



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

DEPARTMENT INFORMATIK

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

# Combining digital and hermeneutic approaches for investigating source quotations in early modern historical texts

**Cristina Vertan**

cristina.vertan@uni-hamburg.de

# Vagueness /Uncertainty in Digital History

Data  
criticism

- ? Which kind of digital object are we using,
- ? What happens with the intrinsic vague/uncertain nature of humanities' objects,

- ? How much „vagueness“ can we afford to encode,
- ? How to encode vagueness and uncertainty and how to manage in a digital environment,

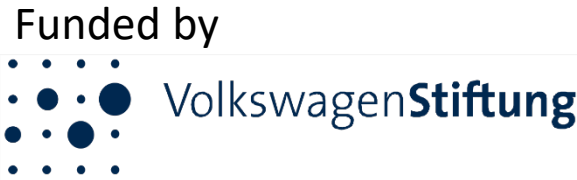
Algorithm  
criticism

- ? Are the tools used managing the encoded vagueness.

Tool  
criticism

# 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

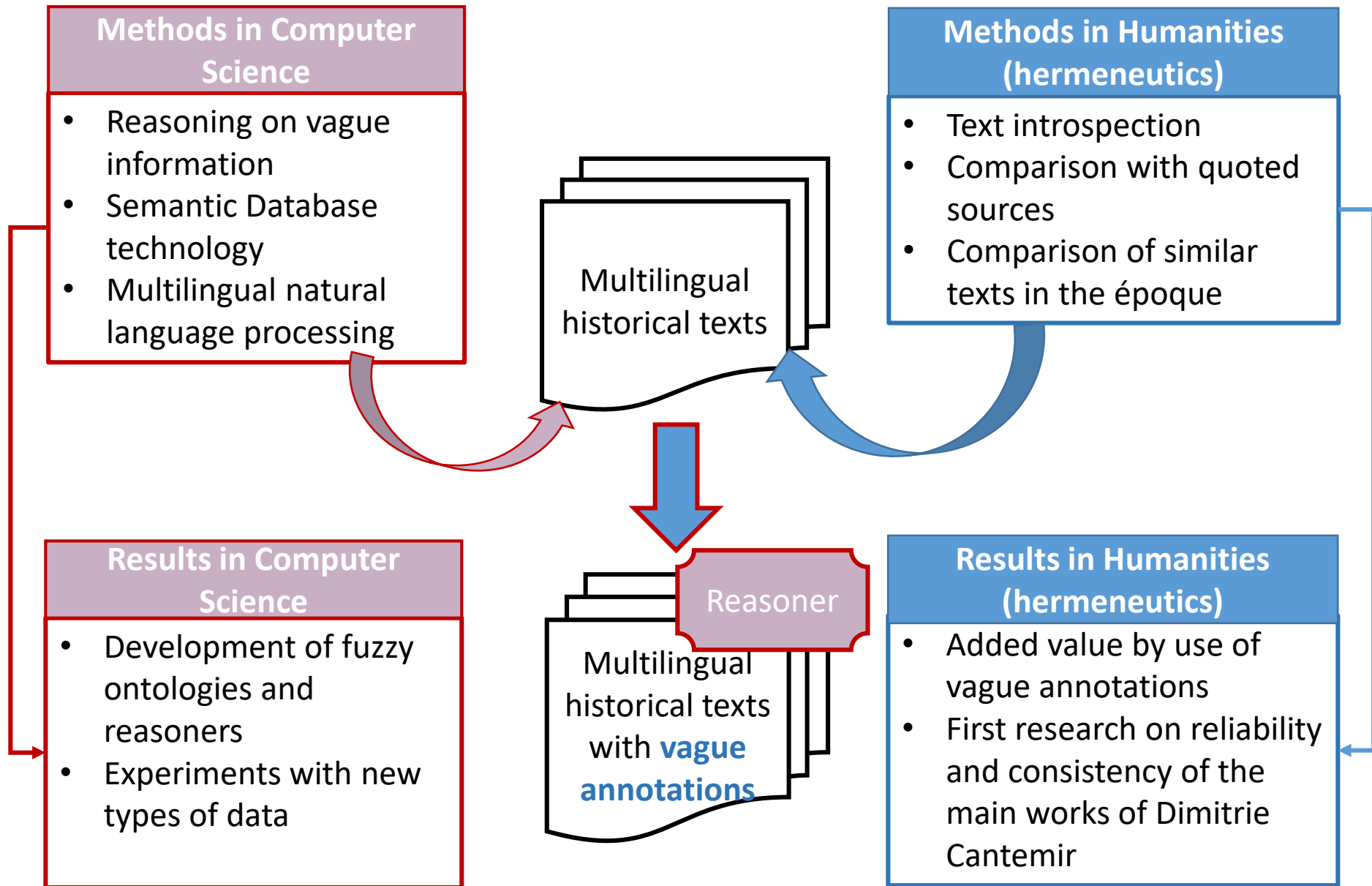
Compare for the first time “original” with translations done in the 18th- 19th century

(In)Validate assumptions about source quotations in original text

CS

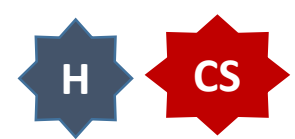
Demonstrate how to include vagueness and imprecision in annotations and interpretations engines

Progress work in automatic recognition of vague expressions





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



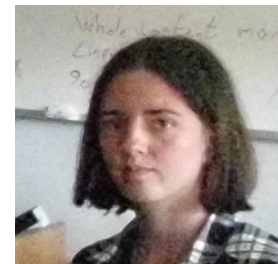
Prof. Dr. Yavuz Köse, UHH  
Turcology



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



Alptug Güney, UHH  
Turcology



Alina Ciobanu, UB  
CL, CS

# Outline

- Rationale of the project
- Corpus' Insight
- System architecture
- Processing “factual uncertainty”

## 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 (Tindal), 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.

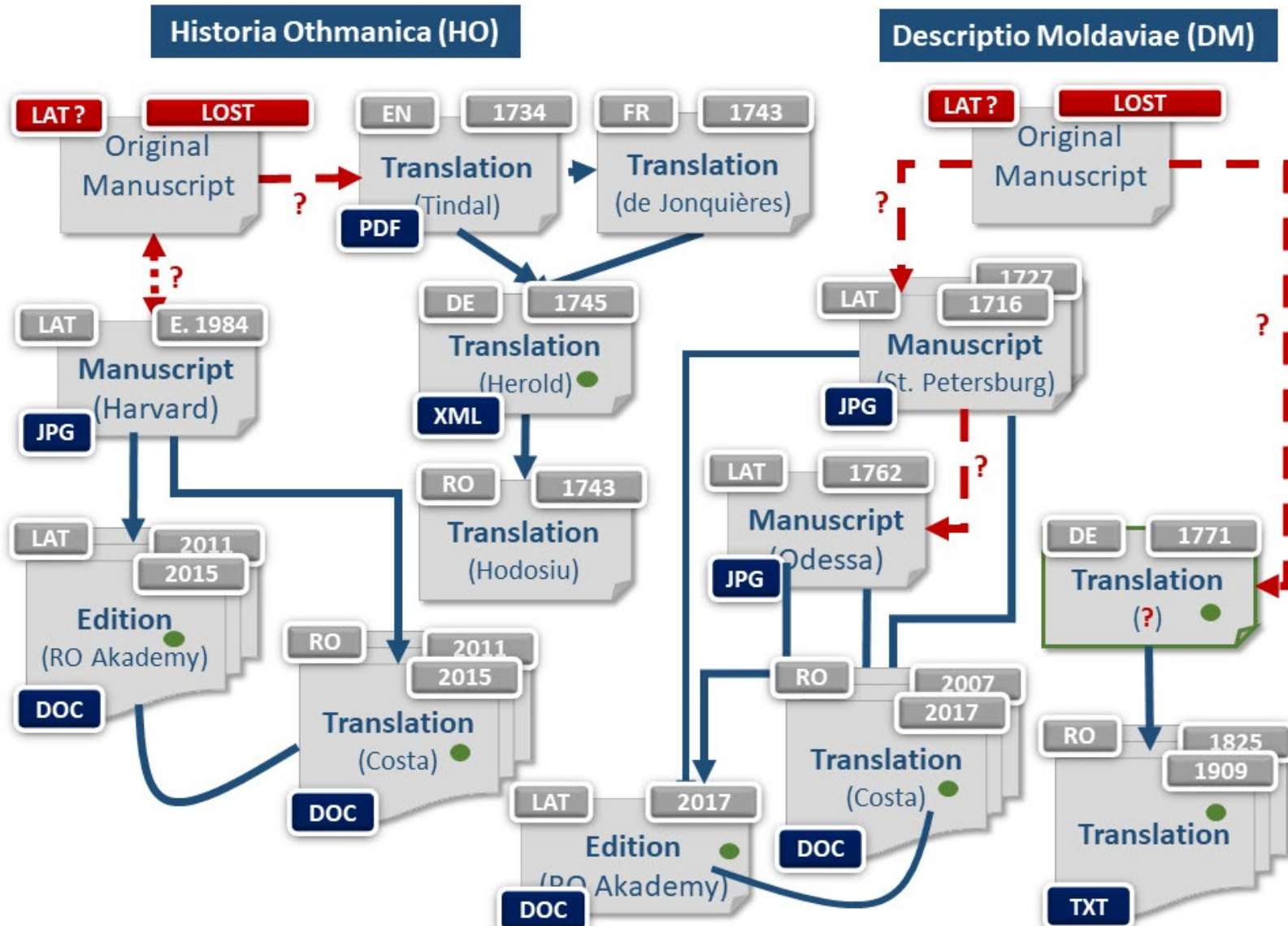
- **Consistency:**

- 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





- p. 7. promittit magnam geographicam Moldaviae  
 p. 20. Cosmasius Constantini institit monasterium Nitran  
 p. 21. De metris Constantini inquit in fides Scythorum  
 p. 26. Damiana quae Moldavia a defensione Constantini patre  
 p. 26. Basilisii in Colnae habet Epitum. augustinus  
 p. 29. Episcopus Pontificus  
 p. 29. Pagus Babiloniarum fundens  
 p. 34. Villa Traiana  
 p. 35. Desfaretia et Didriae unde nomen habent  
 p. 49. Salsignia Moldaviae  
 p. 56. In puellam iura Moldaviae fuit iurimunda p. 66. 70-71.  
 p. 57. 58. Saeculi principum  
 p. 58. Jus Moldaviae principum ecclesiasticum.  
 p. 65. Gaspar sicut Romanus saeculi in Moldaviae inducere  
 p. 71. promittit historiae Moldaviae plenam, quae  
 Confessione praece p. 101. Continatio  
 p. 101. Continuatio  
 p. 101. Continuatio

IB AH CCCP  
 Contop. Brestovick  
 P. 101  
 Cu. 8

IB AH CCCP  
 Contop. Brestovick  
 P. 101  
 Cu. 8



PARS PRIMA  
 GEORGIAPHICA

Caput I.

De antiquis & modernis Moldaviae  
nominiibus.

Totam, quam nunc vocamus Moldaviae, adiacentes  
 que occasum versus regiones, triens fere orbis par-  
 tium hemisphaeres Scythiae primam tenuere, incertis la-  
 men pro patrio more Scythibus. Post plura nomina  
 a succedentibus sibi invicem, horum cohortibus, illis  
 partibus indita. Graecis eius regionis incolae aliquan-  
 do Getae, aliquando Thaci dicti sunt; tandem sub Ro-  
 manorum monarchis, Dacorum appellatio praevaleuit.  
 His Heriae Traiani virtute, devicto rege eorum de-  
 cebato, partim extinctis, partim in diversa sparsis,  
 tota illa, quam possederant, regio, in provincias  
 Romanae formae, redacta, Romanisq; distributa civibus

IB AH CCCP  
 Contop. Brestovick  
 P. 101  
 Cu. 8

Descriptio Moldaviae, Manuscript St. Petersburg 1727

3  
Moldaviae Descriptio  
in tres partes  
distincta

2  
СССР  
ГОС. БИБЛИОТЕКА  
ИМ. Н. И. ЮНКОСКОЙ

HISTORIAE MOLDAVICAE  
PARTES ~~LIBRI~~ TRES.  
PARS I.  
GEOGRAPHICA



CAPVT I.  
De antiquis et hodiernis  
Moldaviae nominibus.

Totam quam nunc vocamus Molda-  
viam, adjacentesque occasum versus re-  
giones, trium fere orbis partium do-  
mitores Scythae primum tenuere, in-  
certis tamen pro patrio more sedi-  
bus. Post plura nomina a succeden-  
tibus sibi invicem horum cohortibus  
illis partibus indita. Graecis eius regi-  
nis incolae aliquando Getae, aliquando  
Daci dicti sunt; tandem sub Romanorum



Descriptio  
Moldaviae  
Odessa  
1762

Profatio. 23

tem, et elegantiam morum denotat, uin e contra, <sup>ترکیجه</sup>  
 Turkidie, rusticitatem, morumq; barbariem inu-  
 at. Hinc et proverbia orta sunt <sup>ترک مدت عمر بنده</sup>  
 Turk muddeti um rinde; Turca per totum vitę sue  
 cursum. sive <sup>ترک جمع زمانده آدم اولز</sup> Turca per  
 totum tempus, vel nunquam homo fiet, id est politi-  
 ores mores Othmaniadum non addisct, incultusq; <sup>manebit</sup>, et indocilis. Quod vivis vocibus arguit Tardu-  
 din (Cum Sultanus) pro augendis copiis cum Mo-  
 sa, Hali, consilium inisset, utriusq; salubris delibera-  
 tio facta est; ut ex filijs rusticorum Turcarum, gas-  
 tores veloces, et agiles seligerentur, et exercitui eque-  
 stri adiungerentur, ut idem historicus cum Usun  
 Hasen improbum, et vilem predicare vellet, eum Turke  
 appellat. <sup>اولان کی طول امرق مفهمده مصدق اولان ترک طبعنت</sup>  
<sup>اولان کی طول امرق مفهمده مصدق اولان ترک طبعنت</sup>  
 melhumine, musadak olan Turk Tabja  
 nside sup. h. e. Ipse

Historia Othmanica  
 Harvard Manuscript

## Vorrede des Verfassers.



vor meine Leser sich zu der folgenden Geschichte des osmanischen Reiches wenden, wird es nöthig seyn, daß ich ihnen einige Sachen besonders erkläre, in denen die christlichen Geschichtschreiber öfters und gröblich geirret zu haben scheinen. Die vornehmsten darunter betreffende Stücke:

- 1, Die Verwandlung der Jahre der Hidschret in die christliche Zeitrechnung;
- 2, Den Namen und das Volk der Türken;
- 3, Den Ursprung des osmanischen Geschlechts, das gegenwärtig den türkischen Thron besizet.

### I. Vergleichung der Hidschret mit der christlichen Zeitrechnung.

Viele Geschichtschreiber, und darunter auch solche, welche in keinem geringen Ansehen der Glaubwürdigkeit stehen, haben, wie ich finde, manche wichtige Begebenheit in verschiedene Jahre gesetzt; und dieses aus Mangel einer richtigen Erkenntniß in der türkischen Zeitrechnung, von den Türken Hidschret, und bey uns Hegira genennet. Hievon will ich ein merkwürdiges Beyspiel anführen. Nichts ist wahrhaftig sint der Aufhebung des kaiserlichen Sitzes zu Constantinopel denkwürdigeres vorgegangen, als die Eroberung dieser Stadt

d 3

von

Historia Othmanica  
Original German Translation,  
Hamburg, 1745

# 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.

serpens, *Kynele*, canis &c. Im plurali setzen sie hinten an die Wörter, die eine lebendige Sache bedeuten, den Artikel *ij*; als: *Saij*, *Oamenij*, equi, homines: leblose Kreaturen aber endigen sich im Plurali auf *ele*, als *Scaunele*, *Vassle*, u. s. w. Auch haben die Moldauer zween Articulos foeminini generis, *e* und *a*, als: *mujere*, *gaina*, mulier, gallina. Wörter, die sich auf *e* endigen, haben im Plurali *ile*, als: *mujerile*, *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*, *Blaster*, *βλασφημιά*, *azyma*, *ἄζυμον*, *Drum*, *δρόμος*, *Pizma*, *πίσμα*. 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“***

## Vagueness/Uncertainty at several levels

- Encoding the manuscript(s) (not well identified letters, not deciphered /ambiguous words, variants across different manuscripts)
- Vague expressions of the author
- Multiple denominations for one place, person etc. (detect and record anaphoric references)
- Uncertain Dates, Places, Persons in Cantemir's text
- Vague expressions of the translator
- Uncertain Dates, Places, Persons in the translator's text
- Translation errors which imply uncertainty

# A Complicated Explicit Example

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

capital is *Kilia*\*, *Lycostomon*, on



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.

was

should  
have  
been

Although our books do not record his descendants, it is a wellknown legend for us that he is coming from the moldavian kings

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 er von gemeiner Herkunft mit einem so hohen Gefolge auf die Jagd (welche die Moldau zu dieser Gelegenheit gegeben,)**

One should trust even more this opinion, as one can hard think that....

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

## 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

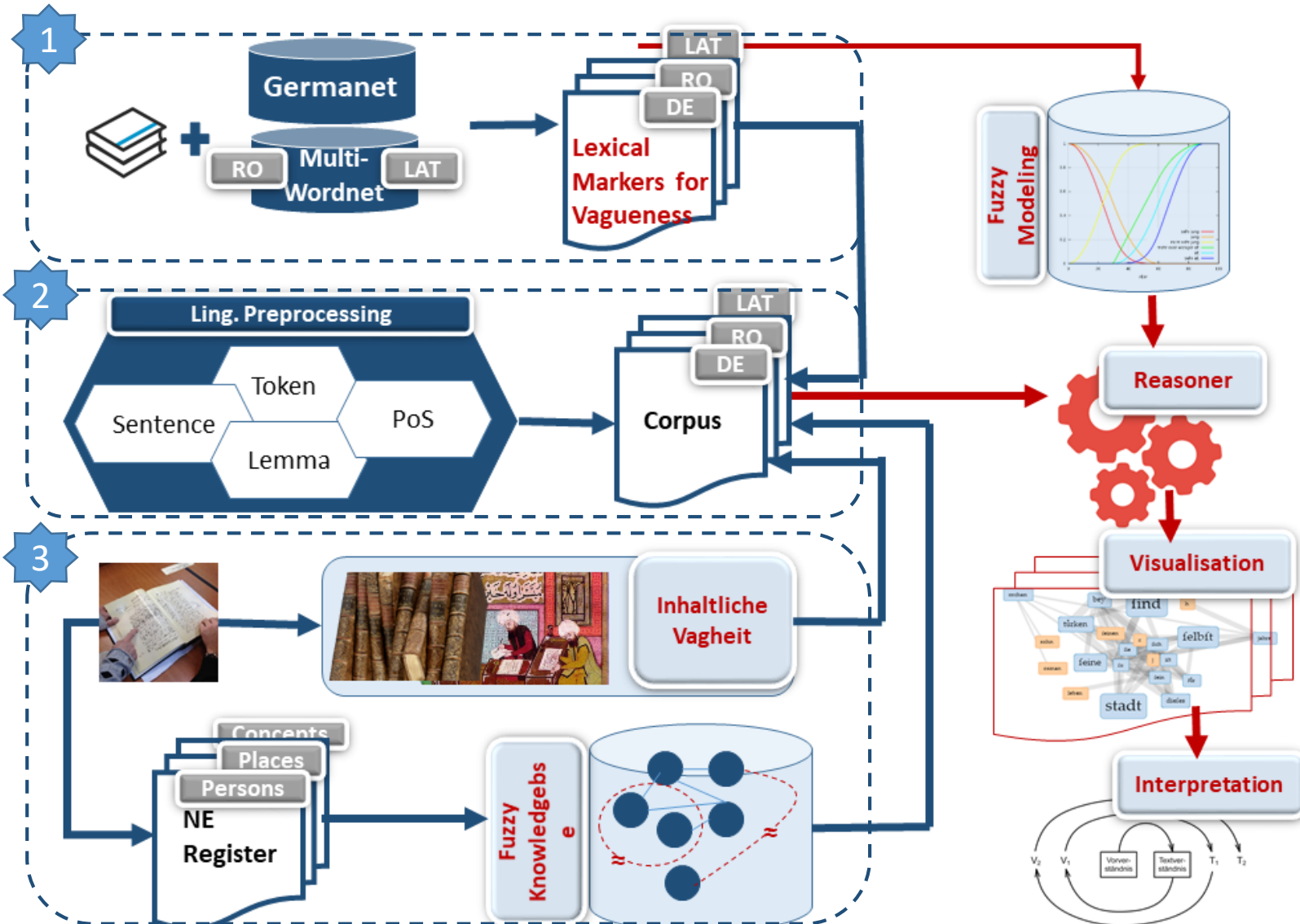
**Vagueness annotation is useful only if it is accompanied by inference rules and adequate ontological knowledge-base**

Rationale

Corpus

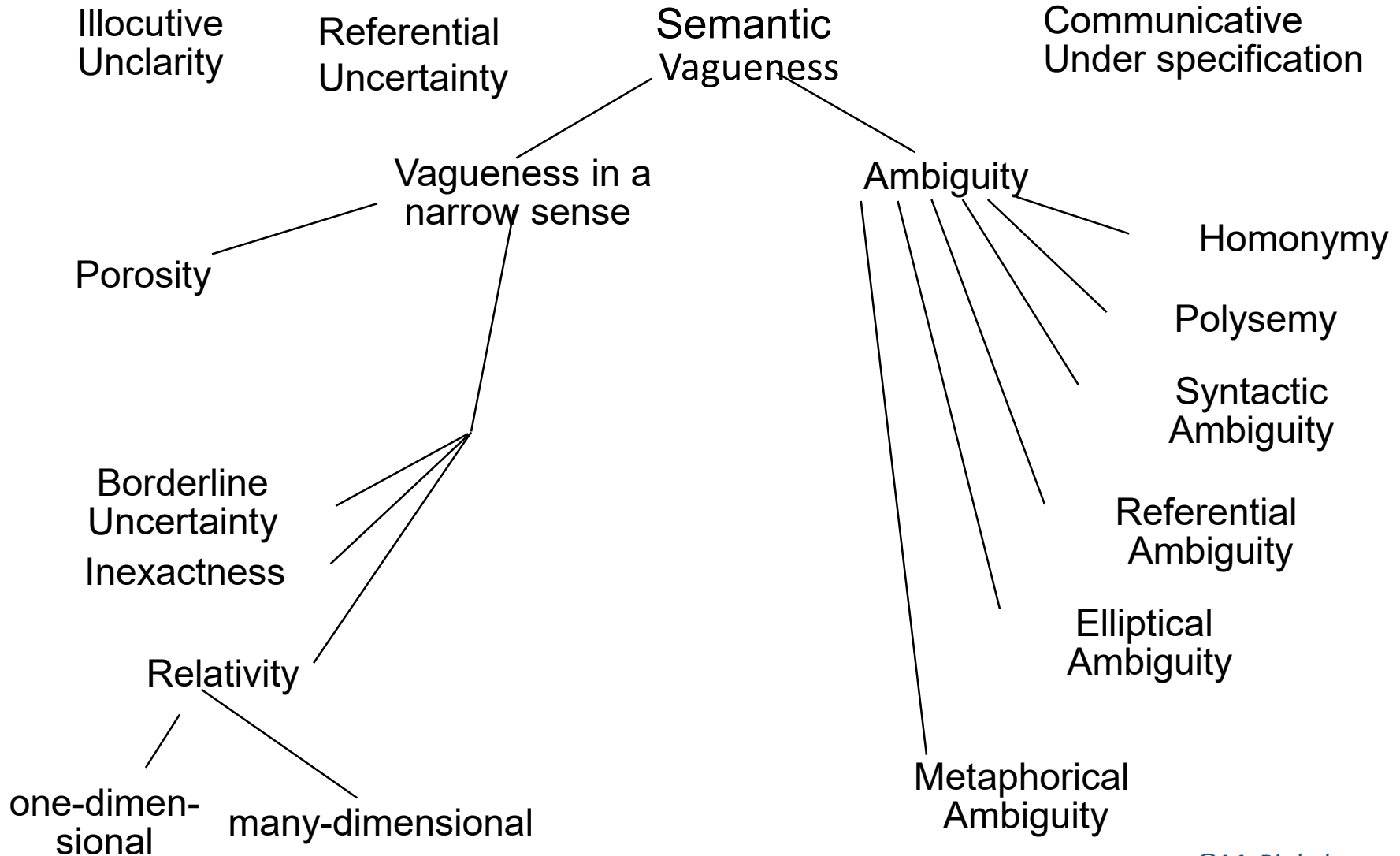
**System architecture**

Factual uncertainty



# 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 eisque, dum in suis sedibus peculiariter peculiaria fuisse. Ita, si quis in Praefatione Principes in duas stirpes fuisse divisos, „Aliothman” unam, et „Ali Dzengiz”<sup>1</sup> alteram, neque ab ipsis horum generum conditionem appellacionem iam apud aliquos valuisse. Vti enim absonum videtur, Aliothmanos Suleimano parentes ab huius nepotum integro post saeculo iis imperabunt fuisse sortitos; ita non minus falso vulgo praedicantur Tartarorum Crimensium Principes ab ipso Dzengizchano „Alidzengiz” appellacionem 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/more”*  
*“größer/bigger”, “älter/older”*
2. non-intersectives e.g. *„vermeintlich/supposed“, „so-genannt/so-called“*
3. Hedges e.g. *„ziemlich/quite“, „einigermaßen/approximately „etwa/about“*
4. inexact measures *„4 Tagereisen/4 days trip“, 10 Fuß /10 feet“*
5. modals (attitudes) e.g. *„vielleicht/maybe“, „hoffentlich/hopefully“*;  
subjunctives verbs
6. lexical quotation markers :*“es wurde gesagt /it is said“*
7. vague quantifiers e.g. *„viele“, „meistens /mostly“*
8. complex quantifiers e.g. *“etwa die Hälfte von den 20-30 tausend Soldaten /  
about a half from the 20-30 thousand soldiers“*
9. numbers
10. range expressions e.g. *“Anfang des 18. Jhds./begin of 18<sup>th</sup> century“*
11. unclear place *„Syrfia“, „Moramor“*
12. unclear person e.g. *„der ehemaligen Herzog / the former duke“*
13. unclear time e.g. *„in alten Zeiten /in old times“*
14. Domain specific e.g. *„Wesir/vizier“ vs. „Wesire/viziers“*

## 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), *Tenkih-i Tevarih-i Mülük* and *Telhisu'l-Beyan fi Kavanin-ı Al-i Osman*
- Âşıkpaşazâde Derviş Ahmet Âşıkî (1400-1484), *Tevarih-i Ali-i Osman* (1478?)
- Nicephorus Gregoras (1295-1360), *Roman History* (1359?)
- Kalkondilas Lanikos
- Şeyh Sadi (1210-1292), *Gülistan* (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 Muhâmmed. [...] 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)

It is absolutely sure and all sources tell unanimously that Bajezid had 4 Sons and none of them had the name 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

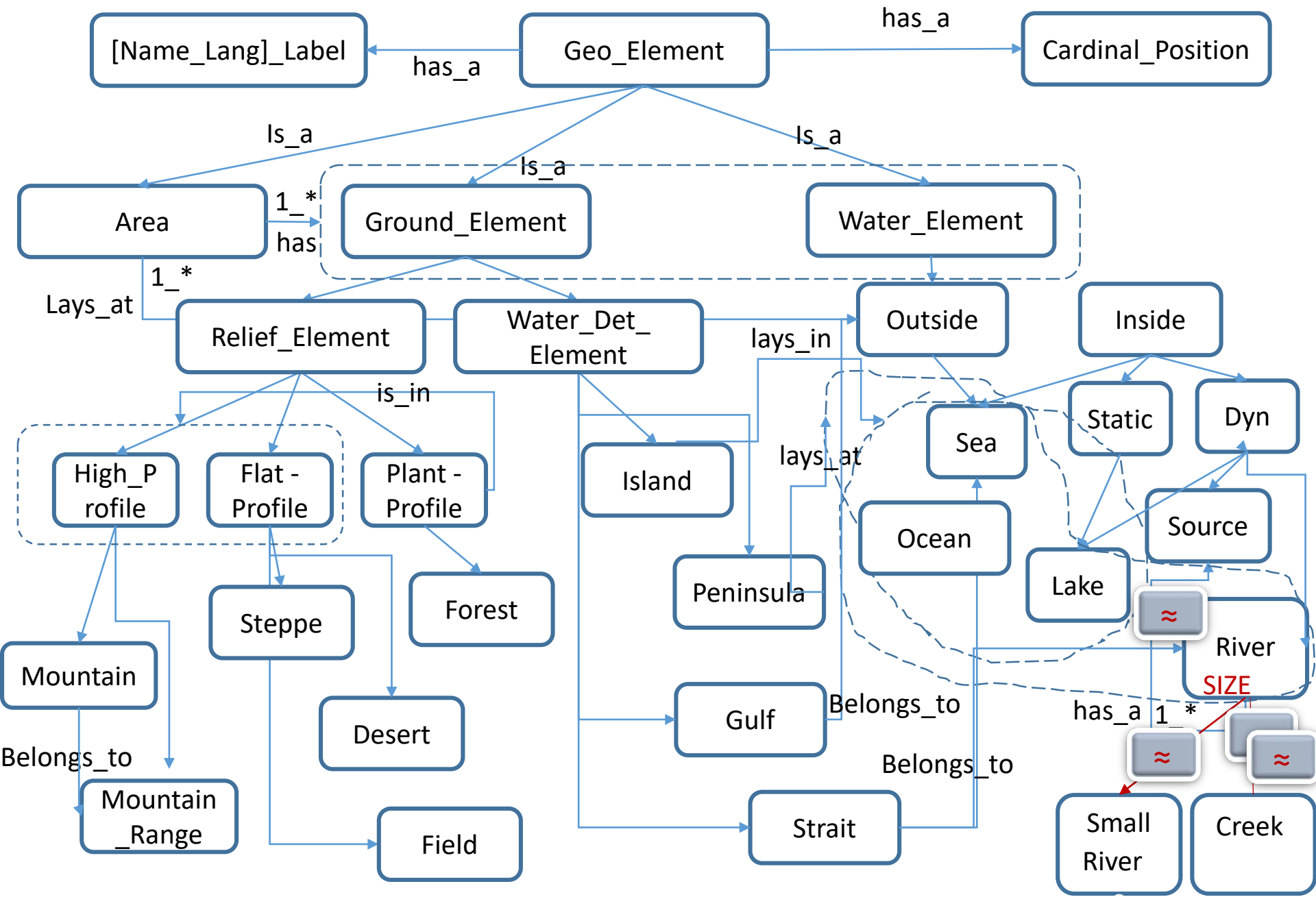
Turkish  
Sources

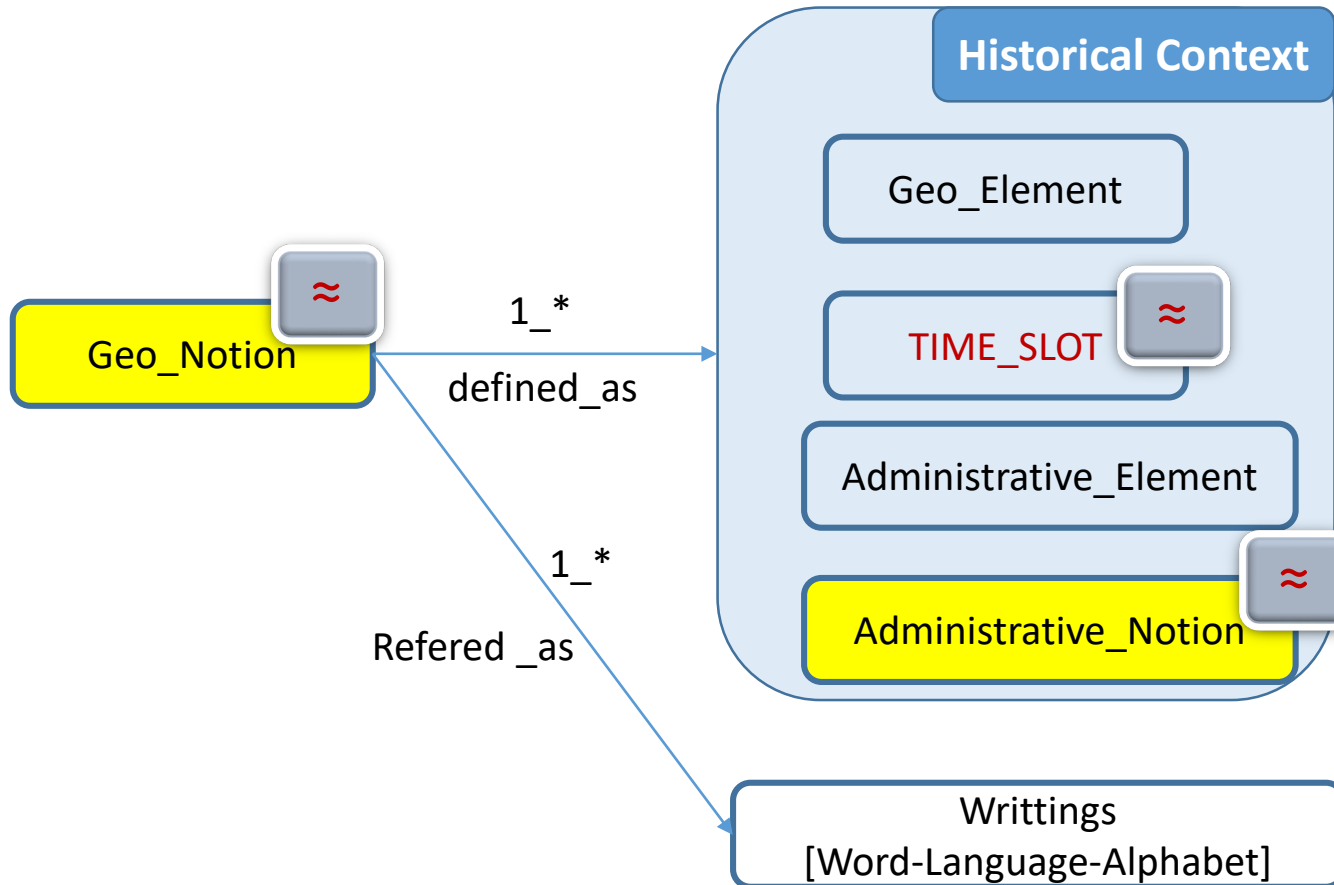
Seraskjer·Sülejman·Pascha	65	Seraskerio·Suleiman·Baszae	38	سراسکر سلیمان پاشا Serasker· Süleyman·Paşaa
Istifan	66	Istefan	38	
Matthias·(Ann.)	66	Mathiam·Corvinum	34	
Bajeβid·Yildirim·(Ann.)	66	Ildirim·Baieزيد	34	ييلدرم بايزيد Yıldırım·Bayezid
Heβarfenn·(Ann.)	66	Hezárffenn	34	
Karaman·Ogli	67	Caramanougly	39	قارمان اوغلي Karamanoğlu
Bajeβid	67	Baieزيد	39	بايزيد Bayezid

Places Database

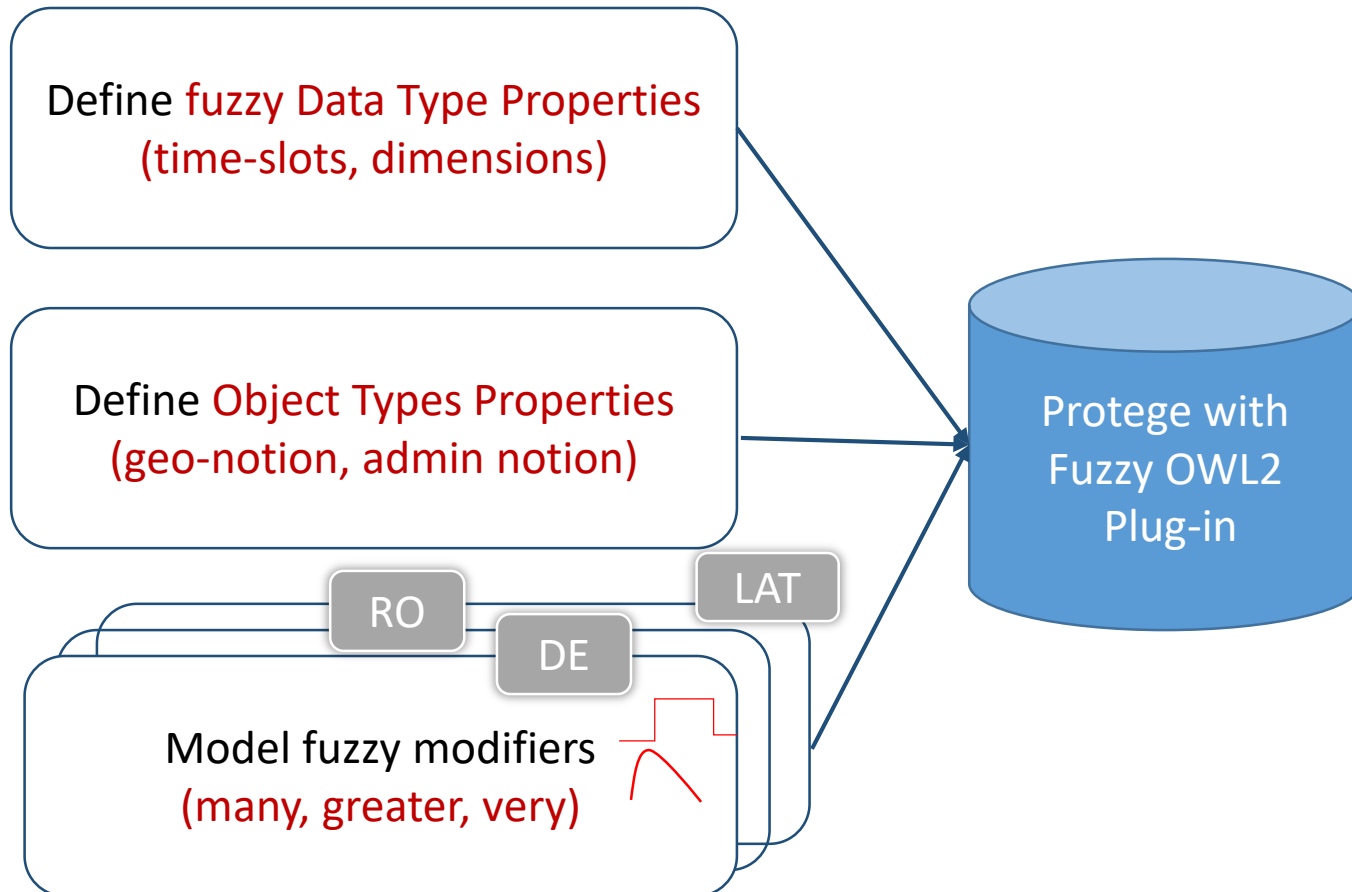
Adrianopol	73	Adrianopol	44	ادرنه·Edirne
Edrene·(Mewlasi)	73	Edrne·(Molasi)	44	ادرنه·Edirne
Edrene·(Mewlasi)·(Ann.)	73	Ederne·(Mollasi) (ادرنه·مولاسي)		ادرنه·Edirne
Adrianopol·(Ann.)	73	Adrianopolis	38	ادرنه·Edirne
Misr·(Mewlasi)·(Ann.)	73	Mýsr·(Mollasi) (مصر·مولاسي)	38	مصر·Mısır
Bursa·(Mewlasi)·(Ann.)	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









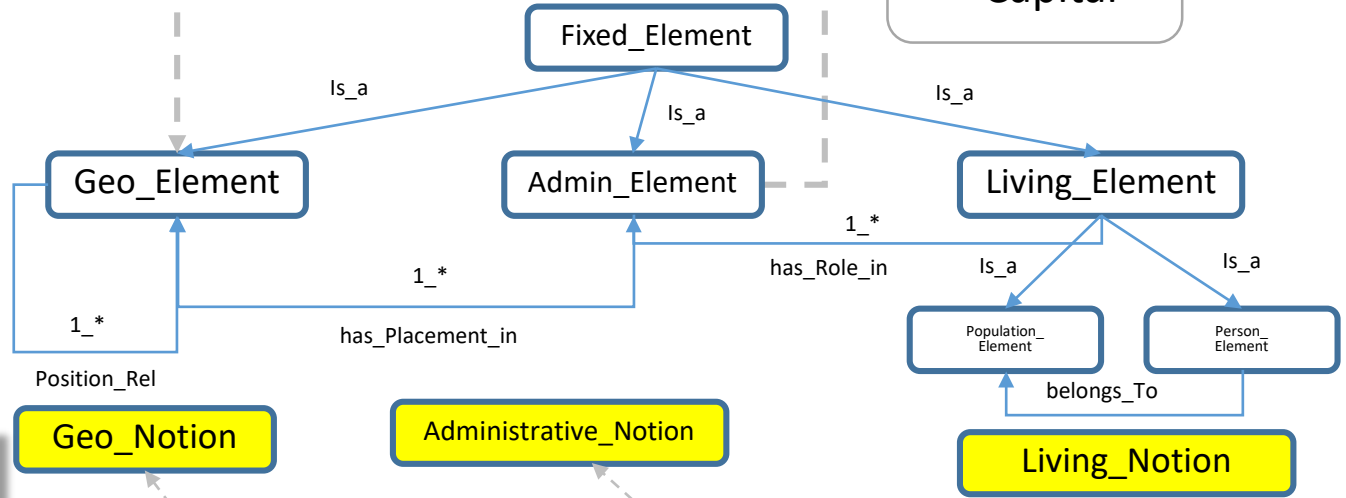
## Example: Annotation of ambiguous places



Ortelius  
Map 1570

- Northern Dobrudja
- Western Macedonia
- Eastern Europe

- Country
- Capital



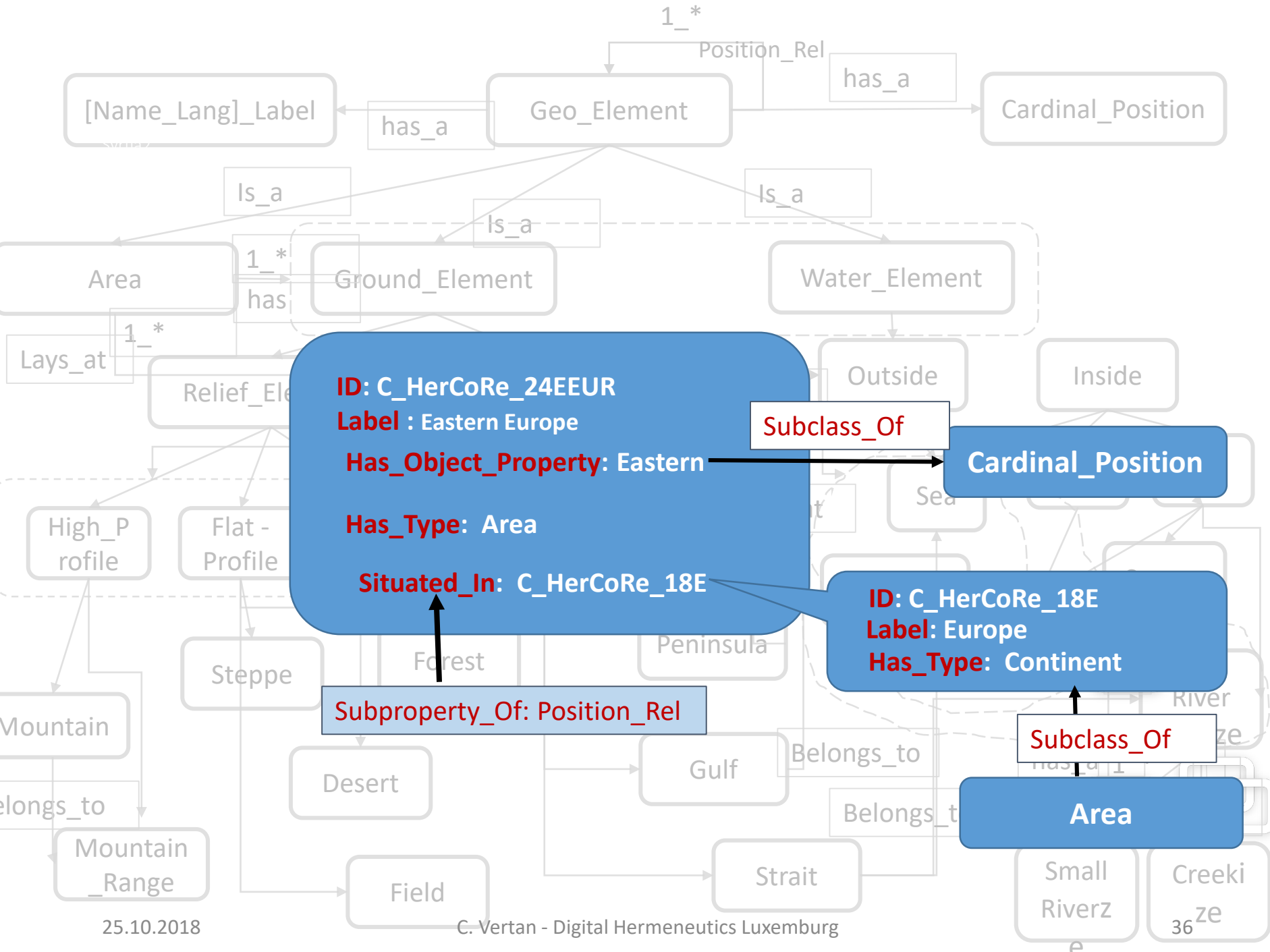
**Syrfia** is the abandoned name of a region in Eastern Europe, used on historical maps until 17th century, designating

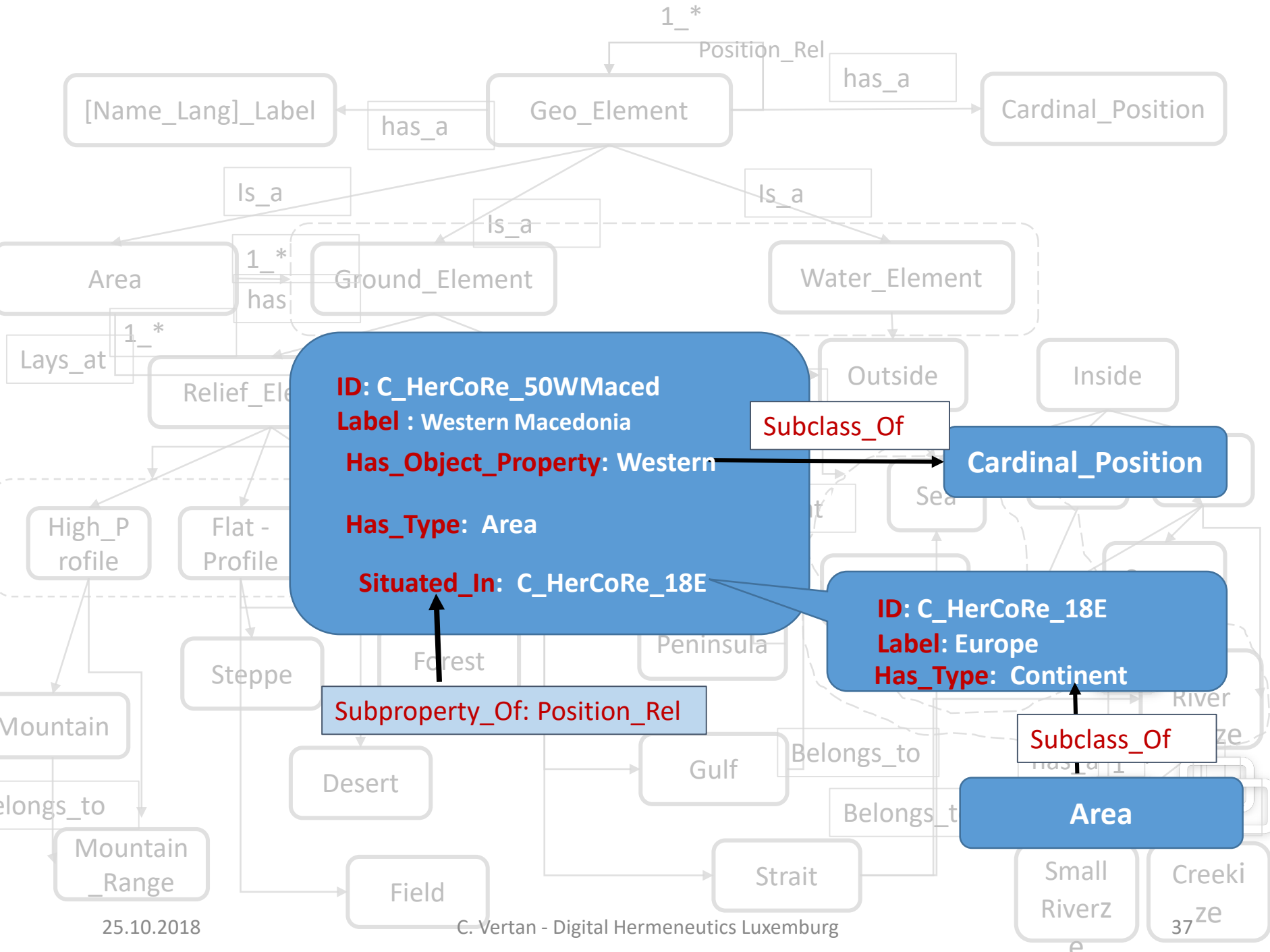
- a part of Northern Dobrudja, coming from the Greek term *Σύρφοι - Syrphoi*, or
- The Cojani region from western Macedonia, today in Greece but in turkish times in the "Serfia sangiac" having the capital *Σέρβια, Servia* ;
- Sârbia, due to phonetic association.

- Cojani Region
  - Sârbia
- Fuzzy Concept

- Greece
  - Serfia sangiac
  - Servia
- Fuzzy Concept

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





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

**Subclass\_Of**

**Cardinal\_Position**

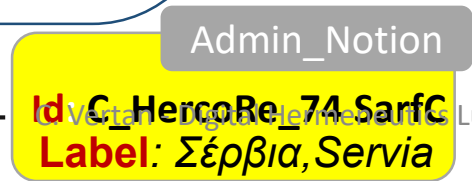
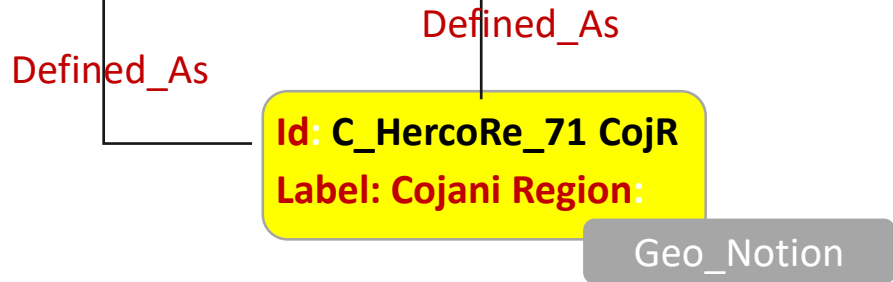
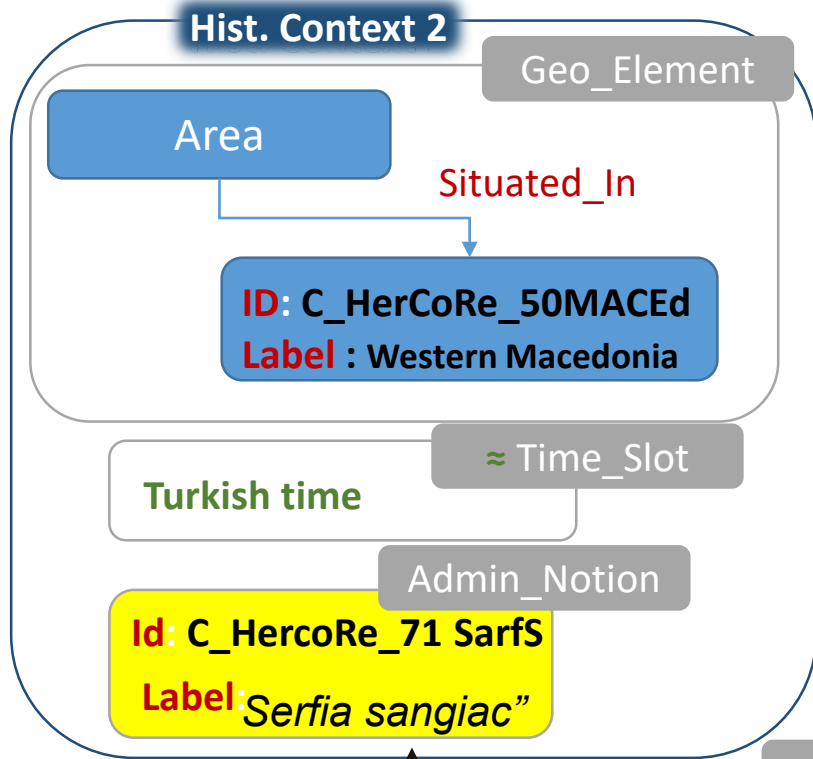
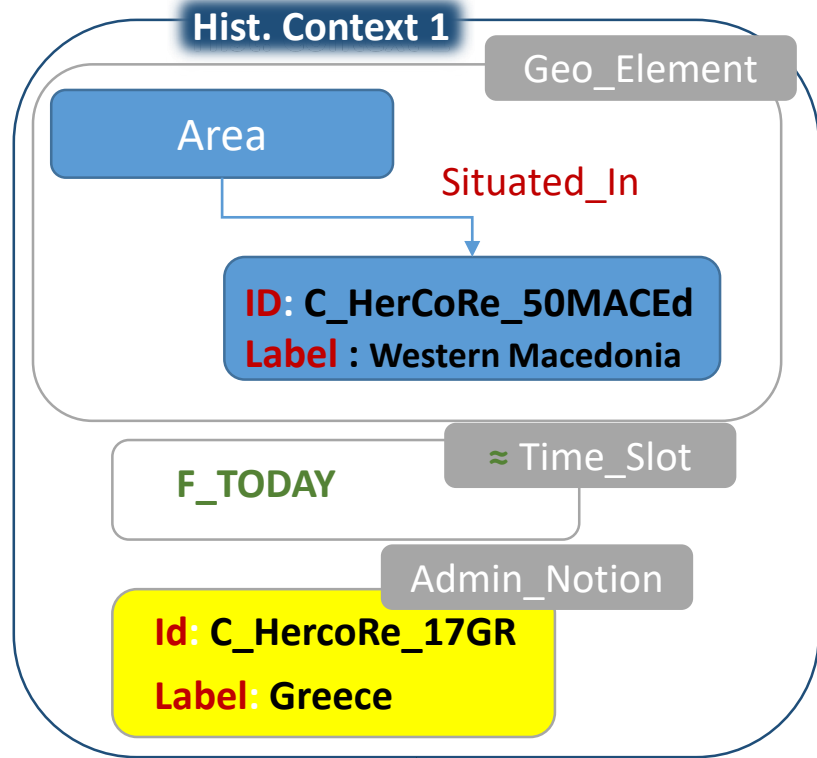
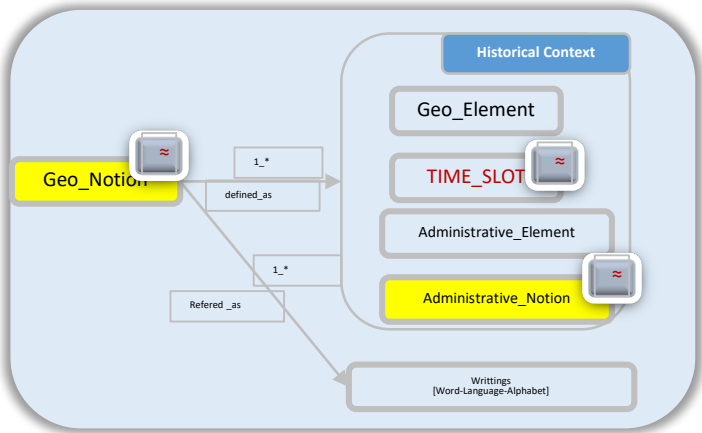
**ID:** C\_HerCoRe\_18E  
**Label:** Europe  
**Has\_Type:** Continent

**Subproperty\_Of: Position\_Rel**

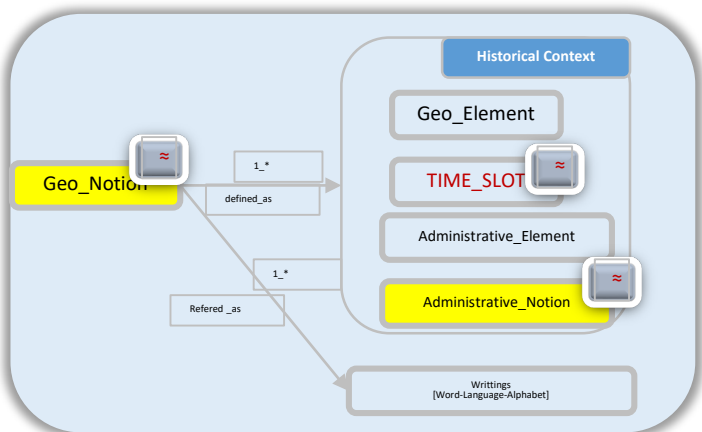
**Subclass\_Of**

**Area**





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



```

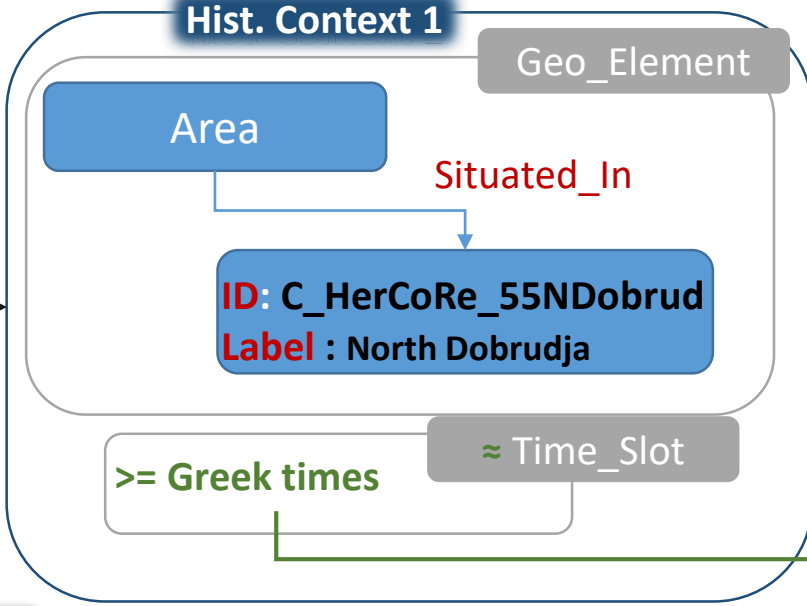
<DatatypeDefinition>
  <Datatype IRI='#GreekTimes' /> <DataIntersectionOf>
    <DatatypeRestriction>
      <Datatype abbreviatedIRI='xsd:integer' />
      <FacetRestriction facet='&xsd;minInclusive'>
        <Literal datatypeIRI='&xsd;integer'>-750</Literal>
      </FacetRestriction>
    </DatatypeRestriction>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

))

### Syrflia 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 "Serflia sangiac" having the capital *Σέρβια*, *Servia* ;
- Sârbia, due to phonetic association.

25.10.2018 Source: Wikipedia

Geo\_Notion

**Id: C\_HercoRe\_70Syrflia**

**Label: Syrflia**

**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**

41



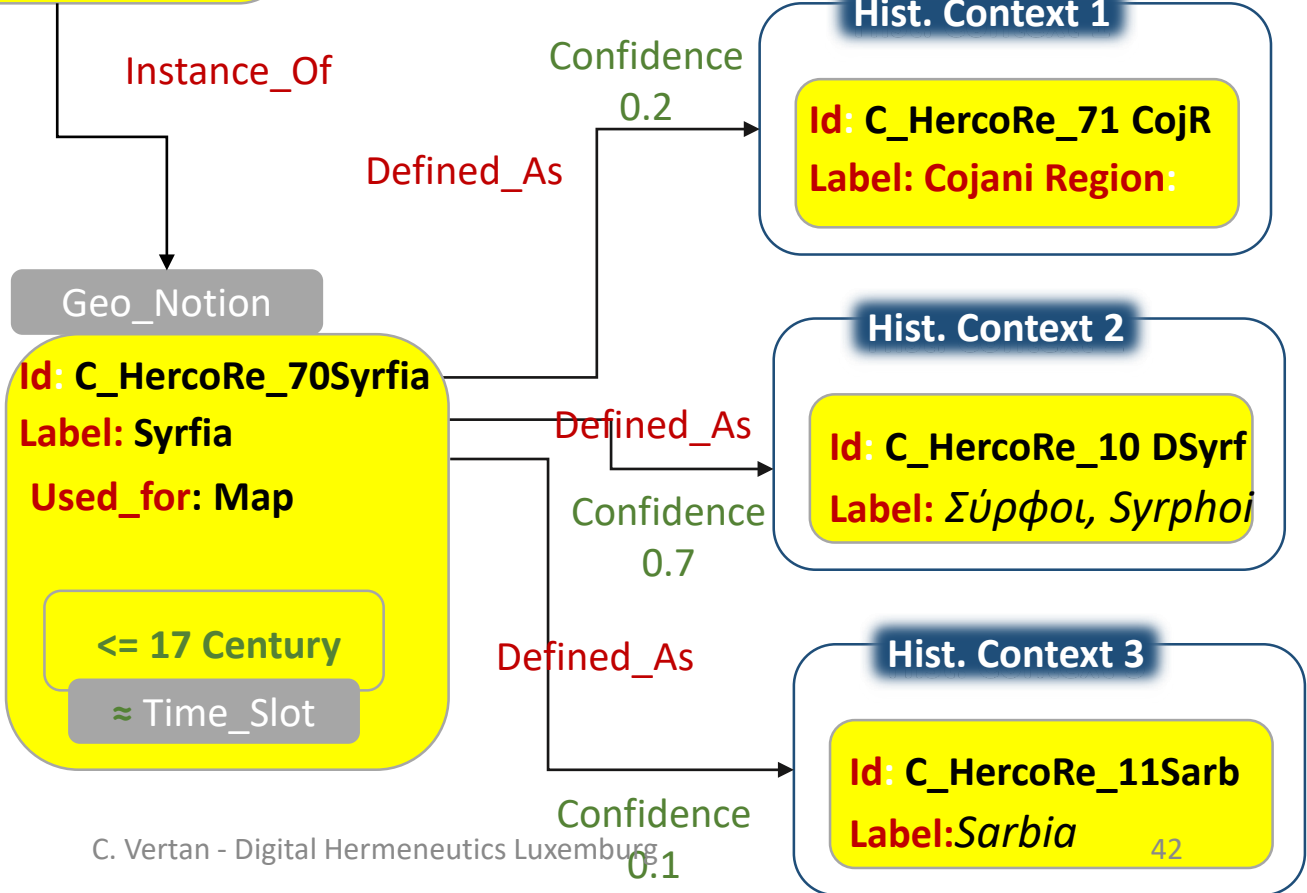
Orteliusmap 1570

**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.

Source: Wikipedia



# Multi-level Annotation

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 .

# Conclusions

- Vagueness /Uncertainty has to be recorded in the digital object. Otherwise researchers work with crisp digital artefacts which differ much from the original
- Any reasoning (automatic) / reflection (human) process on such crisp object will lead to results which cannot correspond to the same process done on the original (physical object).
- Encoding vagueness /uncertainty without using appropriate processing tools is just half-way.
- Vagueness/Uncertainty is present at different levels. Humanists should be able to understand the automatic processes involved, in order to signalise where adaptations/new implementations are required.
- Manual input /annotation is still largely required, but doing this now gives the possibility to have truly accurate data which can be used for qualitative research in history.