



Vagueness and Precision of Historical Objects in DH

Walther v.Hahn

University of Hamburg • Computer Science Department

vhahn@informatik.uni-hamburg.de

Overview

Peculiarities of historical objects in DH,
Why vagueness is interesting,
Wilhelm Dilthey and his theory of humanities,
An example for representing vagueness.

Next generation DH:

1. Non standard-database storage of objects,
2. more specific metadata,
3. more specific linguistic information,
4. inferential expansion of the annotation.

DH Research on Historical Objects

differs in central aspects from research in present time objects:

- incomplete tradition of artefacts,
- fragmentary knowledge of background facts,
- strange semantic encoding (gestures, colours, reference)
- esp. for texts:
 - unfamiliar script,
 - deviating lexical material,
 - divergent writing style,
 - different semantic environment (synonyms, metaphors, co-reference).

Why is Vagueness Interesting?

Vagueness \approx Fuzzyness \approx Uncertainty \approx Unclarity \approx Ambiguity.

- Humanities seemed too imprecise, because of the lack of methods, technologies and verification routines.
- DH applied computational methods for higher precision and better data coverage.
- Along these lines, however, DH neglected the vagueness problem, because it seemed too difficult.
- Therefore, a historic interpretation often looks like a collection of facts instead a collection of hypotheses.
- In general, human symbolic interaction cannot be strictly precise. Otherwise we would need proper names for all single tokens in the world. We wouldn't be able to use "many".
- In image processing, e.g., vagueness can be treated, because you already start with some RGB values, which still are a "colour". In contrast, the binary value "100011110110" of the vague operator "many" is not a "vague operator". You have to give it a semantic description.
- Thus, "seldom" and "rarely" don't have the same binary representation.

Types of Vagueness

Linguistic and logical Vagueness

abstracts	„area“
comparatives	„bigger“, „more“
inexact adjectives	„square“, „near“,
relative adjectives	„big“, „rare“
Inexact measures	„10 feet“
hedges, shields,	„rather“, „more or less“
non-intersectives	„so-called“, „supposed“
modals, attitudes	„probably“, „hopefully“
indexical deixis	„there“, „now“, „you“
continua	„water“, „traffic“,
vague quantifiers	„many“, „most“
complex quantifiers	„roughly half of the 20-30 thousand warriors died in this battle“

Historical obscurity

- ”capire l’antifona”
- “Ai tempi in cui Berta filava”
- “Rimandare qualcosa alle calende greche”

Factual Uncertainty

(yet) unexplored facts	<i>„the moon is 384402,56 m away from the earth“</i>
range expressions	<i>„The beginning of the 18. century“ „Romania in the middle ages“</i>
uncertain definition	<i>„the northern slope of the mountain“</i>
Inexact measures	<i>„a 4 days' journey, 10 feet“</i>
unclear place	<i>„Syrfia“ [on the Ortelius map]</i>
unclear facts	<i>„by order of the sultan“</i>
unclear time	<i>„in prehistoric times“</i>
unclear person	<i>„the former prince“</i>
unclear action	<i>„the submission of the barbarians“</i>

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 .

A Complicated Explicit Example: Ambiguity

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

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

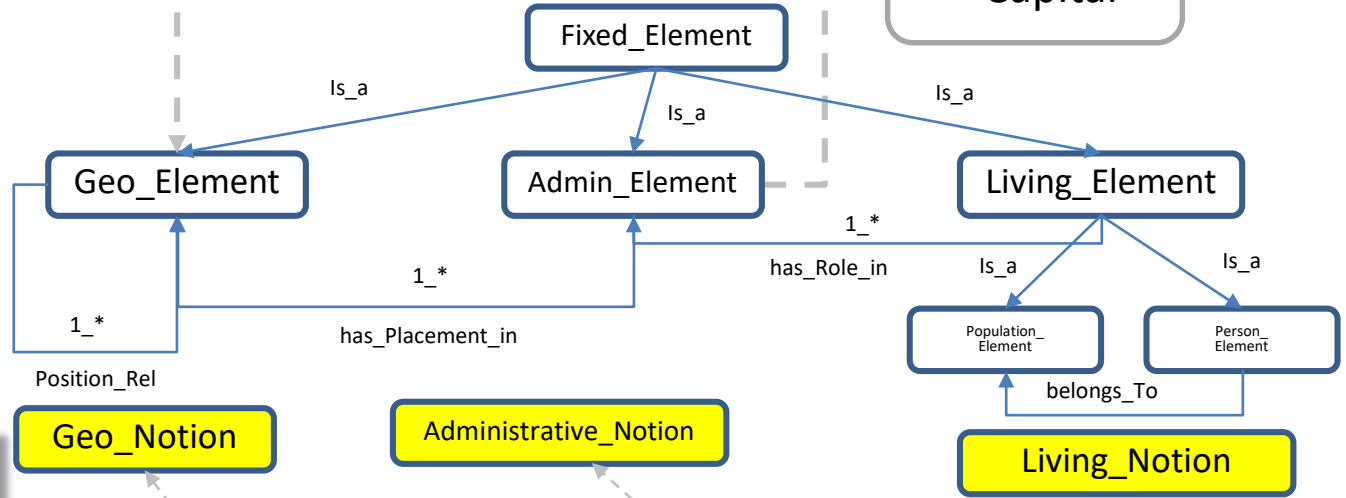


Realistic Example: Ortelius' map from 1570



- Northern Dobrudja
- Western Macedonia
- Eastern Europe

- Country
- Capital



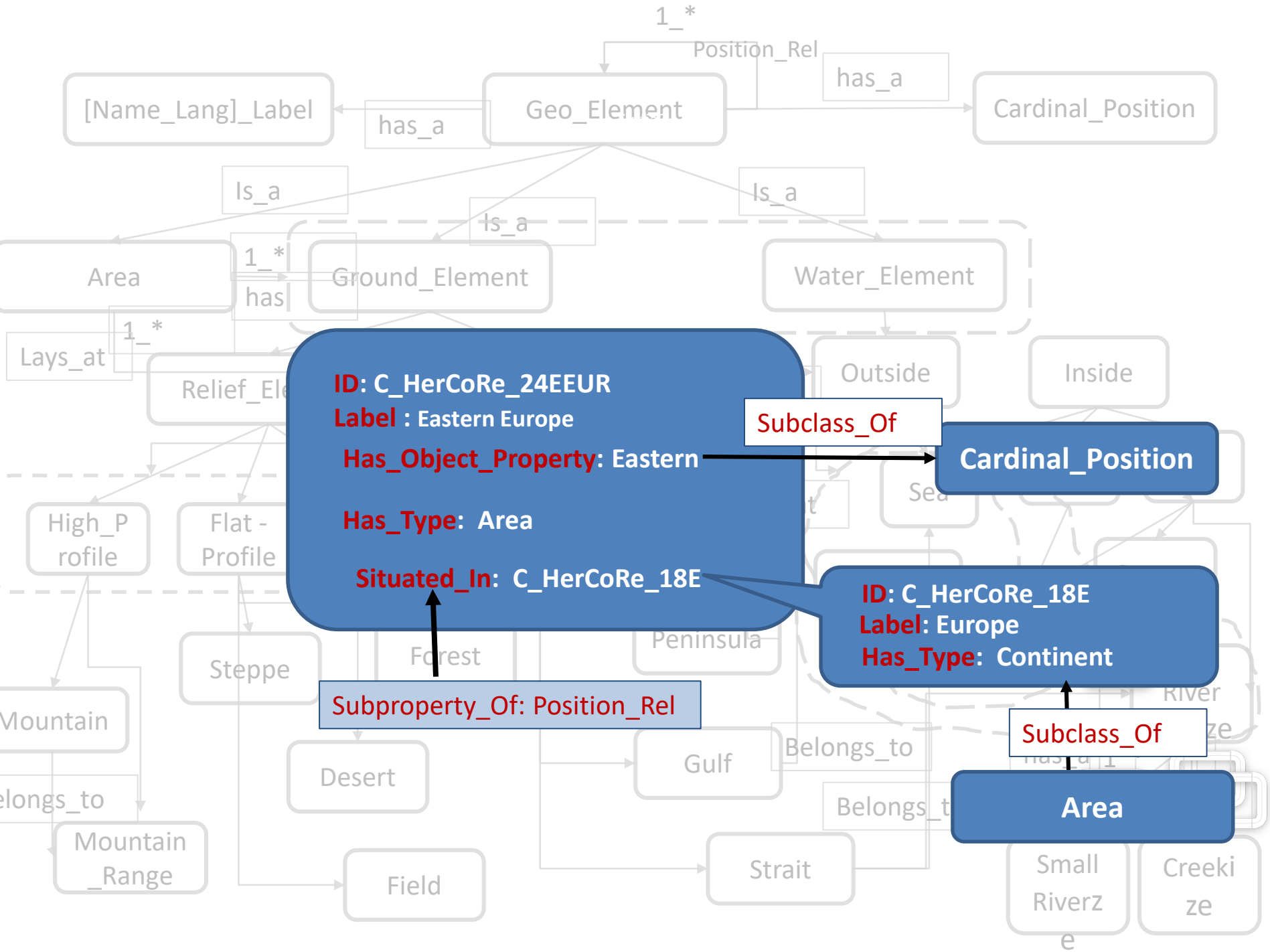
Syrfia is the abandoned name of a region in Eastern Europe, used on historical maps until 17th century, designating

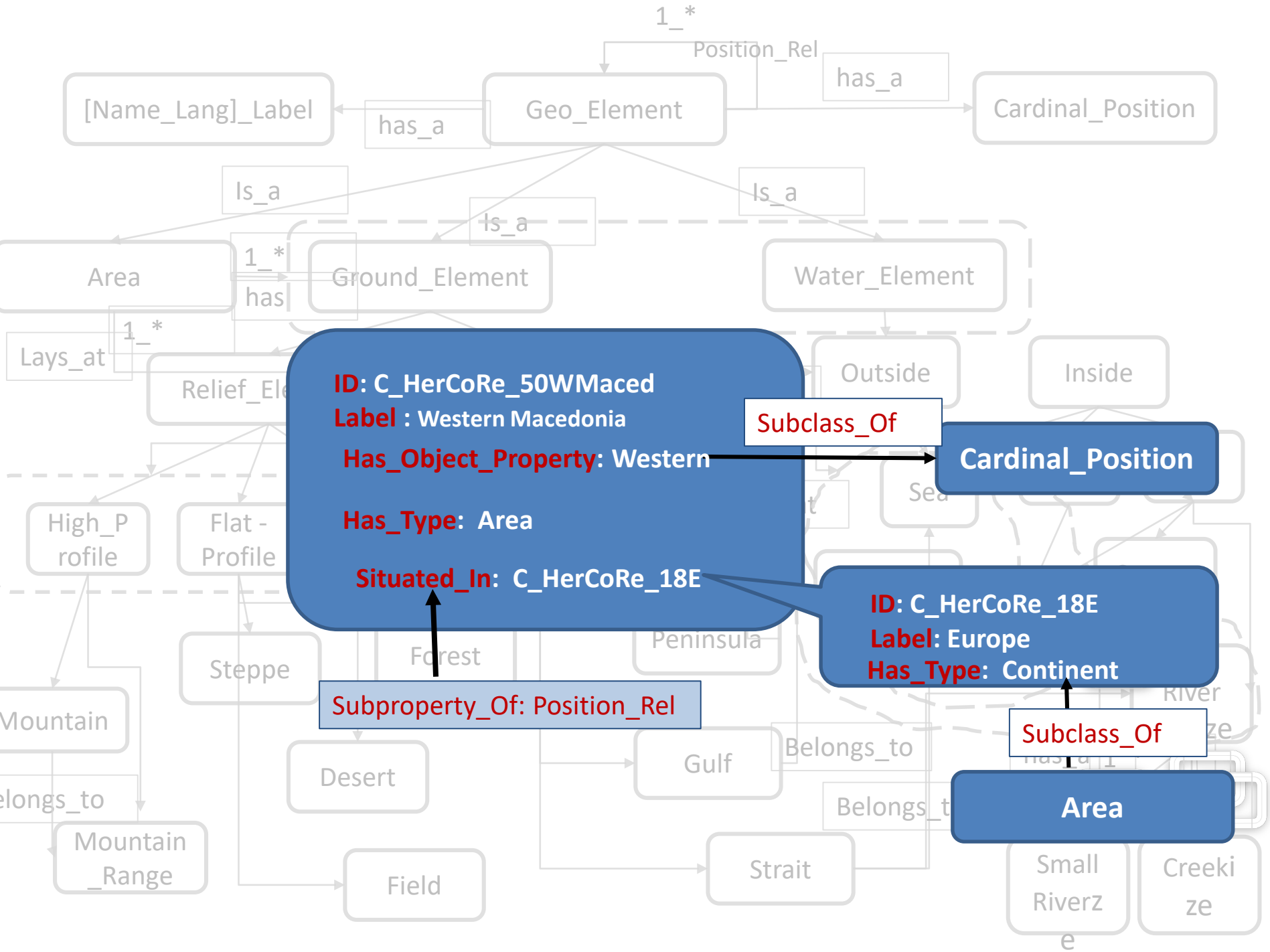
- a part of Northern Dobrudja, coming from the Greek term *Σύρφοι - Syrphoi*, or
- The Cojani region from western Macedonia, today in Greece but in Turkish times in the "Serfia sangiac" having the capital *Σέρβια, Servia*;
- Sârbia, due to phonetic association.

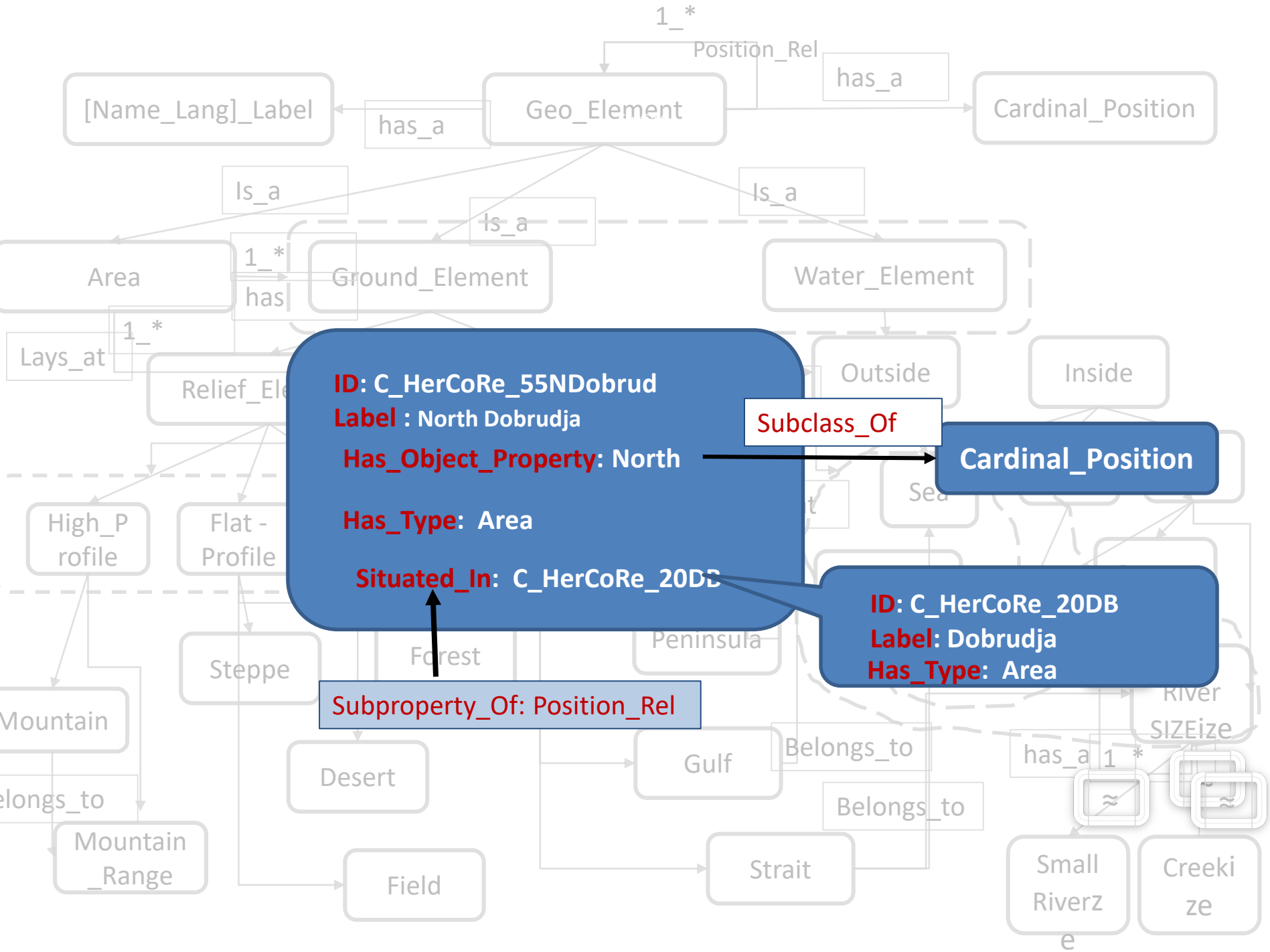
- Cojani Region
 - Sârbia
- Fuzzy Concept

- Greece
 - Serfia sangiac
 - Servia
- Fuzzy Concept

- Turkish Times
 - Greek times
 - 17th century
- Fuzzy Properties







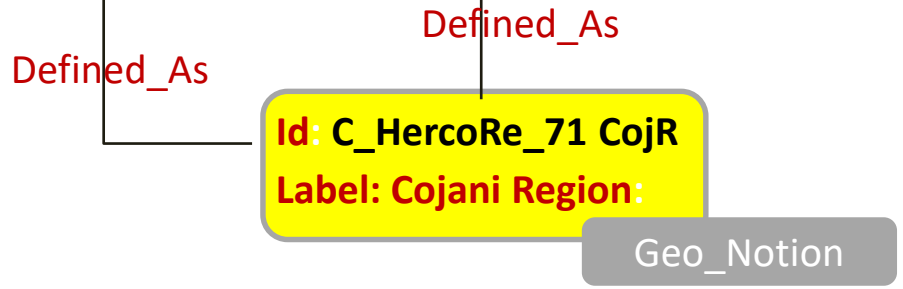
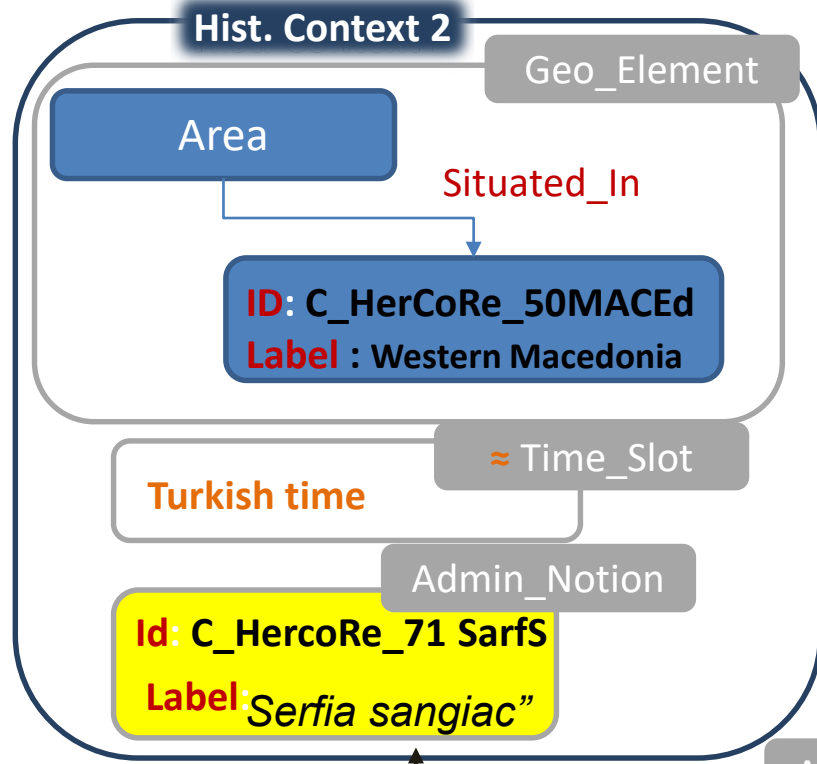
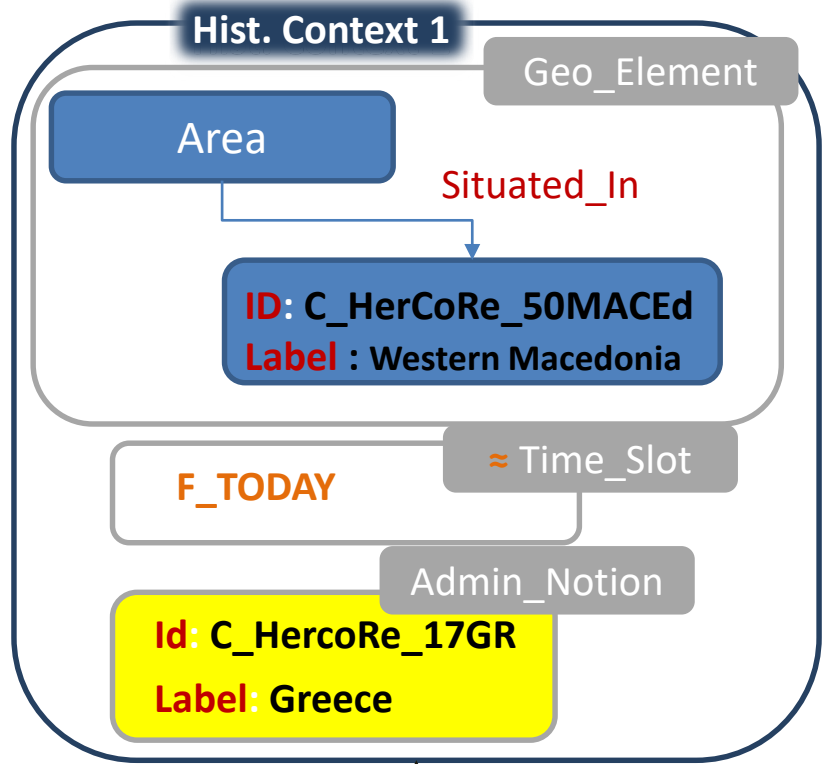
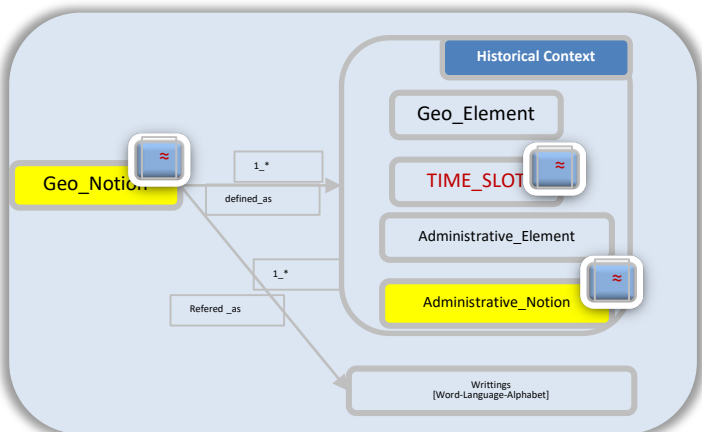
ID: C_HerCoRe_55NDobrud
Label : North Dobrudja
Has_Object_Property: North
Has_Type: Area
Situated_In: C_HerCoRe_20DB

Subclass_Of

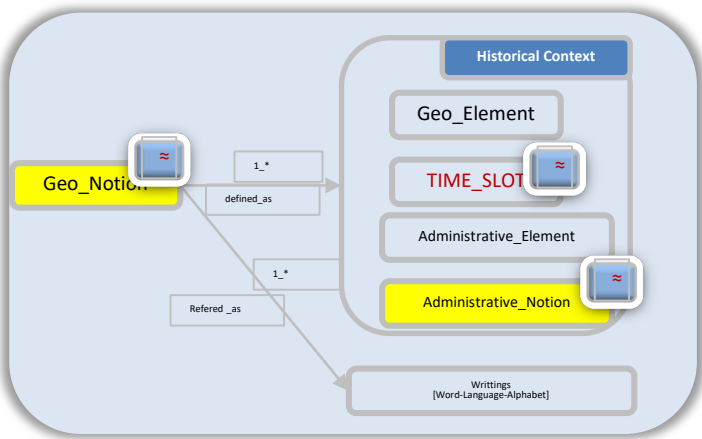
Cardinal_Position

ID: C_HerCoRe_20DB
Label: Dobrudja
Has_Type: Area

Subproperty_Of: Position_Rel



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



```

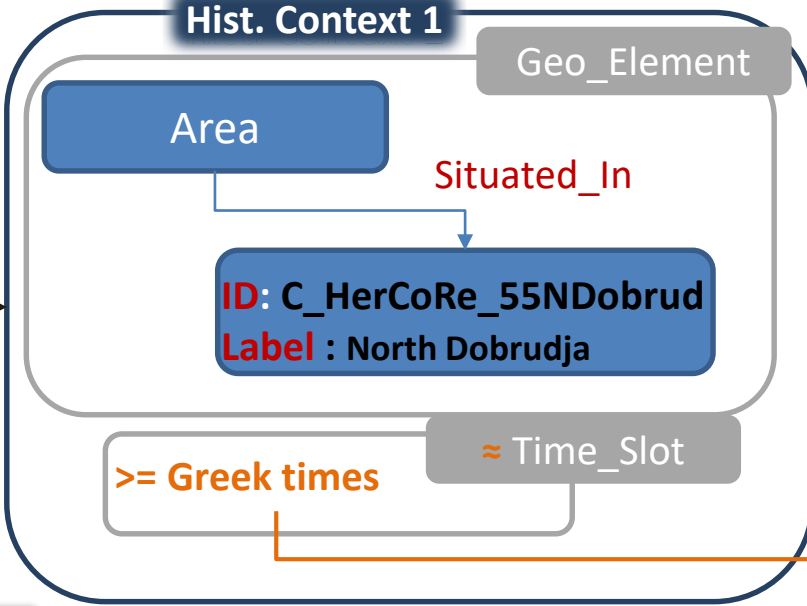
<DatatypeDefinition>
  <Datatype IRI='#GreekTimes' /> <DataIntersectionOf>
    <DatatypeRestriction>
      <Datatype abbreviatedIRI='xsd:integer' />
      <FacetRestriction facet='&xsd;minInclusive'>
        <Literal datatypeIRI='&xsd;integer'>-750</Literal>
      </FacetRestriction>
    </DatatypeRestriction>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

(fuzzyLabel

```
< fuzzyOwl2 fuzzyType =" concept " >
```

```
< Concept type =" weightedSum " >
```

```
< Concept type =" weighted " value ="0.33" base ="C_HercoRe_71CojR " / >
```

```
< Concept type =" weighted " value ="0.33" base =" C_HercoRe_10DSyrf " />
```

```
< Concept type =" weighted " value ="0.33" base =" C_HercoRe_11Srb " />
```

))

Syrfia is

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

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

Source: Wikipedia

Geo_Notion

Id: C_HercoRe_70Syrfia

Label: Syrfia

Used_for: Map

<= 17 Century

≈ Time_Slot

Defined_As

Confidence

0.33

Hist. Context 1

Id: C_HercoRe_71 CojR

Label: Cojani Region:

Defined_As

Confidence

0.33

Hist. Context 2

Id: C_HercoRe_10 DSyrf

Label: Σύρφοι, Syrphoi

Defined_As

Confidence

0.33

Hist. Context 3

Id: C_HercoRe_11Sarb

Label: Sarbia



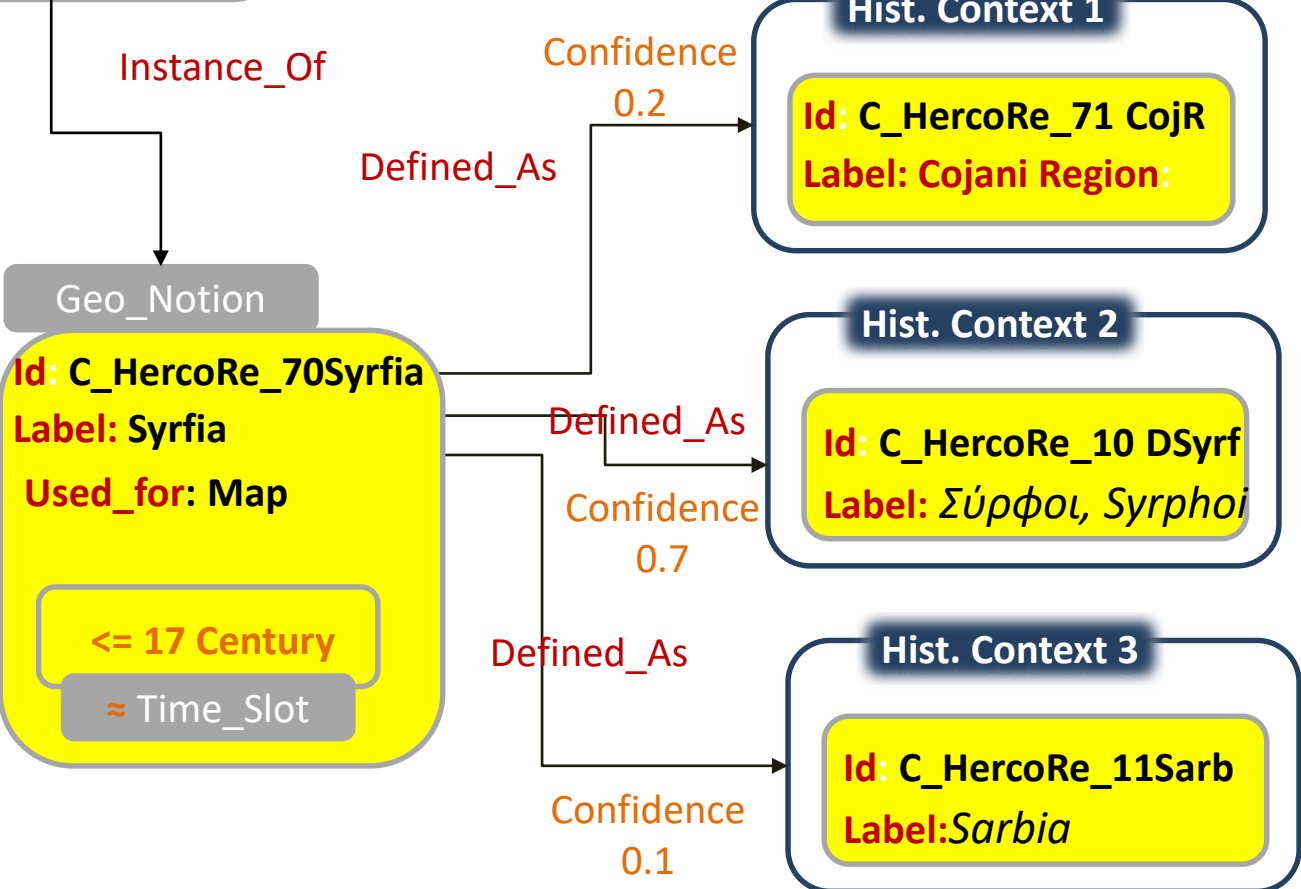
Orteliusmap 1570

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

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



The Precision Fallacy

In DH often the precision of the interpretation result (hermeneutic process), consisting of

1. a research-guiding issue (*„is Cantemir quoting his sources sincerely?“*)
2. defining adequate processes
 - *„every quotation is checked for its author“* or
 - *„every author is checked for his authenticity“* and
3. inference rules (*„if all quotations have the same credibility ...“*)

→ is mixed up with

- the scientific preciseness of scrutinizing all cases with the same formal routines (higher data coverage in keeping a precise formal language), and the
- computational accuracy of retrieval processes.

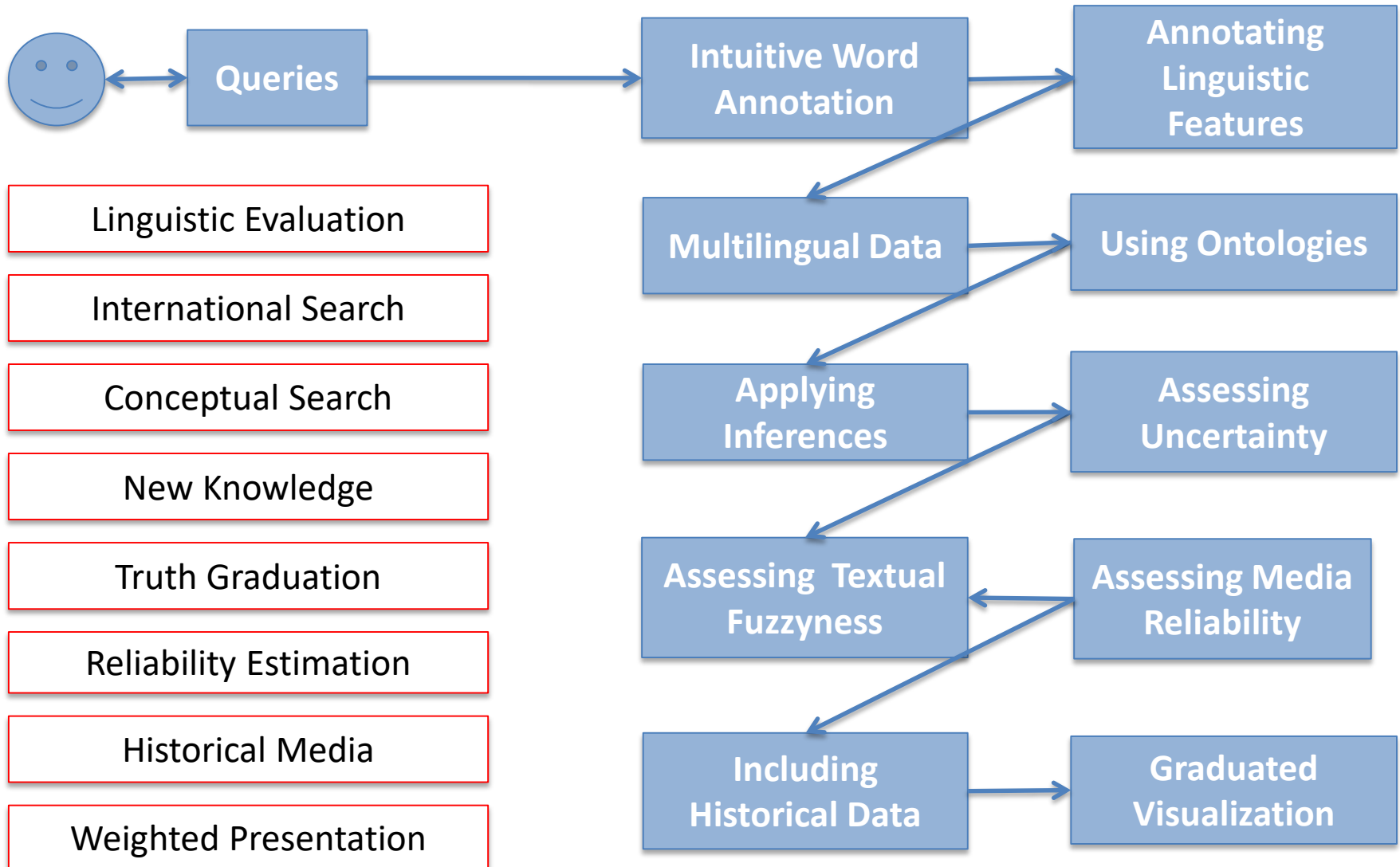
Wilhelm Dilthey and his Theory of Humanities

Wilhelm Dilthey (a philosopher from Leipzig) describes history as “a series of world views.” Man cannot understand himself through reflection or introspection, but only through what “history can tell him ..., never in objective concepts”.

Dilthey emphasizes the “intrinsic temporality of all understanding” i.e., that man’s understanding is dependent on past world views, interpretations, and a shared world.

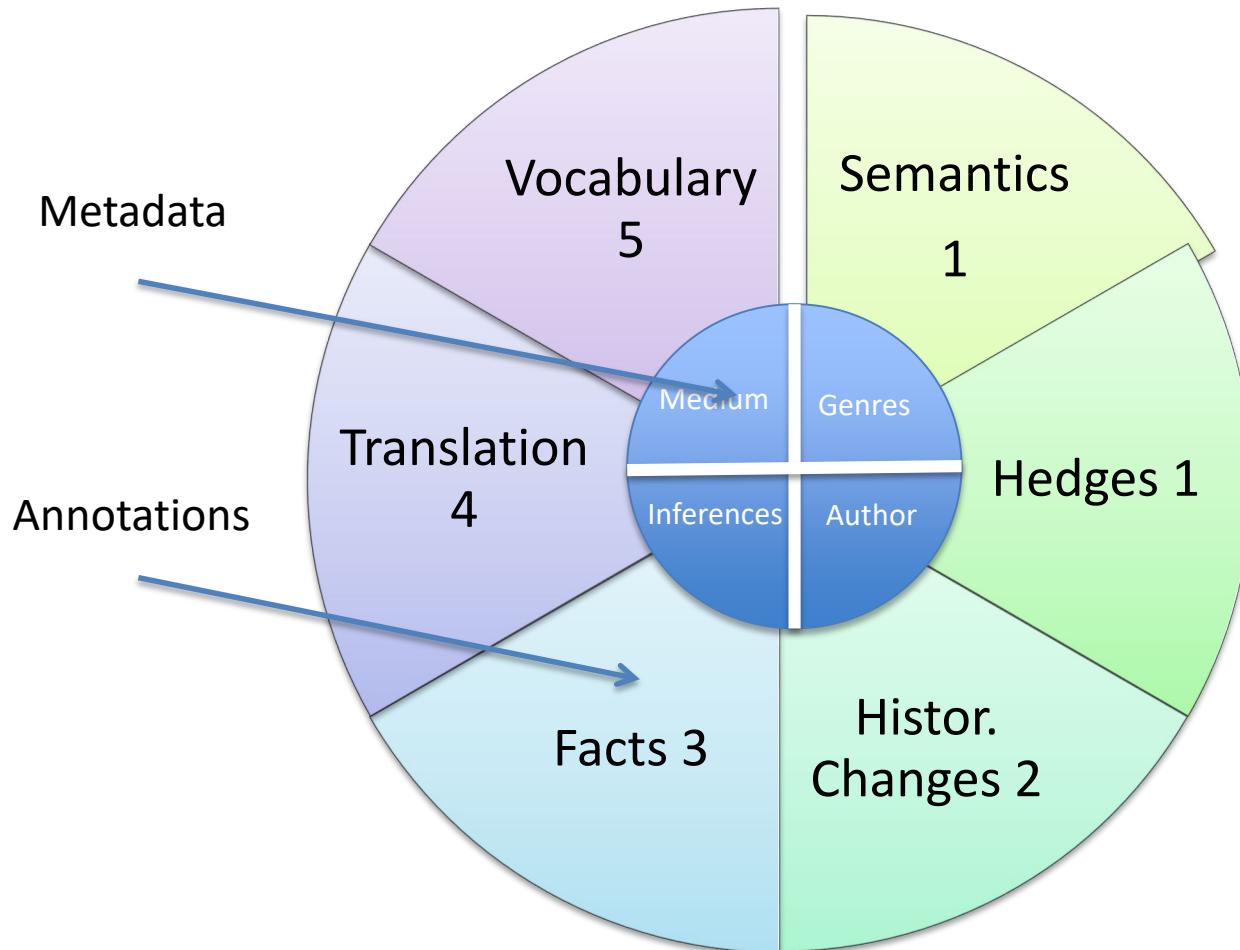
He argues that 'scientific explanation of nature' (*erklären*) must be completed with a theory of how the world is given to human beings through symbolically mediated practices

Best Practice for Including Vagueness



How to Display Vagueness

Vagueness of the material at hand must be always visible to the reader/ researcher (at least in a summarizing way) indicating the ratio of various vagueness types. This will remind the reader of their hermeneutic task.



Criticism 1: Data Base Technology

The storage of objects in traditional (relational) databases, designed for business or natural science applications, pretends to deal with facts, especially, if

- vagueness expressions are omitted from the text fragments or
- factual uncertainty is not annotated.

Nevertheless, relational data bases are used in DH for text storage.

Result:

see the Kilia example:

→ NE “Kilia”. is a capital with alternat.NE “Lycostomon”.

Or:

“Before this the Sireth and Tortuß was believed to be the border of Walachia”

→ Sireth and Tortuss was the border of Walachia”

Descripio Moldaviae, transl., p.41”

Criticism 2: Why are Metadata so General?

Metadata are used mostly for life data of the author, annotation details or further bibliographic details.

However, metadata are the best place to represent the

- structure , genre and style of the object,
 - the credibility and source dependency of the author,
 - the tradition lines of the text, i.e. quotations.
-
- Thus they can be included in an inference chain. .g. the label of Ortelius' map

Criticism 3: Shallow linguistic annotation

Words are not only sequences of characters
they have

- a syntactic function,
- a semantic meaning,
- a given/new role
- a coherence role,
- a discourse marker, and
- a pragmatic value.

As we have shown, they are not necessarily continuous:

Germ. *“der Präsident gab seinen Posten auf”*

“gab ... auf” is a separable verb form, for semantic reasons it must be represented as variant of *“aufgeben”*.

Criticism 4: No Reasoning

Annotations as such are local and don't have a global influence on other sentences, paragraphs or texts.

As they are asserted by the corpus, it seems adequate, to combine truth/fuzzy values of these propositions and thus achieve new knowledge.

Example:

If

Iași was part of Moldavia in 1850 T

∧ Moldavia in 1850 was part of the Ottoman Empire, T

∧ the Ottoman Empire lost Mōldavia in 1878 T

then

Iași was not in the Ottoman Empire after 1878 T

Without reasoning the data base is only an echoing storage and is not really a repository of knowledge.

Conclusions

Next generation DH have the urgent tasks of :

1. Including vagueness in a systematic and formal manner
2. Annotating of deep linguistic features
3. Intensive usage of meta-data
4. Including all levels of annotation (content and meta-data) in reasoning processes (inferences)
5. Better exploiting of media interferences

References

Cimiano, Philipp, Christiane Unger und John McCrae: *“Ontology-based interpretation of natural language”*. Morgan & Claypool 2014.

Dilthey, Wilhelm, *Einleitung in die Geisteswissenschaften. Versuch einer Grundlegung für das Studium der Gesellschaft und der Geschichte*. Leipzig 1883 ff.

Pinkal, Manfred, "Semantische Vagheit: Phänomene und Theorien, Teil I". In: *Linguistische Berichte* Nr. 70, S. 1-26 and 72, S. 1-26. Wiesbaden 1980.

v.Hahn, Walther, *“Vagheit bei der Verwendung von Fachsprachen”*. In: Hoffmann / Kalverkämper /Wiegand: *Fachsprachen*. Band 1. Berlin 1998. S. 383 – 390.

Cristina Vertan and Walther v. Hahn, *On the Annotation and Interpretation of Vague Assertions in Historical Texts*, in *Proceedings of the DCH2017, Interdisciplinary Conference on Digital Cultural Heritage*, Kremers, H. (Ed.), Berlin August 2017.

Zadeh, Lotfi, *Fuzzy sets. Information and Control* 8 (1965): 338–353.

Reste der alten Folien

Cross-Media Interference

Like in the Pavia project, artefacts, images, animation, and text

- explain,
- specify, or
- correct each other.

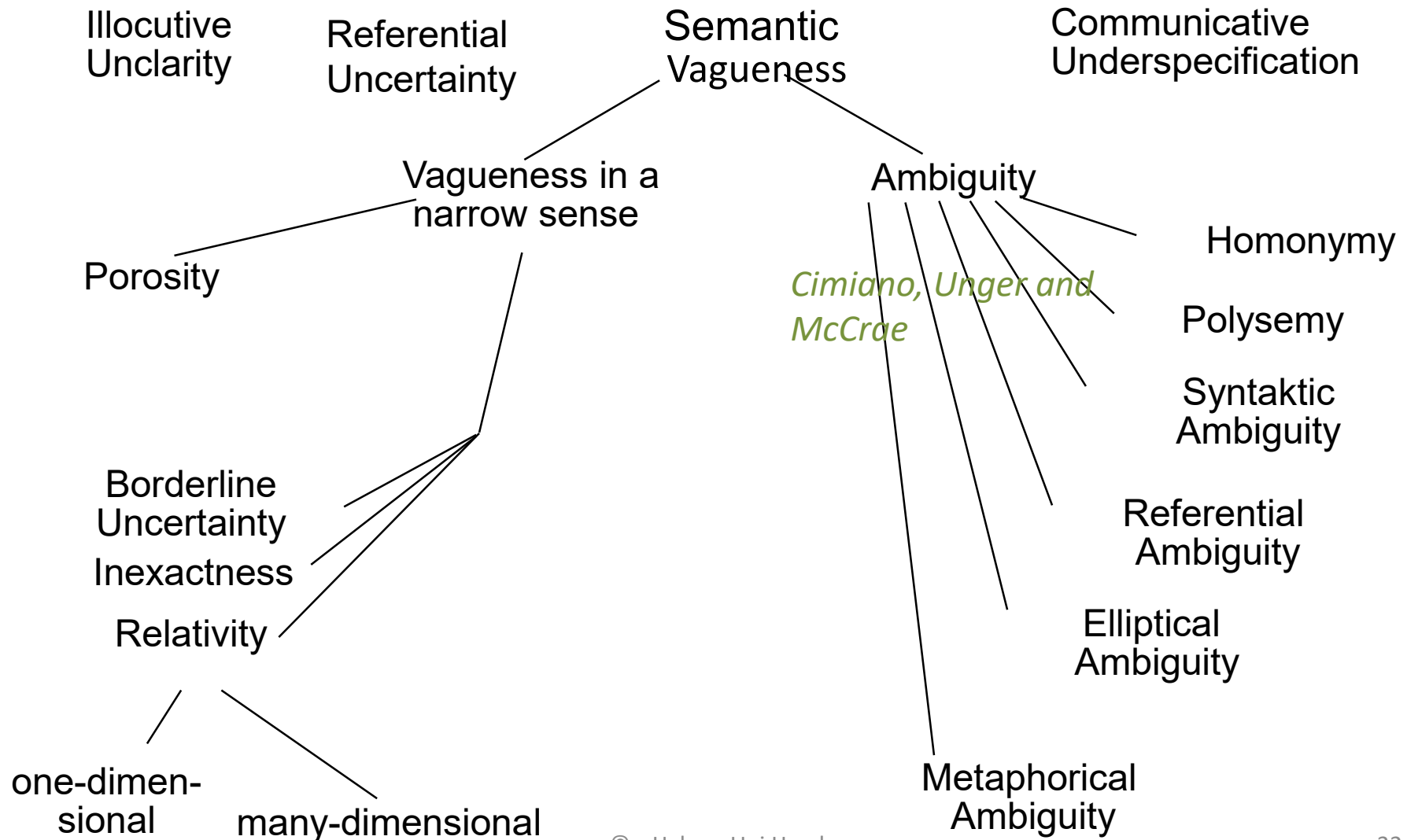
which is self-explaining for remains of architecture, but already pictures follow symbolic rules, e.g.

- the size of persons according to their rank,
 - the vicinity of cities, for political/religious reasons,
- even more in textual sources like historical documents.

They are not realistic, but must be interpreted by humanities.

Even if hard-science material is processed, hermeneutic methods are indispensable.

Manfred Pinkal's Schema of Semantic Vagueness



Linguistic Ambiguity

metaphors

“head” (line), “crane”

several distinct meanings

metonymy

“the sublime porte”

two contiguous meanings

“Constantinopel decided”

abstraction, specification

synecdoche

“the Turkish sultan never...”

homonymy / homography:

“fluke” = 1. the fins on a [whale](#)'s tail. 2. a stroke of [luck](#).

two or few distinct meanings

“übersetzen” = 1. cross a water, 2. translate

polysemy

many contiguous meanings

“Man” = 1. The human species,
2. Males of the human species,
3. Adult human male

Historical Linguistic Change

historical obscurity	"capire l'antifona" "Ai tempi in cui Berta filava" "Rimandare qualcosa alle calende greche"
historical abbreviations	' <i>Karl</i> ' = Charlemagne, "Pius II" = Enea Silvio Piccolomini = The pope in 1500 A.D.
changing meaning	germ. ' <i>wohl</i> ' is not vague, as in modern German, but is the adverb of ' <i>good</i> '
named entities	<i>Istanbul</i> = <i>Constantinopol</i> , <i>Konstantinopel</i>
different writing	"Keniginn" = Königin (queen), "nemlig" = nähmlich"
different script	Cyrillics in Romania, black letters in Germany

Linguistic Vagueness

semantic abstracta	<i>„area“</i>
comparatives	<i>„bigger“, „more“</i>
inexact adjectives	<i>„square“, „hostile“, „near“, „distant“</i>
relative adjectives	<i>„big“, „rare“</i>
Inexact measures	<i>„4 days' journey“, „10 feet“</i>
pp Attachment	<i>„cop kills man with knife“</i>
elliptic ambiguity	<i>„Russia stronger“</i>
question operators	<i>„why“, „how“</i>
hedges, shields,	<i>„rather“, „more or less“</i>
non-intersectives	<i>„so-called“, „supposed“</i>
modals, attitudes	<i>„probably“, „hopefully“</i>

Referential Vagueness

indexical deixis „*there*“, „*now*“, „*you*“

deictic ambiguity „as said before“, „in this picture“, “in this area”

anaphora (left reference)

„*the latter*“, „*done*“

cataphora (right reference)

„*the following:*“, „*namely*“

Logical Vagueness

presuppositions

„I will use the Maserati“

Implikations:

⇒ There is a vehicle called Maserati,

⇒ I have a Maserati”

ellipses

„this too?“

continua

„water“, “traffic“, “the Russian war“

vague quantifiers

„many“, most“

complex quantifiers

„roughly half of the 20-30 thousand warriors died in this battle“

conjunctions

„Pietro and Giulia are married“

among them or to 2 other persons?

scopus ambiguity:

inclusive and exclusive readings,

„All Dacians have an enemy“.

Dacians each have an enemy, or

All Dacians have the same enemy

Introduction

Digital Humanities (DH) is on archiving and presenting material (particularly historical artefacts), but

also on introducing deeper scientific reflexion in humanities by a propagation of computational methods. However, more than ten years of computer-aided research did not lead to an adequate digital modelling of historical objects in a proper hermeneutic sense. In most DH-attempts the main cruces remain

1. the storage of objects in database architectures designed for business or natural science applications,
2. the annotation by only very general metadata,
3. mark-up with merely shallow and local linguistic information,
4. a missing quantitative and inferential analysis.

Consequently images and texts become artificially precise and the mutual illumination of texts and other media loses its traditional hermeneutic power.

Because of these drawbacks many researchers in humanities still investigate digital objects with rather traditional methods.

A paradigm change in modelling and representation of objects with the characteristics of humanities' research is urgently needed.

In this talk I will present current research of our group aiming at modelling vagueness and uncertainty in a way that keeps historical categories and the potential of explanatory relations among media.