Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

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

Semi-automatic Multilevel Annotation of Vagueness in Historical Texts

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-

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

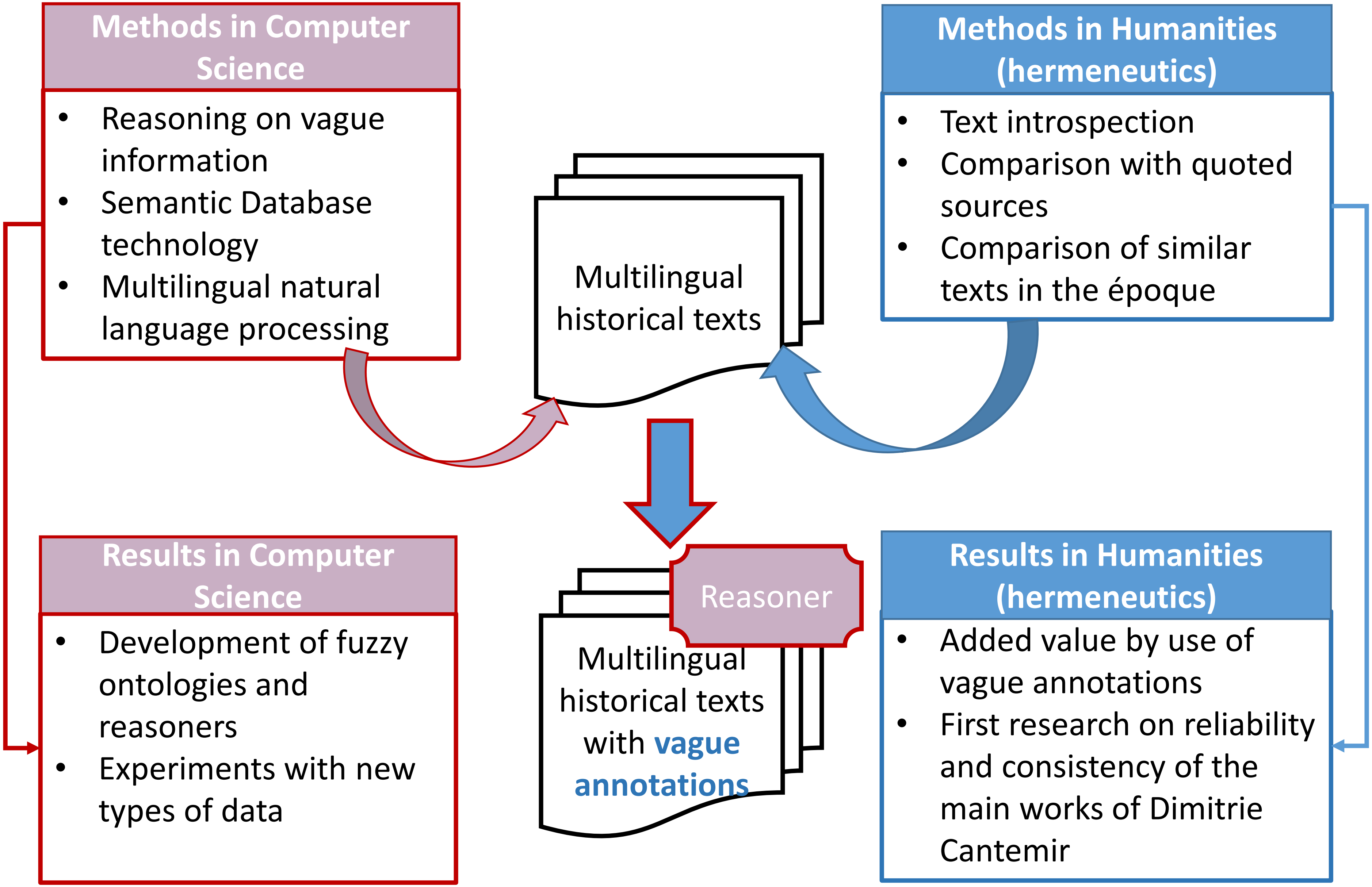
Compare for the first time “original” with translations done in the 18th- 19th century

(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



Methods in Computer Science

- Reasoning on vague information
- 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

CS

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

H CS

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

H

Prof. Dr. Yavuz Köse, UHH
Turcology

H

Alptug Güney, UHH
Turcology

H CS

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

H

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

CS

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

CS

Segiu Nisioi, UB
CL, CS

Outline

- Rationale of the project
- Corpus‘ Insight
- Identifying and annotating vagueness /uncertainty
- 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 (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

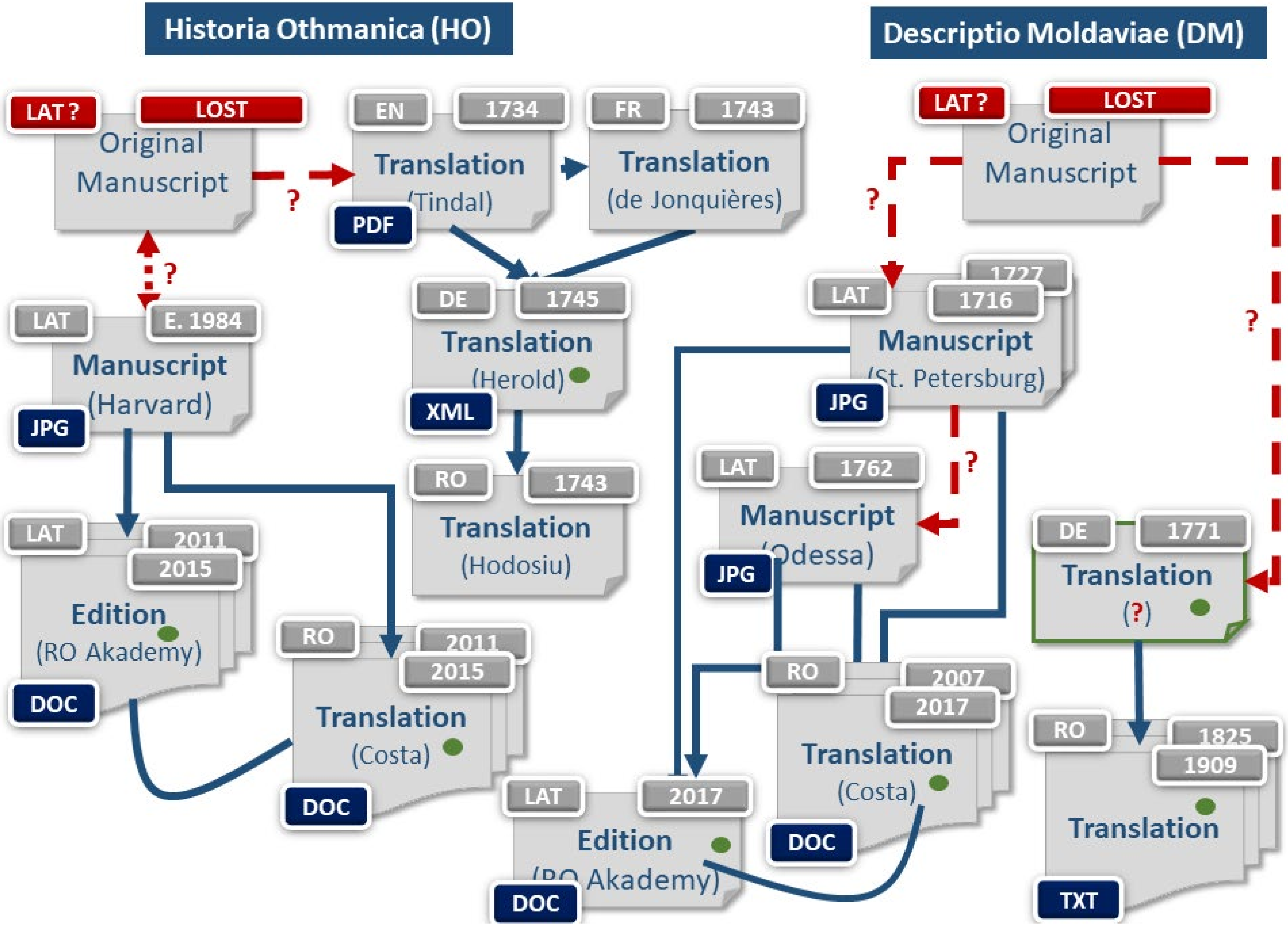
no study about the reliability and consistency of the original and the translations could be performed until now

Computational methods could help in performing this study

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



Historia Othmanica (HO)

LAT ? LOST Original Manuscript

EN 1734 Translation (Tindal) PDF

FR 1743 Translation (de Jonquières)

LAT E. 1984 Manuscript (Harvard) JPG

DE 1745 Translation (Herold) XML

RO 1743 Translation (Hodosiu)

LAT 2011 2015 Edition (RO Akademy) DOC

RO 2011 2015 Translation (Costa) DOC

Descriptio Moldaviae (DM)

LAT ? LOST Original Manuscript

LAT 1727 1716 Manuscript (St. Petersburg) JPG

LAT 1762 Manuscript (Odessa) JPG

DE 1771 Translation (?)

RO 2007 2017 Translation (Costa) DOC

LAT 2017 Edition (RO Akademy) DOC

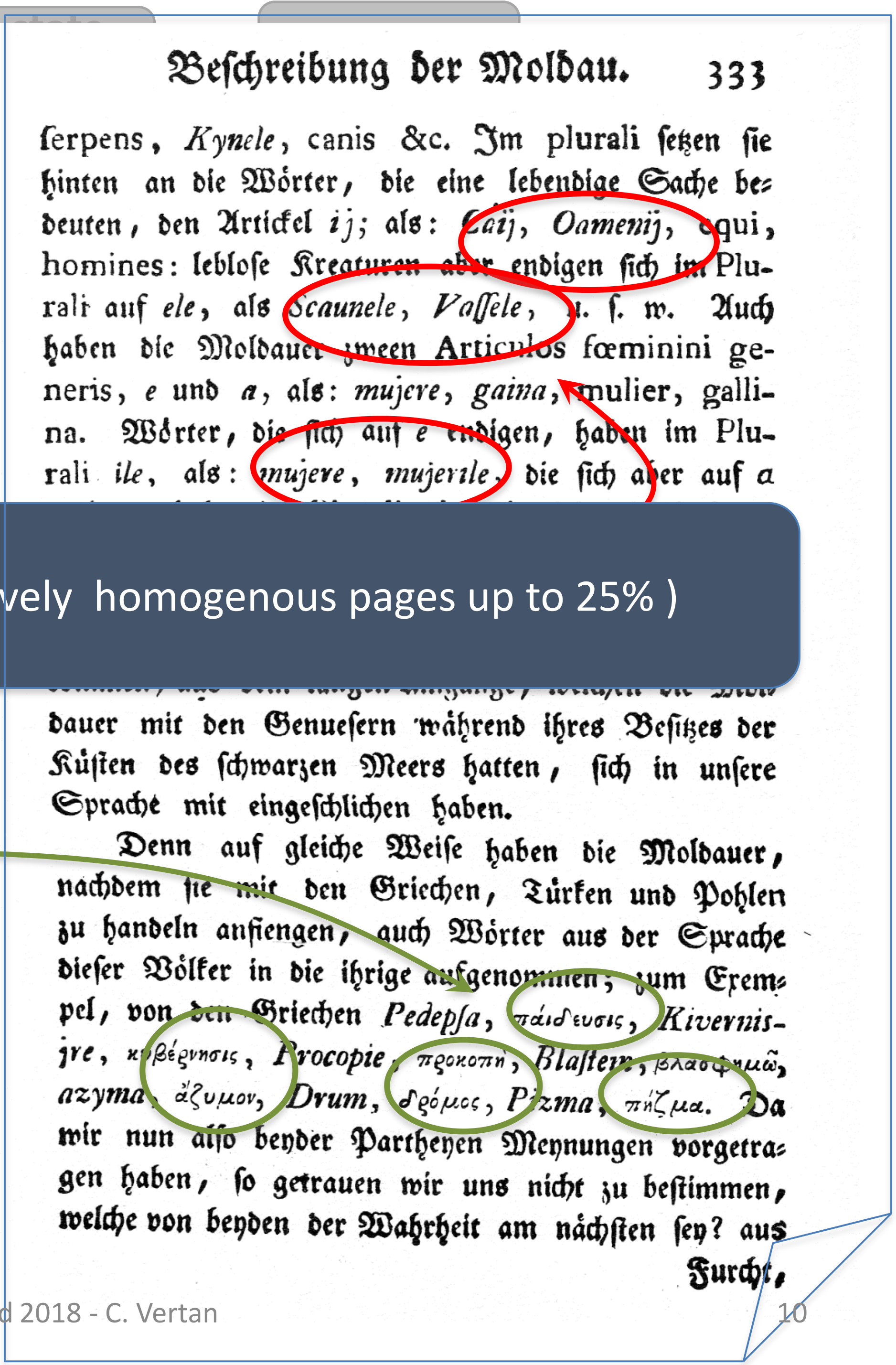
RO 1825 1909 Translation TXT

?

Corpus creation – challenges

- Surface form – level
  - German texts are in black-letter typeface
  - Part of the foreign words are
  - black-letter typeface
  - Mixed typefaces
  - Mixed scripts

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



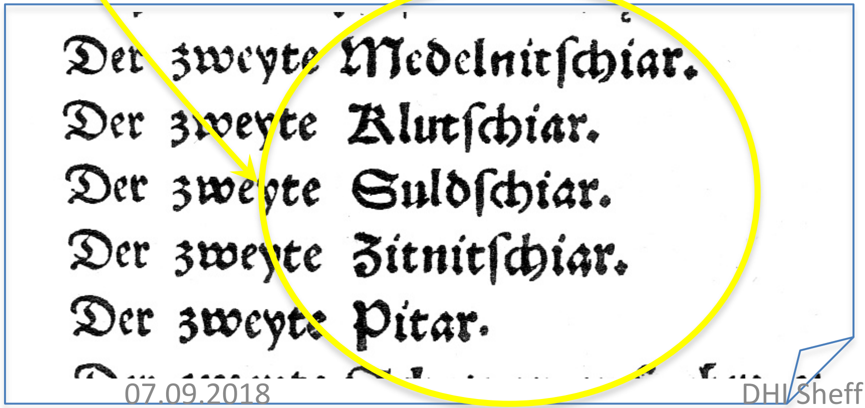
Beſchreibung der Moldau. 333

ſerpens, *Kynele*, canis &c. Im plurali ſetzen ſie hinten an die Wörter, die eine lebendige Sache bedeuten, den Artickel *ij*; als: *Caij*, *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*

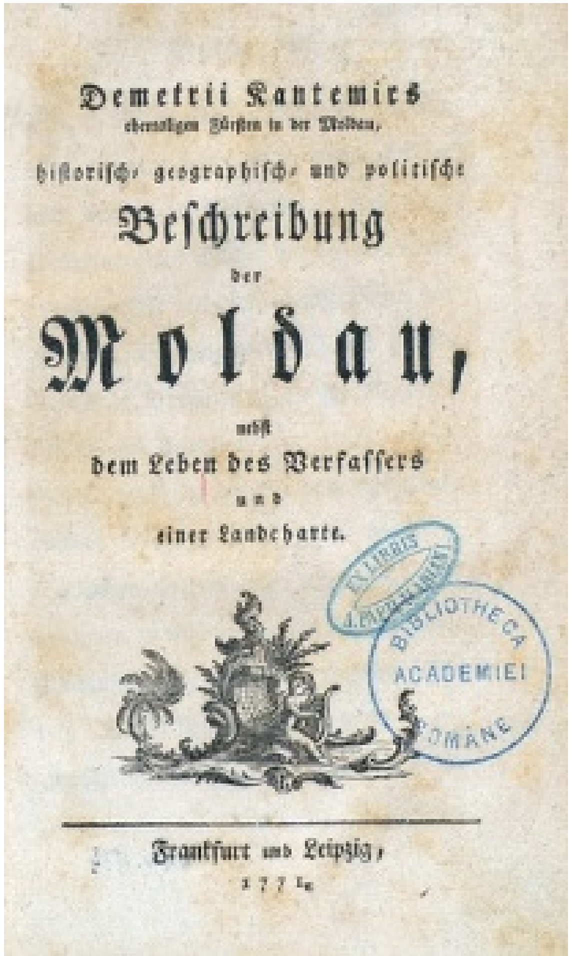
dauer mit den Genueſern während ihres Beſitzes der Küſten des ſchwarzen Meers hatten, ſich in unſere Sprache mit eingeſchlichen haben.

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*, πάιδευσις, *Kivernisjre*, κυβέρνησις, *Procopie*, προκοπὴ, *Blaſtem*, βλασφημῶ, *azyma*, ἄζυμον, *Drum*, δρόμος, *Pizma*, πήζμα. 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,



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



Demetrii Kantemirs
ehemaligen Fürſten in der Moldau,
hiſtoriſch- geographiſch- und politiſche
Beſchreibung
der
Moldau,
nebſt
dem Leben des Verfaſſers
und
einer Landcharte.

Frankfurt und Leipzig,
1771.

„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, 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.)

?

1

capital is Kilia, Lycostomon, on*

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 hronog 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 nu

was

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

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.

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 doch eine beständige Sage bey aus dem alten königlichen moldaui Stamme **gewesen sey**, und den Bogdan zum Vater gehabt habe, welcher ein Sohn des Johannis war, von welchem alle Fürsten den

should have been

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.

Enriched Classical Markup

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 beyzume en, weil man schwerlich glauben kan, da her von gemeiner Herkunft mit einem so** Gefolge auf die Jagd (welche die Molda cken 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 sích 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.)

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 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

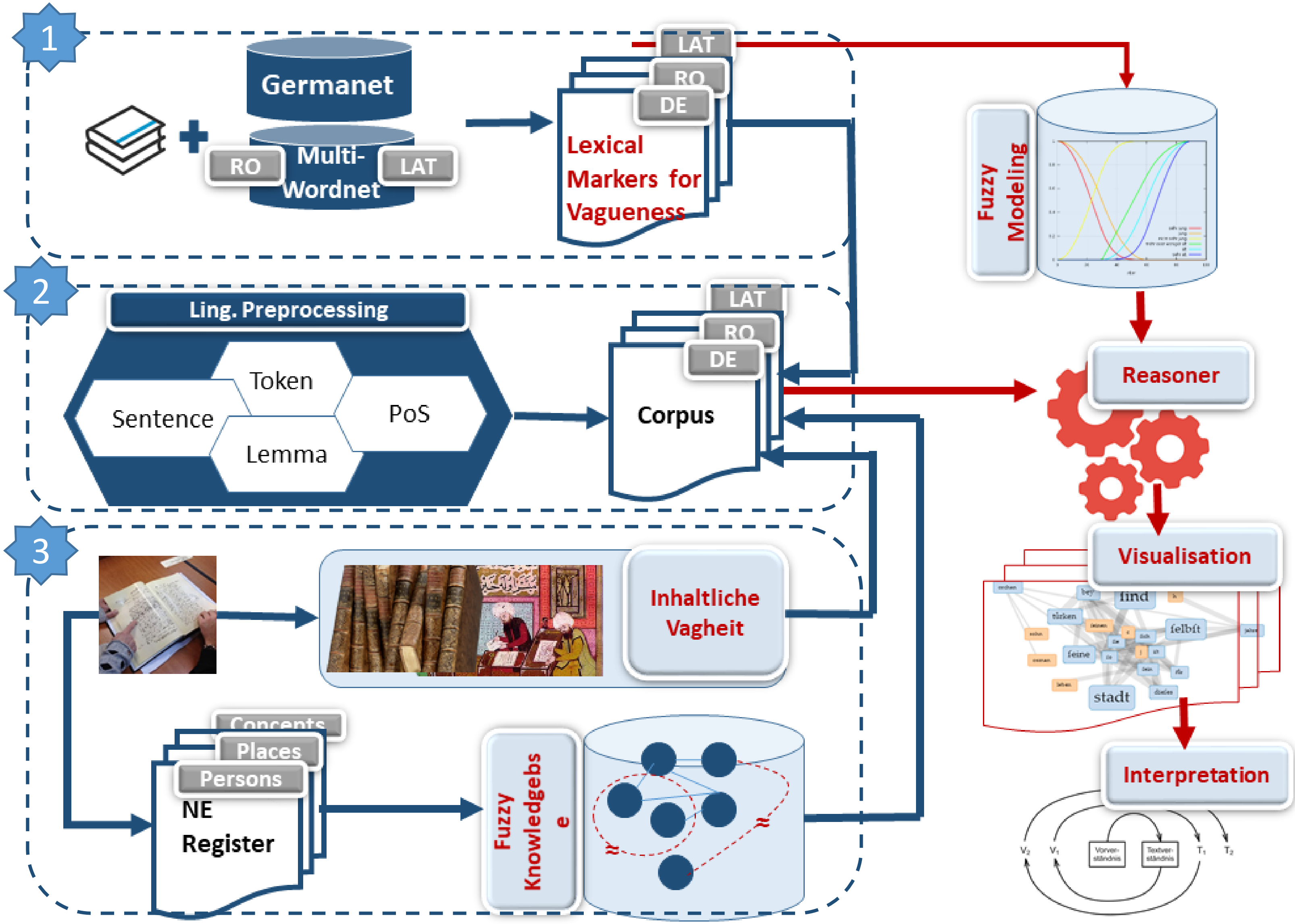
{ } marginals /between line

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



1

Germanet

RO Multi-Wordnet LAT

LAT RO DE Lexical Markers for Vagueness

2

Ling. Preprocessing

Token

Sentence

PoS

Lemma

LAT RO DE Corpus

3

Inhaltliche Vagheit

Concepts Places Persons NE Register

Fuzzy Knowledgebase

Fuzzy Modeling

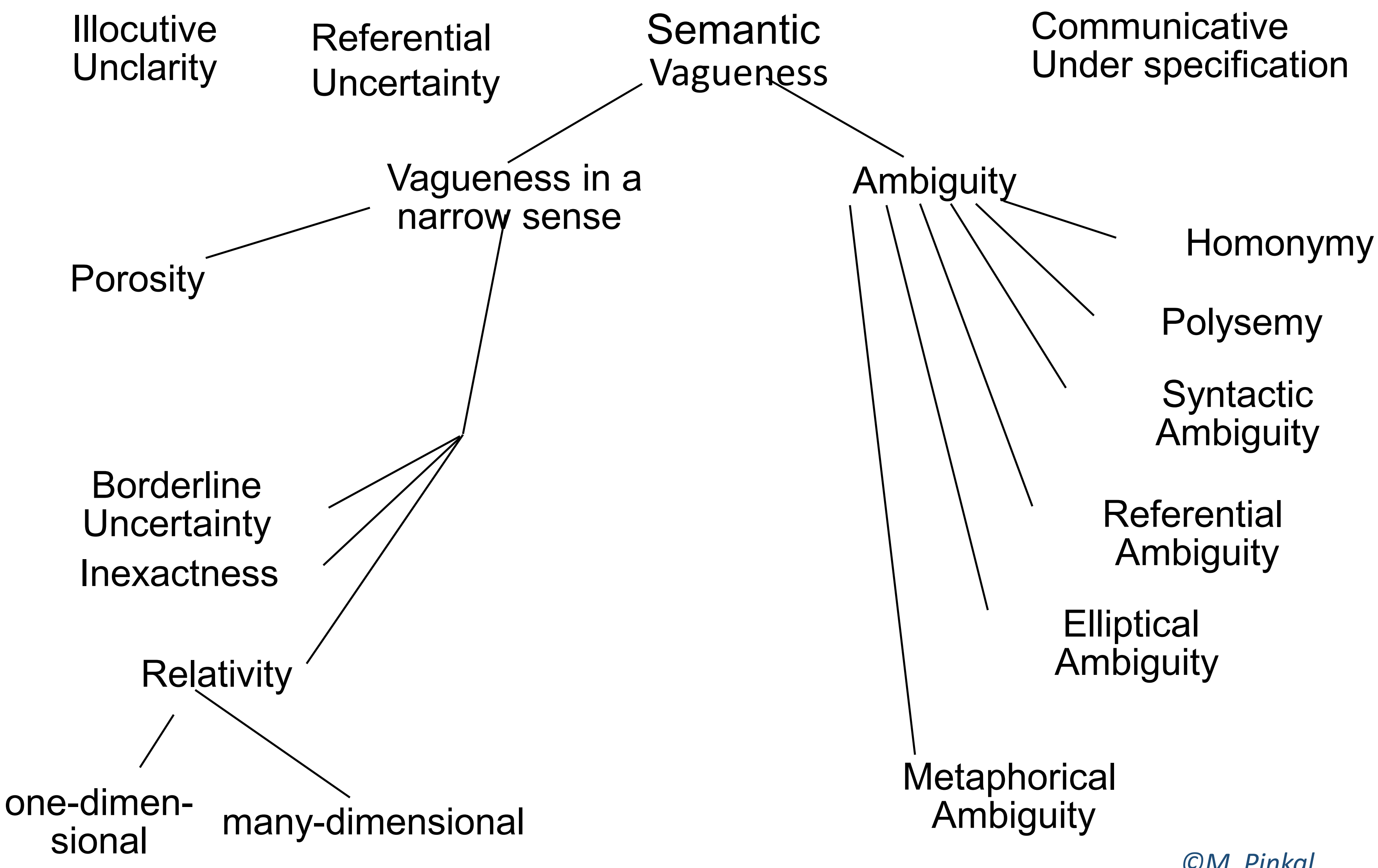
Reasoner

Visualisation

Interpretation

M.Pinkal‘s Schema of Semantic Vagueness

1



Illocutive Unclarity

Referential Uncertainty

Semantic Vagueness

Communicative Under specification

Vagueness in a narrow sense

Ambiguity

Porosity

Homonymy

Polysemy

Syntactic Ambiguity

Borderline Uncertainty

Inexactness

Referential Ambiguity

Elliptical Ambiguity

Relativity

one-dimen-sional

many-dimensional

Metaphorical Ambiguity

©M. Pinkal

Lexical and Syntactic Sources of vagueness

Quotation

More plausible

Would have been…

seems unlikely

equaly false

- Hactenus Gregoras: ad cuius verba observare haud extra propositum erit τὴν πρώτην, quam Gregoras vocat „**Tartari**[illegible] esse, quam hodie vulgo „Ma[illegible] s eiusque incolarum nomina, [illegible] istoricis recenseantur, tamen adscita magis, aut ab exteris indita, quam propria eisque, dum in suis sedibus [illegible] peculiaria fuisse. Ita, si quis [illegible] oni Praefatione legerit O[illegible] ncipes in duas stirpes fuisse divisos, „Aliothman" unam, et „Ali Dzengiz"[1] alteram, ne[illegible] ub ipsis horum generum condi[illegible] appellationem iam apud [illegible] aluisse. Vti enim absonum videtur, Aliothmanos Suleimano parentes ab huius ne[illegible] integro post saeculo iis imperab[illegible] fuisse sortitos; ita non minus falso vulgo praedicantur Tartarorum Crimensium Principes ab ipso Dzengizchano „**Alidzengiz**" appellationem retinuisse.

- Până aici l-am citat pe Gregoras: faţă de cuvintele lui nu va fi nepotrivit să observăm că acea Tartaria „ἡ πρώτη", pe care o numeşte Gregoras, este chiar aceea pe care o numim îndeobşte cea „Mare", iar numele locuitorilor ei, chiar dacă sunt înregistrate de istorici, au fost totuşi mai degrabă împrumutate sau date de străini decât proprii lor, purtate întocmai pe vremea când se aflau în sălaşurile lor. Astfel, dacă va fi citit cineva în Prefaţa pusă înaintea acestui tratat că principii neamului oguzilor au fost împărţiţi 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-şi fi ales numele de la nepotul acestuia, care a domnit peste ei după un secol întreg, la fel de fals se spune îndeobşte că principii tartarilor din Crimea şi-ar fi păstrat denumirea „alidzengiz" chiar de la Dzengizchan

***Historia Othmanica*, I.1.k : Latin original text and Romanian translation**

©Ioana Costa

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“; subjonctives 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 thounsand soldiers”*
9. numbers
10. range expressions e.g. *“Anfang des 18. Jhds./begin of 18th 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, Turcicæ, Arabicæ, Persicæ* (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?

©Alptug Güney

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)

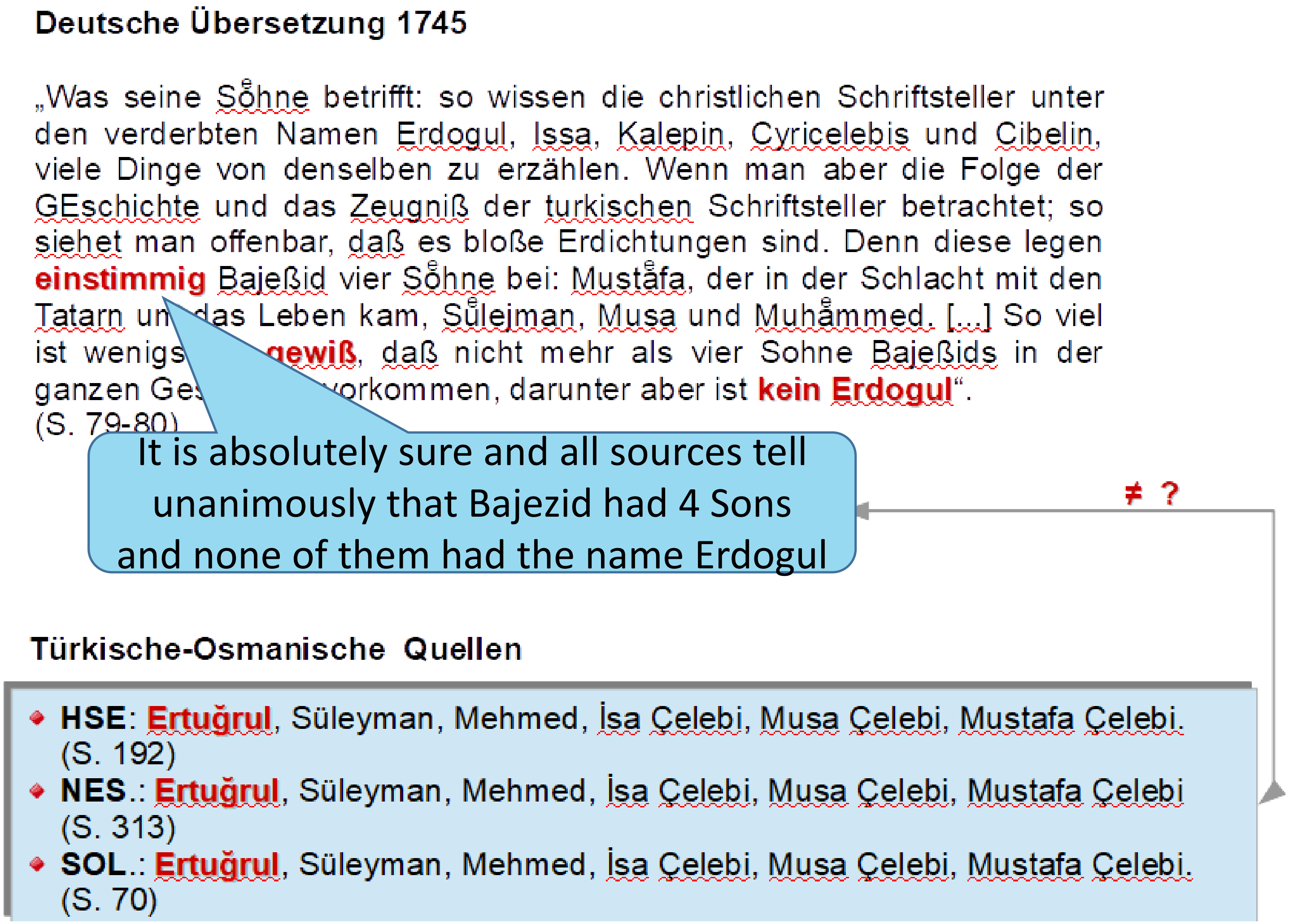
Hezarfen (Hezarfen Hüseyin Efendi) (?-1691/92), Tenkih-i Tevarih-i Mülük: is NOT mentioning this

- 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 Soͤhne 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 Soͤhne bei: Muståfa, der in der Schlacht mit den Tatarn un das Leben kam, Sůlejman, Musa und Muhåmmed. [...] So viel ist wenigs **gewiß**, daß nicht mehr als vier Sohne Bajeßids in der ganzen Ges vorkommen, darunter aber ist **kein Erdogul**".
(S. 79-80)

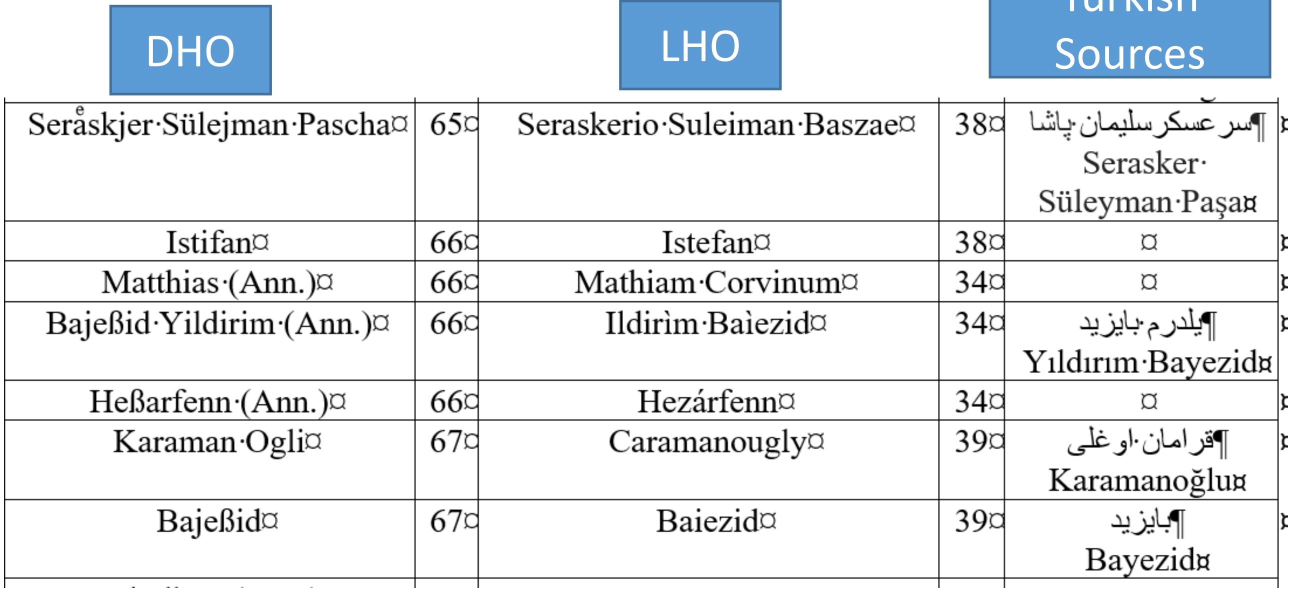
It is absolutely sure and all sources tell unanimously that Bajezid had 4 Sons and none of them had the name Erdogul

≠ ?

Türkische-Osmanische Quellen

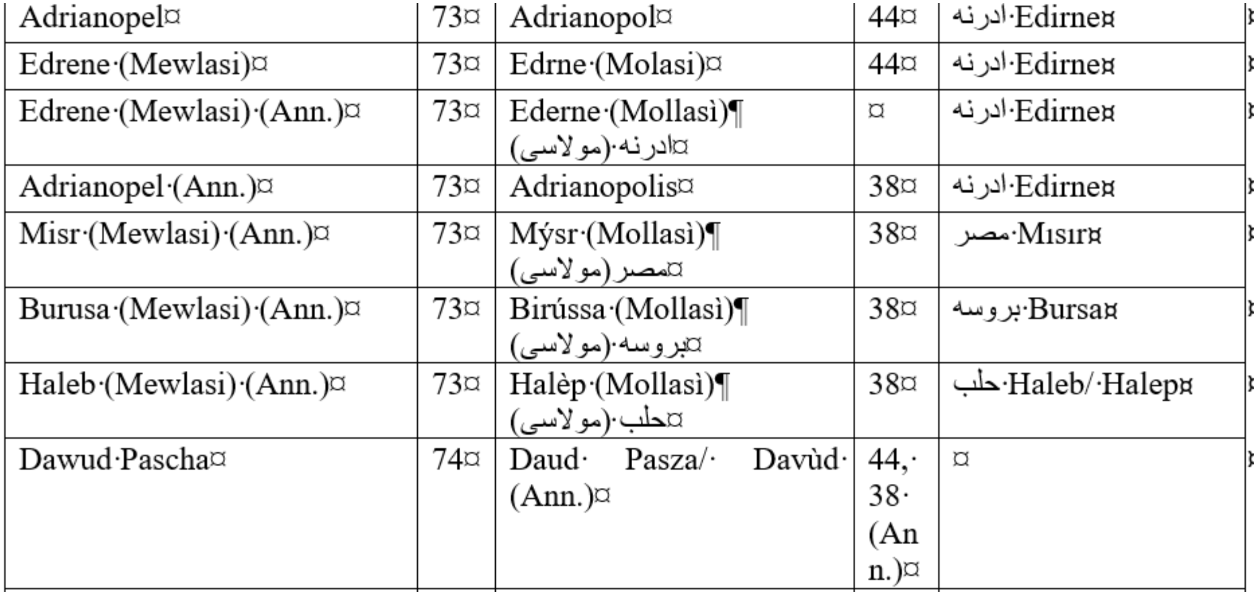
- **HSE**: Ertuğrul, Süleyman, Mehmed, İsa Çelebi, Musa Çelebi, Mustafa Çelebi. (S. 192)
- **NES**.: Ertuğrul, Süleyman, Mehmed, İsa Çelebi, Musa Çelebi, Mustafa Çelebi (S. 313)
- **SOL**.: Ertuğrul, Süleyman, Mehmed, İsa Çelebi, Musa Çelebi, Mustafa Çelebi. (S. 70)

©Alptug Güney



DHO		LHO		Turkish Sources
Seraͤskjer·Sülejman·Pascha¤	65¤	Seraskerio·Suleiman·Baszae¤	38¤	سرعسکرسلیمان·پاشا¶ Serasker· Süleyman·Paşa¤
Istifan¤	66¤	Istefan¤	38¤	¤
Matthias·(Ann.)¤	66¤	Mathiam·Corvinum¤	34¤	¤
Bajeßid·Yildirim·(Ann.)¤	66¤	Ildirìm·Baìezid¤	34¤	يلدرم·بايزيد¶ Yıldırım·Bayezid¤
Heßarfenn·(Ann.)¤	66¤	Hezárfenn¤	34¤	¤
Karaman·Ogli¤	67¤	Caramanougly¤	39¤	قرامان·اوغلی¶ Karamanoğlu¤
Bajeßid¤	67¤	Baiezid¤	39¤	بايزيد¶ Bayezid¤

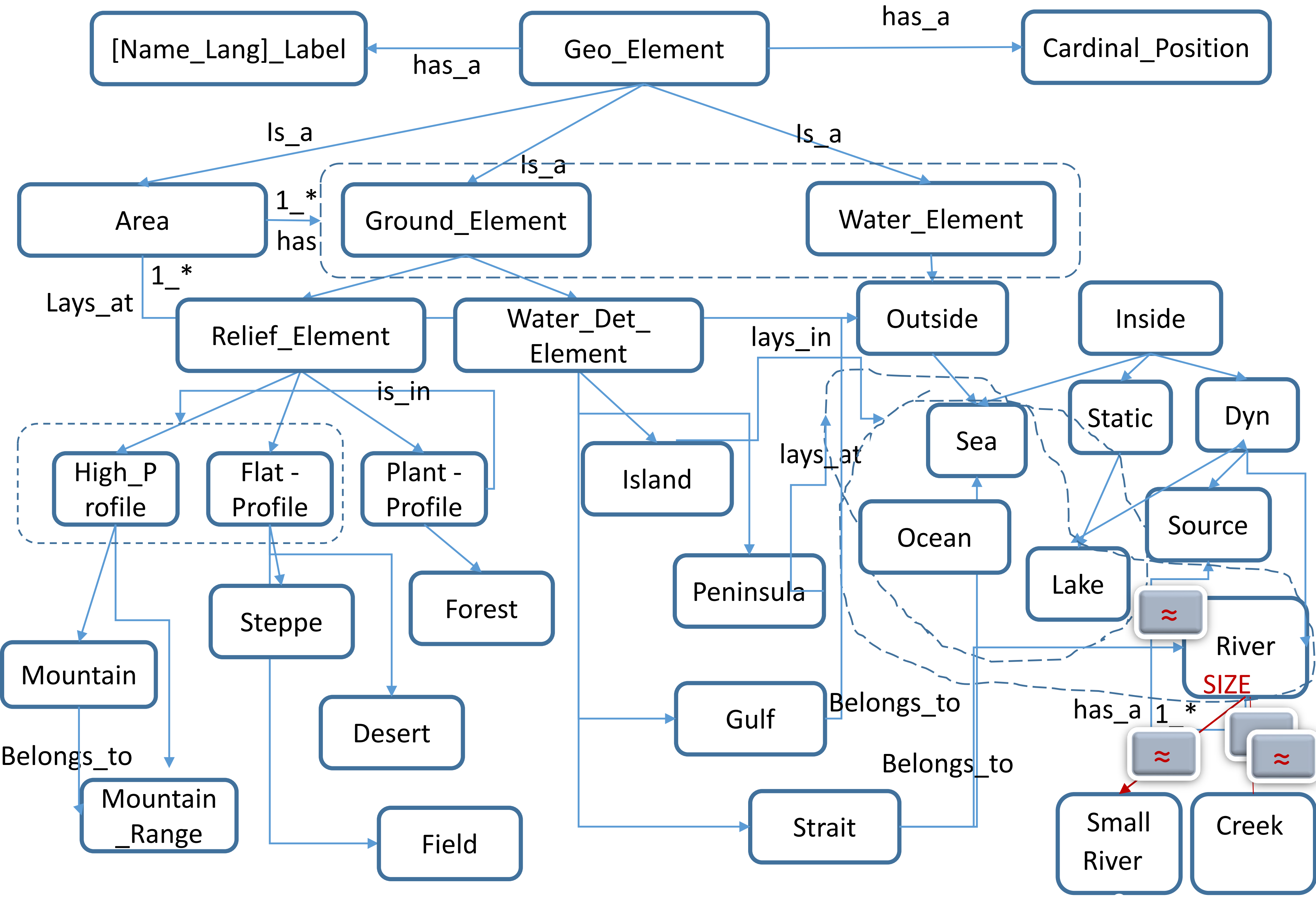
Places Database

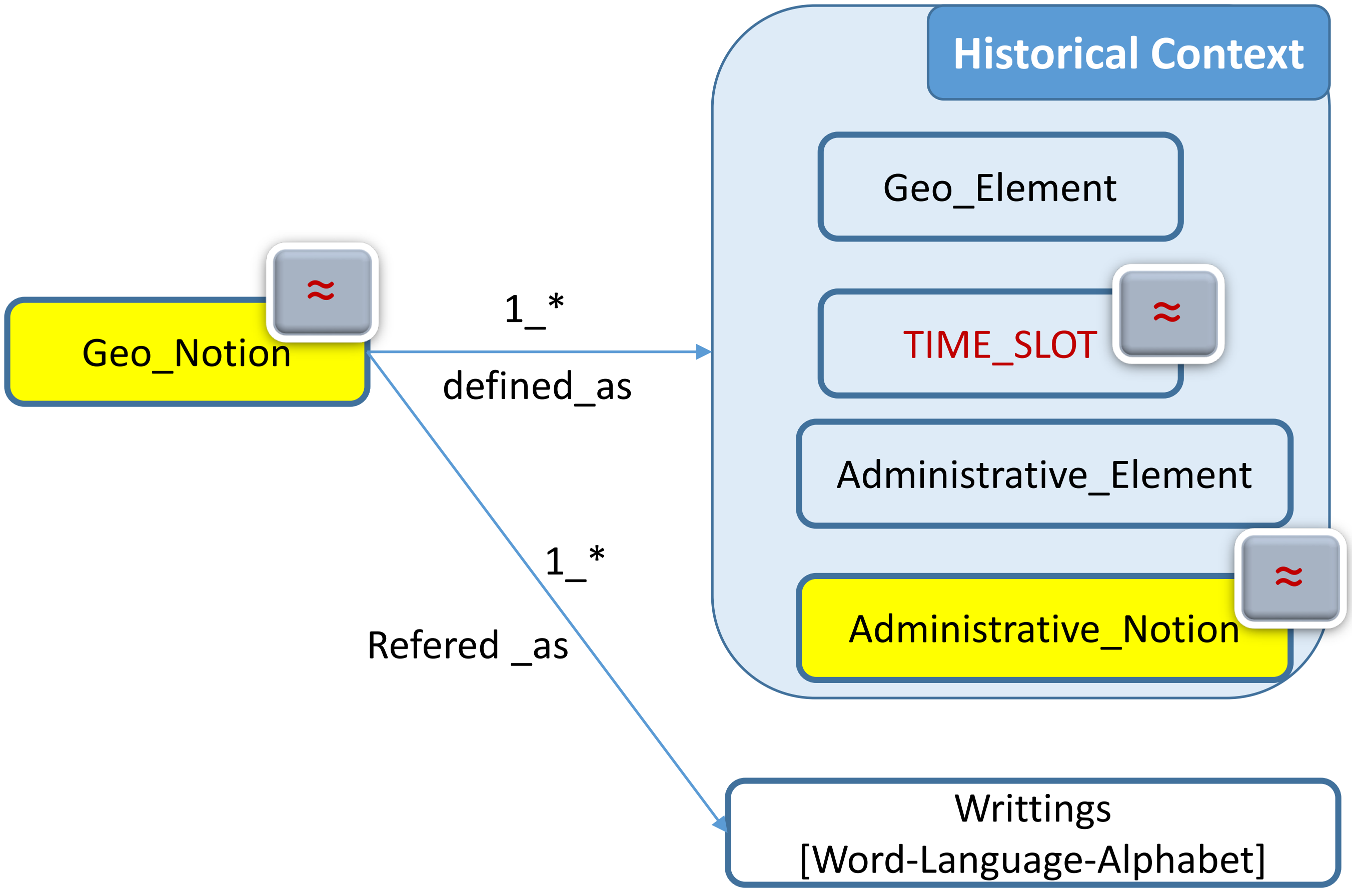


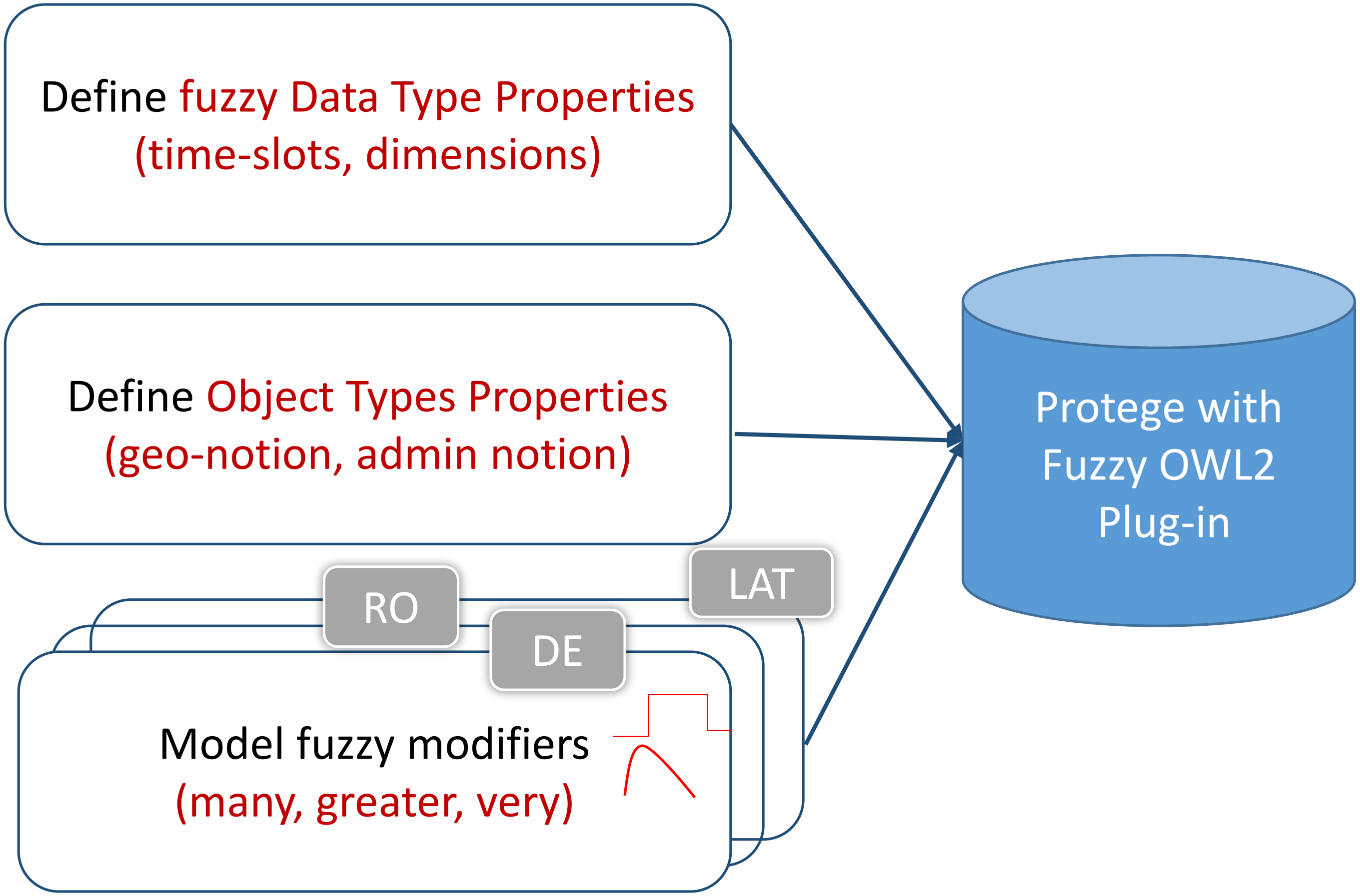
Adrianopel¤	73¤	Adrianopol¤	44¤	ادرنه·Edirne¤
Edrene·(Mewlasi)¤	73¤	Edrne·(Molasi)¤	44¤	ادرنه·Edirne¤
Edrene·(Mewlasi)·(Ann.)¤	73¤	Ederne·(Mollasì)¶ ادرنه·(مولاسی)¤	¤	ادرنه·Edirne¤
Adrianopel·(Ann.)¤	73¤	Adrianopolis¤	38¤	ادرنه·Edirne¤
Misr·(Mewlasi)·(Ann.)¤	73¤	Mýsr·(Mollasì)¶ مصر(مولاسی)¤	38¤	مصر·Mısır¤
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.)¤	¤

Persons Database

©Alptug Güney


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


Historical Context
Geo_Element
TIME_SLOT
≈
Administrative_Element
Administrative_Notion
≈
Geo_Notion
≈
1_*
defined_as
1_*
Refered _as
Writtings
[Word-Language-Alphabet]



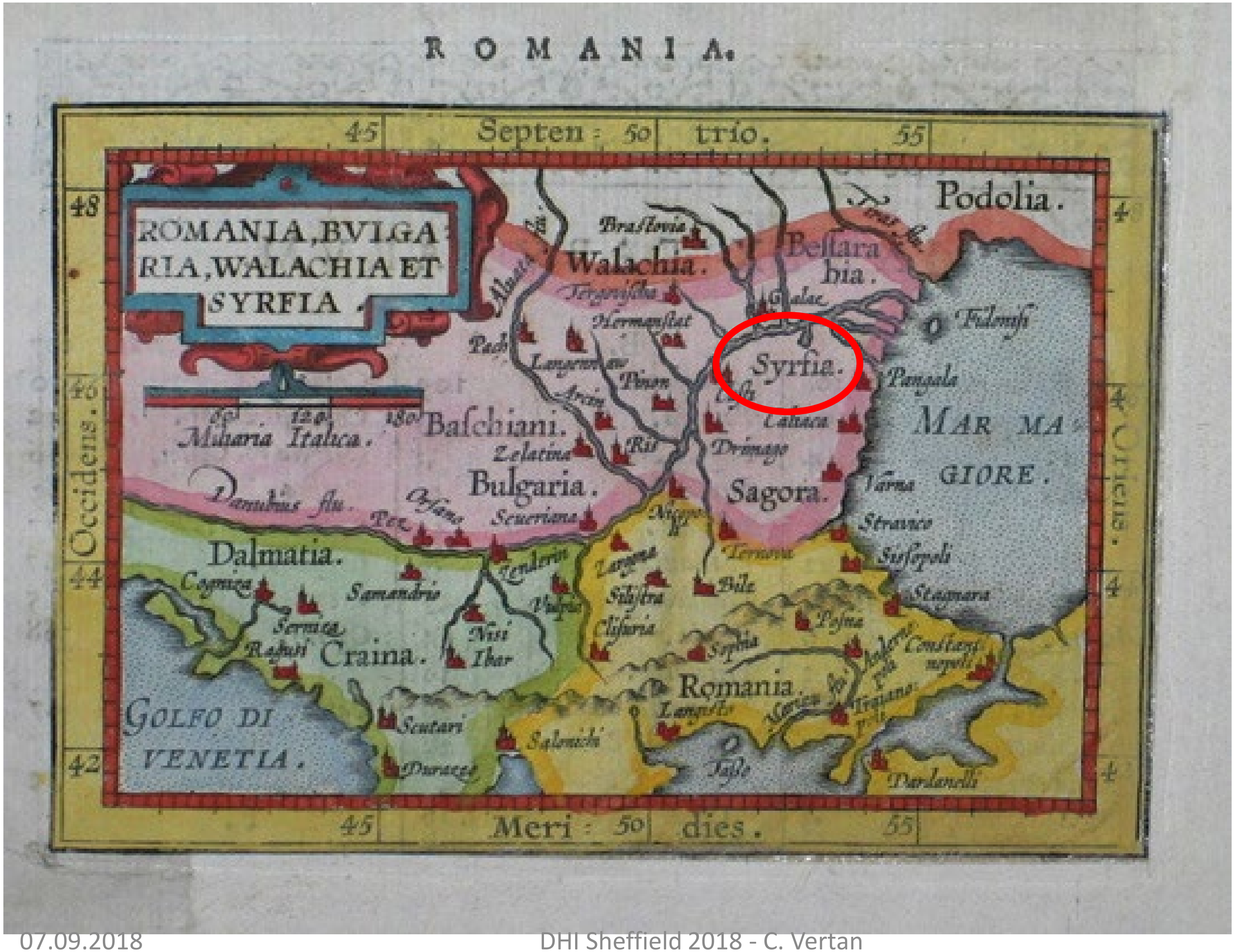
Define fuzzy Data Type Properties (time-slots, dimensions)

Define Object Types Properties (geo-notion, admin notion)

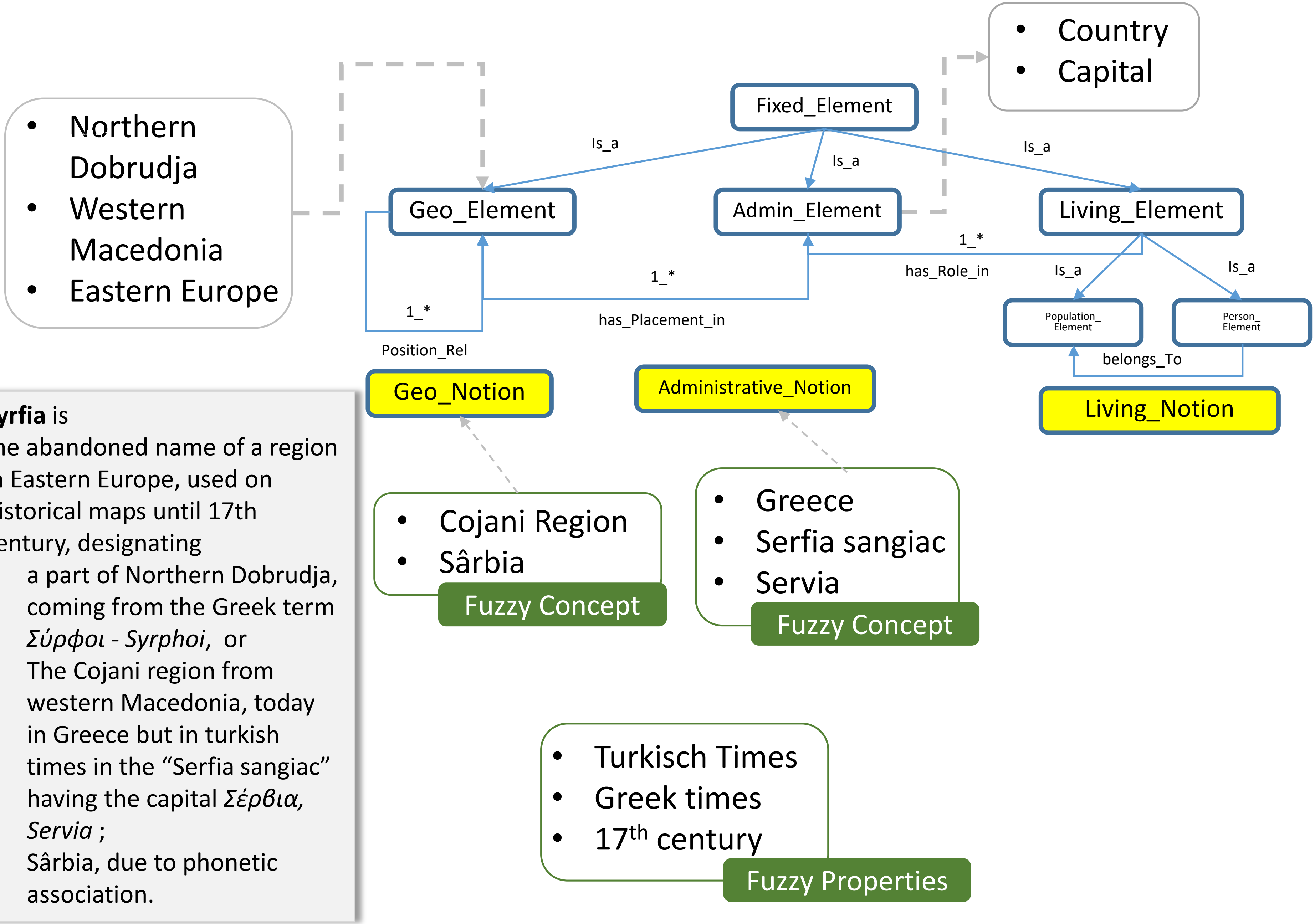
Model fuzzy modifiers (many, greater, very)

Protege with Fuzzy OWL2 Plug-in

Example: Annotation of ambiguous places



Ortelius Map 1570



- Northern Dobrudja
- Western Macedonia
- Eastern Europe

- Country
- Capital

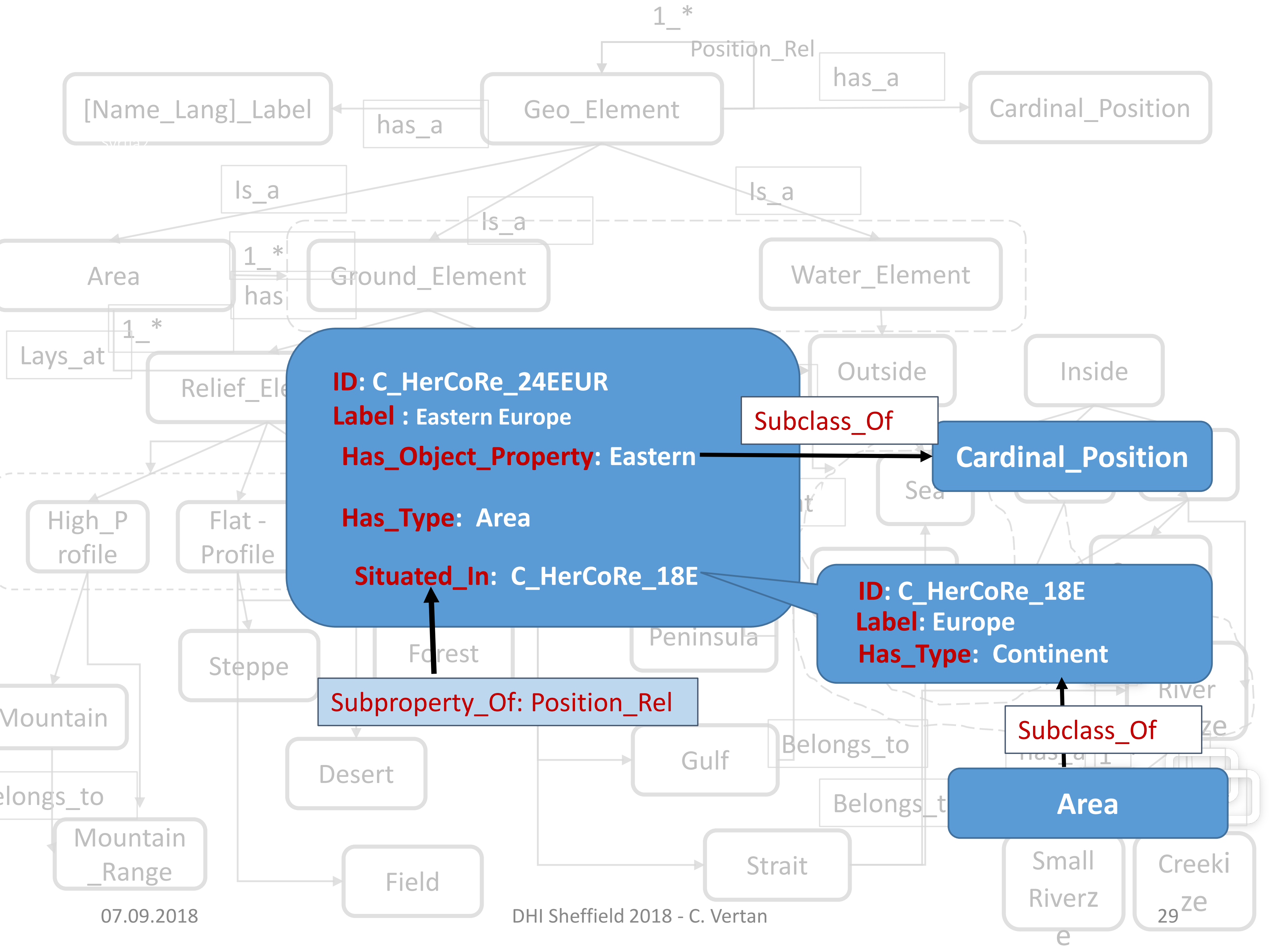
- Cojani Region
- Sârbia

- Greece
- Serfia sangiac
- Servia

- Turkisch Times
- Greek times
- 17th century

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.



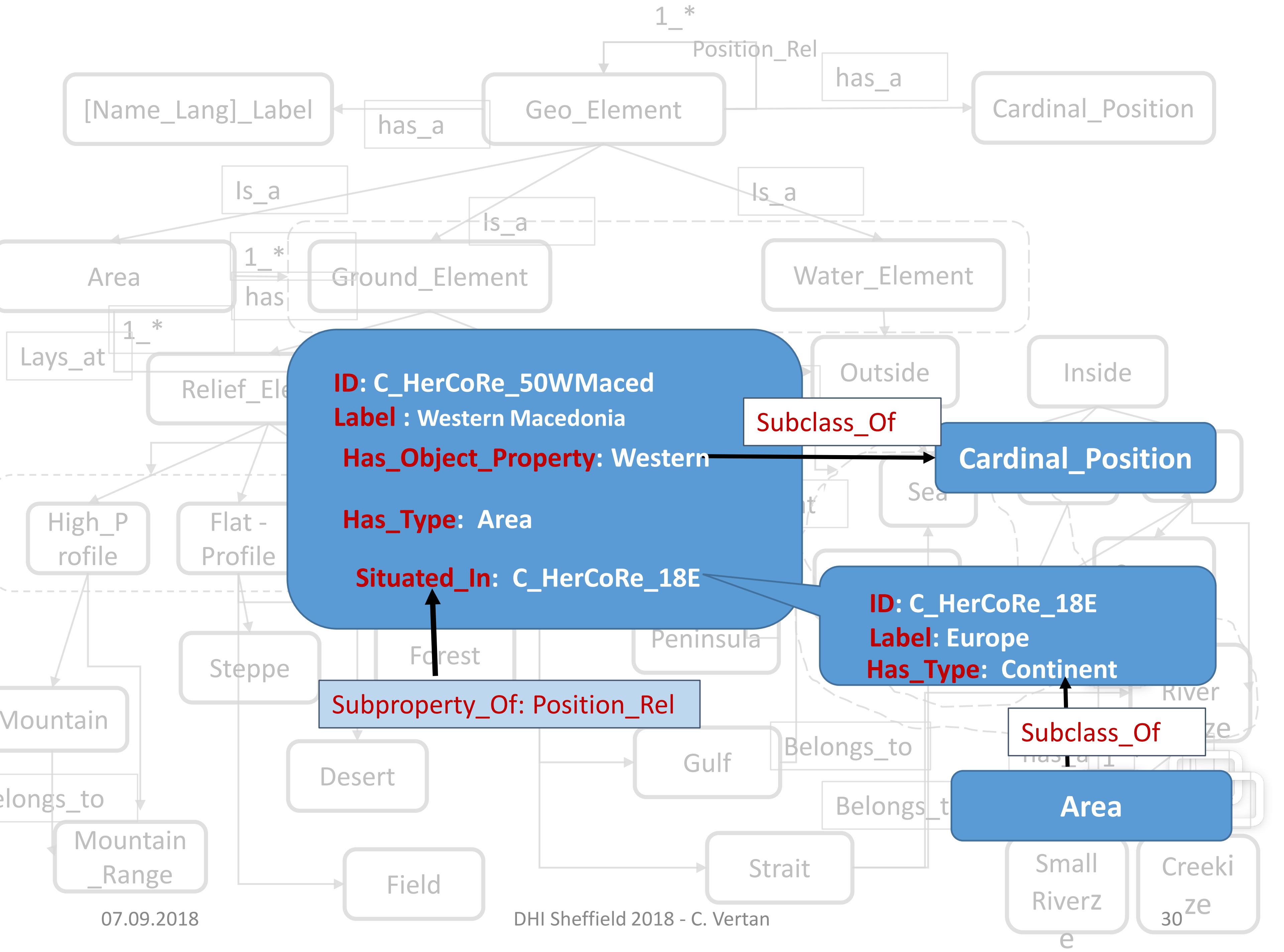
ID: C_HerCoRe_24EEUR
Label : Eastern Europe
Has_Object_Property: Eastern
Has_Type: Area
Situated_In: C_HerCoRe_18E

Subclass_Of
Cardinal_Position

Subproperty_Of: Position_Rel

ID: C_HerCoRe_18E
Label: Europe
Has_Type: Continent

Subclass_Of
Area



ID: C_HerCoRe_50WMaced
Label : Western Macedonia
Has_Object_Property: Western
Has_Type: Area
Situated_In: C_HerCoRe_18E

Subclass_Of

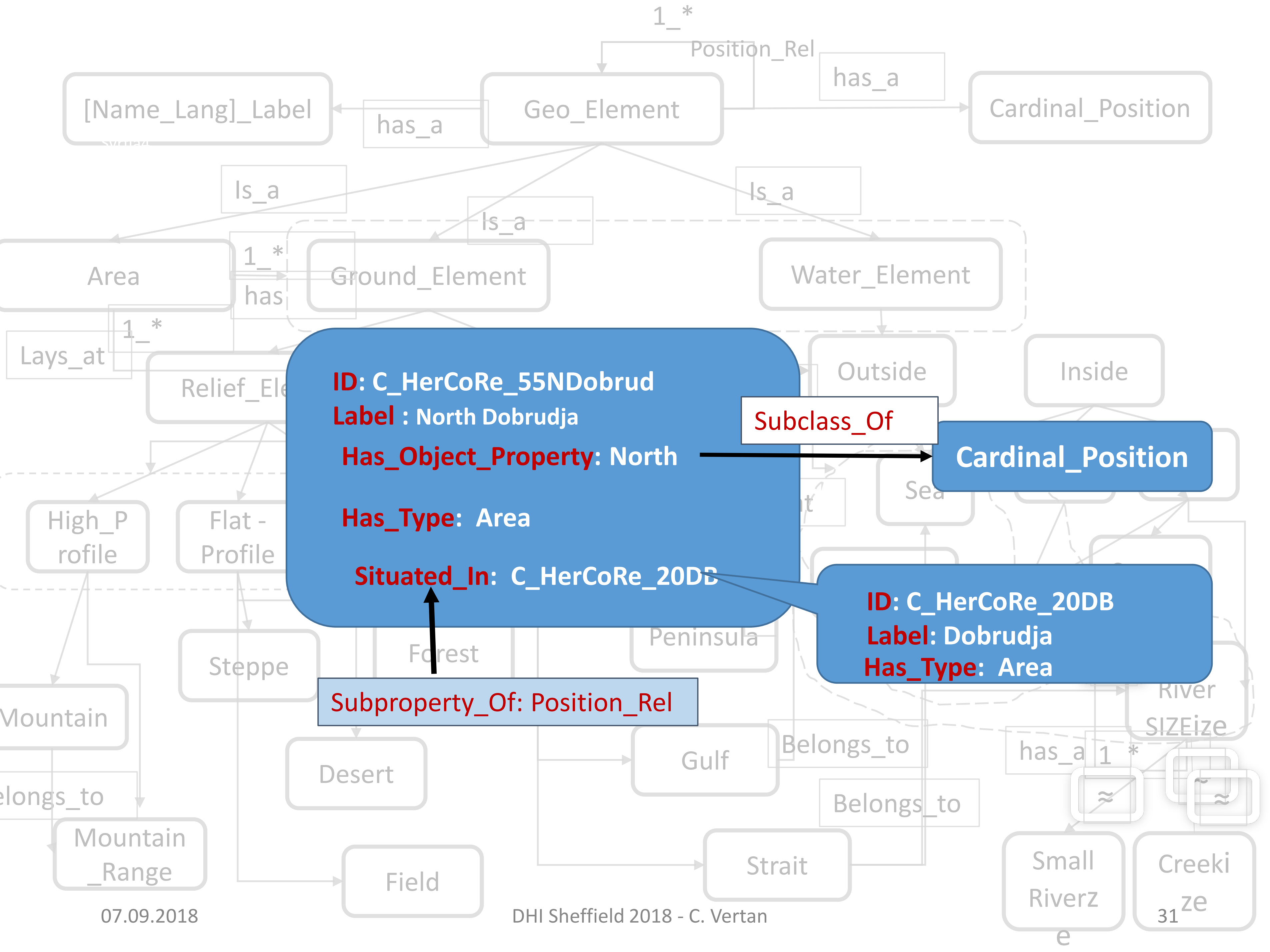
Cardinal_Position

Subproperty_Of: Position_Rel

ID: C_HerCoRe_18E
Label: Europe
Has_Type: Continent

Subclass_Of

Area



ID: **C_HerCoRe_55NDobrud**

Label : **North Dobrudja**

Has_Object_Property: **North**

Has_Type: **Area**

Situated_In: **C_HerCoRe_20DB**

Subclass_Of

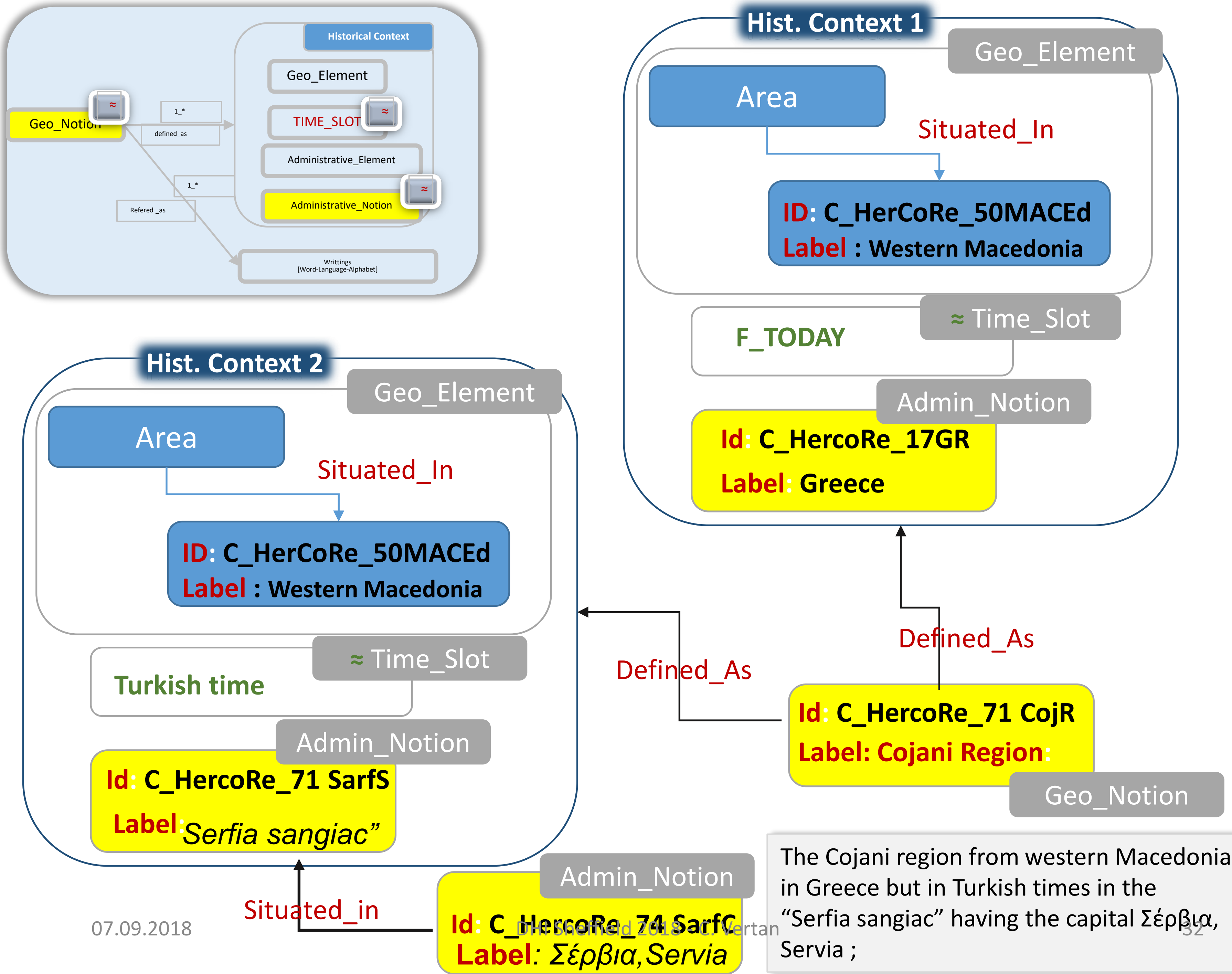
Cardinal_Position

Subproperty_Of: Position_Rel

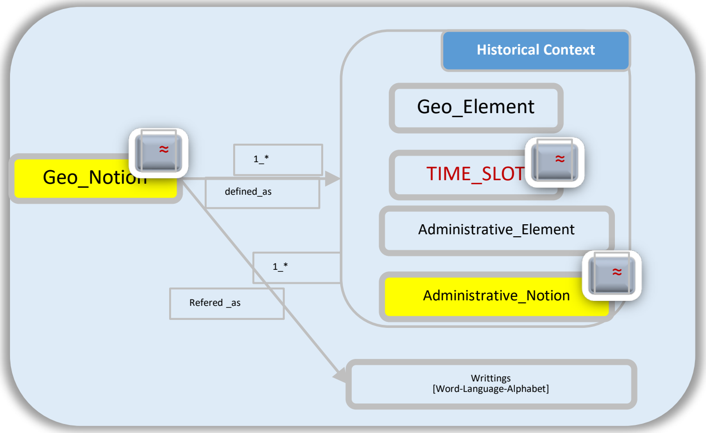
ID: **C_HerCoRe_20DB**

Label: **Dobrudja**

Has_Type: **Area**

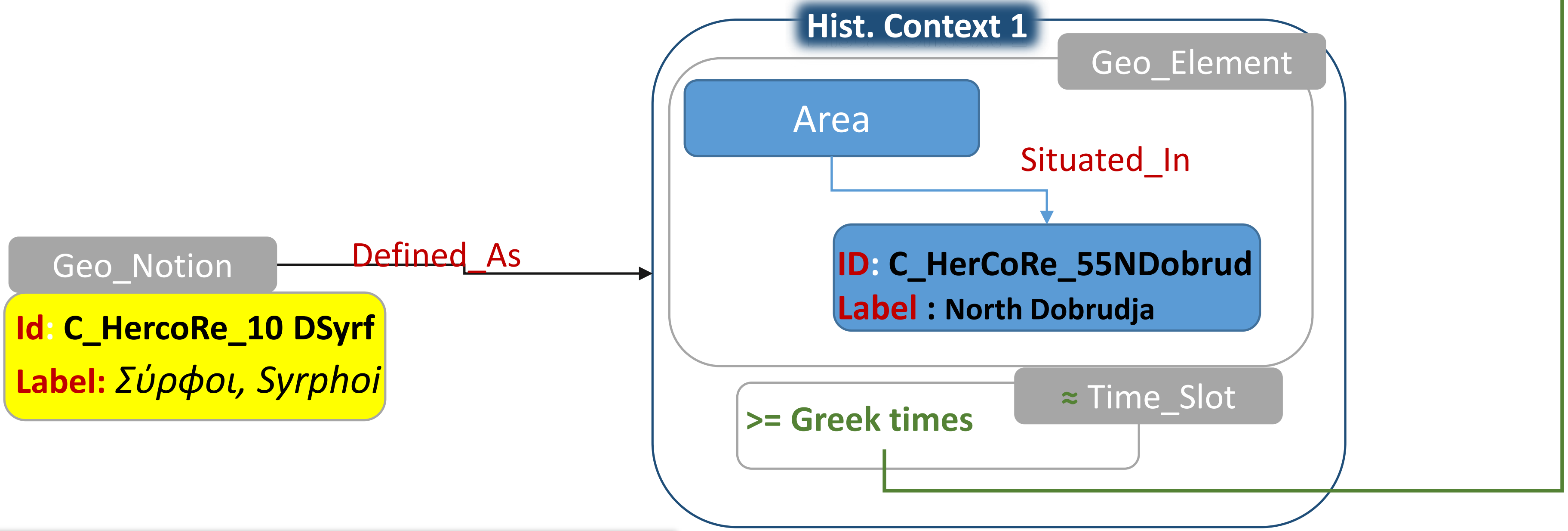


The Cojani region from western Macedonia, today in Greece but in Turkish times in the “Serfia sangiac” having the capital Σέρβια, Servia ;



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


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



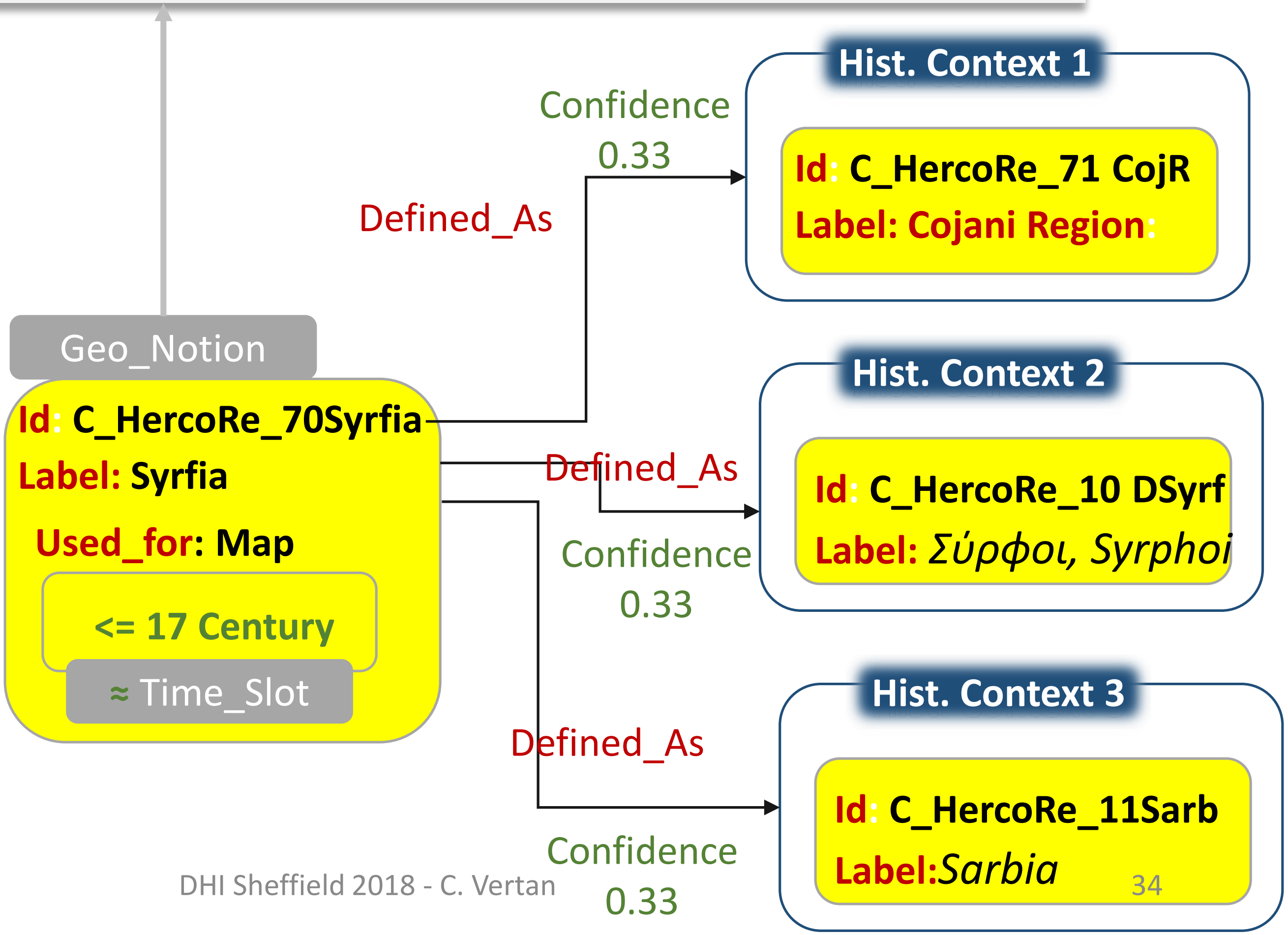
Part of Northern Dobrudja, coming from the Greek term Σύρφοι --Syrphoi;

```
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*

ROMANIA.
ROMANIA, BVLGARIA, WALACHIA ET SYRFIA
Septen trio.
Podolia.
Walachia.
Syrfia.
Bulgaria.
Sagora.
Dalmatia.
Craina.
Romania.
MAR MAGIORE.
GOLFO DI VENETIA.
Meri dies.

Orteliusmap 1570

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

Instance_Of

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

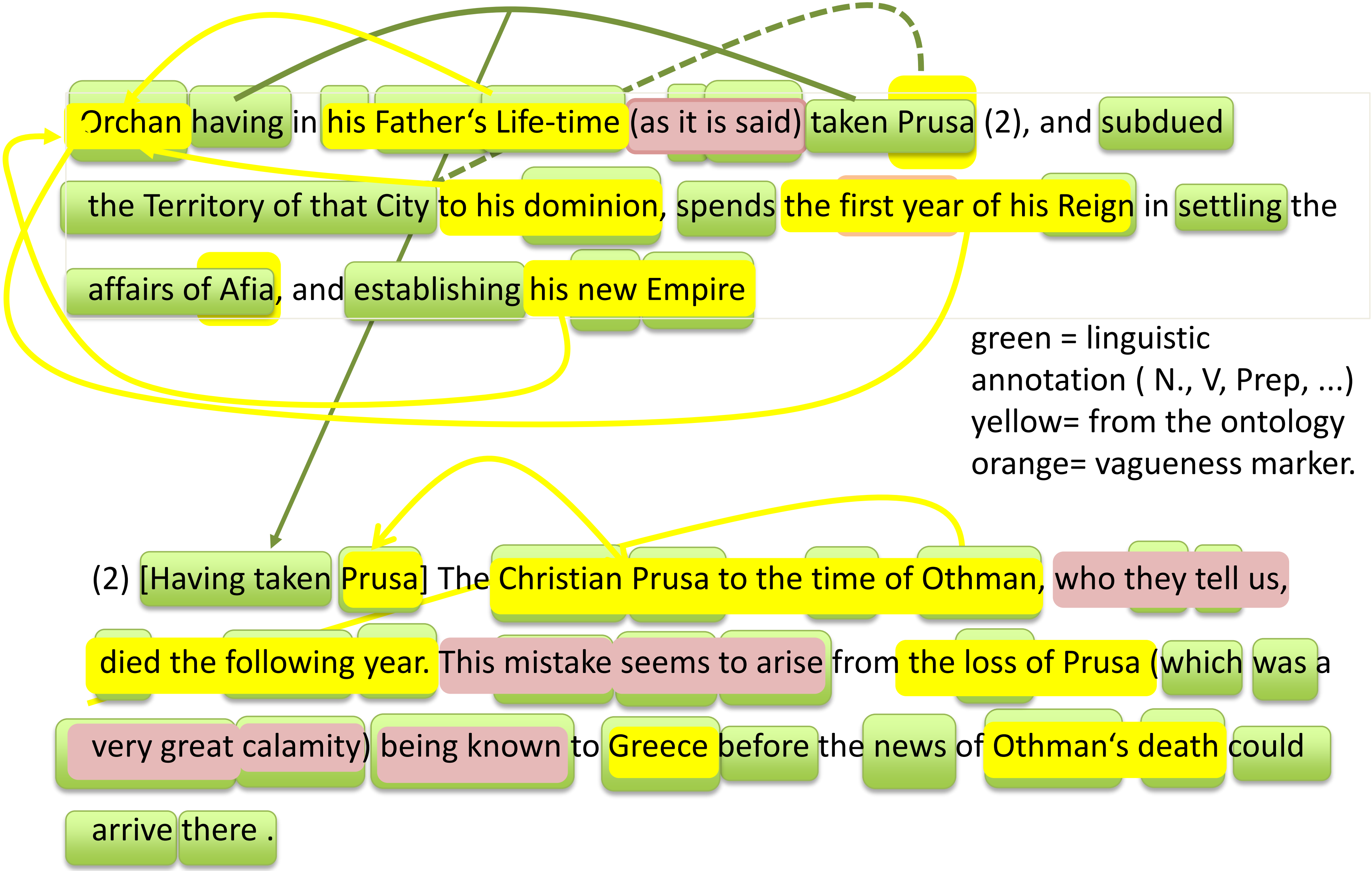
Hist. Context 3

Id: C_HercoRe_11Sarb
Label:*Sarbia*

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



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

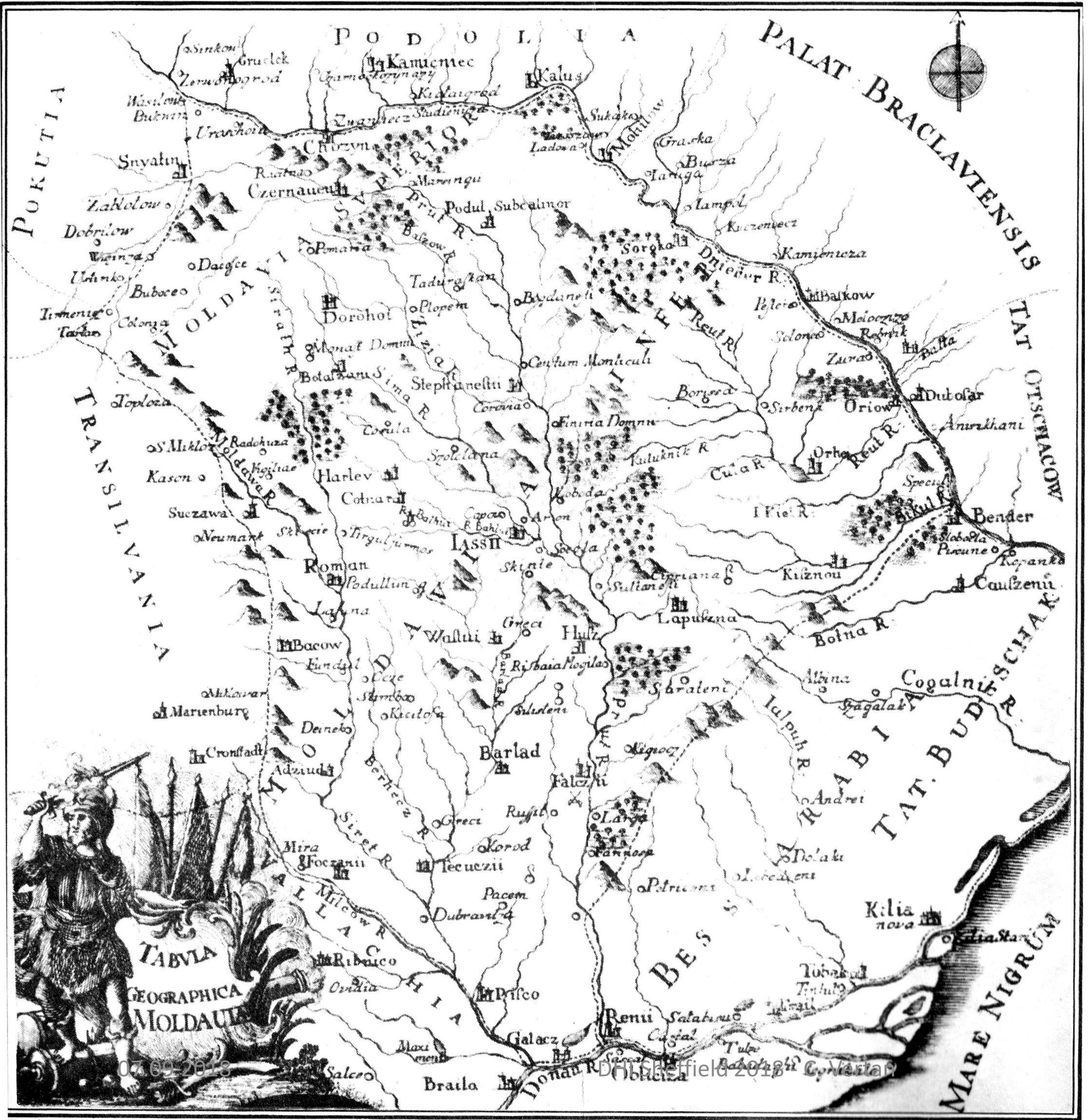
green = linguistic annotation (N., V, Prep, ...)
yellow= from the ontology
orange= vagueness marker.

(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 .

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

Annotation-Environment Functionality

- 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 automatisation 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)



Work ahead

- Semantic Linking between the map and „Geo_notions"
- Formal representation of vagueness markers
- Adaptation of existent fuzzy reasoners
- Specification of user scenarios
- Visualisation of results
- Reuse and enhance the ontology

AND..

- the entire hermeneutic analysis