



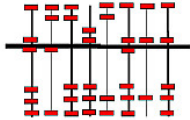
Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG



UNIVERSITATEA
DIN BUCUREȘTI
— VIRTUTE ET SAPIENTIA

Research Center „Digital Humanities“



Research Group
„Computerphilology“

On the annotation of vague expressions -a case study on Romanian historical texts -

Anca Dinu, Cristina Vertan, Walther v. Hahn

{vertan,vhahn}@informatik.uni-hamburg.de

anca_d_dinu@yahoo.com

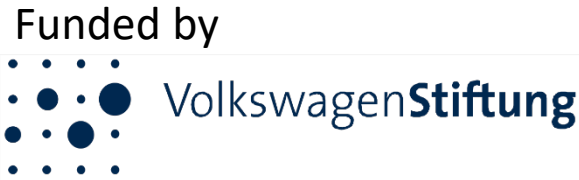
Humanities follow their own rationality (see Dilthey or Habermas), distinct from science and technology.

The integration of vague or heterogeneous information (humanities) with IT – methods still suffers from several drawback as

- use of words instead of concepts and text sections instead of information bits,
- like texts, words are ambiguous or vague on several layers, esp. when being translated, but this ambiguity is not
- simple annotations do not remedy the distortion,
- semantic string-tagging for humanistic interpretation often multiplies ambiguities (esp. in automatic tagging).
- Vague assertions are not marked, or if at all they are interpreted by usual “crisp” processes

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

Combine hermeneutic approaches and methods from computer science for investigating reliability and consistency of original text from 18th century as well as their translations

CS

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

Progress work in automatic recognition of vague expressions

H

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

(In)Validate assumptions about source quotations in original text

Outline

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

Dimitrie Cantemir (1673 -1723)

- Prince of Moldavia (historical province) as well as „universal“ humanist (linguist, ethnographer, musicologist, historian, writer)
- As member of the Royal Academy in Berlin and at the request of this institution wrote two works :
 - Description of his own country („Descriptio Moldaviae“)
 - History of ottoman empire (History of Growth and Decay of Ottoman Empire)
- Original material written in Latin; Both originals were lost already by the end of 18th century
- Several copies were used as basis for translations into German, English, French, Russian and later in Romanian
- Sometimes the translation relies on other translation (e.g. first Romanian translation of “Descriptio Moldaviae” was done after the German version from 1774.



These translations used as reference information about the Ottoman Empire and Romanian provinces until the middle of 19th century, i.e. they give an idea about the reception about this part of the world in Western Europe.

Analysis and interpretation of Cantemir's works

- Already in the 1920'ies, it was demonstrated using selections of texts, that the translations are not respecting the original all the time
 - E.g. Information sources indicated by Cantemir were omitted, because they seemed too unreliable to the translator
- In the XX century researchers claimed that some of the sources, persons and facts quoted by Cantemir were not existing
- 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.

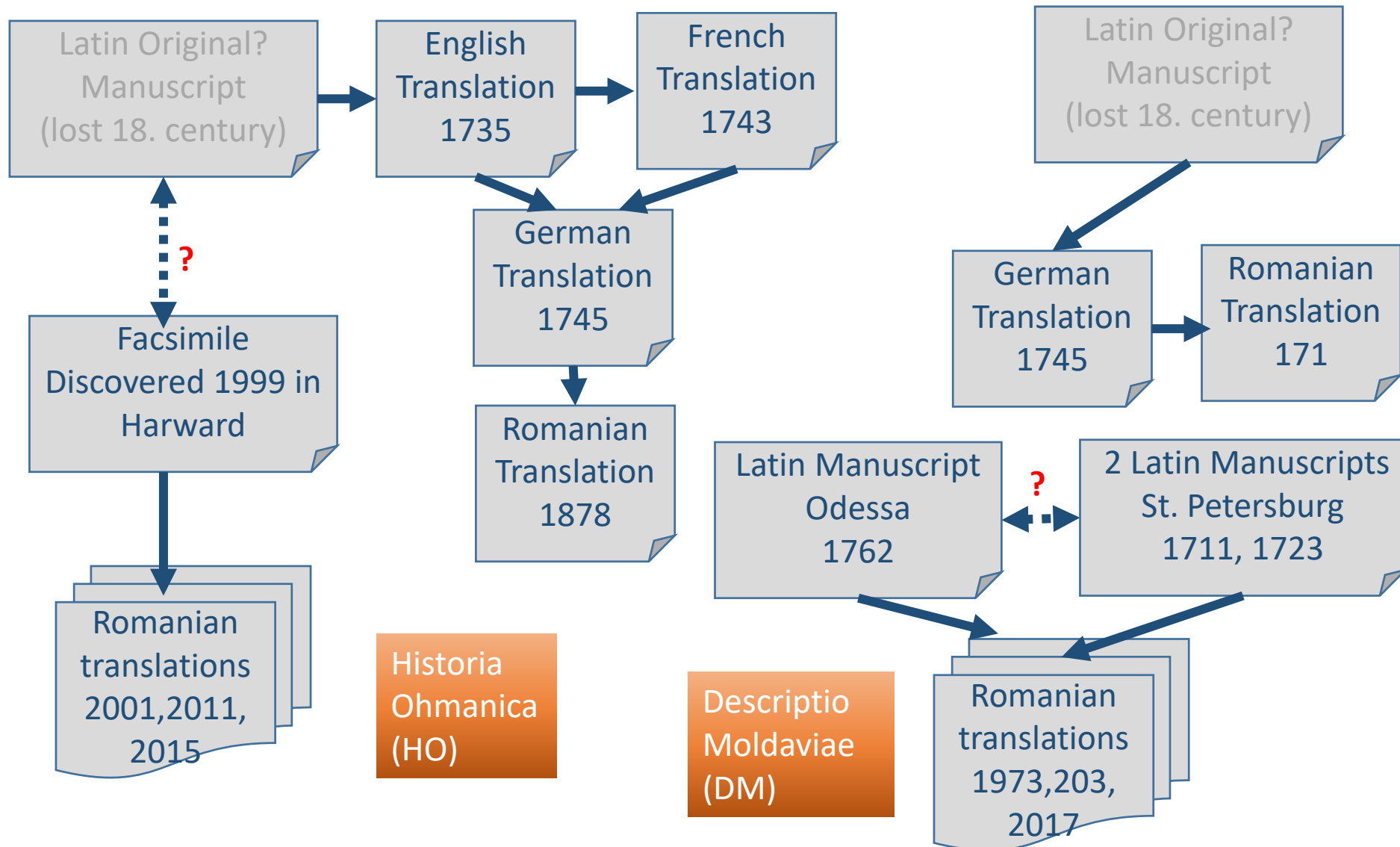
- **Consistency:**

- 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



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 peculiaria fuisse. Ita, si quis Principi Praefatione legerit Principes in duas stirpes fuisse divisos, „Aliothman” unam, et „Ali Dzengiz”¹ alteram, neque sub ipsis horum generum conditionem appellatam iam apud aliquos valuisse. Vti enim absonum videtur, Aliothmanes Suleimano parentes ab huius nepote integro post saeculo iis imperantibus fuisse sortitos; ita non minus falso vulgo praedicantur Tartarorum Crimensium Principes ab ipso Dzengizchano „Alidzengiz” appellationem retinuisse.

More plausible

Would have been...

seems unlikely

equally false

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

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

1. Dragoș și măcar că hronografiile noastre nu arată pentru știința neamului său, dar la noi se zice necontenit, că a fost din neamul cel vechiu al crailor Moldovinești, și a avut tată pe Bogdan fiul lui Ioan, dela carele toți Domnii obișnuiesc a-și pune la iscălitură numele Ioan. Și cuvântul acesta este mai ușor de a se adevăra și pentru aceasta, căci cu greu este de a se crede, că altul din neam mai prost, ar fi putut cu o tovărășie așa mare să meargă la vânat, carele a dat prilej la descoperirea Moldovei și ar fi putut îndemna pe ceilalți patrioți ai săi, ca să vie după dânsul.

*Dimitrie Cantemir, "Descrierea Moldovei", Wikisource
(probably text of the first translation in 19th century after
the German edition*

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

1. Dragoș și măcar că hronografiile noastre nu arată pentru știința neamului său, dar la noi se zice neconținut, că a fost din neamul cel vechiu al crailor Moldovinești, și a avut tată pe Bogdan fiul lui Ioan, dela carele toți Domnii obișnuiesc a-și pune la iscălitură numele Ioan. Și cuvântul acesta este mai ușor de a se adevăra și pentru aceasta, căci cu greu este de a se crede, că altul din neam mai prost, ar fi putut cu o tovărășie așa mare să meargă la vânat, carele a dat prilej la descoperirea Moldovei și ar fi putut îndemna pe ceilalți patrioți ai săi, ca să vie după dânsul.

Dragos	belongs	moldavian kings
Dragos	son_of	Bogdan
Bogdan	son_of	Johannis
Dragos	has_additional_name	Johannis
Bogdan	has_additional_name	Johannis
Drgaos	discovered	Moldau
Dragos	has_acitivity	hunting
Dragos	has_activity	development
Development	takes_place	after Batia invasion

Annotated Text - Dim x
cantemir.easydigital.ro/text.php

Dimitrie Cantemir Semantic Annotation Home Text Resources **Contact**

Entity
 Object
 Agent
 Person
 SocialPerson
 Literator
 Ruler
 Emperor
 King
 Sultan
 Voivode
 Saint
 Theologian
 SocialAgent
 CollectiveAgent
 Group
 FamilyGroup
 NationGroup
 LeaderRelated
 Job
 Personification
 God
 SocialObject
 Place
 City
 Continent

Notice: Below you can read the text written by Dimitrie Cantemir, "Descrierea Moldovei", and you can use the tools on the sidebars to emphasize the annotations. The text is written in Romanian.

[Download annotation](#)

Dimitrie Cantemir

DESCRIEREA MOLDOVEI

I. PARTEA GEOGRAFICĂ

[Click to collapse](#)

CAPITOLUL I

Despre numele cel vechi și cel de acum al Moldovei

[Click to collapse](#)

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

You can click on the desired annotation and extra information about it will be shown below

External Links

- [Imperiul Roman - Wikipedia](#)
- [Imperiul Roman antic - Wikipedia](#)
- [Imperiul Roman de Apus - Wikipedia](#)
- [Imperiul Roman de Răsărit - Wikipedia](#)

Individual

(No information)

Class Information

Imperiul roman belongs to the following classes:

City (ro: Oras)

A place that has all the elementary necessities for its inhabitants. A town.

Relationships

(No information)

PragPresentati...pptx Annotiert101-180.zip Annotiert51-100.zip Annotiert33-50.zip jdk-7u79-docs-....zip jdk-7u79-docs-all.zip jwcdg_precompil...xz deutsch-lexik...lзма Alle Downloads anzeigen...

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

1. Dragosch. Obgleich unsre Jahrbücher sein Geschlechtsregister nicht angeben, so ist es doch eine beständige Sage bey uns, daß er aus dem alten königlichen moldauischen Stamme gewesen sey, und den Bogdan zum Vater gehabt habe, welcher ein Sohn des Johannis war, von welchem alle Fürsten den Namen Johannis in ihrem Titel zu führen pflegen; dieser Meinung ist desto mehr Glauben beyzumessen, weil man schwerlich glauben kan, daß einer von gemeiner Herkunft mit einem so großen Gefolge auf die Jagd (welche die Moldau zu entdecken Gelegenheit gegeben,) habe ausgehen, und seine übrigen Landsleute überreden können, ihn zu folgen.

(*) Diese Stelle bestätigt aufs neue, was wir oben schon vermuthet hatten, daß Dragosch erst nach des Tatars Bathy oder Batu Einfall, d.i. ungefähr nach 1250. aus Siebenbürgen ausgewandert ist; vielleicht aber lassen sich beede Meynungen vereinigen, wenn man zwey Auswanderungen annimmt, die eine in der lezten Hälfte des Zwölften, die andere in der ersten Hälfte des dreyzehenten Jahrhunderts (V.)

*Dimitrie Cantemir, "Beschreibung der Moldau",
Faksimiledruck der Originalausgabe von 1771, Kriterion
Verlag Bukarest, 1973, pag. 111*

Domnul cel dintâi carele după năvălirea lui Batié, 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 Batié Einfall (*) der Moldau ihren vorigen Glanz wieder verschafft hat, war

1. Dragoșch. Obgleich unsre Jahrbücher sein Geschlechtsregister nicht anzeigens, so ist es doch eine beständige Sache uns, daß er aus dem alten königlichen moldauischen Stamme **gewesen sey**, und den Bogdan zum Vater gehabt habe, welcher ein Sohn des Johanniss war, von

Dragosch ≈ belongs_to Moldavian kings
 Dragoșch = son_of Bogdan
 Bogdan = son_of Johanniss
 Dragoșch = has_additional_name Johanniss
 Bogdan = has_additional_name Johanniss

Herkunft mit einem so großen Gefolge auf die Jagd (welche die Moldau zu entdecken Gelegenheit gegeben,) habe ausgehen, und seine übrigen Landsleute überreden können, ihn zu folgen.

was

should
have
been

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

1. Dragosch. **Obgleich unsre Jahrbücher sein Geschlechtsregister nicht angeben, so ist es doch eine beständige Sage bey uns, daß er aus dem alten königlichen moldauischen Stamme gewesen sey,** und den Bogdan zum Vater gehabt habe, welcher ein Sohn des Johannis war, von welchem alle Fürsten den Namen Johannis in ihrem Titel zu führen pflegen; **dieser Meinung ist desto mehr Glauben beyzumessen, weil man schwerlich glauben kan, daß einer von gemeiner Herkunft mit einem so großen Gefolge auf die Jagd (welche die Moldau zu entdecken Gelegenheit gegeben,).**

(*) Diese Stelle bestätigt aufs neue, was wir oben schon vermuthet hatten, daß Dragosch erst nach des Tatars Bathy oder Batu Einfall, d.i. **ungefähr** nach 1250. aus Siebenbürgen ausgewandert ist; vielleicht aber lassen sich beide Meynungen vereinigen wenn man zwey Auswanderungen annimmt, die eine in **der letzten Hälfte des Zwölften**, die andere in der **ersten Hälfte des dreyzehnten Jahrhundert** (V.)

Enriched Classical Markup

Dragosch	≈ belongs	moldavian kings
Dragosch	≈ son_of	Bogdan
Bogdan	≈ son_of	Johannis
Dragosch	≈ has_additional_name	Johannis
Bogdan	≈ has_additional_name	Johannis
Drgaosch	discovered	Moldau
Dragosch	has_acitivity	hunting
Dragosch	has_activity	development
Development	takes_place	after Batia invasion

Dragosch	has_activity	moved
Movement	takes_place	>=1150; <=1200
Movement	takes_place	>=1200; <=1250
Bathy invasion	takes_place	≈ 1250
Bathy	has_alternative_name	Batu
Bathy	is_a	Tatar

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

1. Dragosch. Obgleich unsre Jahrbücher sein

Geschlecht **Legend** es

doc Dragosch **most_probable** belongs moldavian kings

den Dragosch discovered Moldau

gew Dragosch has_acitivity hunting

hab Dragosch has_activity development

wel Development takes_place after Batia invasion

ihre desto mehr Glauben beyzumessen, weil man

schwerlich glauben kann, daß einer von

gemeiner Herkunft mit einem so großen Ge

auf die Jagd (welche die Moldau zu entdeck

Gelegenheit gegeben,) habe ausgehen, und

seine ü **Translator** ne

zu fo Dragosch has_activity moved

Movement takes_place > ≈ 1150; < ≈ 1200

Movement takes_place > ≈ 1200; < ≈ 1250

Bathy invasion takes_place ≈ 1250

Bathy has_alternative_name Batu

Bathy is_a Tatar

HerCoRe – Multiple-faceted fuzzy annotation

Cantemir

Dragosch ≈ son_of Bogdan

Bogdan ≈ son_of Johannis

Dragosch ≈ has_additional_name Johannis

Bogdan ≈ has_additional_name Johannis

thet hatten, daß Dragosch erst nach

athy oder Batu Einfall, D.i. ungefähr

us Siebenbürgen ausgewandert ist;

er lassen sich beide Rechnungen

wenn man zwey Auswanderungen

e eine in der letzten Hälfte des

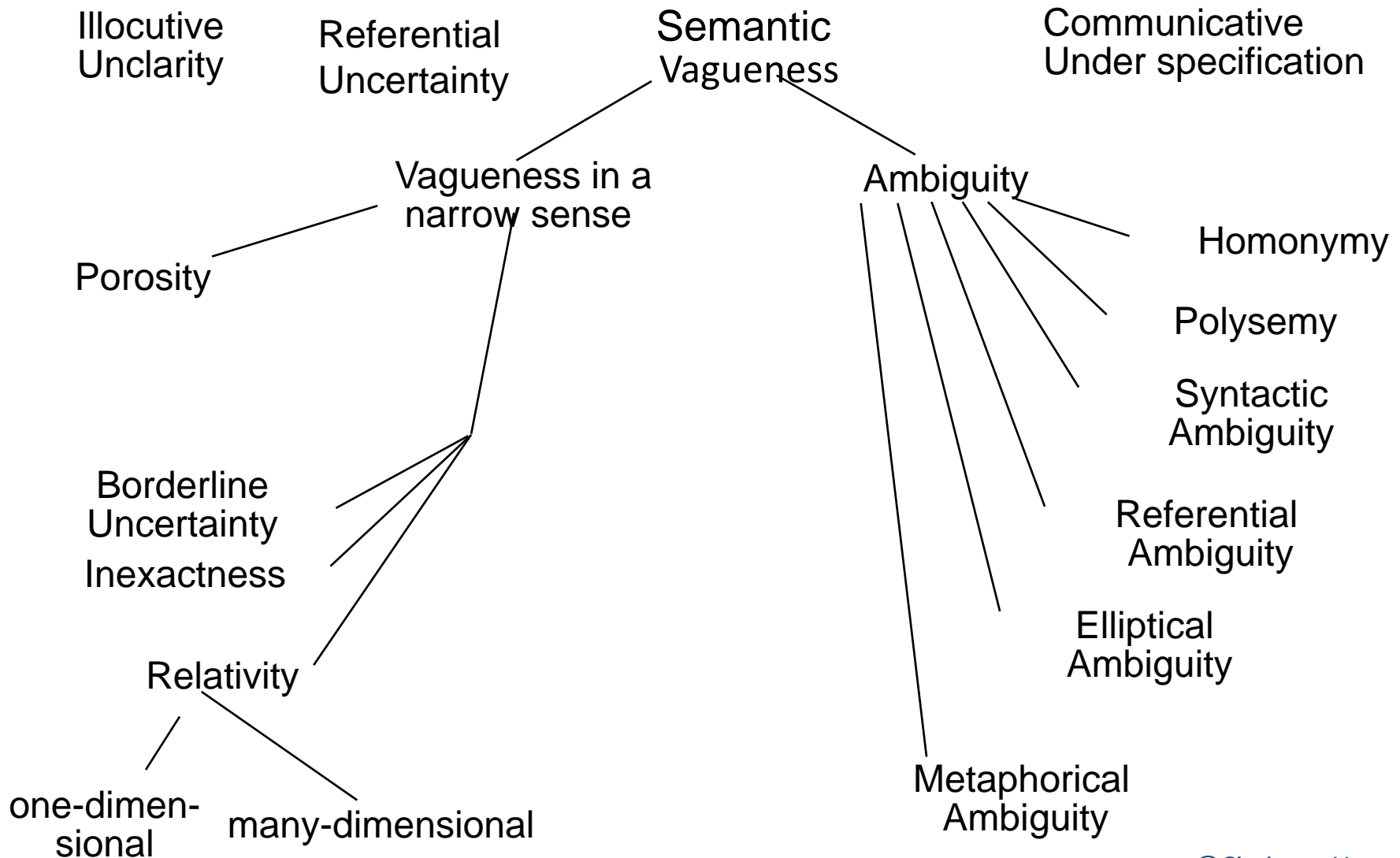
e andere in der ersten Hälfte des

Jahrhundert (B.)

Challenge for DH: Vagueness on several layers

- **Linguistic vagueness,**
- Logical vagueness,
- Fuzzy concepts
 - “Before Stephan the Great, *all mountains around* Moldavia belonged to Transilvania and the country was *narrow* on this side” ...
- Vague or concurrent ontological systems:
 - The Turkish and the Moldavian administration
- Referential vagueness or uncertainty
 - The origin of the hill “Chan Tepesi” or “Mogila Rabuy”
- Naïve History (derived from ‘naive physics’)
 - „The Roman Empire conquered Dacia“
- Historical change
- **Vagueness of the sources**

M.Pinkal's Schema of Semantic Vagueness



Selected markers for linguistic vagueness

1. comparatives, inexact adjectives e.g. *“hostile”, “near”, “distant”*
2. non-intersectives e.g. „supposed“, „so-called“
3. Hedges e.g. „rather“, „more or less“
4. inexact measures *a 4 days’ journey, 10 feet”*
5. modals (attitudes) e.g. probably, hopefully
6. subjunctives v+analyse
7. lexical quotation markers
8. vague quantifiers e.g. „many“, „most“
9. complex quantifiers e.g. *“roughly half of the 20-30 thousand soldiers”*
10. numbers
11. range expressions e.g. *“The beginning of the 18. century”*
12. unclear place „*the former prince*“, „*Moramor*“
13. unclear person e.g. „*the former prince*“
14. unclear time e.g. „*in prehistoric times*“

Other sources /levels of vagueness

2. Factual uncertainty

2.1 References to external publications

2.2 References to external persons

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

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?

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)

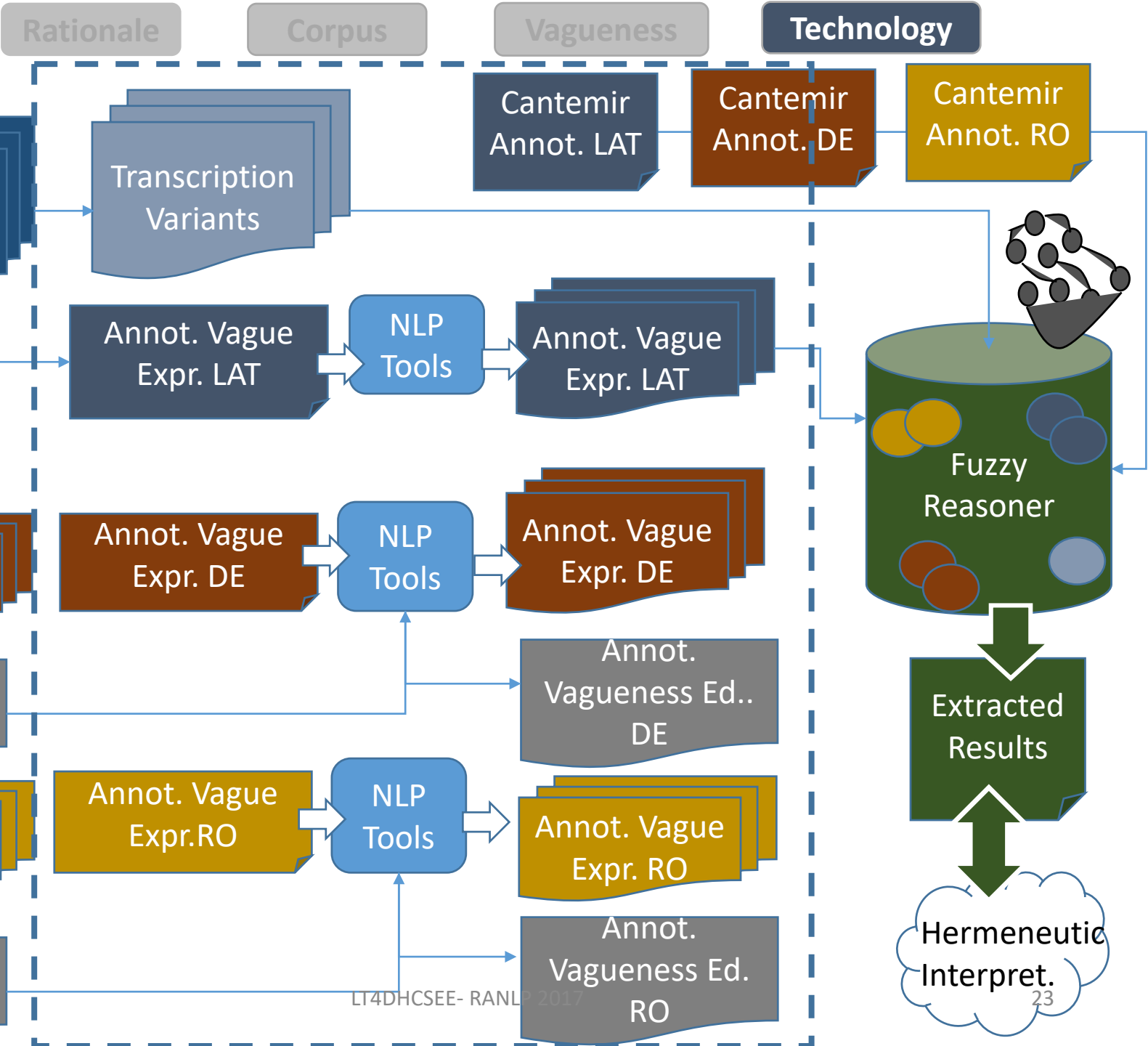
Hezarfen (Hezarfen Hüseyin Efendi) (?-1691/92), *Tenkih-i Tevarih-i Mülük*: **is NOT mentioning this**

- 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:

Workflow



Corpus creation – challenges

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

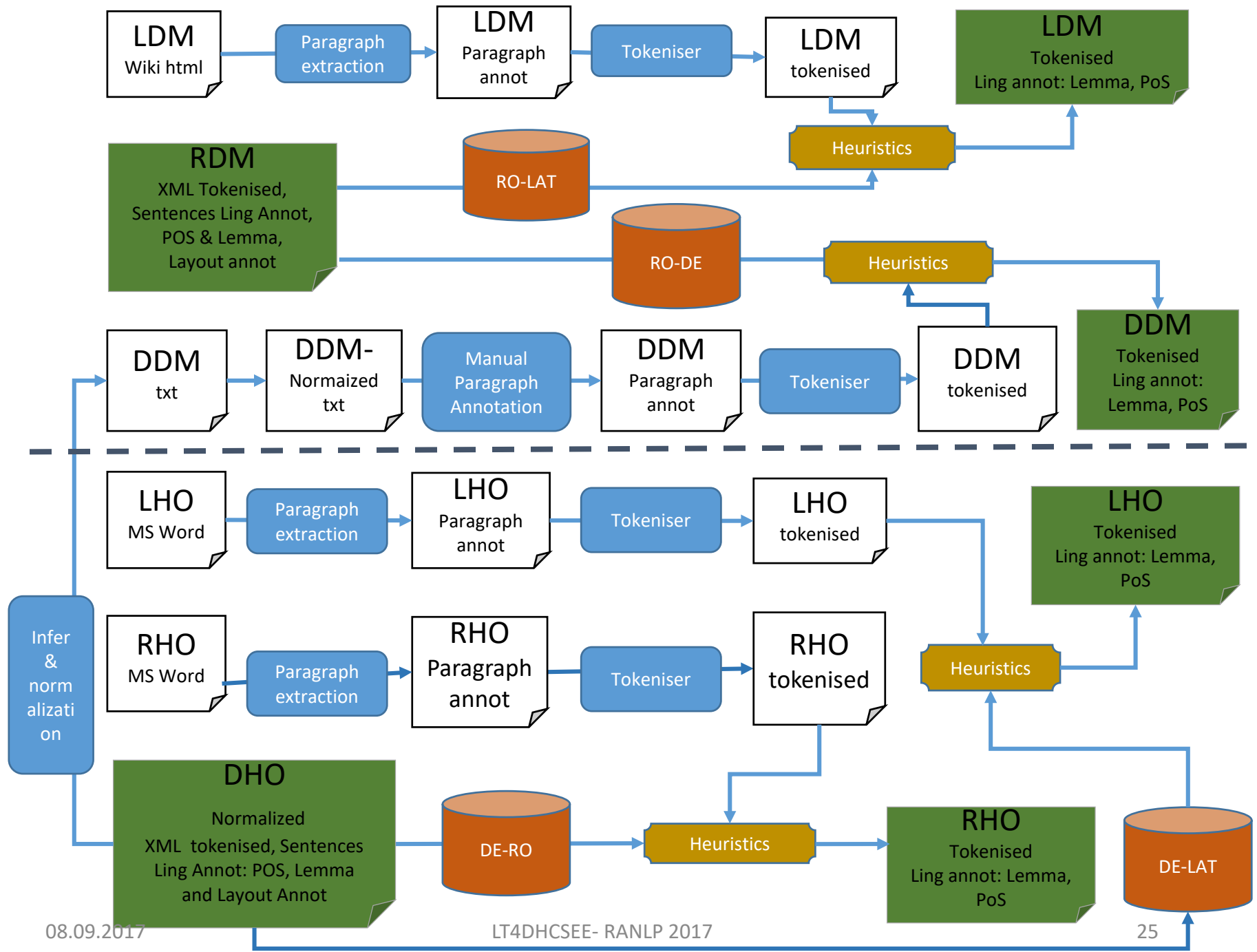
serpens, Kynle, canis &c. Im plurali setzen sie hinten an die Wörter, die eine lebendige Sache bedeuten, den Artikel ij; als: *Łaij, Oamenij*, equi, homines: leblose Kreaturen aber endigen sich im Plurali auf *ele*, als *Scaunele, Vassele*, u. s. w. Auch haben die Moldauer ymeen Articulos foeminini generis,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

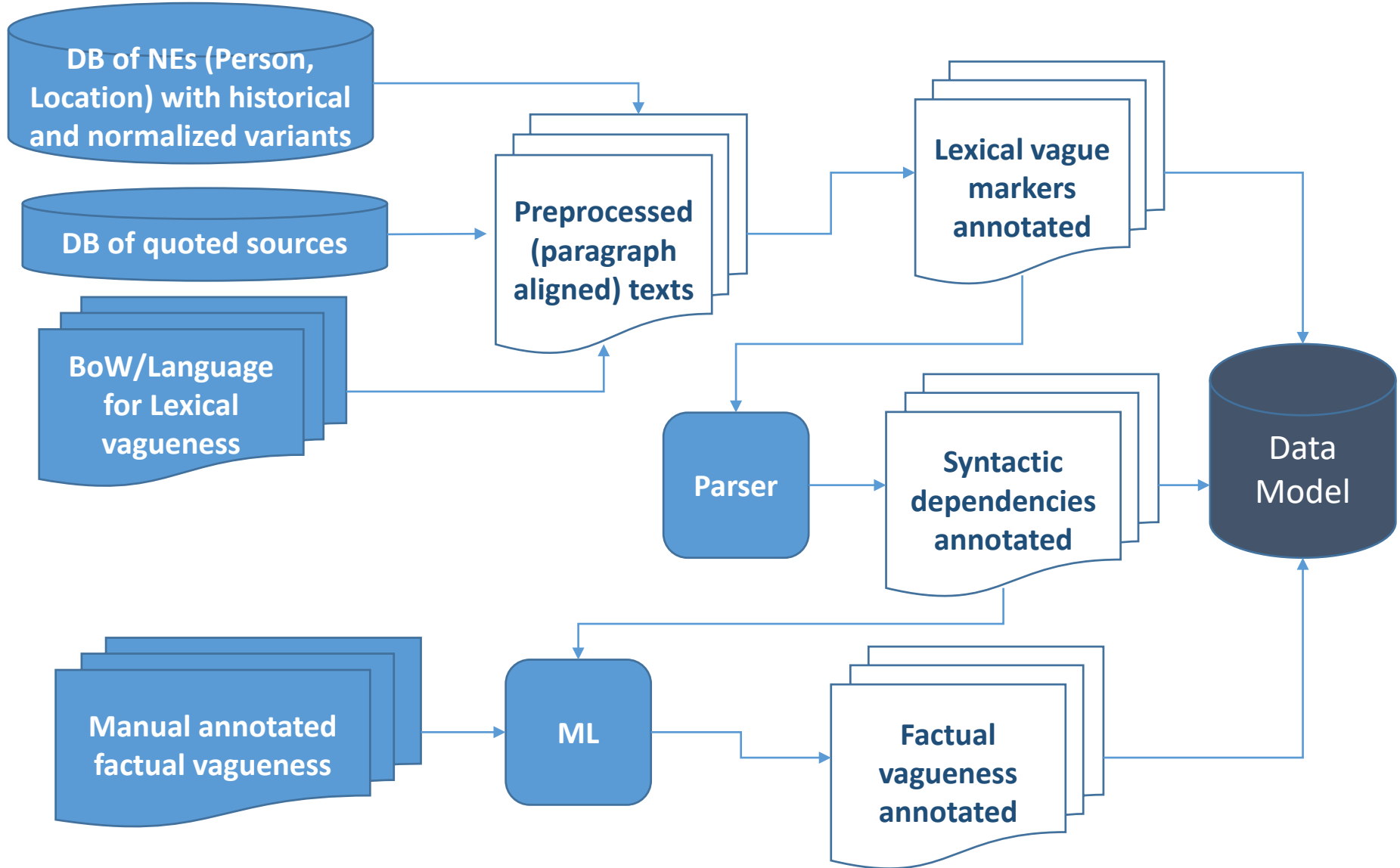
3) Kan man vielleicht wahrscheinlicher muthmaßen, daß diejenigen Wörter, welche mehr mit der Italiänischen als mit der alten römischen Sprache übereinkommen, aus dem langen Umgange, welchen die Moldauer mit den Genuesern während ihres Besizes der Küsten des schwarzen Meers hatten, sich in unsere Sprache 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*, *προκοπιη*, *Blaster*, *βλασφημιω*, *azyma*, *ἀζυμον*, *Drum*, *δρόμος*, *Pizma*, *πιζμα*. 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,

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





First steps: annotating vague expressions -1-

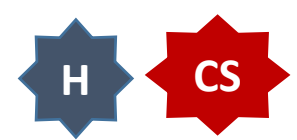
- To annotate vague expressions, we first need to (semi-automatically) identify them.
- Since there are two types of clues for uncertainty and vagueness (explicit lexical-semantic clues and implicit syntactic clues), two separate strategies are needed.
- The explicit lexical clues are words or expressions such as *some, most of, a few, probably, possibly, somewhere*, etc. and verbs like *to believe, think, prefer*, etc.
- The implicit syntactic clues are mainly verb moods such as conditional-optative for Romanian, conjunctive mood or imperfect/pluperfect for Latin, all of them indicating a non-reality (doubt, hear-say, possibility, etc.)

First steps: annotating vague expressions -2

- To automatically identify (mark up in text) the explicit lexical-semantic clues, our strategy is the following:
 - One manually create a list of words and expressions that are possible indicators of vagueness for the three languages (Latin, Romanian and German), from selected parts from both texts
 - Pre-processing the texts: chunking, lemmatizing, PoS tagging
 - Automatically finding and marking all the (inflected forms of) explicit vagueness terms
 - Manually checking the marking for a short part of text for evaluation (feedback and slight improvement)
- The automatic identification of syntactic clues is a much more difficult/complex task. There is an inherent ambiguity in the text between vagueness and plain quotation (often intentionally created by the author) that is difficult to decide upon even for a human annotator, and thus impossible for the machine. A possible strategy to be investigated is:
 - To use machine learning techniques (may be the power of deep learning) on a training set of positive examples obtained from explicit clues and negative examples of certain text.



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



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



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



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



Prof. Dr. Yavuz Köse, UHH
Turcology



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



Alptug Güney, UHH
Turcology



Segiu Nisioi, UB
CL, CS



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

(Dimitrie Cantemir, History of Moldavia, 1752)