Representing Vagueness of Historical Objects in Digital Humanities

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. quantitative and inferential analysis.

DH Research on Historical Objects

differs in central aspects from research on present time objects:

- incomplete tradition of media,
- fragmentary knowledge of background facts,
- strange semantic encoding (symbols, colours, reference)
- esp. for texts:
    - unfamiliar script (blackletters),
    - variational lexical material,
    - different orthography,
    - non-normalized writing,
    - deviating syntax,
    - divergent discourse style,
    - different semantics and semantic environment (synonyms, metaphors, co-reference).

Processing tools

- For technical projects based on measurements - like the Pavia reconstruction -, geometry or shadowing of buildings is the same for any historical period as for present time. Even if metrics change over time (from foot to cm), the size of a building or its foundation walls do not change.

- Modern arithmetic and graphical tools can be used, names can be replaced.

- Humanities, which are directly or indirectly based on texts, rely on symbolic objects like words, phrases, dialogues.

- Even the creation of a machine readable text (e.g. digitalization) is error-prone and thus hand work is required,

- Linguistic tools often have a very high error rate:

- Tokenizers, parsers, pos-taggers, lexicon tools, morphological or semantic analysers, used for fast pre-processing of big-data in information extraction, don't deliver enough accurate data for humanities because:
    - They are too shallow and/or
    - Erroneous.

Why is Vagueness Interesting?

There are two variants of "ignorance" (to put it as Smithson 1989 in Piotrowski 2010)

- Incomplete or unavailable information about facts: uncertainty ≈ unclarity ≈ ambiguity (**factual or ontological vagueness**)

- *"Oswald von Wolkenstein lived in the beginning of the 15th century".*

- Incomplete symbolic communication about knowledge, attitudes or beliefs: vagueness in a narrow sense ≈ fuzziness, inaccuracy (**epistemological or communicative vaguenenss**)

- *"Perhaps Oswald von Wolkenstein wanted to participate in the crusades"*

- This last type of vagueness can stem from

  – 1. an attitude of Oswald's to appear a good Christian, or

  – 2. a later interpretation of Oswald's behaviour by another author, or

  – 3. an incomplete tradition.

- In technical projects it is not always necessary to cope with vagueness, because you have to make certain assumptions about facts, but they do not propagate in the project later on, however, if you explicitly represent vagueness and draw inferences, the reliability of the whole project decreases by every step in the inference chain.

DH misunderstandings

- Humanities try to document even the least remains of a historical period and interpret (try to understand) them by combining bits of information for finding its relevance for the time of the researcher (hermeneutics).

- additionally, DH claims to apply computational methods "for higher precision",

- Often a DH "interpretation" intentionally should look like a collection/ combination of facts rather than of a collection of hypotheses.

- However, theoretically, even standard objects (artefacts, texts) are intentionally vague, because human symbolic interaction cannot be strictly precise. Otherwise we would need proper names for all single tokens in the world (cf. the theory of language).

A Bit of Theory on Humanities

According to Wilhelm Dilthey, (*1833 in Biebrich, †1911 in *Siusi allo Sciliar*) the most prominent German theoretician on humanities: humanities partake in a hermeneutic paradigm for establishing social hypotheses. Historical documents often consist either of texts mirroring attitudes, allegations, beliefs, etc., or are reactions of test subjects to verbal stimuli. Such material cannot be treated as facts like numbers or positive formal propositions.

On the other hand, analysing only formal features (e.g. syntactic structures) in the material does rarely contribute to the hermeneutic aims of the historical quest.

In his volume (*Einleitung in die Geisteswissenschaft* 1922):

Dilthey describes history as "a series of world views." Man can only understand himself 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.

# Data Storage vs. Hermeneutic Interpretation in DH

- Humanities (and thus „Digital Humanities") tend to collect data as propositions (without modalities) in a (relational) data base. In most cases vagueness gets lost in the course of annotating or inferencing.

- A better solution is a semi-automatic annotation of vagueness. It can be achieved by
  - metadata about the credibility of texts, genres and authors as well as
  - by lexical annotations of modalities and
  - lexical annotations of vagueness markers. Occurrences of *"perhaps", "mostly"* or *"to a certain extent"* (to name only obvious examples) may support a fair social interpretation,
  - Annotations of the (discontinuous) logical focus.

- Moreover, annotation supports semantic qualifications and allows for reasoning over vague features.

- quantitative methods support humanities' qualitative research, but do not replace them. The main hermeneutic task is left open.

Current DH - Approach

University of Pavia 2020

Including Vagueness

University of Pavia 2020
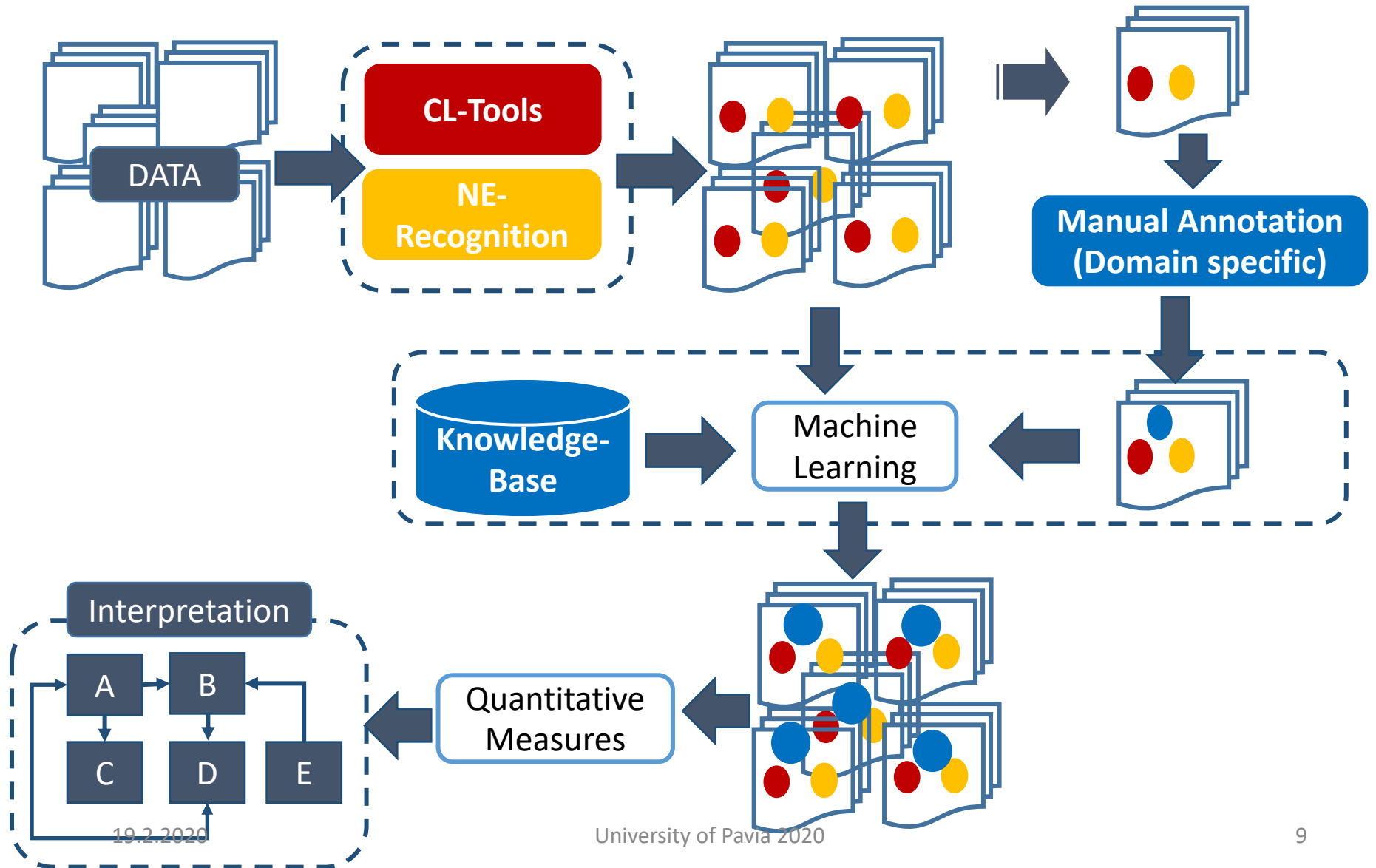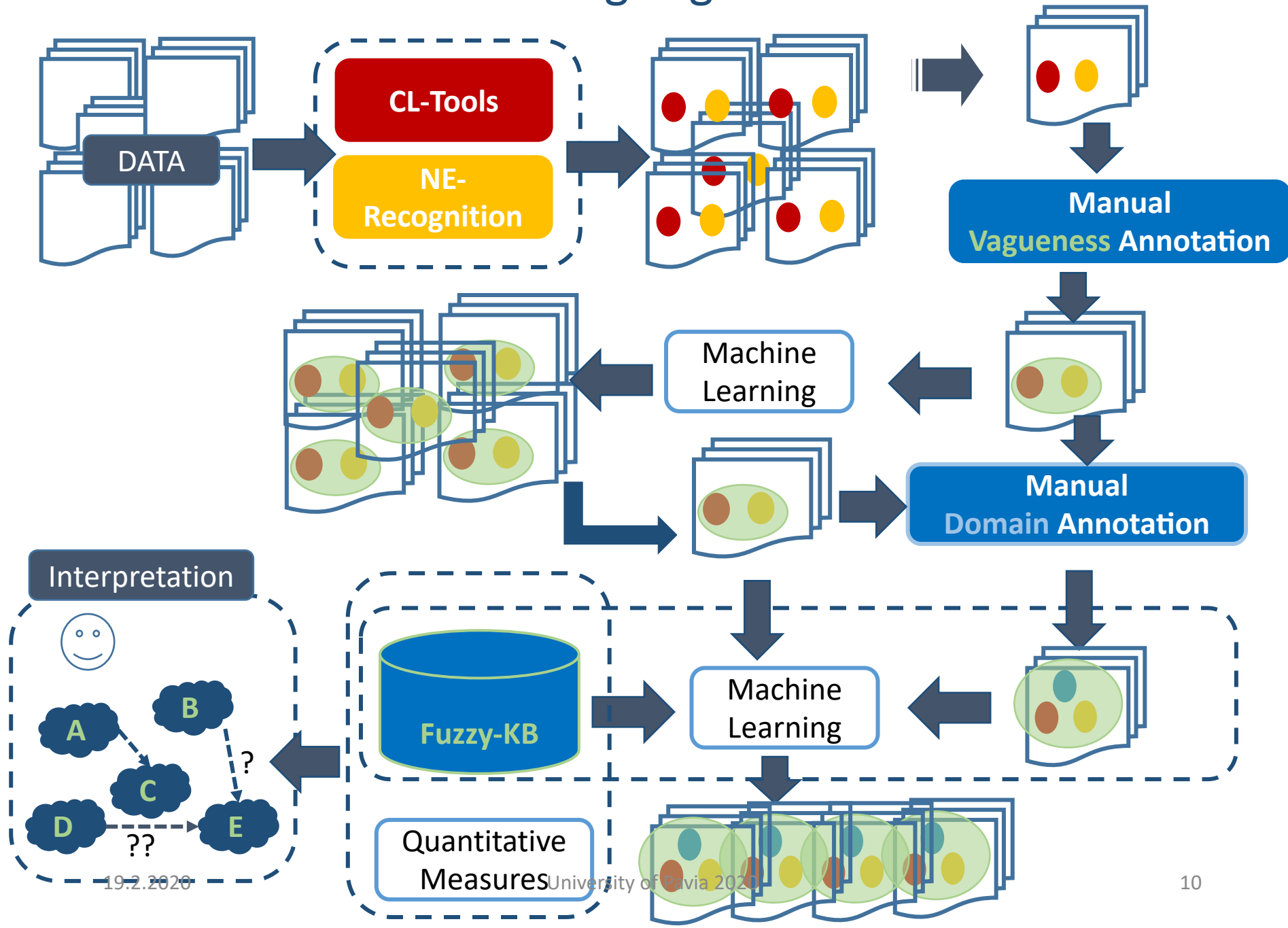
Types of Communicative Vagueness
(„It is meant to be vague")

<u>Linguistic and logical Vagueness</u>

abstracts	*„area"*
isolated comparatives	*„bigger", „more"*
inexact adjectives	*„square", „near",*
relative adjectives	*„big", „rare"*
hedges, shields,	*„rather", „more or less"*
non-intersectives	*„so-called", „supposed"*
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"*
modals, attitudes	*„probably", „hopefully"*
vague quantifiers	*„many", most"*
complex quantifiers	*„roughly half of the king's army died in this battle"*

Factual Uncertainty
(in principle it is precise, but we don't know)

(yet) unexplored facts *„the moon is 384402,56 m away from the earth"*

range expressions *„In the beginning of the 18. century"*
„Italy in the middle ages"

uncertain definition *„the northern slope of the mountain"*

Inexact measures *„a 4 days' journey, 10 feet"*

unclear place *„Syrfia" [on the Ortelius map]*

unclear facts *„by order of the sultan"*

unclear time *„in prehistoric times"*

unclear person *„the former prince of the country"*

unclear action *„the submission of the barbarians"*

Range expressions "in the beginning of 10th century"

continua „water", "traffic",

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 relations and annotations (N., V, Prep, ...)
yellow= data from the ontology
orange= vague expressions

(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

A Complicated Explicit Example

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

?

3. *The Kilian countryside; its capital is Kilia*, formerly called Lycostomon, on on the midnight-side mouth of the Danube, which the greek bargees used to give the same name, because it seems to spout its water like a wolf's gorge. …*

? ?

1

2

3

4

*Its name Kilia was normally used by the Moldavians, by the Turks Kili, by recent Greeks Lycostomon (others call it - as I think, erroneously - Lytrostomon and Lyththrostomon, as you will find it in in Leunclavii Pandect. 146), Bonfinius takes with some others as Achilleam; Stanislaus Sarnicius assumes, it was Tomos, the city, which became famous as Ovidius Naso's exile).

5

6

7

9

8

10

It is impossible for me to find out, which out of so many different beliefs of the writers is the correct one

Uncertainty & Vagueness in Computational (Linguistics) Modelling View

P1: „Tomorrow it will rain"

Uncertainty

P1=true or P1=false, each with a certain probability

P2: „Yesterday it was heavy rain"

P2.1: „Yesterday it rained."

∧

P2.2: „The rain was heavy".

Vagueness

prob(P2.1=true)=1

For P2.2 the degree of truth is „true" or „false" or something in between, „heavy" depends of context

P3: Tomorrow there will be heavy rain

P3.1: „Tomorrow will rain."

∧

P3.2: „The rain will be heavy".

Uncertainty +Vagueness

P3.1 is uncertain
P3.2 is vague
P3.2 presuposes P3.1

Probability theory for modelling uncertainty
Fuzzy logic for modelling vagueness

Mathematical modeling

According to weather forecasts:
- Light rain under 2.7 mm
- Moderate rain btw. 2.7 and 7.6 mm/hour
- Heavy rain more than 7.7 mm/hour

X={set of all days}
A={set of days with heavy rain}
amount(x) rain of day x
$$A \subseteq X$$
How to define $belongs_A(x)$ with $x \in X$?

Crisp approach:
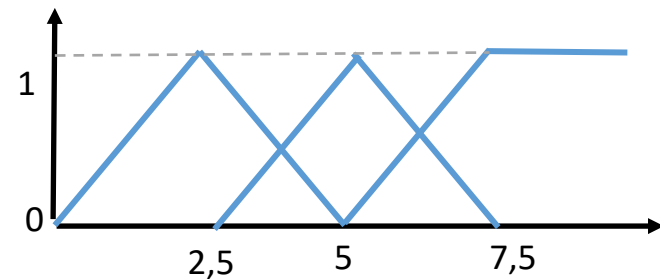7.6 mm/hour -> moderate rain
7.7 mm/hour -> heavy rain

$$belongs_A(x) = \begin{cases} 0 \ if \ amount(x) \leq 7.6 \\ 1 \ if \ amount(x) \geq 7.7 \end{cases}$$
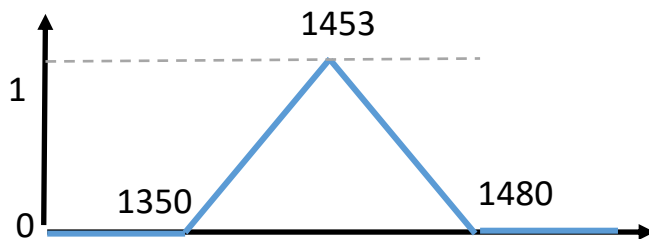
Fuzzy approach:

$$belongs_A(x) = \begin{cases} 1 \ if \ amount(x) \geq 7.5 \\ \dfrac{x-5}{2.5} \ if \ amount(x) \in [5,7,5) \\ 0 \ otherwise \end{cases}$$
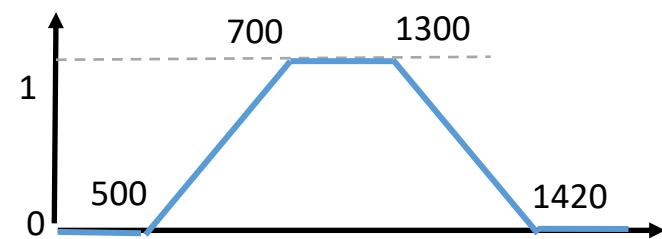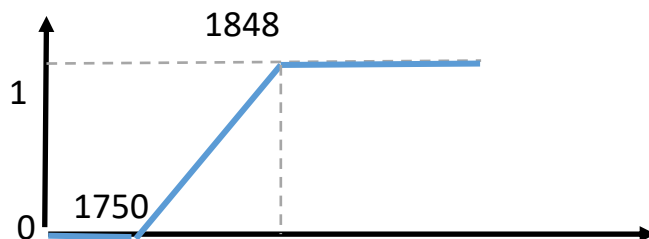
Fuzzy membership functions

- Depends on the context and may be subjectiv: one may have a different function if the rain measured in Hamburg or in Palermo
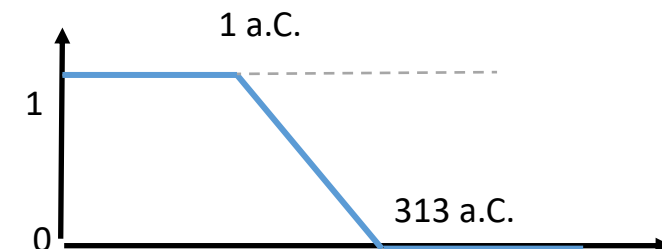
- Shape may vary

- Most used shapes:

Triagular e.g. „Constantinopole's conquest"

Trapez e.g. „middle ages"

Right-shoulder: e.g. „modern times"

left-shoulder e.g. „pre-christian"

Uncertainty & Vagueness and Local Context

P1.1: „**I think** that tomorrow will rain"

P1.2: „**I believe** that tomorrow will rain"

P1.3: „**I heard** that tomorrow will rain"

P1.4: „**The national Meteorological Institute announced**
that tomorrow will rain"

prob(P1.i=true) depends on:
1. The lexical semantics of verbs („think", „believe","heard")
2. The contextual degree of trust of the subject („I", „the national Meteorological Institute)

Most Current Text Mining approaches extract „tomorrow"+"rain" and map tomorrow on the current calendar
- Uncertainty due to lexical semantics 1) can be inferred from linguistic ressources, e.g. Wordnet. One must assign uncertainty factors to synsets (synonym clusters)
- For 2) ontological knowledge is required

Uncertainty & Vagueness and Historical Context
- classical approach -

P1: „Dio Cassius mentions the ash cloud which probably killed the people in Pompey in August 79 A.D."

P1.1: <ash_cloud><killed><inhabitantsPompey>
Prob(P1.1)=0.85

P1.2: <P1.1><happened><August 79 A.D.>
P1.3: <DioCassius><mentions><P1.1>

P2: Pliny the Younger describes the ash cloud which probably killed the people in Pompey in August 79 A.D."

P2.1: <ash><killed><inhabitantsPompey>
Prob(P2.1)=0.85

P2.2: <P2.1><happened><August 79 A.D.>
P2.3: <PlinyYounger><describes><P2.1>

P3: „Excavations demonstrated that the ash cloud killed the people in Pompey in August 79 A.D."

P3.1: <ash><killed><inhabitantsPompey>
Prob(P3.1)=0.85

P3.2: <P3.1><happened><August 79 A.D.>
P3.3: <Excavations><demonstrate><P3.1>

P1.1, P1.2, P2.1,P2.2,P3.1 and P3.2 are all **true**

P1: „Dio Cassius mentions the ash cloud which probably killed the people in Pompey in August 79 A.D.

P1.1: <ash_cloud><killed><inhabitantsPompey>
Prob(P1.1)=0.85

P1.2: <P1.1><happened><August 79 A.D.>

P1.3: <DioCassius><mentions><P1.1>

P2: Pliny the Younger describes the ash cloud which probably killed the people in Pompey in August 79 A.D.“

P2.1: <ash><killed><inhabitantsPompey>
Prob(P2.1)=0.85

P2.2: <P2.1><happened><August 79 A.D.>

P2.3: <PlinyYounger><describes><P2.1>

P3: „Excavations demonstrated that the ash cloud killed the people in Pompey in August 79 A.D.“

P3.1: <ash><killed><inhabitantsPompey>
Prob(P1.1)=0.85

P3.2: <P3.1><happened><August 79 A.D.>

P3.3: <Excavations><demonstrate><P3.1>

P4: „**Researchers found out that most people died in Pompey before the ash cloud came**“
P5: „**Date of Pompey destruction has to be revised, probably to October 79 A.D:** “

P1.1, P1.2, P2.1,P2.2,P3.1 and P3.2 are all **false** and P4 and P5 are **true**

P1: „Dio Cassius mentions the ash cloud which probably killed the people in Pompey in August 79 A.D."

P1.1: <ash_cloud><killed><inhabitantsPompey>

Prob(P1.1)=0.85, value(P1.1)= { 0,60 / **0,40**

P1.2: <P1.1><happened><August 79 A.D.>

Value(P1.2)= { 0,60 / **0,40**

P1.3: <DioCassius><**mentions**><P1.1>

P2: Pliny the Younger describes the ash cloud which probably killed the people in Pompey in August 79 A.D."

P2.1: <ash><killed><inhabitantsPompey>

Prob(P2.1)=0.85, value(P2.1)= { 0,80 / **0,60**

P2.2: <P2.1><happened><August 79 A.D.>

Value(P2.2)= { 0,80 / **0,60**

P2.3: <PlinyYounger><**describes**><P2.1>

P3: „Excavations demonstrated that the ash cloud killed the people in Pompey in August 79 A.D."

P3.1: <ash><killed><inhabitantsPompey>

Prob(P3.1)=0.85 value= { 0,90 / 0,95

P3.2: <P3.1><happened><August 79 A.D.>

Value(P3.2)= { 0,90 / **0,95**

P3.3: <Excavations><demonstrate><P3.1>

P1: „Dio Cassius mentions the ash cloud which probably killed the people in Pompey in August 79 A.D."

P1.1: <ash_cloud><killed><inhabitantsPompey>

Prob(P1.1)=0.85, value(P1.1)= $\begin{cases} 0,60 \\ \mathbf{0,40} \end{cases}$

P1.2: <P1.1><happened><August 79 A.D.>

Value(P1.2)= $\begin{cases} 0,60 \\ \mathbf{0,40} \end{cases}$

P1.3: <DioCassius><mentions><P1.1>

P2: Pliny the Younger describes the ash cloud which probably killed the people in Pompey in August 79 A.D."

P2.1: <ash><killed><inhabitantsPompey>

Prob(P2.1)=0.85, value(P2.1)= $\begin{cases} 0,80 \\ \mathbf{0,60} \end{cases}$

P2.2: <P2.1><happened><August 79 A.D.>

Value(P2.2)= $\begin{cases} 0,80 \\ \mathbf{0,60} \end{cases}$

P2.3: <PlinyYounger><describes><P2.1>

P3: „Excavations demonstrated that the ash cloud killed the people in Pompey in August 79 A.D."

P3.1: <ash><killed><inhabitantsPompey>

Prob(P3.1)=0.85 value(P3.1)= $\begin{cases} 0,90 \\ 0,95 \end{cases}$

P3.2: <P3.1><happened><August 79 A.D.>

Value(P3.2)= $\begin{cases} 0,90 \\ \mathbf{0,95} \end{cases}$

P3.3: <Excavations><demonstrate><P3.1>

P4: „Researchers **found out** that most people died in Pompey before the ash cloud came"

Value(P4)= $\begin{cases} 0,80 \\ \mathbf{0,90} \end{cases}$

P5: „Date of Pompey destruction has to be revised, probably to October 79 A.D: "

Value(P5)= **0,90**

P1: „Dio Cassius mentions the ash cloud which probably killed the people in Pompey in August 79 A.D."

P1.1: <ash_cloud><killed><inhabitantsPompey>

Prob(P1.1)=0.85, value(P1.1)= 0,60 / **0,40**

P1.2: <P1.1><happened><August 79 A.D.>

Value(P1.2)= 0,60 / **0,40**

P1.3: <DioCassius><mentions><P1.1>

P2: Pliny the Younger describes the ash cloud which probably killed the people in Pompey in August 79 A.D."

P2.1: <ash><killed><inhabitantsPompey>

Prob(P2.1)=0.85, value(P2.1)= 0,80 / **0,60**

P2.2: <P2.1><happened><August 79 A.D.>

Value(P2.2)= 0,80 / **0,60**

P2.3: <PlinyYounger><describes><P2.1>

P3: „Excavations demonstrated that the ash cloud killed the people in Pompey in August 79 A.D."

P3.1: <ash><killed><inhabitantsPompey>

Prob(P3.1)=0.85 value(P3.1)= 0,90 / 0,95

P3.2: <P3.1><happened><August 79 A.D.>

Value(P3.2)= 0,90 / **0,95**

P3.3: <Excavations><demonstrate><P3.1>

P4: „Researchers **found out** that most people died in Pompey before the ash cloud came"

Value(P4)= 0,80 / **0,90**

P5: „Date of Pompey destruction has to be revised, probably to October 79 A.D: "

Value(P5)= **0,90**

Vague and Uncertain objects

- Vague objects are identical only with themselves (reflexivity)
- Between 2 vague objects can exist a similarity relation,
- Uncertainty can be reduced or increased during the hermeneutic process. Vagueness remains constant,
- Not only in texts: vague elements in pictures are objects which lack on clarity e.g. clouds, dunes
- Vague elements in historical non-textual objects
  - Backgrounds (fantasy cities, fantasy castles)
  - Fantasy characters in paintings of H. Bosch or romanesc capitals,
  - Places /areas on maps
- Uncertain objects in historical non-textual objects:
  - Places on maps
  - Historical or biblical figures not 100% identifiable
  - Damaged frescoes

TEI P5

- the <note> element: the user can write unstructured text, targeting the degree and scope of the identified vague aspect;
- the <certainty> element: offers a possibility to structure the information about precision. It can refer to:
  - the name of the annotation tag considered uncertain (e.g. a person or a place name),
  - the position in text where the annotation tag starts, or a value of an attribute contained in the annotation tag.
  - Through the attribute @degree it is possible to refine the level of certainty.
- The <precision> element, which can be applied for any numerical value (a date, or a measure). It indicates the numerical accuracy associated to some aspects of a text mark-up

```
<certainty target="#ESX" locus="name"
 assertedValue="placeName" cert="low">
<desc>It is unlikely, but possible, that this refers to the place rather than the person.</desc>
</certainty>a
```

```
<date notBefore="1632" notAfter="1642">Not more than ten years before the start of the Civil War
<precision match="@notBefore" precision="low"/>
 <precision match="@notAfter" precision="high"/>
</date>
```

TEI Criticism

- The elements refer to the uncertainty of the annotator, not the one of the text,
- The metrics of quantifying uncertainty is arbitrary and has no formal semantics,
- All these references are made through pointers, so any annotation and/or processing is extremely cumbersome.
- There is nothing about vagueness.
- Discontinuous elements cannot be represented,
- Different sources of uncertainty cannot be distinguished.

However:

TEI-P5 specifications claim that *"The certainty element allows for indications to be structured with at least as much detail and clarity as appears to be currently required in most ongoing text projects"* (https://www.tei-c.org/release/doc/tei-p5-doc/zh-TW/html/CE.html)

Representation of Uncertainty & Vagueness beyond TEI

- In order to be processed, both uncertainty and vagueness have to be embedded in a conceptual system which allows for reasoning, e.g. an ontology.

- Several layers of vagueness have to be considered and the evaluation must be realised at each layer. One cannot mix results.

- Example of layers:
  - Metadata
  - Linguistic
  - Factual (epistemic)
  - Source credibility

Data and information workflow

Cantemir's Texts

extract

<Subject><Predicate><Object>
+
    ?<Quotation Marker>

Fuzzy-Ontology

A

map

- Vague Time
- Vague Locations
- Historical contexts for one entity

+ Lexical Vagueness/Uncertainity Markers

+ Vagueness /Uncertainty Patterns (e.g. „it is said")

B

C

Reliability Score:

A

A+B

A+B+C
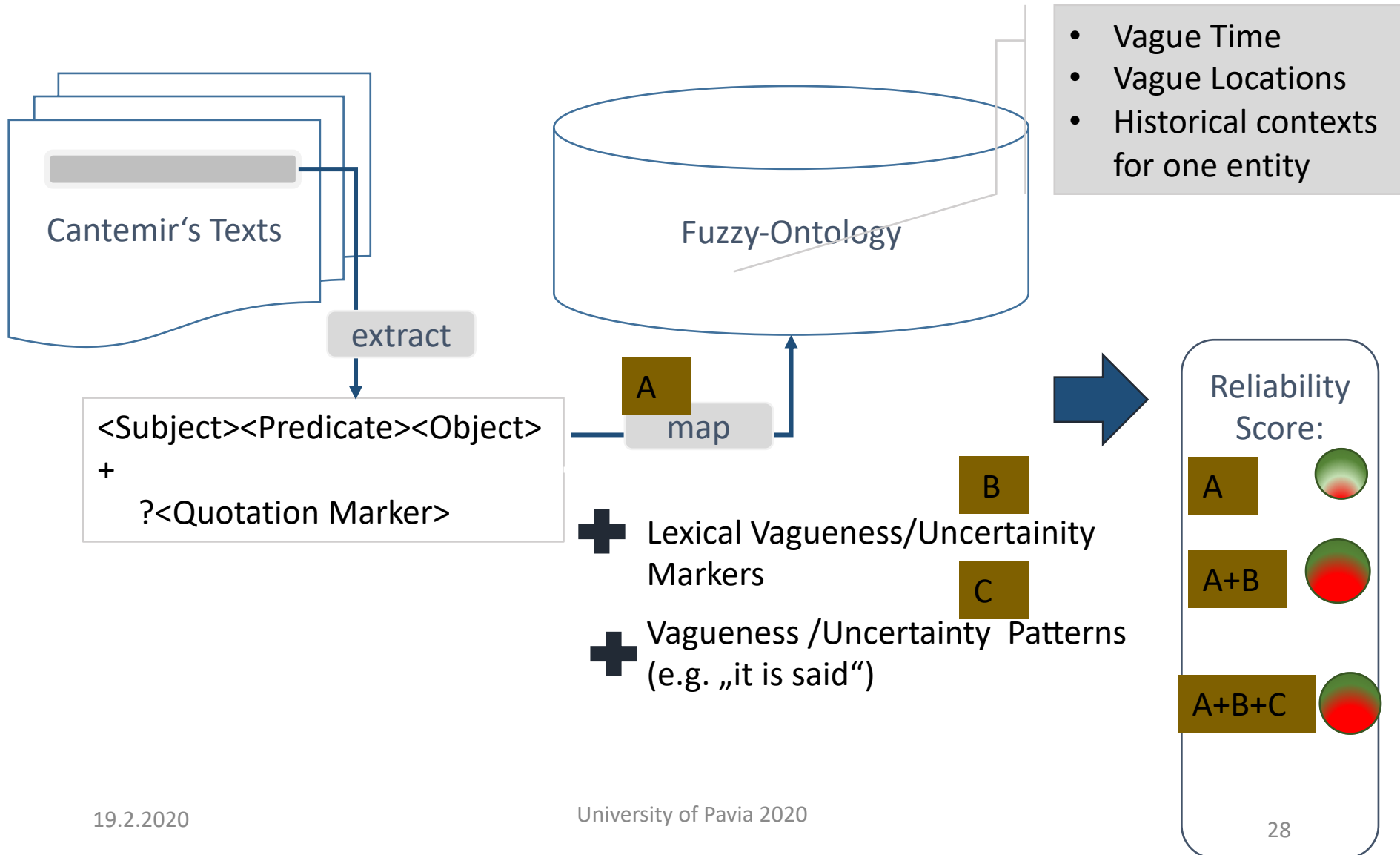
Relevant Metadata information for Vagueness

genre:

- official document,

- letter

- fiction,

- fairy tale

- legend,

- folk tradition

credibility of author

- politician

- journalist

- fiction writer

historical distance

- modern

- historical

decreasing
reliability

decreasing
credibility

University of Pavia 2020

Representation of vague time
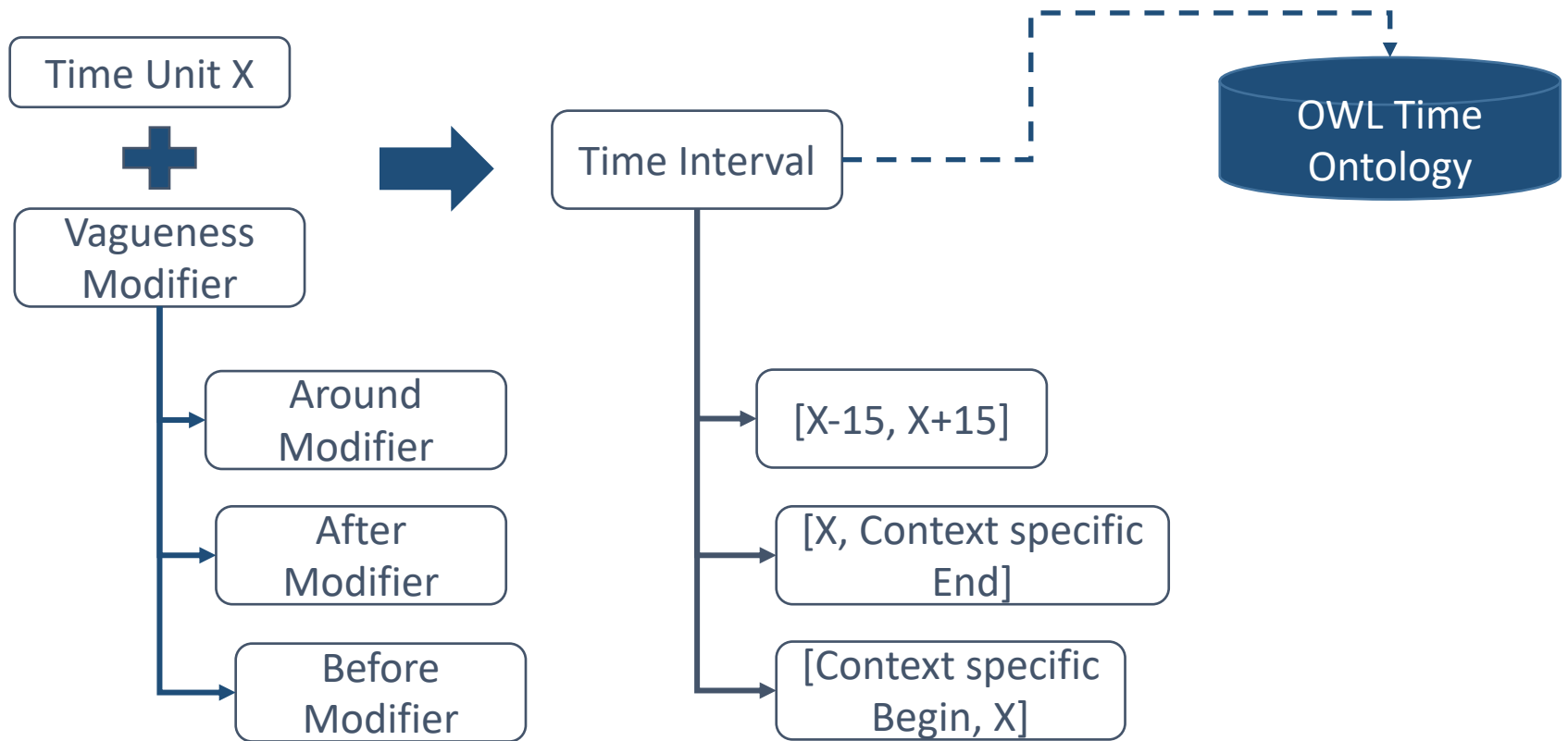
Fuzzy Representation of Time

FuzzyBefore   FuzzyAfter   FuzzyAround

Year1   Year2

FuzzyRelInterval   FuzzyRelInterval

Time Interval   Time Interval   Time Interval

e.g. before, after, around

Time IntervalInstance

TimeIntervalRelation
(e.g. lifeTime, reignTime

Concept X (e.g. Person)
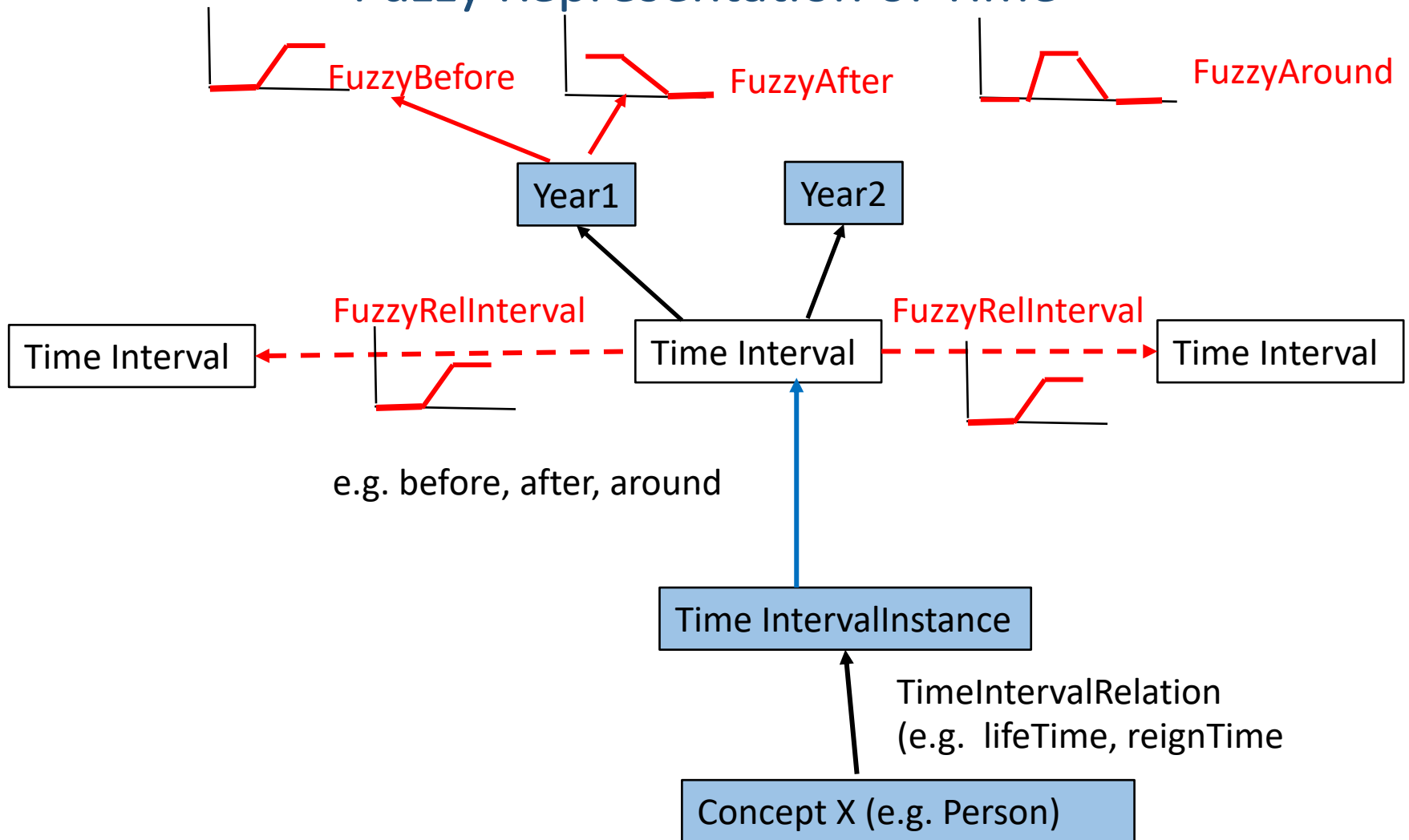
All Fuzzy Datatypes and Datatype properties are generated automatically

How to represent vague properties e.g „is important pilgrim place"
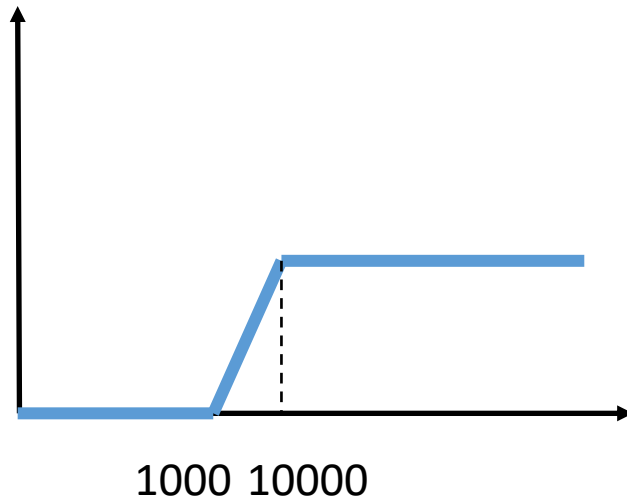
pilgrim place (fuzzy property)

important pilgrim place

Define a data type „vistor _number"

(xsd:double[>= "0.0"^^xsd:double])

Define a concept class „pilgrim place "
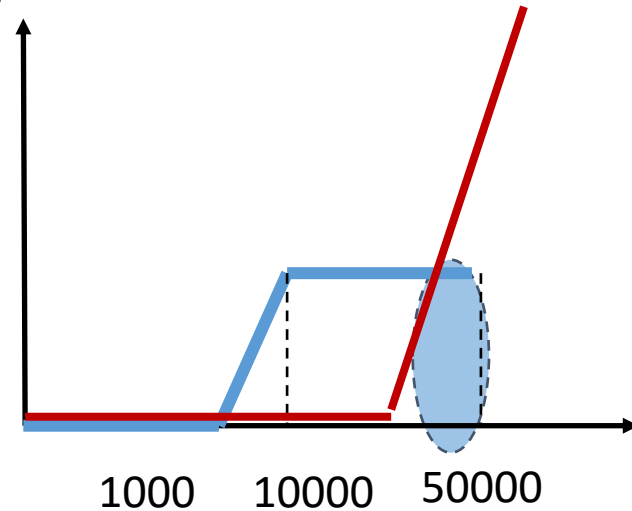
Define a fuzzy datatype „pilgrim_vistor_number"

```
<fuzzyOwl2 fuzzyType=\"datatype\">
<Datatype type=\"rightshoulder\" a=\"1000\"
b=\"10 000\" />
</fuzzyOwl2>"
```

```
<owl:Class rdf:about="http://hercore.uni-hamburg.de/onto#PilgrimPace">
    <rdfs:subClassOf rdf:resource=" http://hercore.uni-hamburg.de/onto#Place/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="=" http://hercore.uni-hamburg.de/onto#hasVisitors"/>
        <owl:someValuesFrom rdf:datatype="=" http://hercore.uni-hamburg.de/onto#pilgrim_visitor_number"/>
      </owl:Restriction>
    </rdfs:subClassOf>
```

Representation of Vague Location

Fix Geographical Location
(e.g. Carpaths, Black Sea, Asia)

+

Cardinal Direction (S, N, E, W, NE, SW, NW, SE) (e.g. north to..)

+

Measure Unit
(e.g. 5 Miles north)

+

Vagueness Modifier (e.g. approximatelly)
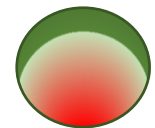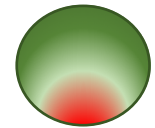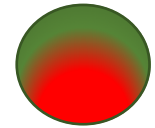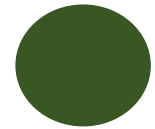
Complex Example: Ortelius' map from 1570

- Country
- Capital

- Northern Dobrudja
- Western Macedonia
- Eastern Europe
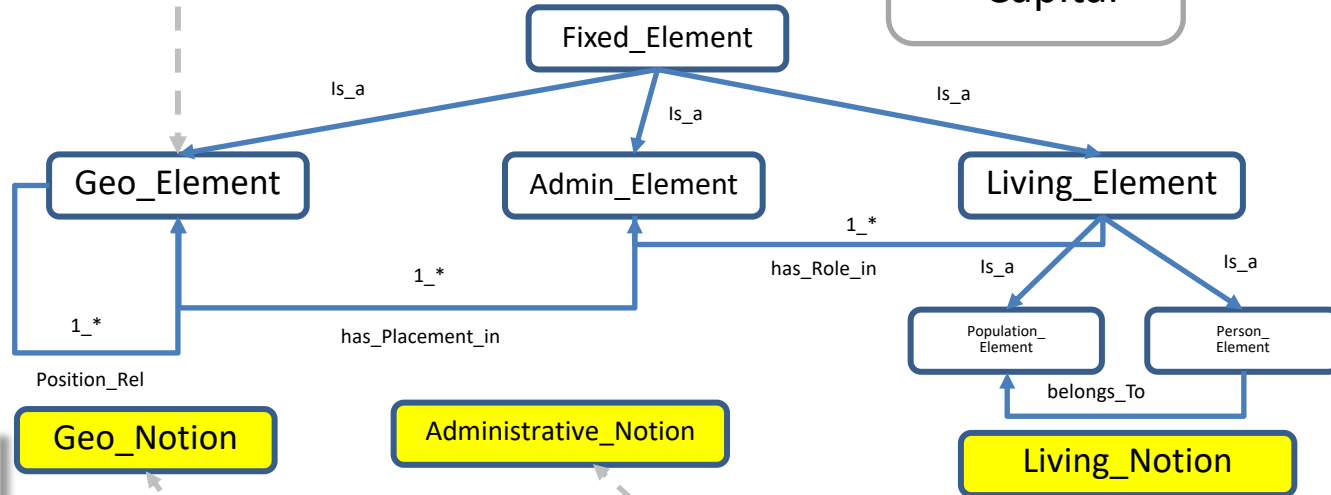
Fixed_Element

Is_a → Geo_Element
Is_a → Admin_Element
Is_a → Living_Element

Living_Element
Is_a → Population_Element
Is_a → Person_Element

has_Role_in 1_*

has_Placement_in 1_*

Position_Rel 1_*

belongs_To

Geo_Notion

Administrative_Notion

Living_Notion

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

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

- Cojani Region
- Sârbia

Fuzzy Concept

- Greece
- Serfia sangiac
- Servia

Fuzzy Concept

- Turkisch Times
- Greek times
- 17th century

Fuzzy Properties

ID: **C_HerCoRe_24EEUR**
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

1_*
Position_Rel
has_a
[Name_Lang]_Label
Geo_Element
Cardinal_Position
has_a
Is_a
Is_a
Is_a
Area
Ground_Element
Water_Element
1_*
has
Lays_at
1_*
Relief_Ele
Outside
Inside

ID: **C_HerCoRe_55NDobrud**
Label : **North Dobrudja**
Has_Object_Property: **North**

Subclass_Of

Cardinal_Position

Sea

Has_Type: **Area**

Situated_In: **C_HerCoRe_20DB**

ID: **C_HerCoRe_20DB**
Label: **Dobrudja**
Has_Type: **Area**

High_P rofile
Flat - Profile
Peninsula

Subproperty_Of: Position_Rel

Steppe
Forest

River SIZEize

Mountain
Desert
Gulf
Belongs_to
has_a 1_*
≈
≈

elongs_to
Belongs_to

Mountain _Range
Field
Strait
Small Riverz e
Creeki ze

Historical Context

Geo_Element

TIME_SLOT ≈

Administrative_Element

Geo_Notion → defined_as 1_* ≈

Administrative_Notion ≈

Refered _as 1_*

Writtings
[Word-Language-Alphabet]

Hist. Context 1

Geo_Element

Area

Situated_In

ID: C_HerCoRe_50MACEd
Label : Western Macedonia

F_TODAY

≈ Time_Slot

Admin_Notion

Id: C_HercoRe_17GR
Label: Greece

Hist. Context 2

Geo_Element

Area

Situated_In

ID: C_HerCoRe_50MACEd
Label : Western Macedonia

≈ Time_Slot

Turkisch time

Admin_Notion

Id: C_HercoRe_71 SarfS
Label: *Serfia sangiac"*

Defined_As

Defined_As

Id: C_HercoRe_71 CojR
Label: Cojani Region:

Geo_Notion

Situated_in

Id: C_HerCoRe_74 SarfC
Label: *Σέρβια,Servia*

Admin_Notion

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

Geo_Element

TIME_SLOT

Administrative_Element

Geo_Notion

defined_as

1_*

Refered _as

1_*

Administrative_Notion

Writtings
[Word-Language-Alphabet]

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


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

Hist. Context 1

Geo_Element

Area

Situated_In

ID: C_HerCoRe_55NDobrud
Label : North Dobrudja

Geo_Notion

Defined_As

Id: C_HercoRe_10 DSyrf

Label: Σύρφοι, Syrphoi

≈ Time_Slot

>= Greek times

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 " />
 ))
```
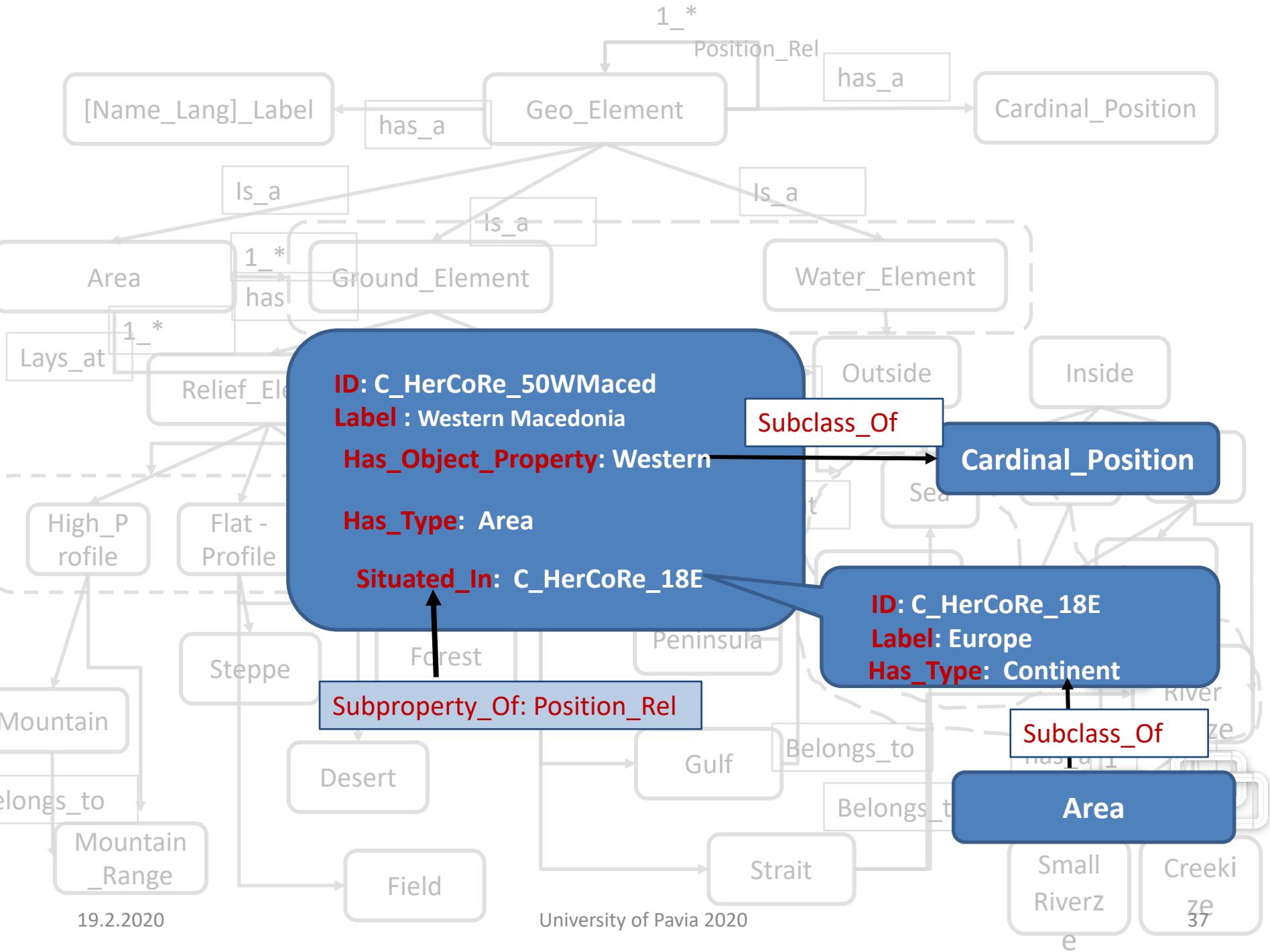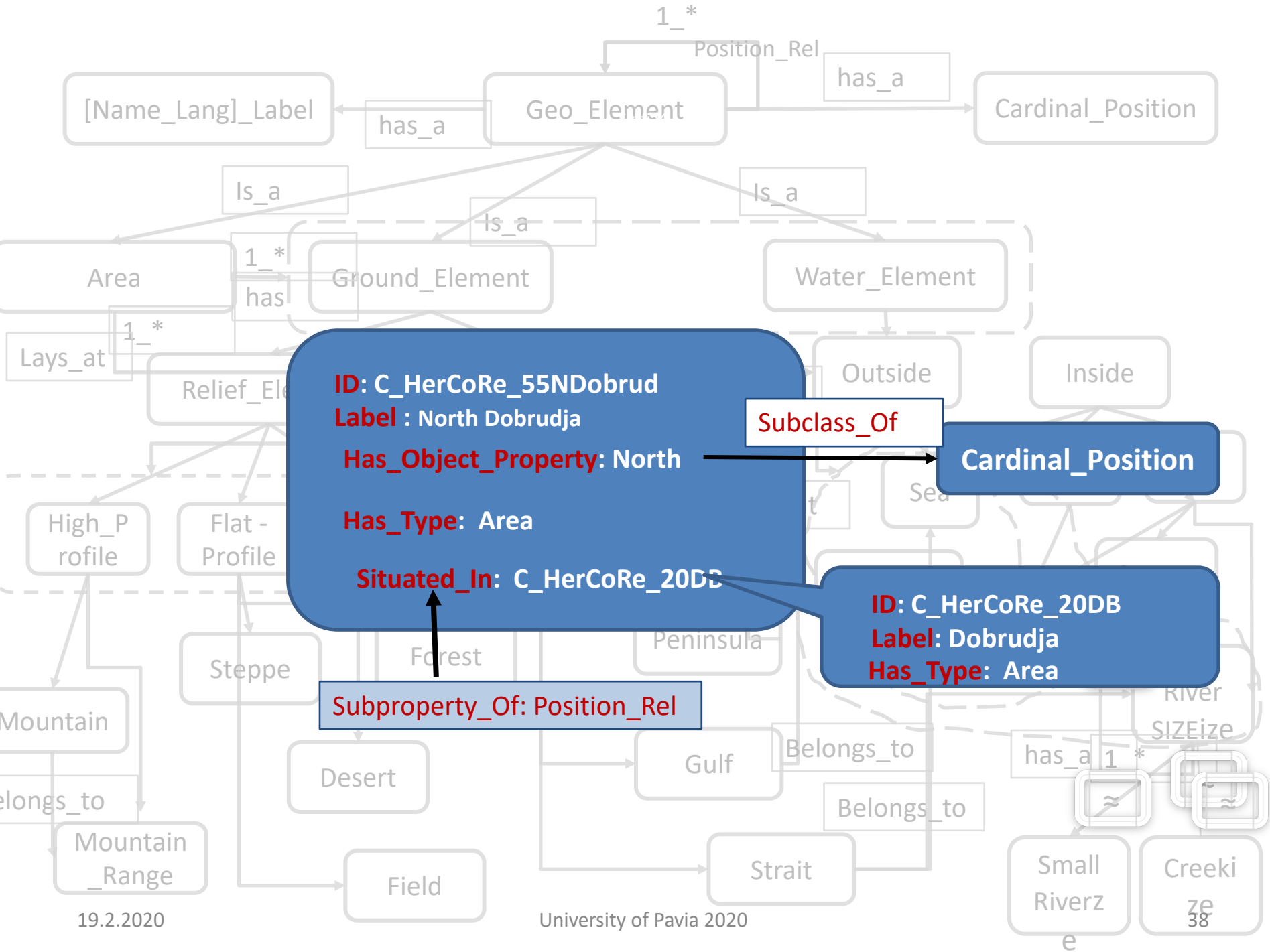
Syrfia is
the abandoned name of a region in Eastern Europe, used on historical maps until 17th century, designating
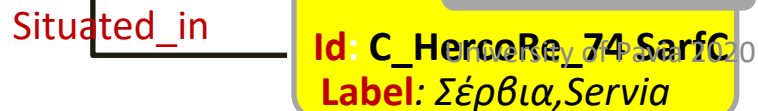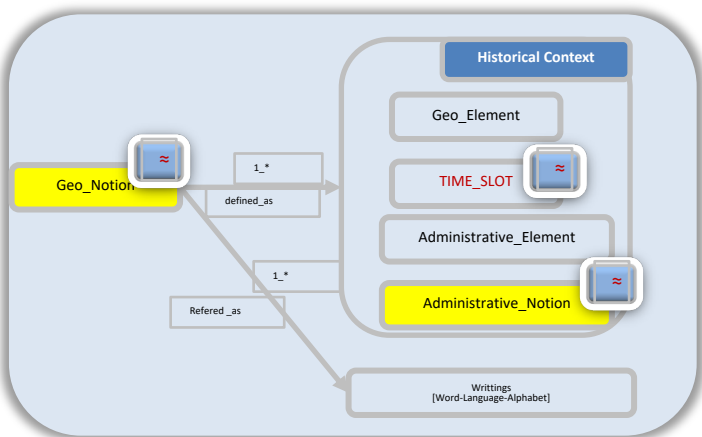- a part of Northern Dobrudja, coming from the Greek term *Σύρφοι - Syrphoi*, or
- The Cojani region from western Macedonia, today in Greece but in turkish times in the "Serfia sangiac" having the capital *Σέρβια, Servia* ;
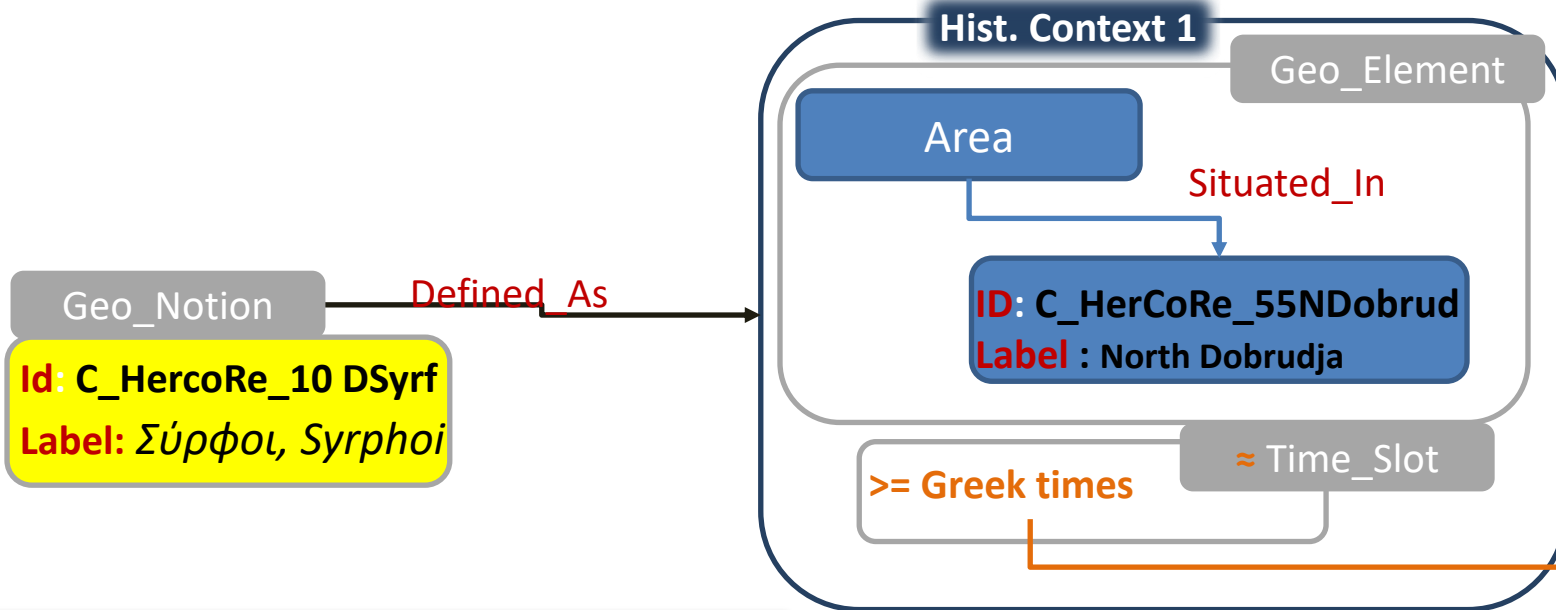- Sârbia, due to phonetic association.

Geo_Notion

Id: **C_HercoRe_70Syrfia**
Label: Syrfia
Used_for: Map

<= 17 Century
≈ Time_Slot

Confidence 0.33

Defined_As

Hist. Context 1

Id: **C_HercoRe_71 CojR**
Label: Cojani Region:

Defined_As

Hist. Context 2

Id: **C_HercoRe_10 DSyrf**
Label: *Σύρφοι, Syrphoi*

Confidence 0.33

Defined_As

Confidence 0.33

Hist. Context 3

Id: **C_HercoRe_11Sarb**
Label: *Sarbia*

ROMANIA.

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.

Instance_Of

Geo_Notion

Id: C_HercoRe_70Syrfia

Label: Syrfia

Used_for: Map

<= 17 Century

≈ Time_Slot

Defined_As

Confidence 0.2

Defined_As

Confidence 0.7

Defined_As

Confidence 0.1

Hist. Context 1

Id: C_HercoRe_71 CojR

Label: Cojani Region:

Hist. Context 2

Id: C_HercoRe_10 DSyrf

Label: *Σύρφοι, Syrphoi*

Hist. Context 3

Id: C_HercoRe_11Sarb

Label:*Sarbia*

Source: Wikipedia   University of Pavia 2020   42

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 (by annotation of named entities in a precise formal language, e.g.), and the

- computational accuracy of retrieval processes

Best Practice for Including Vagueness

Queries

Linguistic Evaluation

International Search

Conceptual Search

New Knowledge

Truth Graduation

Reliability Estimation

Historical Media

Weighted Presentation

Intuitive Word Annotation

Annotating Linguistic Features

Multilingual Data

Using Ontologies

Applying Inferences

Assessing Uncertainty

Assessing Textual Fuzzyness

Assessing Media Reliability

Including Historical Data

Graduated Visualization

How to Display Vagueness?

Vagueness of the material at hand must be presented to the reader/researcher at least in a summarizing way indicating the ratio of vagueness types. This will remind the reader of their hermeneutic task and avoids to present numeric values.
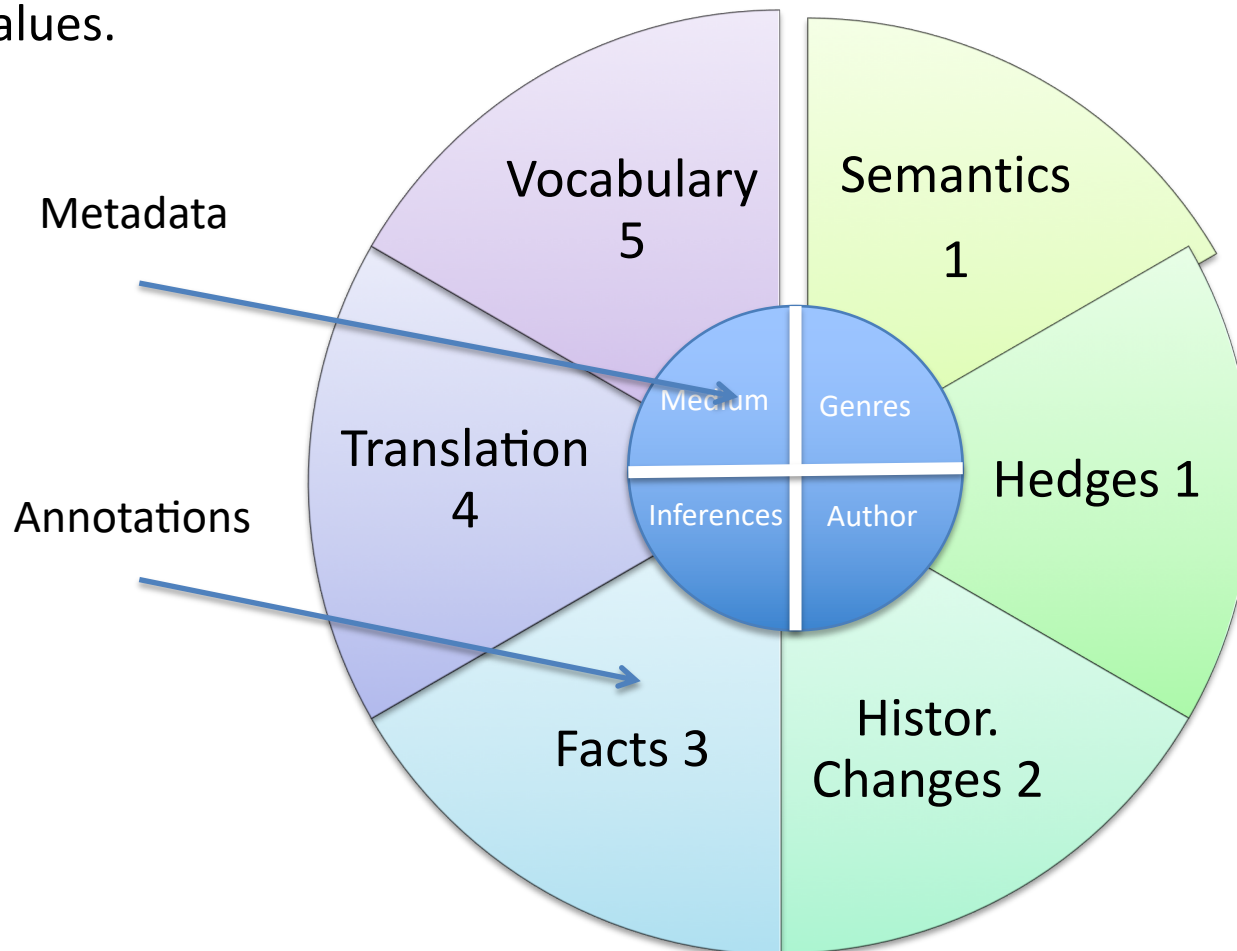
Criticism 1: Data Base Technology

The storage of objects in traditional (relational) databases, designed for business or natural science applications, pretends to deal with facts, (Piotrowski: „the illusion of factuality")

especially, if

– vagueness markers and logical modalities of the propositions are omitted or

– factual uncertainty is not annotated.

Nevertheless, relational data bases are still used in DH for text storage because of the existin information extraction paradigm.

Result:

"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: Metadata are too General

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

However, metadata are the best place to represent the

- genre of a document and style of the text,

- the credibility and source dependency of the author,

- the tradition paths of the text,

- et cetera.

Only by richer metadata the researcher can include global text features in an inference chain.

Criticism 3: Shallow linguistic annotation

(Groups of words are not only sequences of characters

they have

- a syntactic (logical) function,

- a semantic (e.g. referential) meaning,

- a given/new role (informational value)

- a coherence role (context dependency),

- a discourse marker, and

- a pragmatic (action) value.

Moreover, they are not necessarily continuous:

Germ. *"der Präsident gab seinen Posten auf"*

"gab … auf" for semantic reasons must be represented as *"aufgeben".*

Criticism 4: No Reasoning

Annotations as such are local and don't have a global influence on 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 Moldava in 1850 ⊤
  ∧ Moldava in 1850 was part of the Ottoman Empire, ⊤
  ∧ the Ottoman Empire lost Moldava in 1878 ⊤
then
    Iași was not in the Ottoman Empire after 1878 ⊤
```

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

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.

Simon Parsons: Qualitative approaches for reasoning under uncertainty. Cambridge, MA 2001.

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

Michael Piotrowski: Accepting and Modeling Uncertainty. In: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. 2019 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4).

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.

Grazie per la vostra attenzione

Cross-Media Interference

Like in your 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 on maps, for political/clerical reasons,

even more in textual sources like historical texts.
They might not be factual, but must be interpreted by humanities.

Coming back to the use of tools in Science and humanities (cf. slide 4):

For this reason, even when science data are processed, hermeneutic methods are indispensable.

Manfred Pinkal's Schema of Semantic Vagueness

Illocutive
Unclarity

Referential
Uncertainty

Semantic
Vagueness

Communicative
Underspecification

Vagueness in a
narrow sense

Ambiguity

Porosity

*Cimiano, Unger and
McCrae*

Homonymy

Polysemy

Syntaktic
Ambiguity

Borderline
Uncertainty

Inexactness

Referential
Ambiguity

Relativity

Elliptical
Ambiguity

one-dimen-
sional

many-dimensional

Metaphorical
Ambiguity

Linguistic Ambiguity

metaphors   *"head" (line), "crane"*   several distinct meanings

metonymy   *"the sublime porte"*   two contiguous meanimgs

                     "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 males

Historical Linguistic Change

historical obscurity   *"capire l'antifona"*
     "Ai tempi in cui Berta filava"
     "Rimandare qualcosa alle calende greche"

historical abbreviations   '*Karl*' = Charlemagne,
     '*Pius II*'   = "Enea Silvio Piccolomini"
                     = "The pope in 1500 A.D."

changing meaning   Germ. '*wohl*' is not vague, as in modern German,
     but is the adverb of '*good*'

named entities   '*Istanbul*' *= Constantinopol, Konstantinopel*

different writing   '*Keniginn*' = Königin (queen), "nemlig" = nähmlich"

different script   Cyrillics in Romanian,
     black letters in German,
     Arabic letters in Othman tukish

Logical Vagueness

presuppositions   *„I will use the Maserati"*

                                        Implications:

                                        $\Rightarrow$ There is a vehicle called Maserati,

                                        $\Rightarrow$ 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.