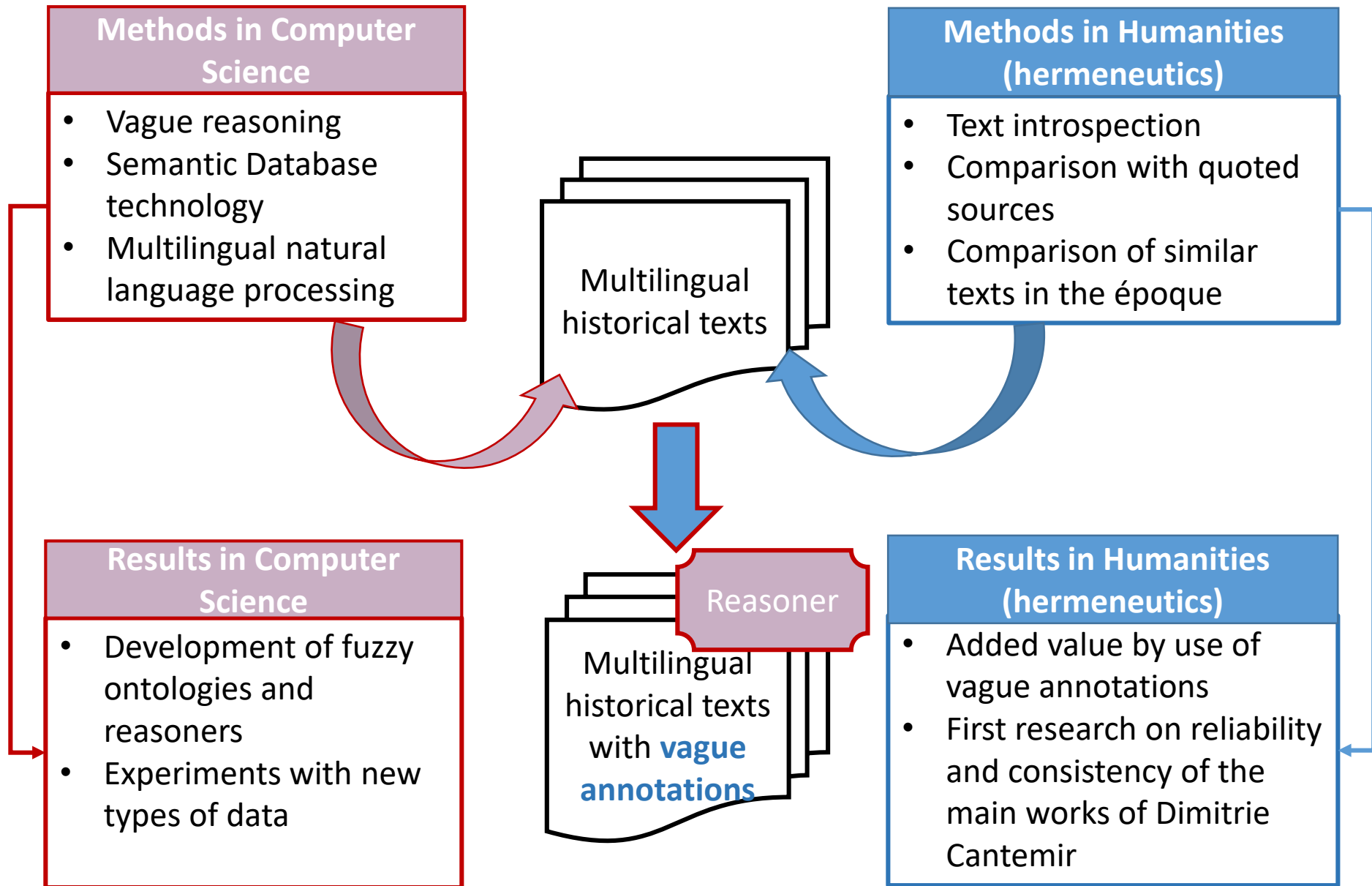
Compare for the first time “original” with translations done in the 18th- 19th century

(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



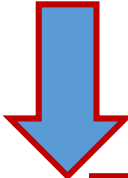
**Methods in Computer Science**

- Vague reasoning
- Semantic Database technology
- Multilingual natural language processing

**Methods in Humanities (hermeneutics)**

- Text introspection
- Comparison with quoted sources
- Comparison of similar texts in the époque

Multilingual historical texts



Reasoner

Multilingual historical texts with **vague annotations**

**Results in Computer Science**

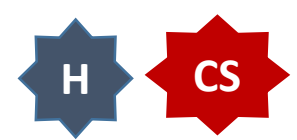
- Development of fuzzy ontologies and reasoners
- Experiments with new types of data

**Results in Humanities (hermeneutics)**

- Added value by use of vague annotations
- First research on reliability and consistency of the main works of Dimitrie Cantemir



Dr. Cristina Vertan, UHH  
Project coordination,  
DH, CL, CS



Dr. Anca Dinu, UB  
Team Leader UB,  
Linguistics, CL



Prof. Dr. Walther v. Hahn,  
UHH  
Vagueness, CL, DH  
German Linguistics,



Prof. Dr. Ioana Costa, UB  
Cantemir Translations,  
Classical philology



Prof. Dr. Yavuz Köse, UHH  
Turcology



Prof. Dr. Liviu Dinu, UB  
Fuzzy Logic, CL, CS



Alptug Güney, UHH  
Turcology



Segiu Nisioi, UB  
CL, CS

# Outline

- Rationale of the project
- Corpus' Insight
- First steps on identifying and annotating vagueness
- Technological approach

## 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.**

## Analysis and interpretation of Cantemir's works

- Already in the 1920'ies, it was demonstrated using selections of texts, that the translations are not respecting the original all the time
  - E.g. Information sources indicated by Cantemir were omitted, because they seemed too unreliable to the translator
- In the XX century researchers claimed that some of the sources, persons and facts quoted by Cantemir were not existing (e.g. Babinger).
- BUT given the:
  - Geographic distribution of material (originals in libraries in USA and Russia; translations and copies across Europe; most part of the quoted sources in Turkey),
  - The multilingual character of the materials to be investigated (Latin, German, Romanian, English, Turkish at least) and
  - The volume of data which has to be processed in parallel

no study about the reliability and consistency of the original and the translations could be performed until now

## Directions of investigation

- **Reliability:**

- Of the original: are the quotations made by Cantemir grounded? Is there a concordance between his degree of trust in these sources and the current knowledge about them (e.g. is there any evidence that a person which Cantemir claims to have spoken to, really lived in that time?)
- Of the translation against the original; Here an important role have the inserted editorial annotations.

- **Consistency:**

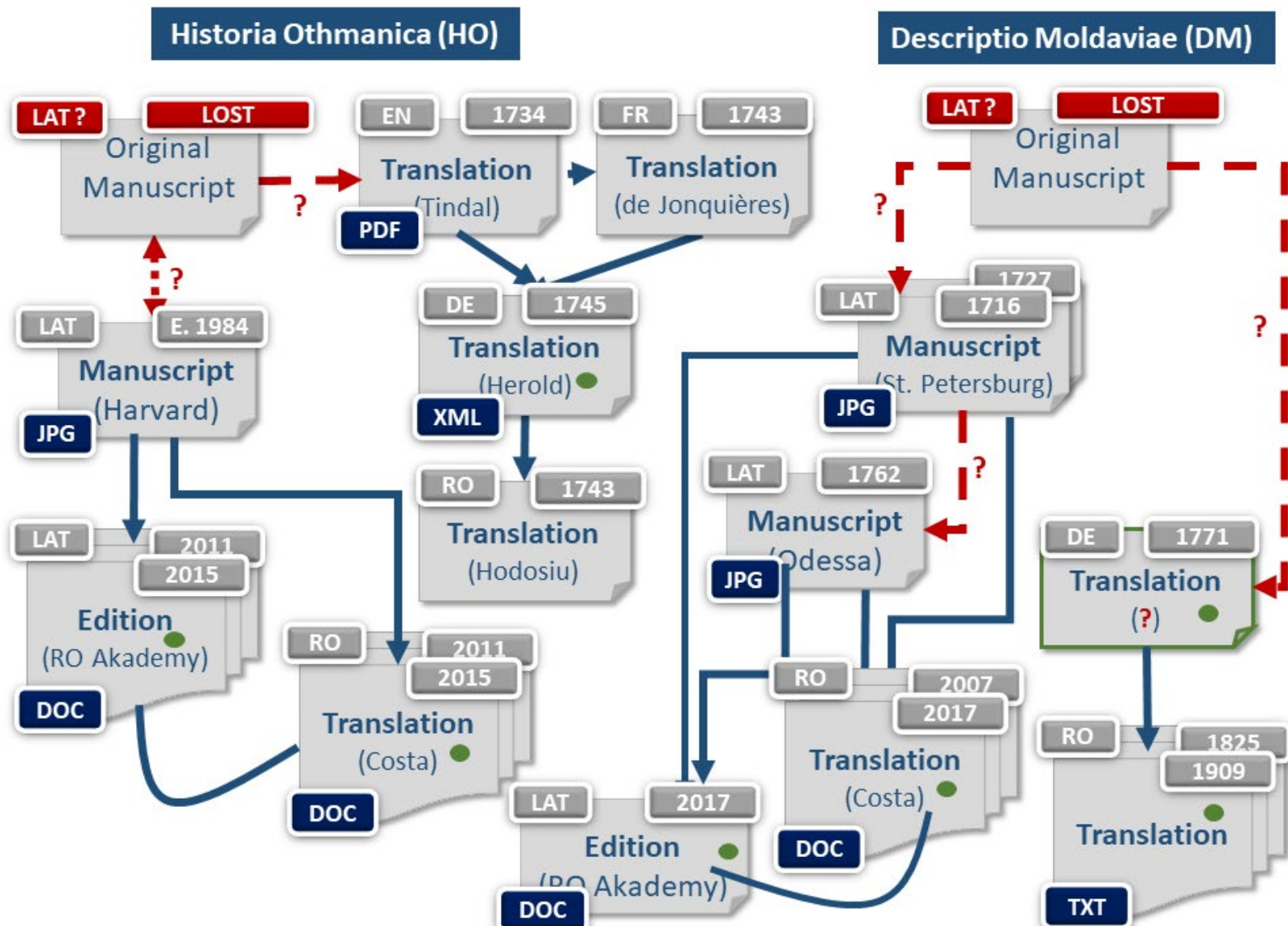
- Within the original: keeps Cantemir a constant opinion about persons, events, facts across the text? (see his own annex with annotations vs. the text)
- Across the 2 “originals”: Are common persons and events described similarly?
- Between original and translation: does the translation preserve the degree of vagueness /certainty stated by Cantemir?

- **Vagueness**

- Political or tactical reasons for imprecise expressions



# Manuscripts, editions, translations- unclear tradition



# Corpus creation – challenges

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

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

- Mixed typefaces
- Mixed scripts

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

## Beschreibung der Moldau.

333

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

dauer mit den Genuesern während ihres Besizes 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*, *παιδευσις*, *Kivernisjre*, *κιβέρνισις*, *Procopie*, *προκοπη*, *Blastern*, *βλασφημῶ*, *azyma*, *ἄζυμον*, *Drum*, *δρόμος*, *Pizma*, *πίζμα*. Da wir nun also bender Partheyen Meinungen vorgetragen haben, so getrauen wir uns nicht zu bestimmen, welche von beyden der Wahrheit am nächsten sey? aus

Furcht,



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

(Dimitrie Cantemir, Descrierea stării Moldaviei în vechime și azi, traducere Ioan Costa 2017)

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

# A Complicated Explicit Example: Ambiguity

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

capital is *Kilia*\*, ? 1 Lycostomon, 2



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

1. Dragoș și măcar că hronografiile noastre nu arată pentru știința neamului său, dar la noi se zice neconținut, 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 adevăra 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 19<sup>th</sup> century after  
the German edition*

Domnul cel dintâi carele după năvălirea lui Batie, a agonisit iarăși strălucirea cea mai dinainte 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 obișnuiesc a-și pune la iscălitură numele Ioan. Și cuvântul acesta este mai ușor de a se adevăra 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		

- 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 is written in Romanian.

Download annotation

Dimitrie Cantemir

## DESCRIEREA MOLDOVEI

### I. PARTEA GEOGRAFICĂ

Click to collapse

### CAPITOLUL I

#### Despre numele cel vechi și cel de acum al Moldovei

Click to collapse

Toată țara pe care o numim astăzi **Moldova**, precum și ținuturile învecinate dinspre asfințit, au fost stăpânite la început de sciți, care cuceriseră aproape trei părți ale lumii, cu toate că, după obiceiul lor strămoșesc, ei nu aveau așezări statornice. Pe lângă feluritele numiri date de hoardele care s-au perindat pe cursul vremii, **grecii** i-au numit pe locuitorii acestor ținuturi când geți, când daci. Sub stăpânirea romanilor s-a statornicit numirea de daci. După ce acest popor pierdu pe regele **Decebal**, biruit de viteazul **Nerva Traian**, și a fost parte nimicit, parte risipit încoace și încolo, întreaga țară pe care o locuia a fost prefăcută în provincie romană, pământul fiind împărțit cetățenilor romani, după care s-a deosebit în trei părți: marginașă, de mijloc și muntoasă. În cea dintâi se cuprindea o parte din **Ungaria** de astăzi și **Valahia**, în a doua **Transilvania**, iar în cea de pe urmă, cea mai mare parte a **Moldovei** noastre, așezată între **Dunăre** și **Pрут**, împreună cu ținuturile de la hotar ale **Valahiei**. În epoca următoare, când **Imperiul roman** a început să decadă, barbarii — și anume sarmații, hunii și goții — au pus în **Moldova** de mai multe ori și au silit pe coloniștii romani să fugă în munți, ca să caute loc de adăpost împotriva cruzimii lor, în partea muntoasă a **Maramureșului**. După ce au viețuit acolo câteva sute de ani, apărați ca într-o cetate naturală, cu voievozii și legile lor, deoarece populația s-a înmulțit peste măsură, **Dragoș**, un fiu al voievodului lor **Bogdan**, se hotărî în cele din urmă către anul () să cuteze un marș peste munți, spre răsărit, însoțit la început numai de trei sute de oameni, ca și cum ar pleca la vânatoare. Pe drum dădu din întâmplare peste un bou sălbatic, numit de moldoveni zimbru, și tot gonindu-l, ajunse la poalele munților. Când căteaua lui de vânatoare, căreia-i zicea **Molda** și pe care o iubea foarte mult, se repezi întârâtată asupra fiarei, bourul se azvârli într-un râu, unde săgețile îl uciseră; dar și căteaua, care sărise în apă după fiara fugărită, fu luată de undele repezi. Întru pomenirea acestei întâmplări, **Dragoș** fu cel dintâi care numi acest râu **Moldova**, iar locul unde se petrecuse ră acestea îi dădu numele de **Roman**, după numele semînției sale și luă ca stemă a noului său principat capul bourului. După aceea, când cercetă ținuturile învecinate și găsi câmpuri roditoare cu ape îmbelșugate, târguri, cetăți 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** își urmară **domnul** de bunăvoie și, cu bucurie, trecură munții împreună cu el, în cete mari, așezându-se după aceea în ținuturile aflate într-un chip atât de minunat, iar pe **Dragoș**, cel care le aflase, îl numiră cel dintâi domn al țării noi. Întoarsă în acest chip stăpânilorilor de odinioară, țara își pierdu, odată cu legile trase din dreptul civil **roman**, și numele dacic și latinesc și atât străinii, cât și locuitorii ei înșiși 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 țările vecine ocupate de dânzii în **Europa**, au dat moldovenilor mai întâi numele de **akvlach**. Apoi, când **Bogdan**, după voia cea din urmă a tatălui său **Ștefan cel Mare**, le-a închinat țara turcii, obisnuiti să numească țările supuse după numele domnitorilor, au început să-i zică **Bogdan**; totuși 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)

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 bestätigt 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 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 Batiu, a agonisit iarăși strălucirea cea mai dinainte a Moldovei a fost:

1. Dragoș și măcar că hronograful nostru 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

Dragos = belongs\_to Moldavian kings  
 Dragos = son\_of Bogdan  
 Bogdan = son\_of Johanniss  
 Dragos = has\_additional\_name Johanniss  
 Bogdan = has\_additional\_name Johanniss

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

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

1. Dragoș. Obgleich unsre Jahrbücher sein Geschlechtsregister nicht angeben, doch eine beständige Sage bey uns aus dem alten königlichen moldauischen Stamme **gewesen sey**, und den Bogdan zum Vater gehabt habe, welcher ein Sohn des Johanniss war, von welchem alle Fürsten den

Dragosch ≈ belongs\_to Moldavian kings  
 Dragoș = son\_of Bogdan  
 Bogdan = son\_of Johanniss  
 Dragoș = has\_additional\_name Johanniss  
 Bogdan = has\_additional\_name Johanniss

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

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 bestätigt 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 beide Meynungen vereinigen wenn man zwey Auswanderungen annimmt, die eine in **der letzten Hälfte des Zwölften**, die andere in der **ersten Hälfte des dreyzehnten 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

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

1. Dragosch. Obgleich unsre Jahrbücher sein

Geschlecht **Legend** es

doc Dragosch **most\_probable** belongs moldavian kings  
den Dragosch discovered Moldau  
gew Dragosch has\_acitivity hunting  
hab Dragosch has\_activity development  
wel Development takes\_place after Batia invasion  
ihre

desto mehr Glauben bezumessen, 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 **Translator** ne

zu fo Dragosch has\_activity moved

Movement takes\_place > ≈ 1150; < ≈ 1200

Movement takes\_place > ≈ 1200; < ≈ 1250

Bathy invasion takes\_place ≈ 1250

Bathy has\_alternative\_name Bat

Bathy is\_a Tatar

## HerCoRe – Multiple-faceted fuzzy annotation

**Cantemir**

Dragosch ≈ son\_of Bogdan  
Bogdan ≈ son\_of Johannis  
Dragosch ≈ has\_additional\_name Johannis  
Bogdan ≈ has\_additional\_name Johannis

thet hatten, daß Dragosch erst nach

athy oder Batu Einfall, D.i. ungefähr

us Siebenbürgen ausgewandert ist;

er lassen sich beide Rechnungen

wenn man zwey Auswanderungen

e eine in der letzten Hälfte des

e andere in der ersten Hälfte des

Jahrhundert (B.)

## Sources /levels of vagueness to be annotated

### 1. Linguistic markers for vagueness

### 2. Factual uncertainty

2.1 References to external written materials (publications)

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

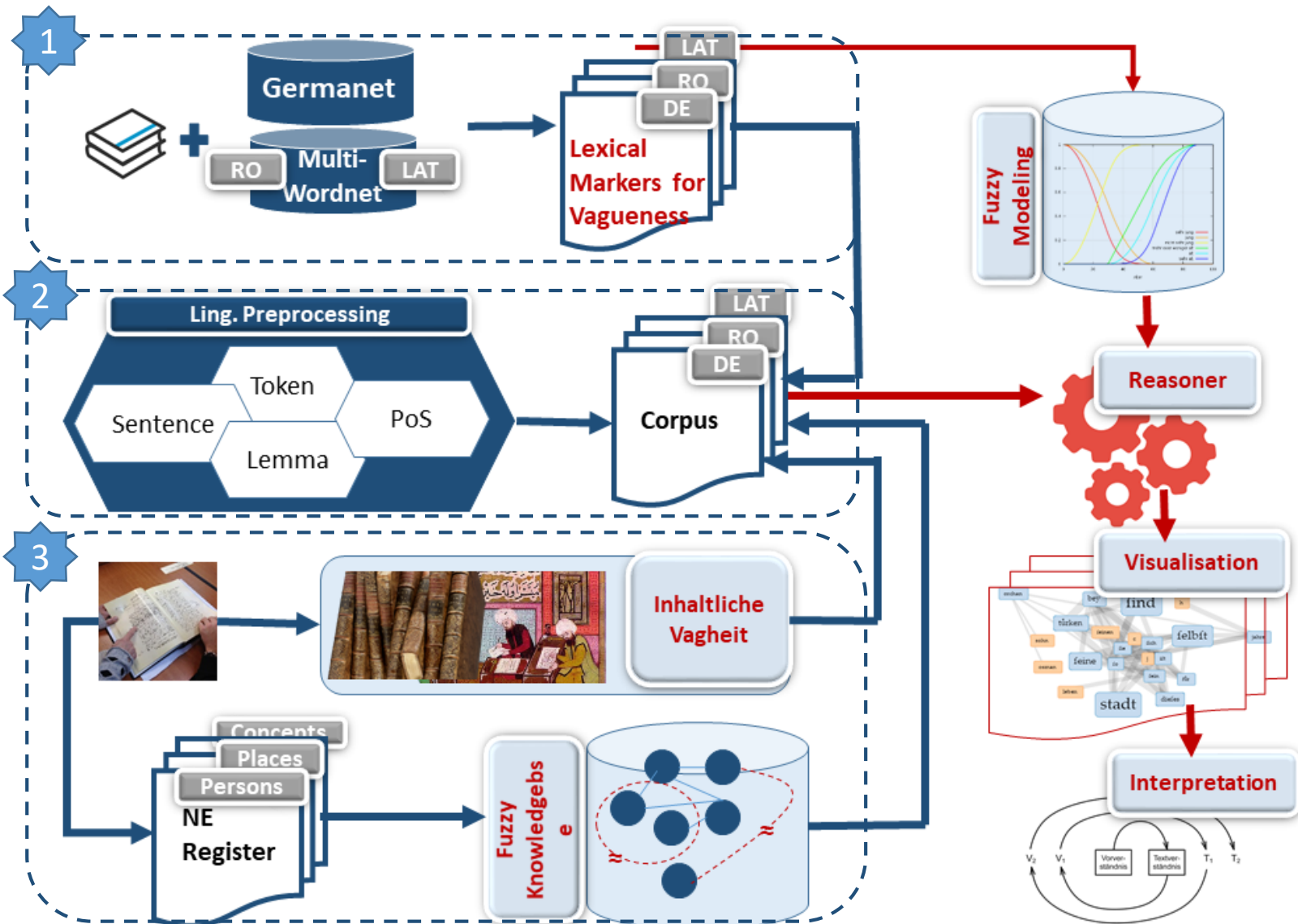
4.1 genre

4.2. author

4.3 translation

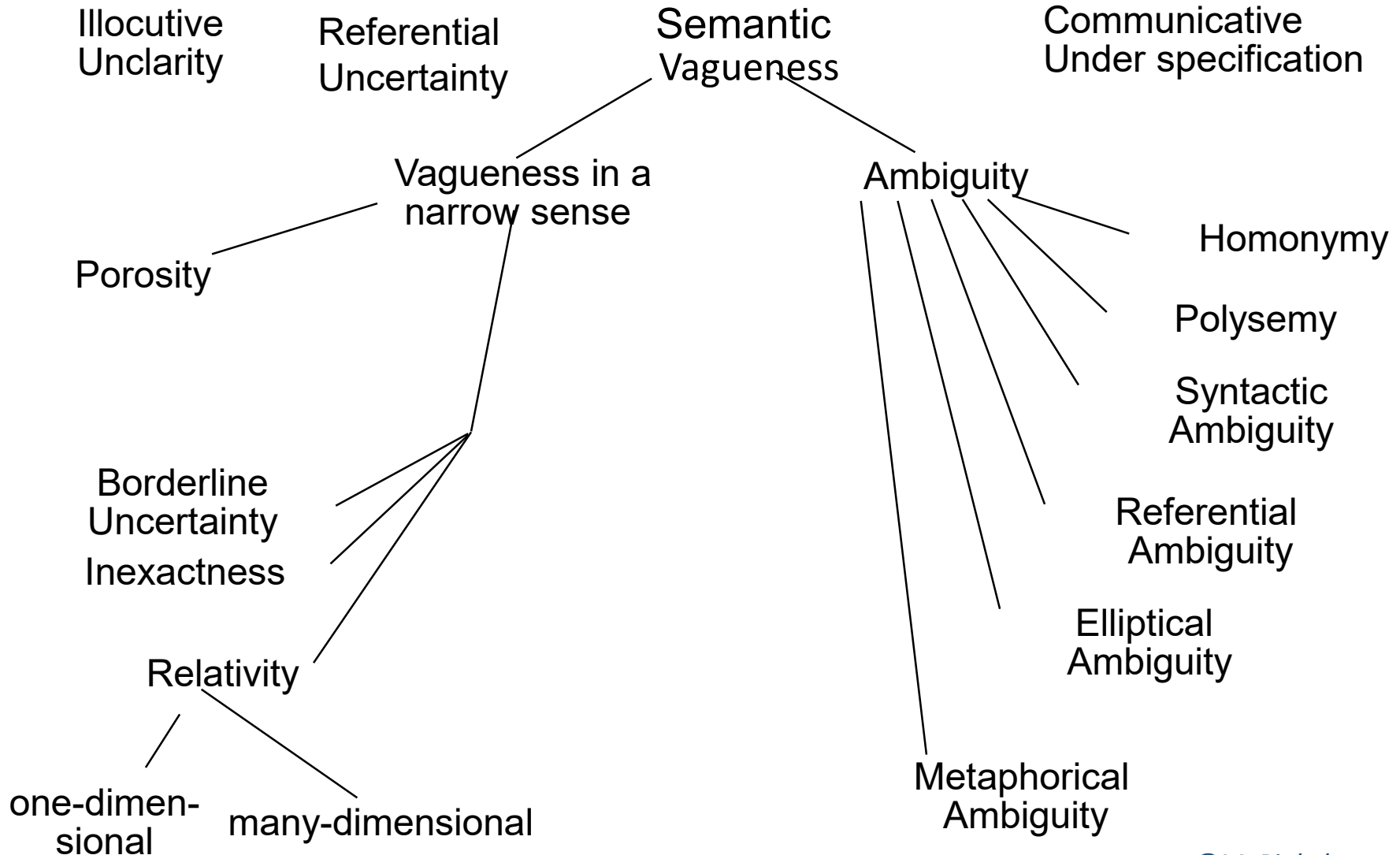
4.4. tradition path

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



# M. Pinkal's Schema of Semantic Vagueness

1



## Quotation

## Lexical and Syntactic Sources of vagueness

Hactenus Gregoras: ad cuius verba observare haud extra propositum erit τὴν πρώτην, quam Gregoras vocat „Tartaria” esse, quam hodie vulgo „Mare” appellat, eiusque incolarum nomina, historicis recenseantur, tamen adscita magis, aut ab exteris indita, quam propria eis, dum in suis sedibus peculiaria fuisse. Ita, si quis principi Praefatione legerit Principes in duas stirpes fuisse divisos, „Aliothman” unam, et „Ali Dzengiz”<sup>1</sup> alteram, neque ab ipsis horum generum conditionem appellationem iam apud aliquos valuisse. Vti enim absonum videtur, Aliothmanos Suleimano parentes ab huius nepote integro post saeculo iis imperabunt fuisse sortitos; ita non minus falso vulgo praedicantur Tartarorum Crimensium Principes ab ipso Dzengizchano „Alidzengiz” appellationem retinuisse.

More plausible

Would have been...

seems unlikely

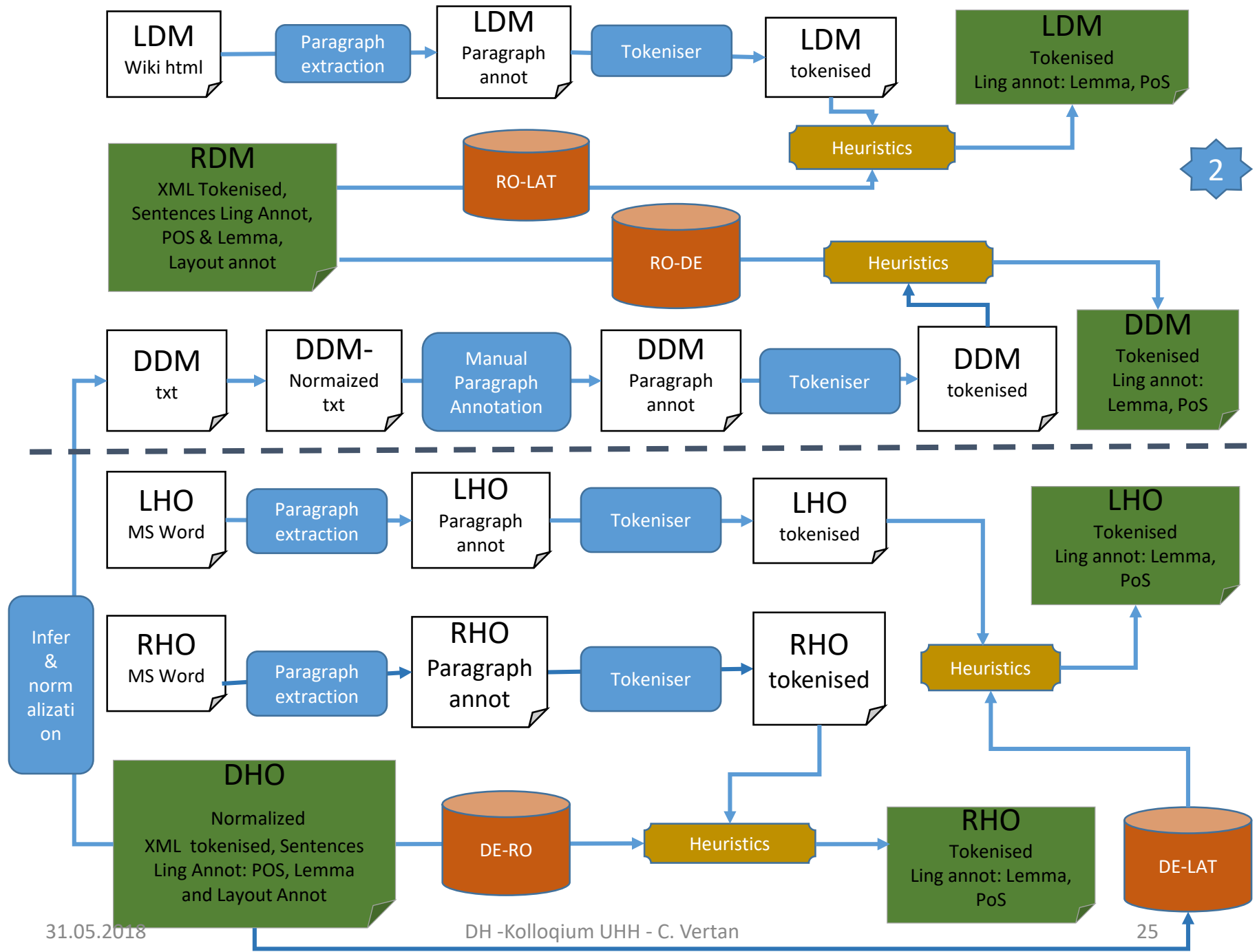
equally false

- Până aici l-am citat pe Gregoras: față de cuvintele lui nu va fi nepotrivit să observăm că acea Tartaria „ἡ πρώτη”, pe care o numește Gregoras, este chiar aceea pe care o numim îndeobște cea „Mare”, iar numele locuitorilor ei, chiar dacă sunt înregistrate de istorici, au fost totuși mai degrabă împrumutate sau date de străini decât proprii lor, purtate întocmai pe vremea când se aflau în sălașurile lor. Astfel, dacă va fi citit cineva în Prefața pusă înaintea acestui tratat că principii neamului oguzilor au fost împărțiți în două stirpe, una „aliothmană”, cealaltă „alidzengiză”, să nu creadă că denumirea aceasta era de-acum valabilă pentru întemeietorii acestor neamuri. Căci, după cum pare nepotrivit ca aliothmanizii care i se supun lui Suleiman să-și fi ales numele de la nepotul acestuia, care a domnit peste ei după un secol întreg, la fel de fals se spune îndeobște că principii tartarilor din Crimeea și-ar fi păstrat denumirea „alidzengiz” chiar de la Dzengizchan

## Selected markers for linguistic vagueness

1. comparatives, inexact adjectives e.g. *“mehr” “größer”, “älter”*
2. non-intersectives e.g. *„vermeintlich“, „so-genannt“*
3. Hedges e.g. *„ziemlich“, „einigermaßen“, „etwa“*
4. inexact measures *„4 Tagereisen“, 10 Fuß“*
5. modals (attitudes) e.g. *„vielleicht“, „hoffentlich“* subjunctives v+analyse
6. lexical quotation markers
7. vague quantifiers e.g. *„viele“, „meistens“*
8. complex quantifiers e.g. *“etwa die Hälfte von den 20-30 tausend Soldaten”*
9. numbers
10. range expressions e.g. *“Anfang des 18. Jhds.”*
11. unclear place *„Syrfia“, „Moramor“*
12. unclear person e.g. *„der ehemaligen Herzog“*
13. unclear time e.g. *„in alten Zeiten“*
14. Domain specific e.g. *„Wesir“ vs. „Wesire“*





## Historians and Sources Mentioned by Cantemir

- Giovanni Battista Riccioli (on Muslim calendar)
- Franciscus a Mesgnien Meninski (1598-1671), *Thesaurus Linguarum Orientalium, Turcicae, Arabicae, Persicae* (1680)
- Philipp Lonicerus (1532-1599), *Chronicorum Turcicorum* (1584)
- Hoca Saadeddin Efendi (1536/7-1599), *Tac üt-Tevarih* (1520?)
- Solakzade Mehmet Hemdemi Efendi (1590-1657), *Tarih-i Solakzade* (1660?)
- Mehmed Neşrî (Hüseyin bin Eyne Bey?) (?-1520), *Kitâb-ı Cihannümâ* (1485?)
- İbrahim Peçevî (1572-1650), *Tarih-i Peçevi* (1640?)
- Hezarfen (Hezarfen Hüseyin Efendi) (?-1691/92), *Tenkîh-i Tevarih-i Mülük* and *Telhisu'l-Beyan fi Kavanin-ı Al-i Osman*
- Âşıkpaşazâde Dervîş Ahmet Âşıkî (1400-1484), *Tevarih-i Ali-i Osman* (1478?)
- Nicephorus Gregoras (1295-1360), *Roman History* (1359?)
- Kalkondilas Lanikos
- Şeyh Sadi (1210-1292), *Gülîstan* (1258)
- Seyyid Nimetullah Efendi (Nakibzâde), *Tuhfe-i Ni'meti* (Persian dictionary) (1637)
- **Lexicon Persicorum Turcicum?**

## Manual Annotation of Factual uncertainty

- [...] 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 uncertainty

## 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 türkischen Schriftsteller betrachtet; so siehet man offenbar, daß es bloße Erdichtungen sind. Denn diese legen **einstimmig** Bajeßid vier Söhne bei: Mustafa, der in der Schlacht mit den Tatarn um das Leben kam, Süleyman, Musa und Muhammed. [...] So viel ist wenigstens **gewiß**, daß nicht mehr als vier Söhne Bajeßids in der ganzen Geschichte 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)

DHO

LHO

Turkish  
Sources

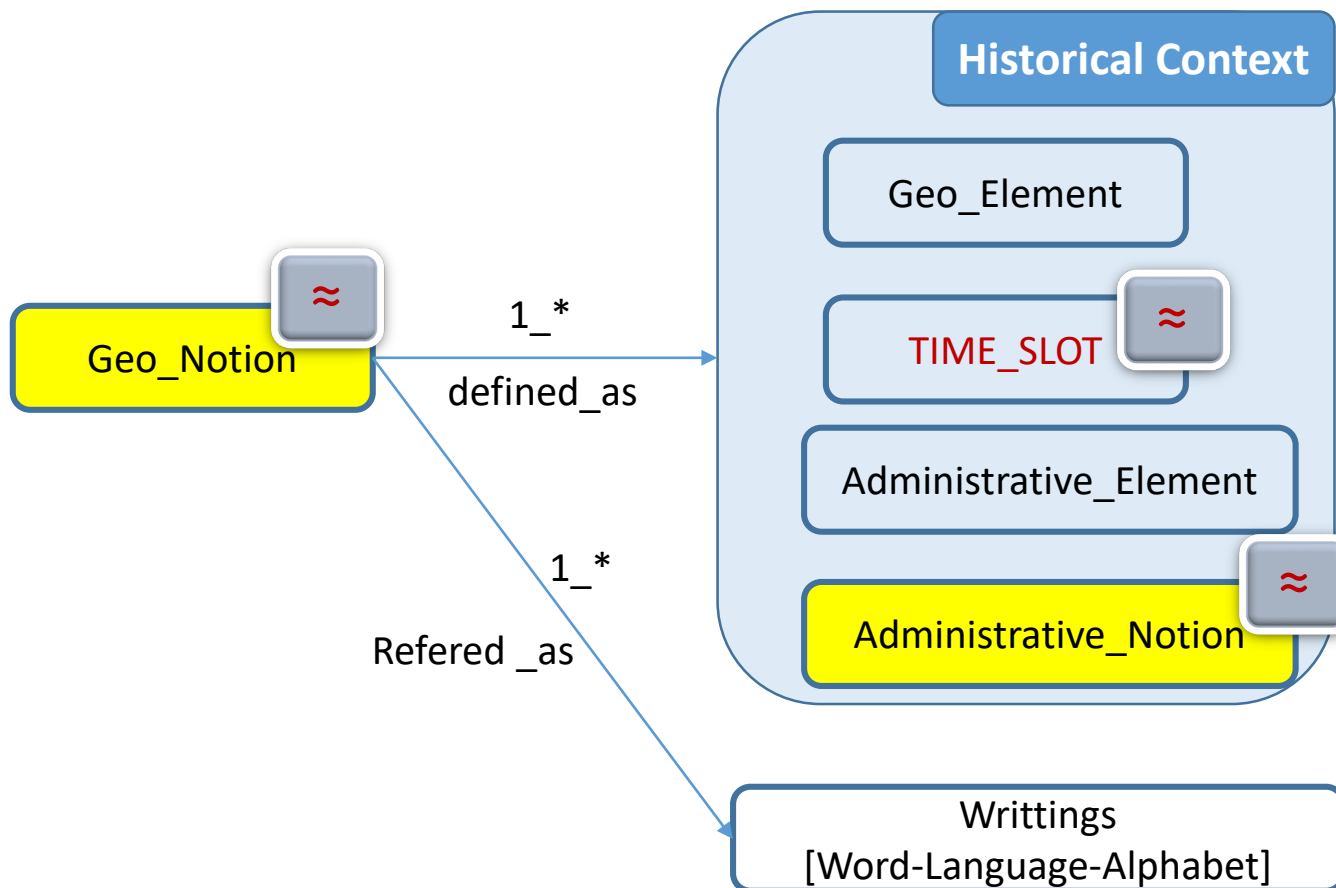
Seraskjer·Sülejman·Pascha	65	Seraskerio·Suleiman·Baszae	38	سر عسكر سليمان پاشا Serasker· Süleyman·Paşaa
Istifan	66	Istefan	38	
Matthias·(Ann.)	66	Mathiam·Corvinum	34	
Bajeβid·Yildirim·(Ann.)	66	Ildirim·Baiezig	34	ييلدرم بايزيد Yıldırım·Bayezid
Heβarfenn·(Ann.)	66	Hezáfenn	34	
Karaman·Ogli	67	Caramanougly	39	قرا مان او غلي Karamanoğlu
Bajeβid	67	Baiezig	39	بايزيد Bayezid

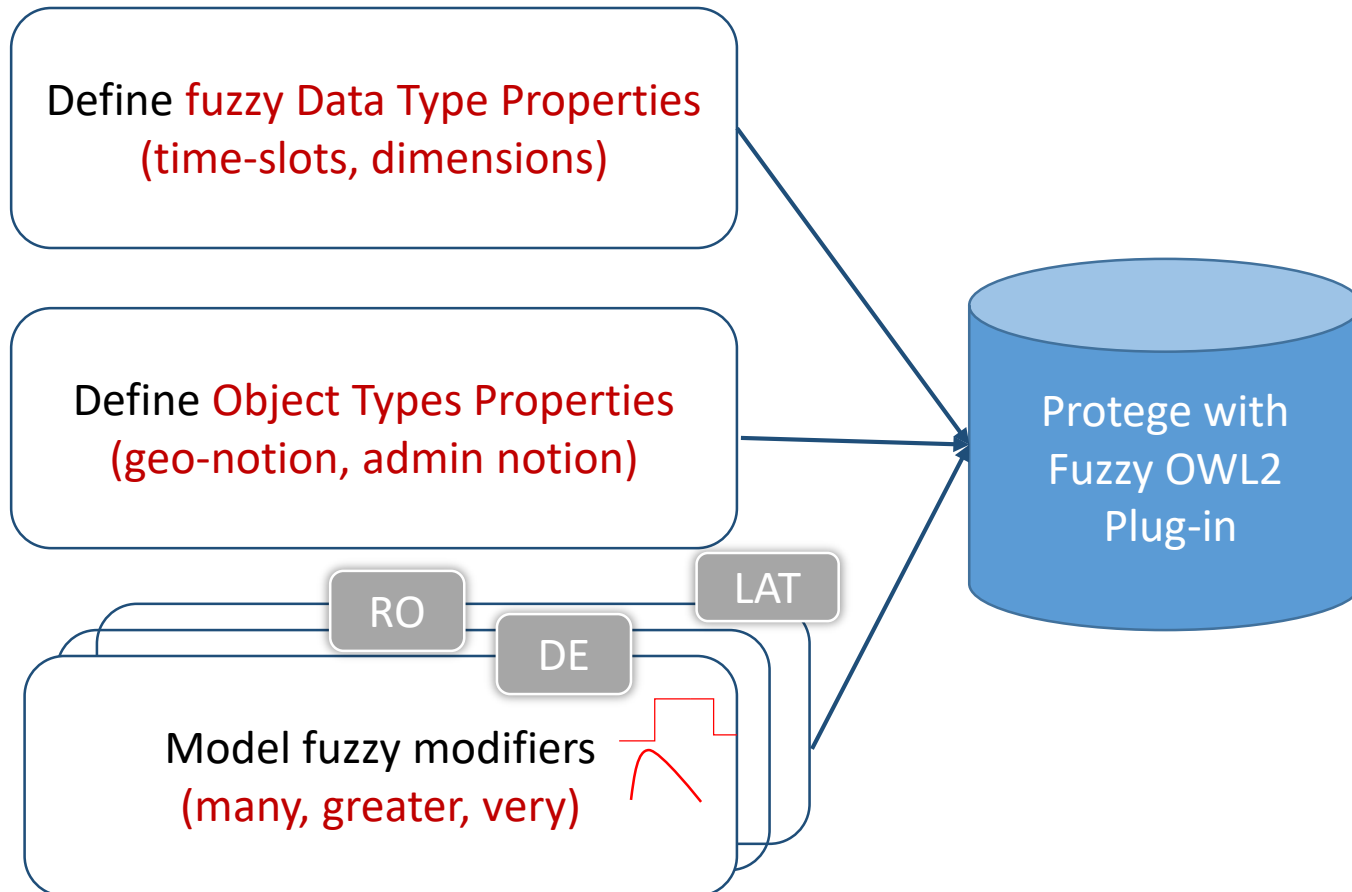
Places Database

Adrianopol	73	Adrianopol	44	ادرنه·Edirne
Edrene·(Mewlasi)	73	Edrne·(Molasi)	44	ادرنه·Edirne
Edrene·(Mewlasi)·(Ann.)	73	Ederne·(Mollasi) (ادرنه·مولاسي)		ادرنه·Edirne
Adrianopol·(Ann.)	73	Adrianopolis	38	ادرنه·Edirne
Misr·(Mewlasi)·(Ann.)	73	Mýsr·(Mollasi) (مصر·مولاسي)	38	مصر·Mısır
Bursa·(Mewlasi)·(Ann.)	73	Birússa·(Mollasi) (بئروسه·مولاسي)	38	بئروسه·Bursa
Haleb·(Mewlasi)·(Ann.)	73	Halèp·(Mollasi) (حلب·مولاسي)	38	حلب·Haleb·/·Halep
Dawud·Pascha	74	Daud· Pasza·/· Davùd· (Ann.)	44· 38· (An n.)	

Persons Database









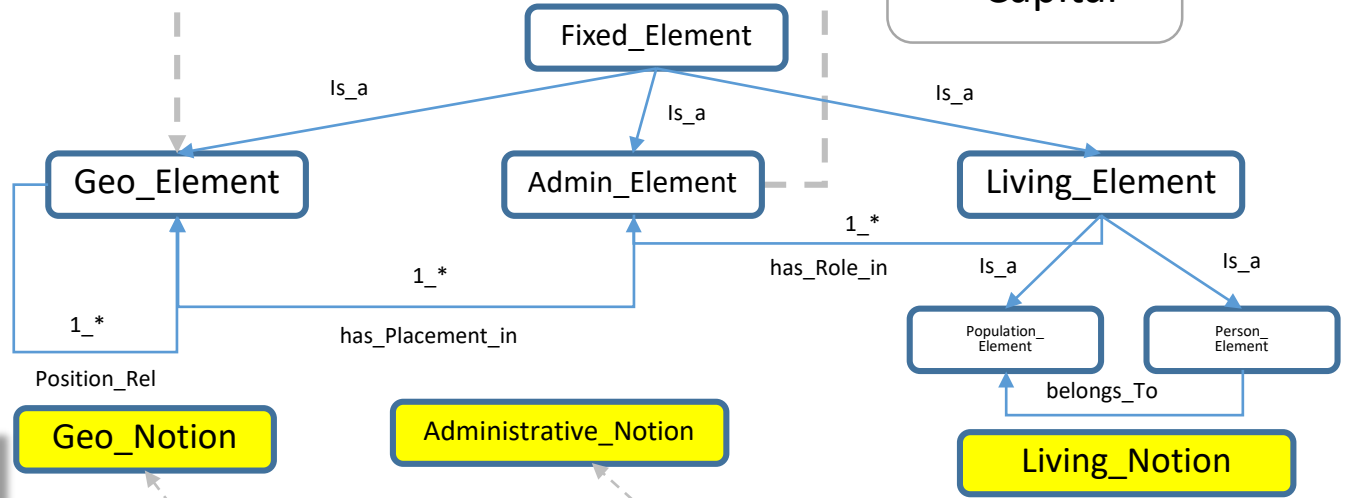
## Example: Annotation of ambiguous places



Ortelius  
Map 1570

- Northern Dobrudja
- Western Macedonia
- Eastern Europe

- Country
- Capital



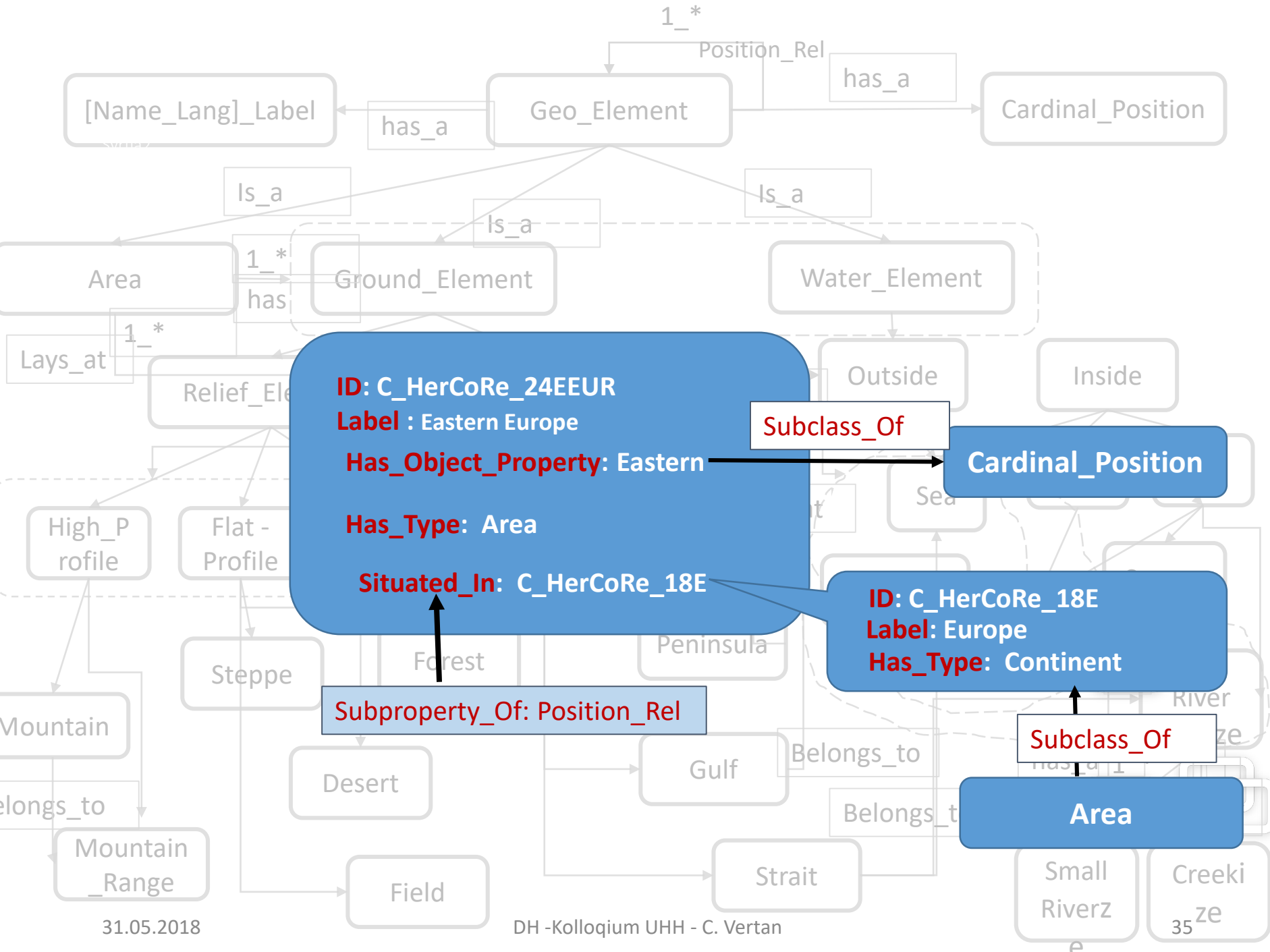
**Syrfia** is the abandoned name of a region in Eastern Europe, used on historical maps until 17th century, designating

- 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

- Turkish Times
  - Greek times
  - 17<sup>th</sup> century
- Fuzzy Properties



**ID:** C\_HerCoRe\_24EUR  
**Label :** Eastern Europe  
**Has\_Object\_Property:** Eastern  
**Has\_Type:** Area  
**Situated\_In:** C\_HerCoRe\_18E

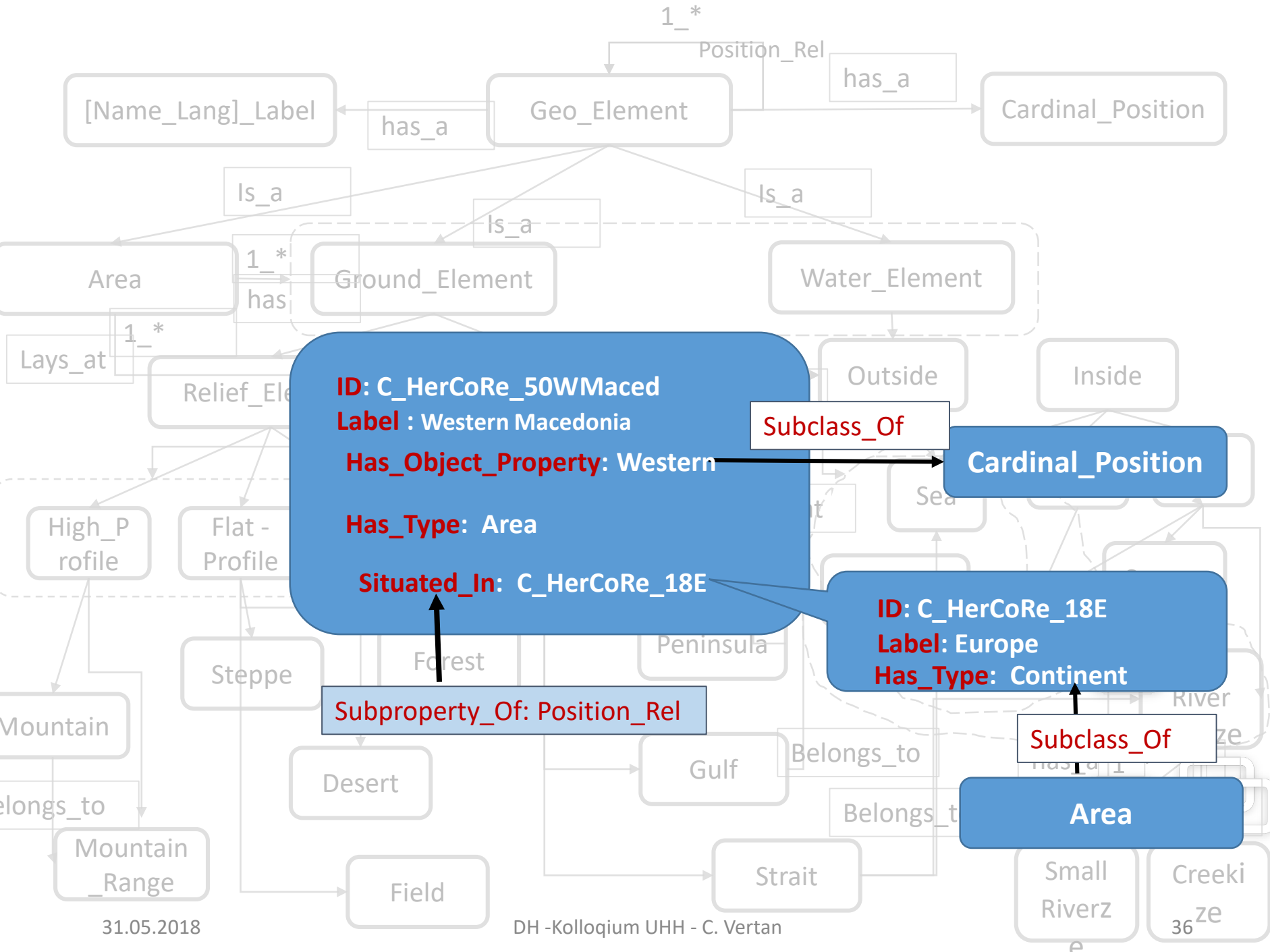
**Subclass\_Of**

**Cardinal\_Position**

**ID:** C\_HerCoRe\_18E  
**Label:** Europe  
**Has\_Type:** Continent

**Subproperty\_Of: Position\_Rel**

**Subclass\_Of**  
**Area**



**ID:** C\_HerCoRe\_50WMaced  
**Label:** Western Macedonia  
**Has\_Object\_Property:** Western  
**Has\_Type:** Area  
**Situated\_In:** C\_HerCoRe\_18E

Subclass\_Of

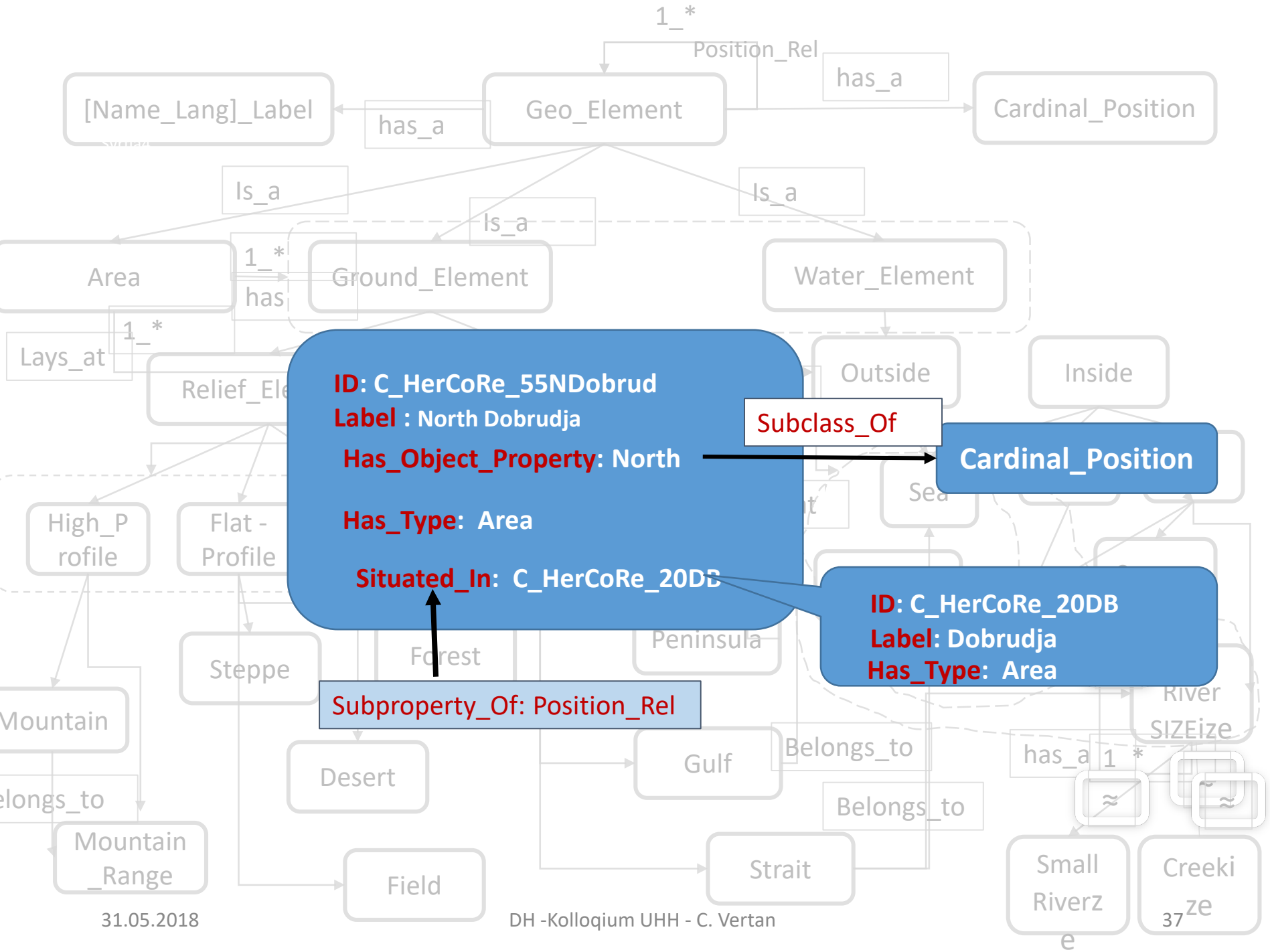
**Cardinal\_Position**

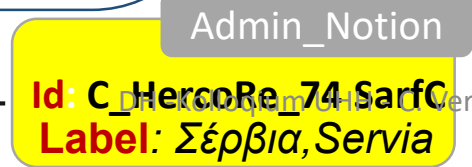
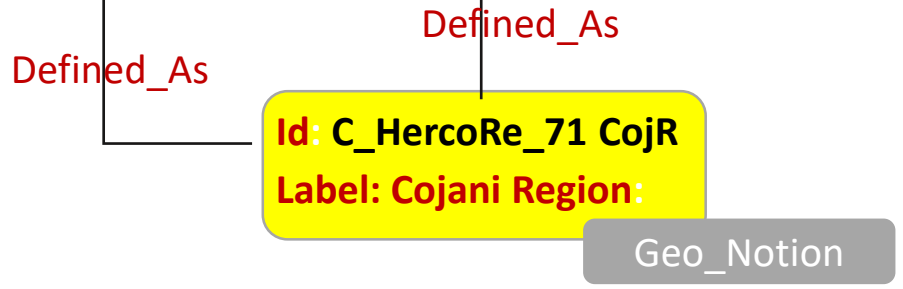
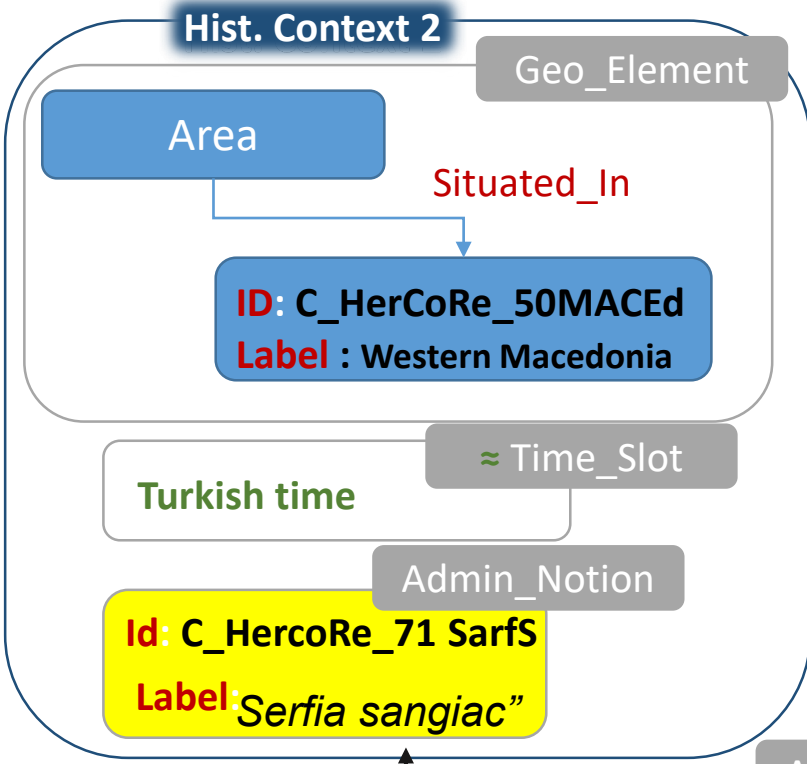
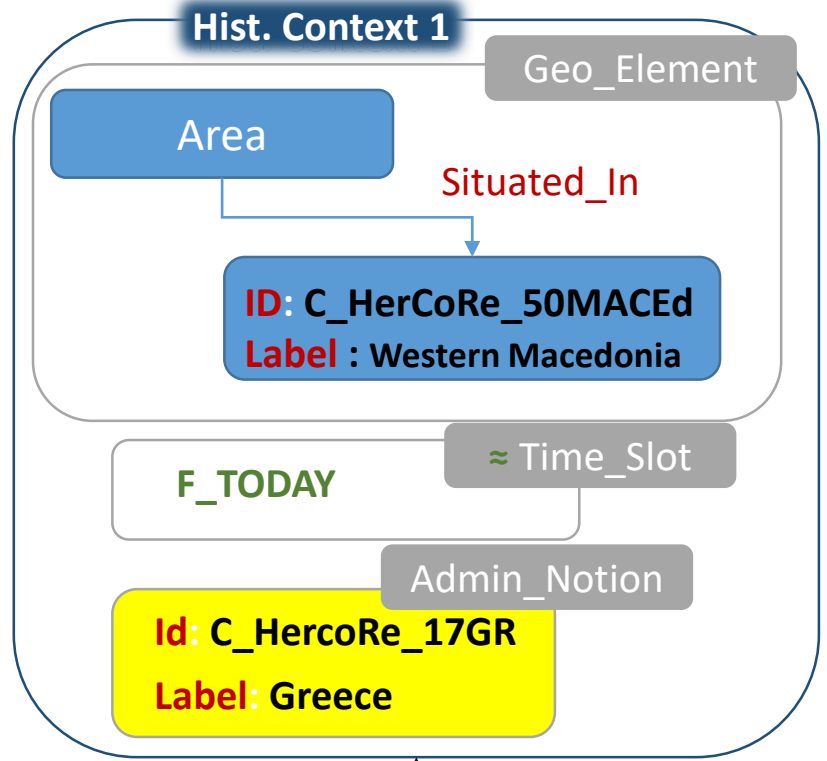
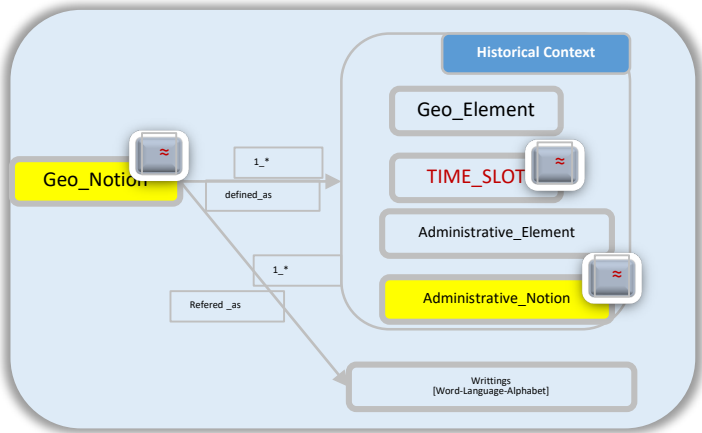
**ID:** C\_HerCoRe\_18E  
**Label:** Europe  
**Has\_Type:** Continent

Subproperty\_Of: Position\_Rel

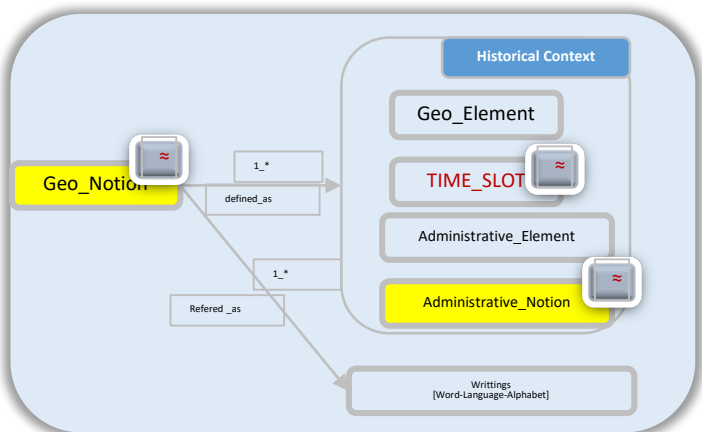
Subclass\_Of

**Area**





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



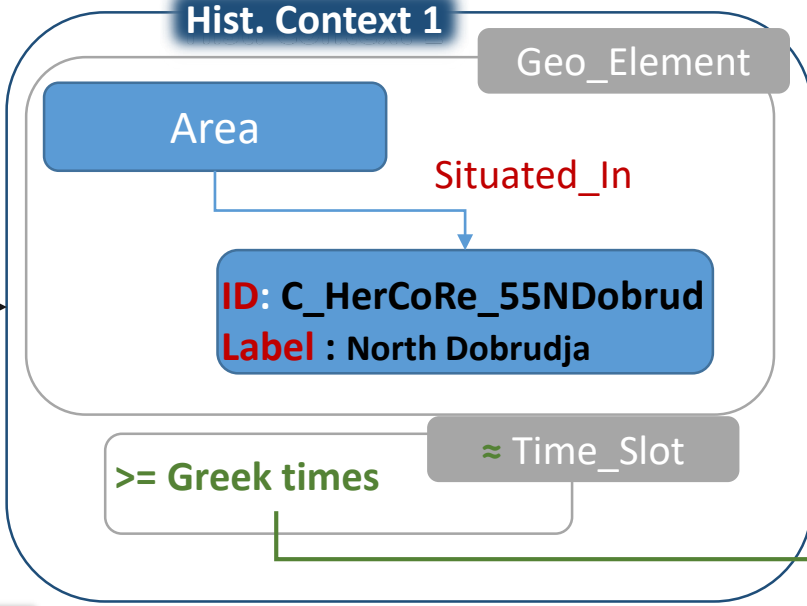
```

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

Geo\_Notion

**Id: C\_HerCoRe\_10 DSyrf**  
**Label: Σύρφοι, Syrphoi**

Defined\_As



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

## Class ( Syrflia 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.

31.09.2018 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**

40





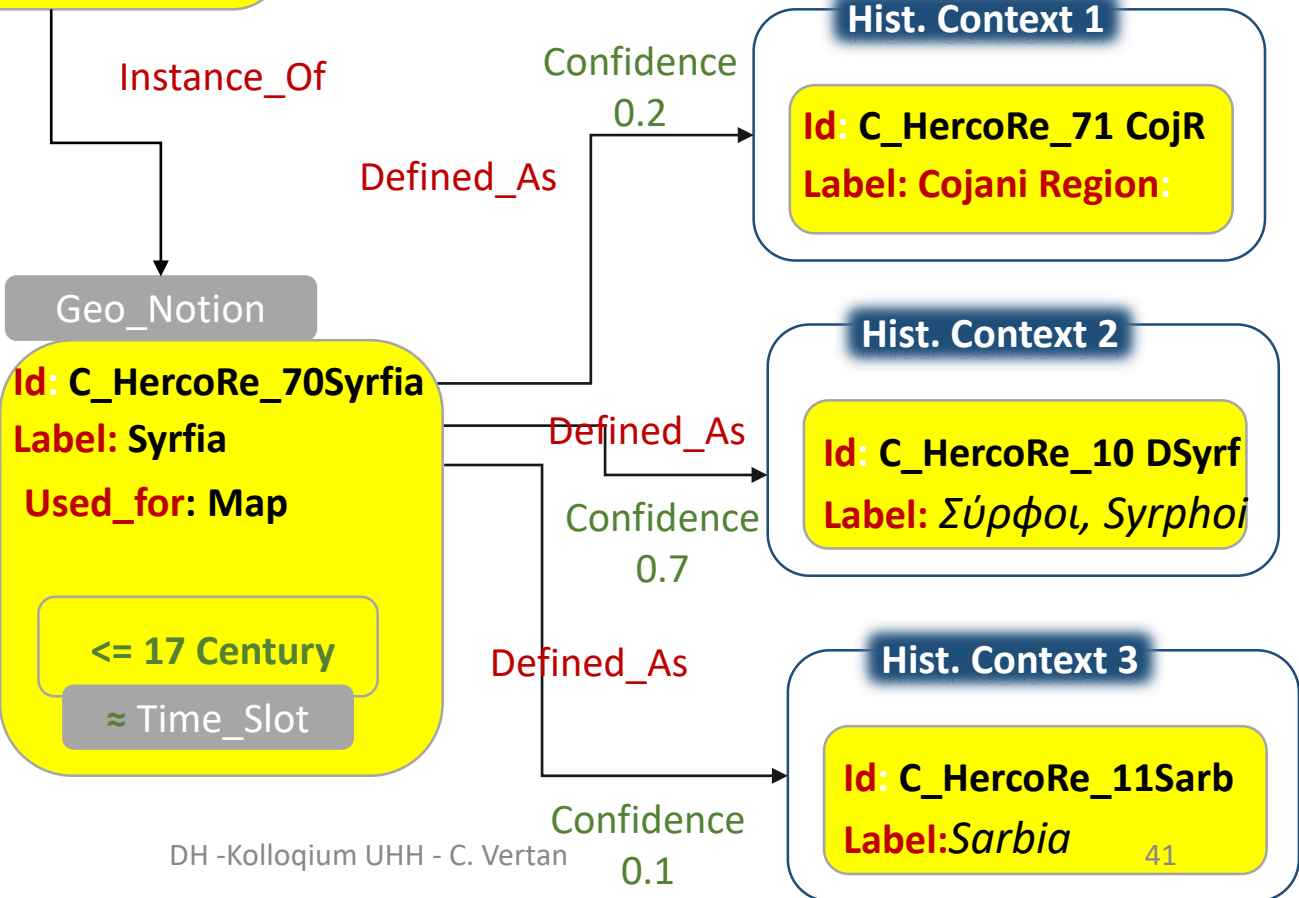
Orteliusmap 1570

**Id:** I\_C\_HerCoRe\_Sy  
**Label:** Syrfia\_Ortelius  
**Used\_for:** Map\_Ortelius  
**Time\_slot:** 1570

**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.

31.09.2018 Source: Wikipedia



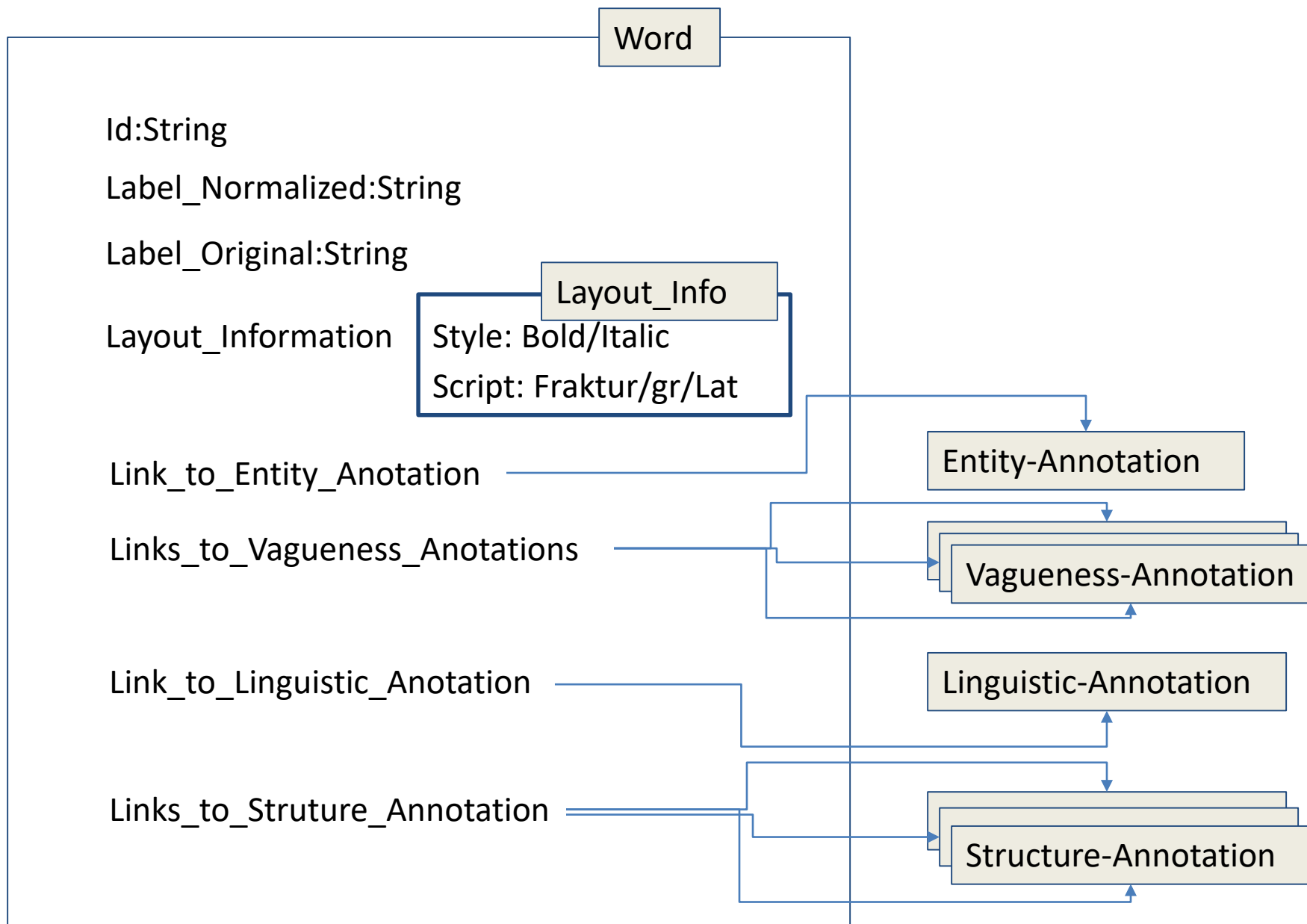
Orchan having in his Father's Life-time (as it is said) taken Prusa (2), and subdued the Territory of that City to his dominion, spends the first year of his Reign in settling the affairs of Afia, and establishing his new Empire

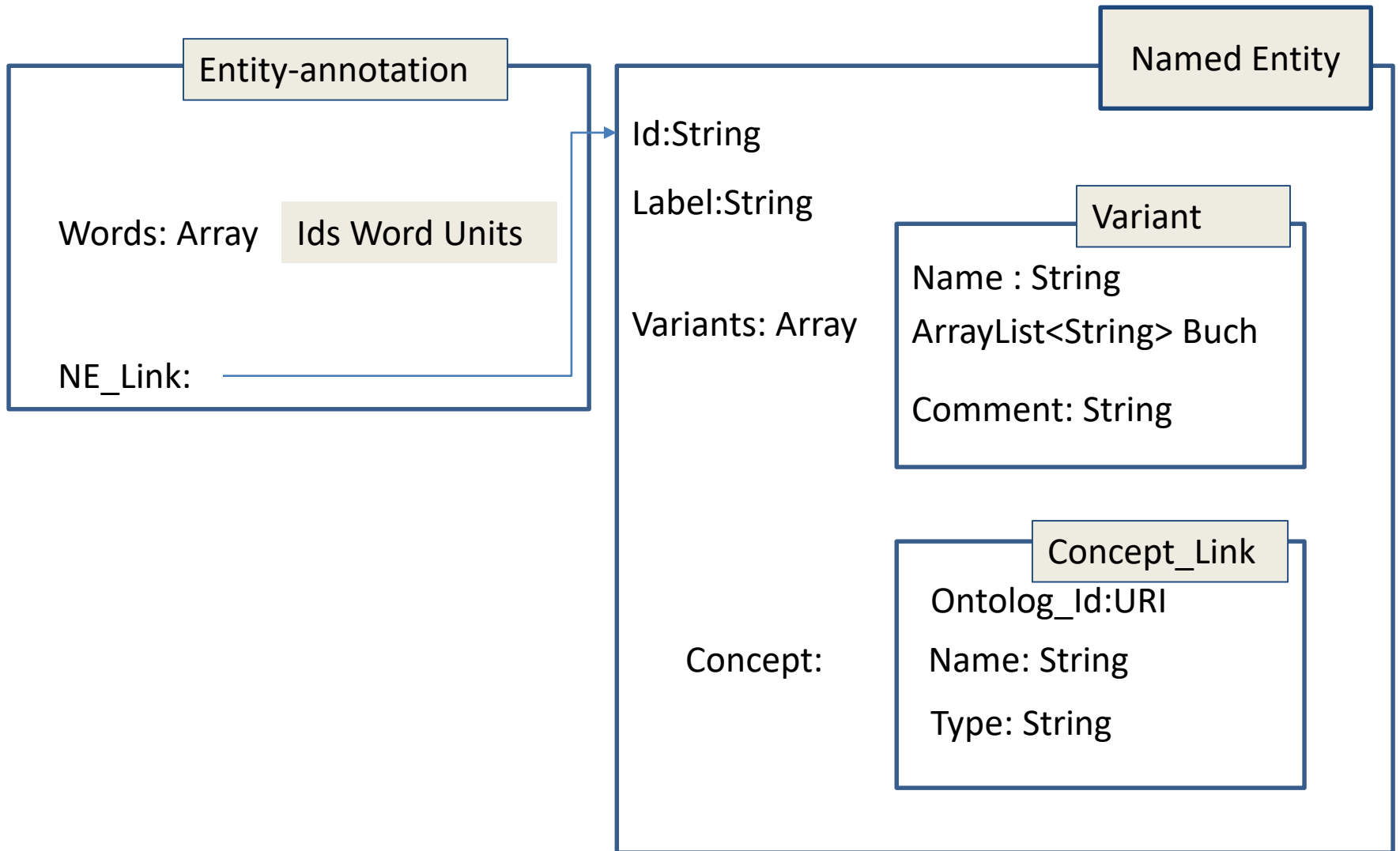
green = linguistic annotation ( N., V, Prep, ...)  
yellow = from the ontology  
orange = vagueness marker.

(2) [Having taken Prusa] The Christian Prusa to the time of Othman, who they tell us, died the following year. This mistake seems to arise from the loss of Prusa (which was a very great calamity) being known to Greece before the news of Othman's death could arrive there .

# Annotation-Tool Requirements

- Several Layers of Annotation (e.g. linguistic, editorial, text structure, domain specific).
- Annotation layers are interconnected
- Synchronisation between different text variants (original, translation editorial remarks)
- Discontinuous annotation segments
- Controlled automatisisation of manual annotation
- Need of user-friendly annotation interfaces
- Modular Architecture flexible at changes (new layers, new annotation categories)
- Import of automatic annotation (basic linguistic, basic vagueness)





## Vagueness-annotation

Id:String

Type: String (Quotation /Linguistic / Edition /Geo/Genre)

Subtype: String (dependent on each type)

Words: Array Word\_Id

Confidence: String (low/medium/high)

## Linguistic-annotation

Id:String

Lemma: String

PoS: String

Morpho\_Features: String

Words: Array Word\_Id

## Structure-annotation

Id:String

Level: String (Chap/Paragraph/Sentence)

Type: String (Author/Editor)

Words: Array Word\_Id

S1  
Level: Sentence  
Type: Author  
Words: W1,W2,W3,W4,W5,W6,W7,W8,W9,W10,W11

W1 Die Die n <b>frak</b>	W2 Moldau Moldau b <b>frak</b>	W3 hat hat n <b>frak</b>	W4 nicht nicht n <b>frak</b>	W5 zu zu n <b>frak</b>	W6 allen allen n <b>frak</b>	W7 Zeiten Zeiten n <b>frak</b>	W8 eben eben n <b>frak</b>	W9 dieselben dieselben n <b>frak</b>	W10 Grenzen Gränzen n <b>frak</b>	W11 gehabt gehabt n <b>frak</b>
--------------------------------------	--	--------------------------------------	--	------------------------------------	--	--	--	--	---	---

Ling Ann :L2  
Lemma:Moldau  
PoS :N  
Morpho:  
Words: W2

Vagueness Ann :V1  
Type: ling  
Subtype : time expression  
Confid: low  
Words: W5,W6,W7

Ent Ann :E1  
Words: W2  
NE20

NE20  
Label: Moldova  
Var: Moldau (DDM)  
URI: https://..  
Name: placeName  
Type: country

Ling Ann :L3  
Lemma:haben  
PoS :Verb  
Morpho: 3.sg. Perf.  
Words: W3,W11

# Work ahead

- Semantic Linking between the map and „Geo\_notions“
  - Formal representation of vagueness markers
  - Adaptation of existent fuzzy reasoners
  - Specification of user interface
  - Visualisation of results
  - Reuse and enhance the ontology
- AND..
- the entire hermeneutic analysis

