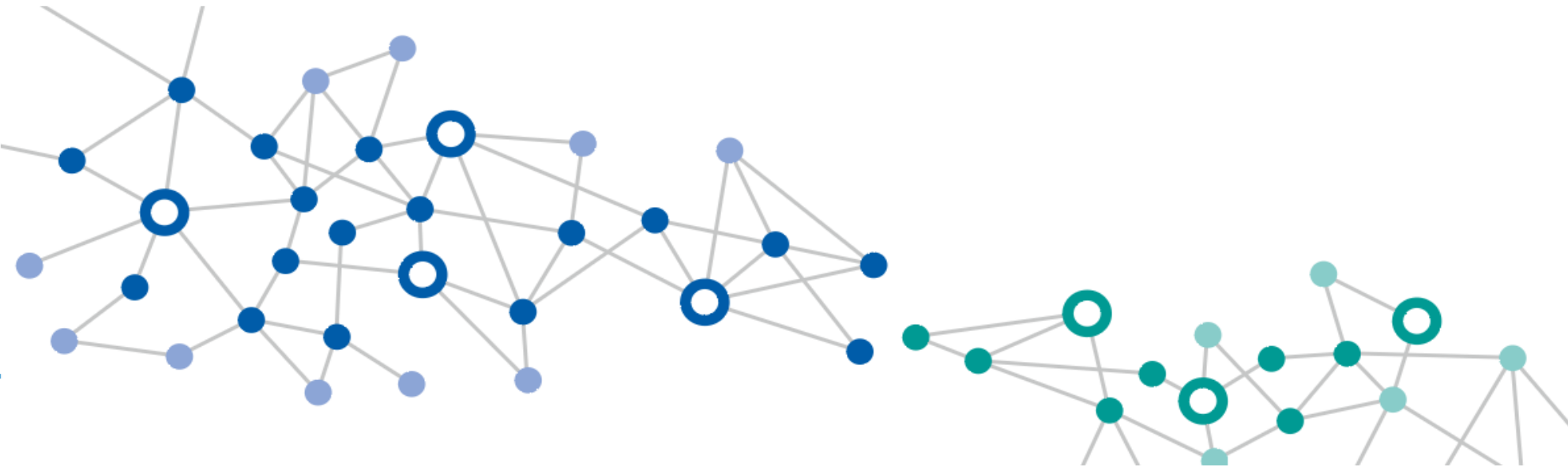


Mehrdeutigkeit und/oder Widerspruch. Herausforderungen bei der Modellierung literaturhistorischer Aussagen

Maria Hinzmann und Christof Schöch

mit Beiträgen von Anne Klee, Johanna Konstanciak, Julia Röttgermann und Moritz Steffes

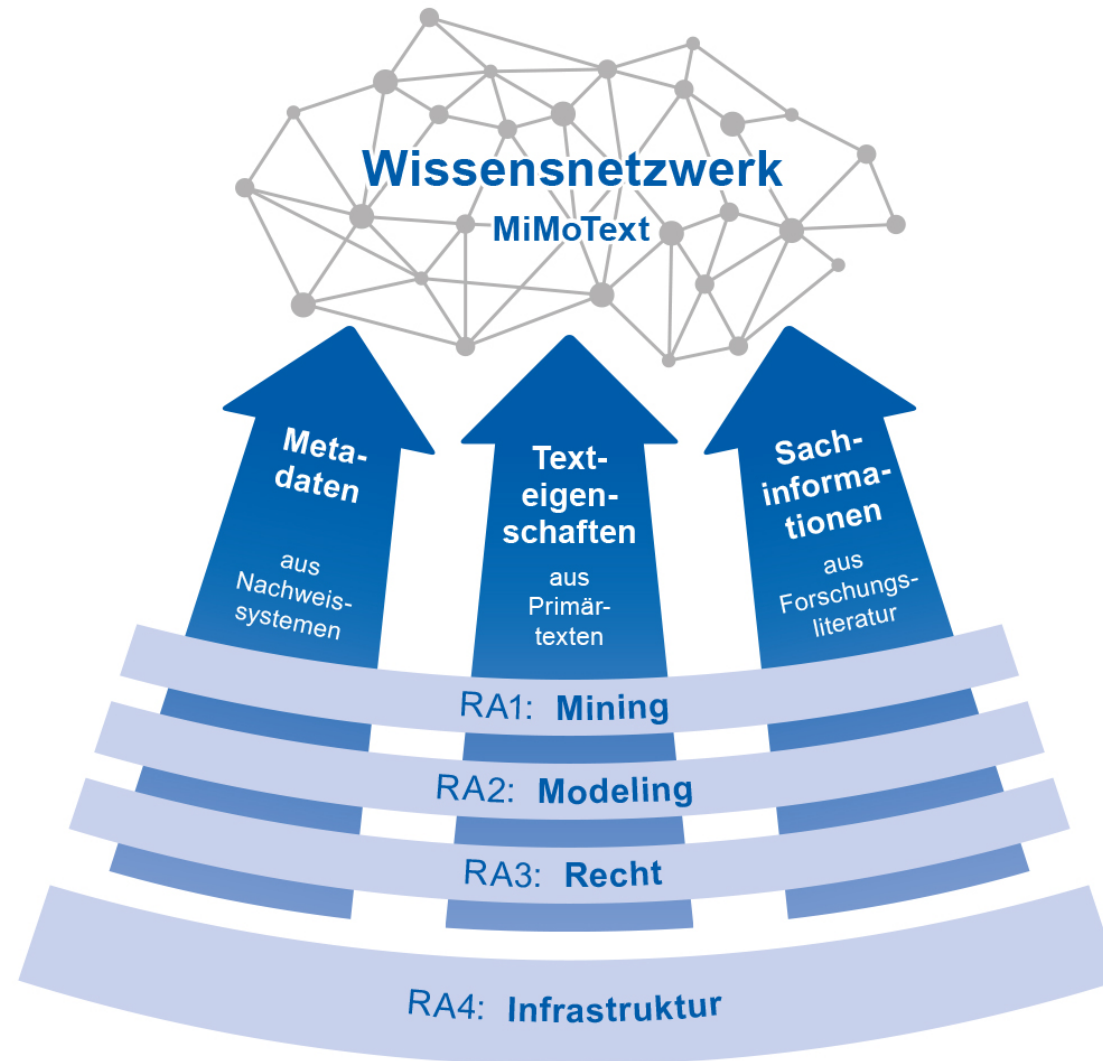


Gliederung

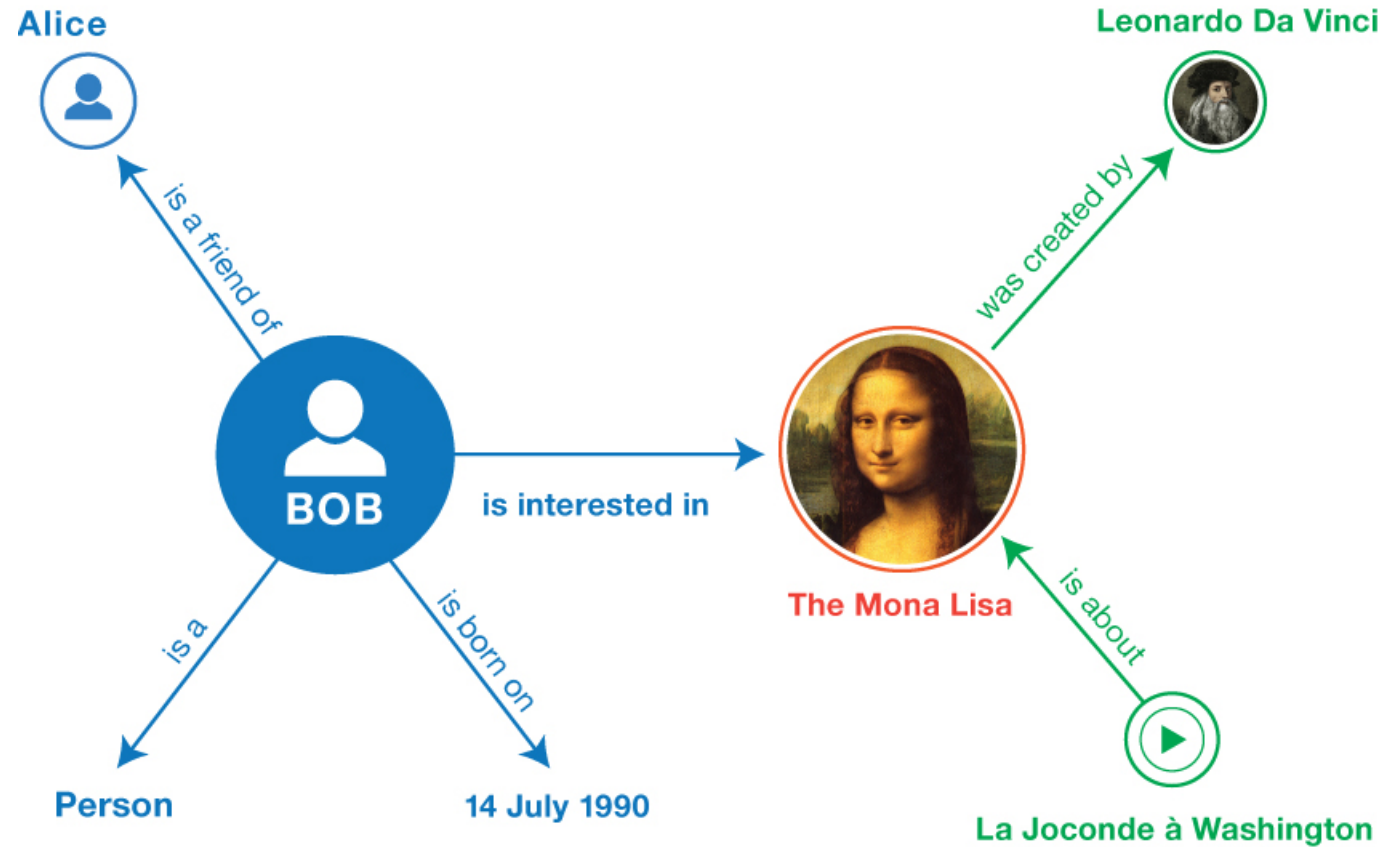
1. Projektüberblick „Mining and Modeling Text“ (MiMoText)
2. Informationsextraktion in MiMoText
3. Aussagenmodellierung in MiMoText
4. Theorieangebote der Digital Humanities
5. Systematisierungsversuch
6. Fazit

1. Projektüberblick „Mining and Modeling Text“

1. Projektüberblick



Modellierungsansatz: Linked Open Data (LOD)



Tripel-Struktur in der MiMoTextBase

Item Discussion Read View history Search MiMoText

Candide (Q1022)

No description defined

▼ In more languages
Configure

Language	Label	Description	Also known as
English	Candide	No description defined	

All entered languages

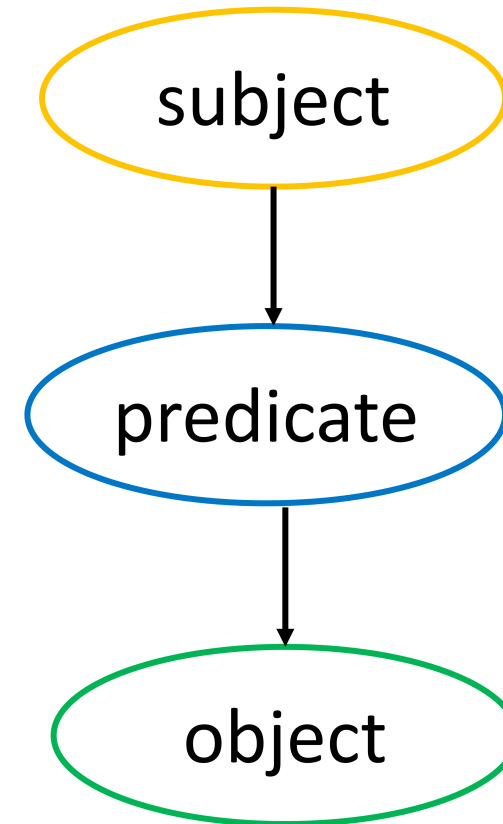
Statements

about mourning
▶ 2 references

narrative location Venice
▶ 1 reference

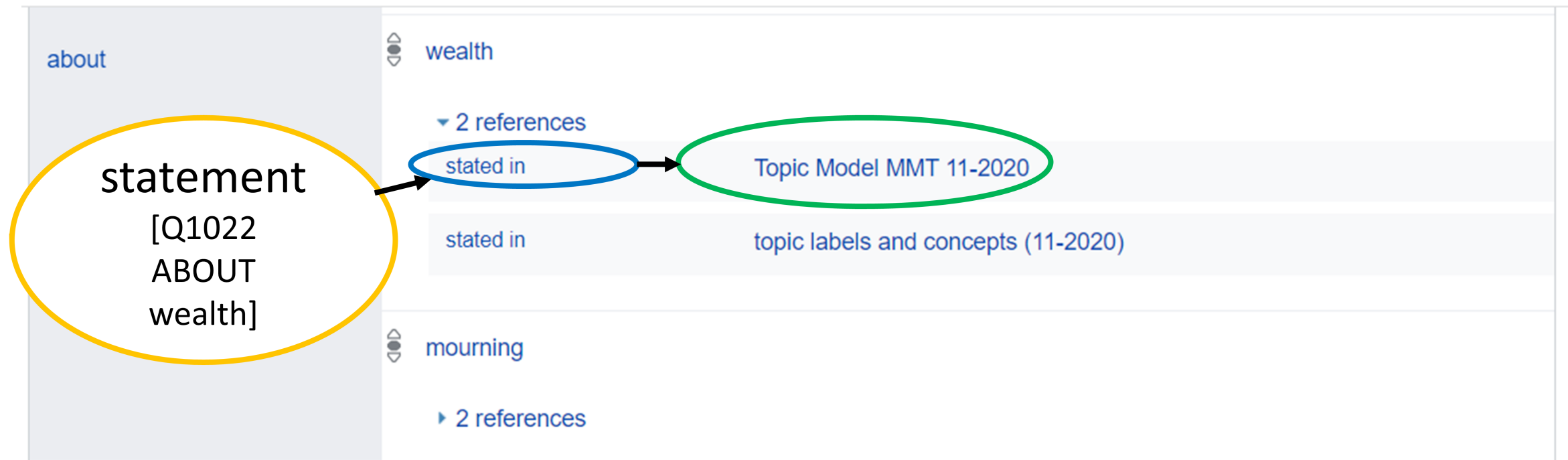
Constantinople
▶ 1 reference

full work available at https://github.com/MiMoText/roman18/blob/master/plain/files/Voltaire_Candide.txt



Aussagen über Aussagen (reification)

=> zentrale Option für die Referenzierung von Quellen und die Modellierung von Informationsqualitäten (Unsicherheit, Zuverlässigkeit etc.)



2. Informationsextraktion in „Mining and Modeling Text“

Was sind literaturhistorisch relevante und extrahierbare Informationen?

1. Metadaten aus Nachweissystemen

- Metadaten: Autoren, Werke, Verlage etc.
- Keywords in der Bibliographie: Erzählperspektive, Handlungsorte, Figuren, Themen

2. Primärtexte

- Themen / Topics (basierend auf Topic Modeling)
- Handlungsorte und Figurennamen (basierend auf Named Entity Recognition)
- Textlänge, Anteile direkter Rede uvm.

3. Fachliteratur

- Beziehungen zwischen Autor:innen und Werken
 - Inhalte, Themen, Formen, Handlungen, Gattungszugehörigkeit von Werken
 - Wertungen von Autor:innen
-

97.21 DIDEROT, Denis
Jacques le fataliste et son maître, par Diderot
Paris, Buisson, an cinquième de la République, 2t. in-8,
xxii + 286, 320p.
BM BN†
JP 8 vend. V
Gay Q Saint-Fargeau Tchemerzine
Selon Tchemerzine, certains exemplaires portent: *Jacques
la fataliste...*
Cet ouvrage a paru à la fin de 1796, date que donnent les
bibliographes. Nous le classons ici en raison de la date ré-
volutionnaire que porte la page de titre.
*3e personne, avec dialogues, récits intercalés 1re personne,
intervention du narrateur: Jacques, son maître, personnages
rencontrés sur le chemin; voyage, aventures bouffonnes,
galantes, romanesques; thèmes philosophiques, mise en cause
des techniques du romancier.*

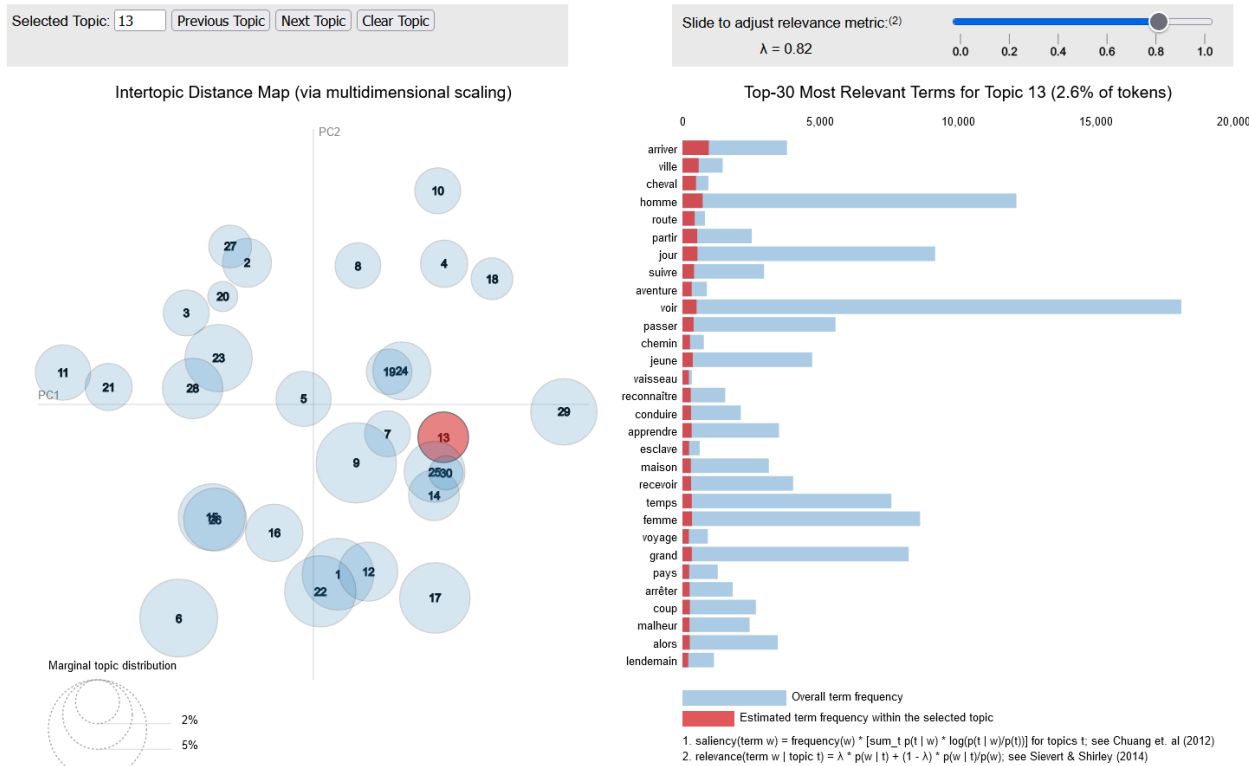
Keywords (kursiv):

- Erzählform
- Handlungsort (fehlt)
- Figuren
- Themen/Plotstrukturen
- Stil/Tonalität

Extrahierte Themen:

- Reise
- Philosophie
- Galanterie

Informationsextraktion Romane: Themen / Topic Modeling



humain société seul mal droit
homme
peuple principe raison vice
nature
loi préjugé
vertu point
chose devenir
naître action

Herausforderungen:

- Labeling der Topics (hier: philosophy | Philosophie | philosophie)
- Mapping auf Themenkonzepte



The screenshot shows a text editor window with a toolbar at the top containing icons for download, undo, redo, play/pause, and other editing functions. The main text area contains German text with several entities highlighted in green boxes: "Denis Diderot", "Supplément au voyage de Bougainville", "Voltaire", and "(Sub)". A search overlay is visible on the right side of the editor, displaying a list of search results for the term "candide". The results are numbered [1], [2], and [3], each showing a Wikidata URL and a brief description. Below the list, it says "50 items found". At the bottom of the overlay, there is a search input field with the text "candide" and a close button (X).

Vielleicht hängt damit die Tatsache zusammen, daß die "großen" Aufklärer

Ausnahme von Denis Diderot (Supplément au voyage de Bougainville) die Utopie kaum gepflegt und sie allenfalls zuweilen in ihre Werke inkorporiert haben, Montesquieu die historische Gesellschaftstheorie der "Histoire des Troglodytes" in die Lettres persanes von 1721 (Briefe XI-XIV) oder Voltaire die im Kontext der Erzählung fragwürdige Utopie von Eldorado in seinem Candide von 1759 (Kap. XVII-XVIII) oder wie der Marquis de Sade in seinem Briefroman Aline et Valcour. In der zweiten Jahrhunderthälfte wird die literarische Utopie häufig als "

[1] **Candide**
<http://www.wikidata.org/entity/Q215894>
1759 book by Voltaire

[2] **Candide**
<http://www.wikidata.org/entity/Q44703489>
fictional character from the book 'Candide' by Voltaire

[3] **Candide**
<http://www.wikidata.org/entity/Q450360>
Wikimedia disambiguation page

50 items found

candide X

- Deutschsprachige Fachliteratur des 19. bis 21. Jahrhunderts
- Generierung von Trainings- und Testdaten durch manuelle Annotation in INCEpTION
- Training von ML-Verfahren zur weiteren Datenextraktion
- Möglichkeiten der Disambiguierung durch Verknüpfung mit Wikidata und MiMoTextBase

3. Aussagenmodellierung in „Mining and Modeling Text“ (Beispiel: Themen)

Aussagen in der MiMoTextBase

Item

Discussion

Jacques le fataliste et son maître (Q1088)

No description defined

▼ In more languages

Configure

Language	Label	Description
English	Jacques le fataliste et son maître	No description defined

All entered languages

Statements

BGRF ID	97.21
▼ 0 references	
author	DIDEROT, Denis
▼ 0 references	

about

travel

▼ 4 references

stated in	Bibliographie du genre romanesque français
stated in	BGRF_matching-table (##-####)
stated in	Topic Model MMT 11-2020
stated in	topic labels and concepts (11-2020)

philosophy

► 2 references

aliment

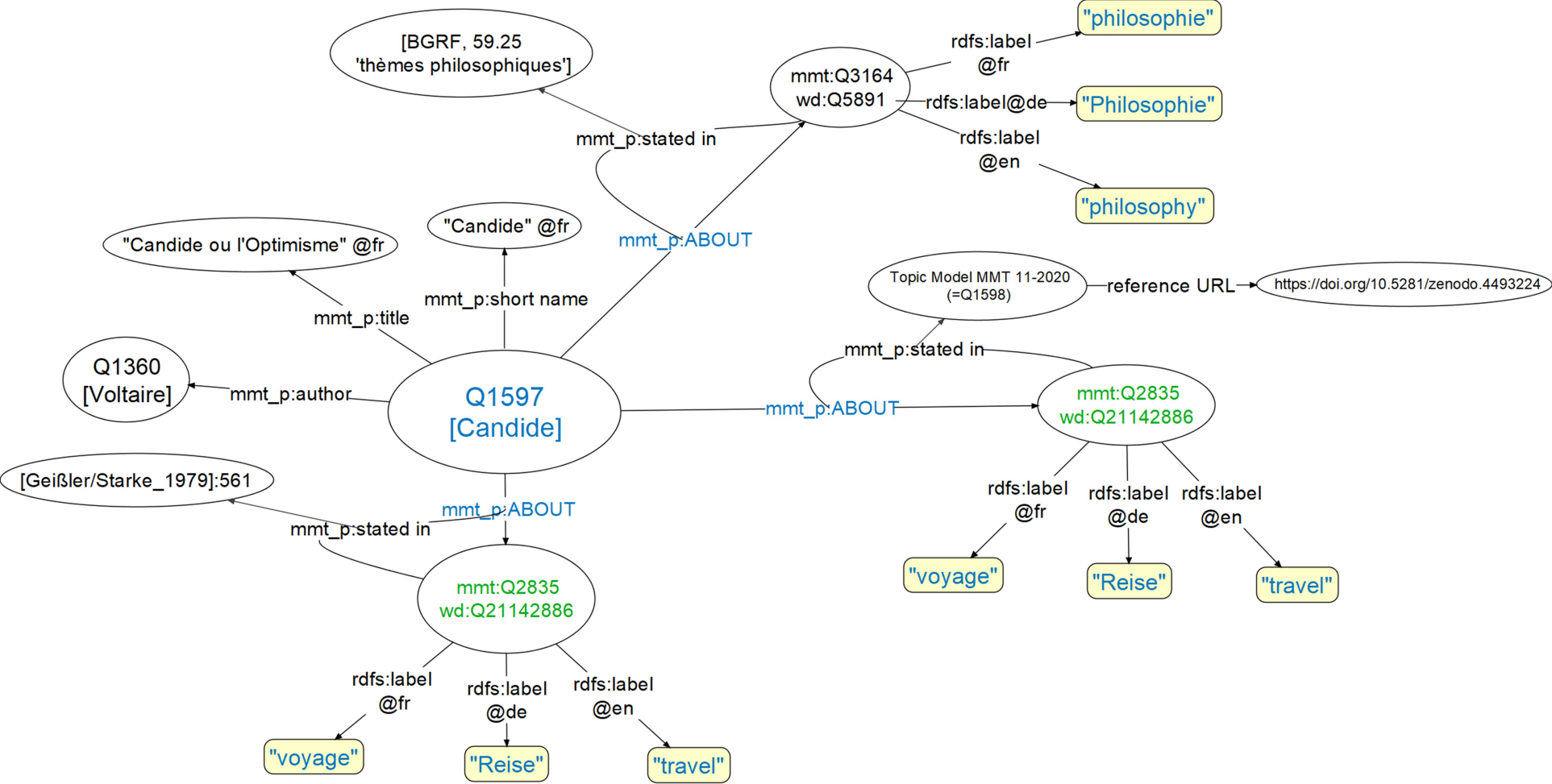
► 2 references

sociability

► 2 references

night

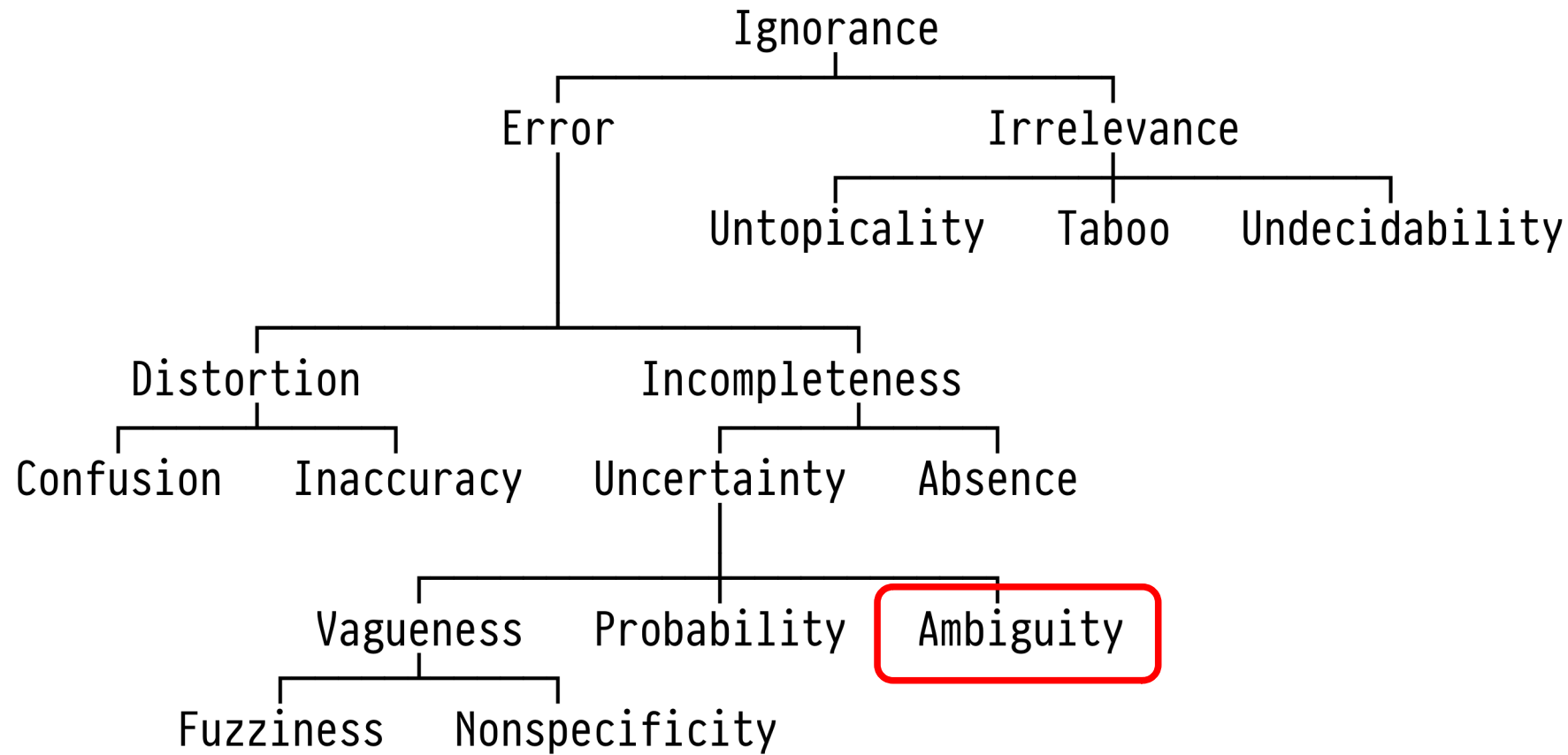
Modellierung von Thementaussagen



4. Theorieangebote der Digital Humanities

Kontext Informationssysteme:

Ambiguität als untergeordneter Teilaspekt



Pietrowski (2019) nach Smets (1997)

Kontext Annotation: Annotation von Mehrdeutigkeit

& Widersprüche beim Annotieren

1.2 Ein Best Practice-Vorschlag



Ambiguitäts-Datenbank TInCAP

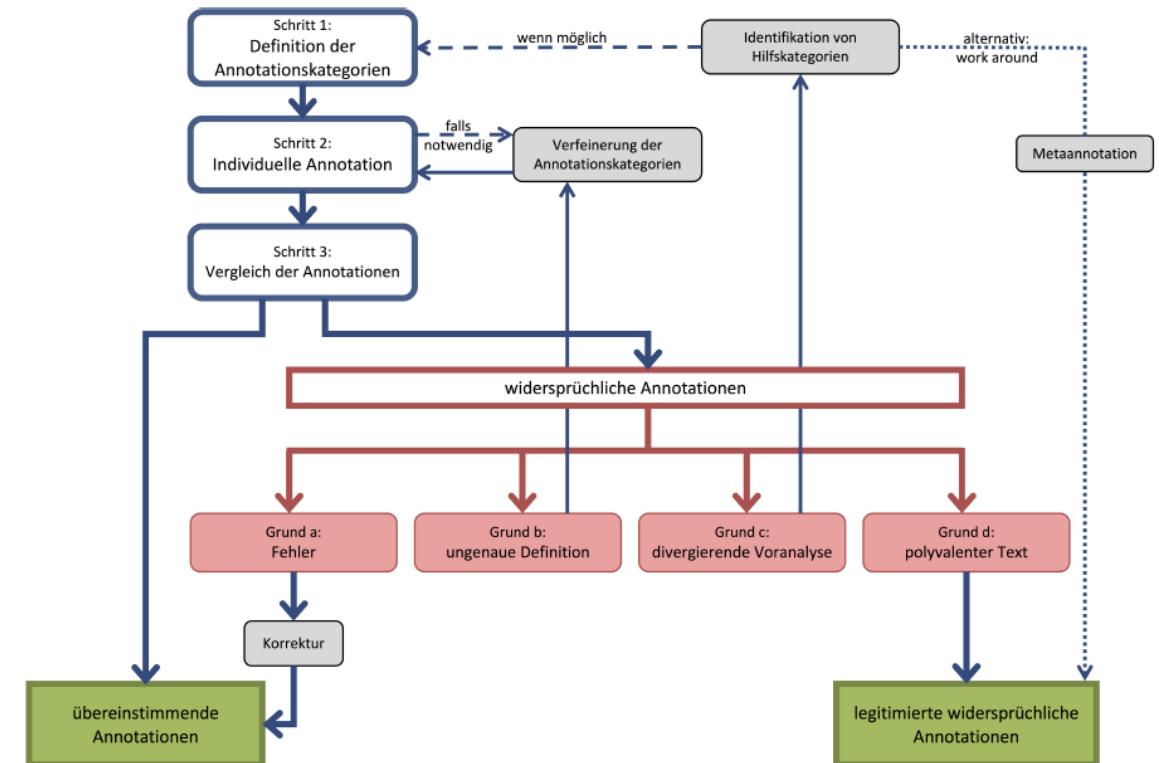


Abbildung 1: Ablaufschema kollaborativer literaturwissenschaftlicher Annotationen

Forschungskontext: Modellierung von Graphdatenbanken in den Digital Humanities



Sonderband der ZfdG (2019): *Die Modellierung des **Zweifels** – Schlüsselideen und -konzepte zur graphbasierten Modellierung von **Unsicherheiten***

- Wagner: „**Ambiguität** und **Unsicherheit**: Drei Ebenen eines Datenmodells“
- Kasper/Kuczera: „Modellierung von **Zweifel** – Vorbild TEI im Graphen“
- Unold/Thierry/Mees: „Academic Meta Tool. Ein Web-Tool zur Modellierung von **Vagheit**“
- Thomas Efer: „Graphbasierte Modellierung von Faktenprovenienz als Grundlage für die Dokumentation von **Zweifel** und die Auflösung von **Widersprüchen**“

Forschungskontext: Kunsthistorische Provenienzforschung

VISU

Vagueness: approximation of spatial or temporal data

Incompleteness: gap in historical reconstruction

Subjectivity: hypotheses to fill a historical gap

Uncertainty: confidence in formulating a hypothesis

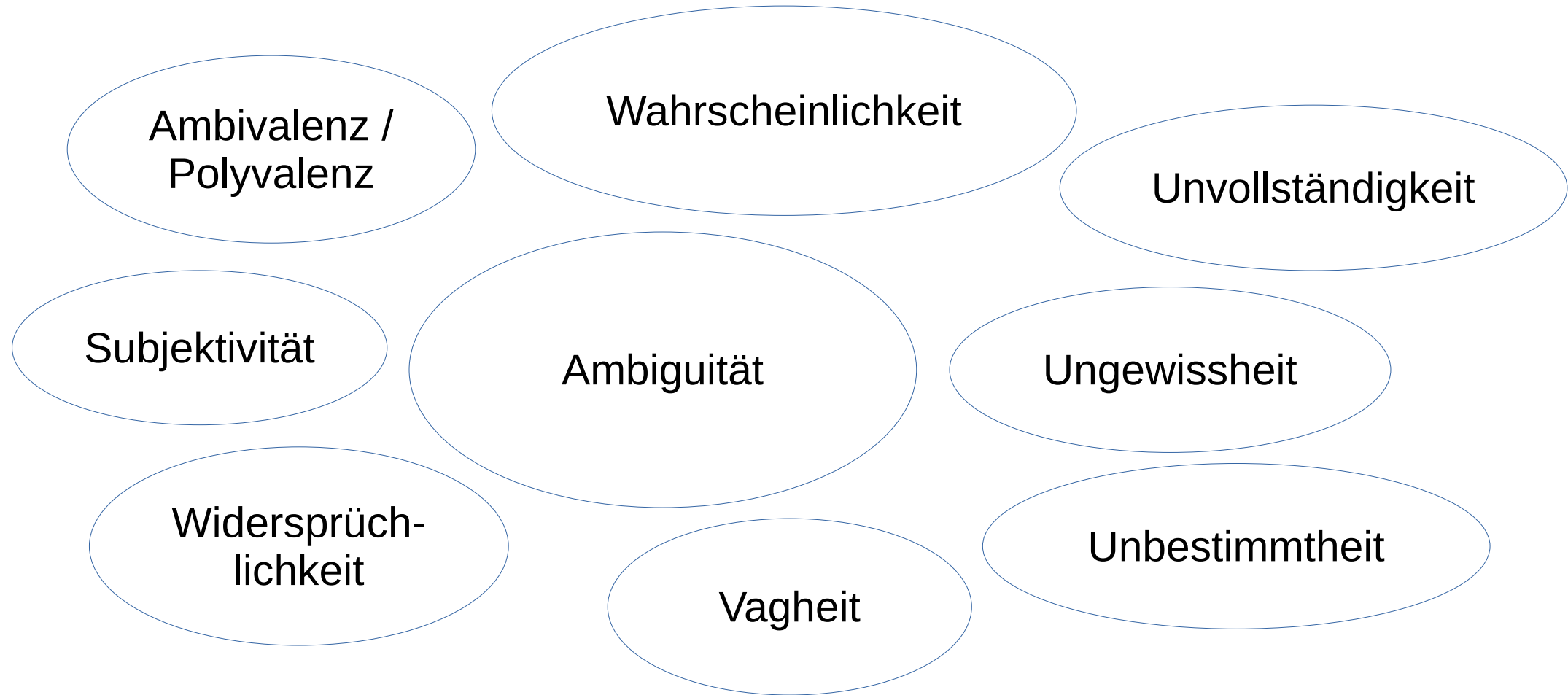
VISU, from Latin *de visu*, "with your own eyes" → Human supervision is required.

Without VISU we run the risk of making clear, complete, objective and certain what is not.



Fabio Mariani. 2022. "Probably Sold to Paalen, Possibly by Exchange": Vagueness, Incompleteness, Subjectivity, and Uncertainty in Digital Art Provenance. COMHUM 2022.

5. Systematisierungsversuch



Konzept-Ebene vs. Tripel-Ebene

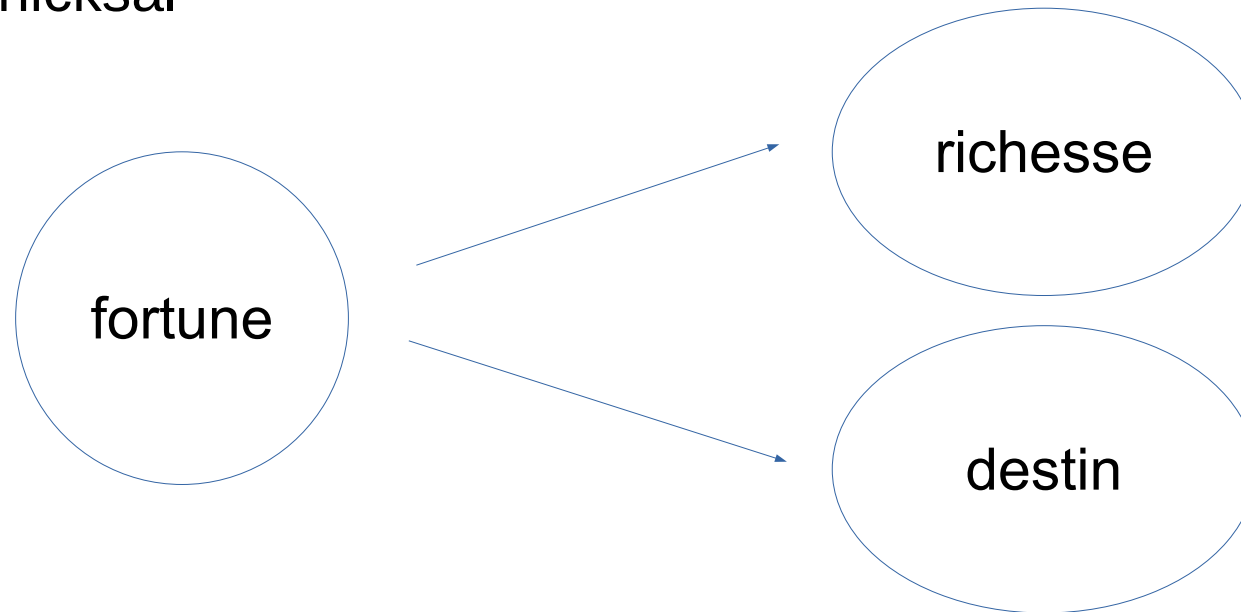
Konzept-Ebene (1): Aufbau Themenvokabular

Explizitmachen von Mehrdeutigkeit/en auf Konzept-Ebene

- Mehrsprachigkeit als Herausforderung
- Interpretationsbedarf

"fortune" hat zwei Konnotationen => Aufsplittung in zwei Themenkonzepte:

- richesse | wealth | Reichtum
- destin | destiny | Schicksal



Konzept-Ebene (2a): Named Entities (allg.)

▼ All			▼ Column 1	▼ bgrf	▼ LOC1	▼ Wikidata Label fr	▼ URL	▼ Quantity LOC
☆	🗨	1.	Abbes_Voyage	58.5	Palais	palais	https://www.wikidata.org/wiki/Q16560	3
☆	🗨	2.	Anonym_Suzon	83.9	Couvent	couvent	https://www.wikidata.org/wiki/Q1128397	21
☆	🗨	3.	Anonyme_Zoloe	00.37				
☆	🗨	4.	Arnaud_Epoux	83.15	Paris	Paris	https://www.wikidata.org/wiki/Q90	27
☆	🗨	5.	Arnaud_Matinees	99.43				
☆	🗨	6.	Arnaud_Sentiment	70.21	Nancy	Nancy	https://www.wikidata.org/wiki/Q40898	111
☆	🗨	7.	Barthelemy_Voyage	88.27	Grèce	Grèce	https://www.wikidata.org/wiki/Q41	440

Reconciliation von Ortsnamen mit Open Refine:
Mappen der Strings aus Named Entity Recognition mit Wikidata-IDs
= Disambiguierung
(Ebenso: Werktitel und eponyme Figur, bspw. Candide, Julie, Justine etc.)

Konzept-Ebene (2b): Ortsnamen

de ses différents maîtres : Je vous demande cette grâce, madame, avec d'autant plus de confiance, que vous m'avez annoncé que je pouvais y avoir quelque droit. Je serai ravi à vous les renvoyer, ainsi que toutes vos lettres, par la première occasion, si vous le jugez à propos. Vous auriez la bonté de me les envoyer par le carrosse de voiture de **Caen LOC**, qui loge au **Grand-Cerf LOC**, **rue Saint-Denis LOC**, à **Paris LOC**, et part tous les lundis.

Ainsi finit l'histoire de l'infortunée sœur **Suzanne Saulier PER**, dite **Simonin PER** dans son histoire et dans cette correspondance. Il est bien triste que les mémoires de sa vie n'aient pas été mis au net; ils auraient formé une lecture très-intéressante. Après tout, M. le marquis de Croismare doit savoir gré à la perfidie de ses amis de lui avoir fourni une occasion de secourir l'infortune avec une noblesse, un intérêt, une simplicité vraiment dignes de lui : le rôle qu'il joue dans cette correspondance n'est pas le moins touchant du roman.

On nous blâmera, peut-être, d'avoir inhumainement hâté la fin de sœur **Suzanne PER**, mais ce parti était devenu nécessaire à cause des avis que nous reçûmes du château de **Lasson PER**, qu'on y meublait un appartement pour recevoir **Mlle de Croismare PER**, que son père voulait faire sortir du couvent, où elle avait été depuis la mort de sa mère. Ces avis ajoutaient qu'on attendait de **Paris LOC** une femme de chambre, qui devait en même temps jouer le rôle de gouvernante auprès de la jeune personne, et que **M. de Croismare PER** s'occupait d'ailleurs à pourvoir la bonne qui avait été jusqu'alors auprès de sa fille. Ces avis ne nous laissèrent pas le choix sur le parti qui nous restait à prendre; et ni la jeunesse, ni la beauté, ni l'innocence de sœur **Suzanne PER**, ni son âme douce, sensible et tendre, capable de toucher les cœurs les moins enclins à la compassion, ne purent la sauver d'une mort inévitable. Mais comme nous avons tous pris les sentiments de **Mme Madin PER** pour cette intéressante créature, les regrets que nous causa sa mort ne furent guère moins vifs que ceux de son respectable protecteur.

Herausforderungen

- Ambiguität (z.B. Homonymie)
- Fiktionalität
- Historizität

Tripel-Ebene (1): Vagheit / Granularität

Bibliographie-Daten

- “Italie (XVIe siècle)”
- “Italie antique”
- “Vendée (Nantes, au moment de la Révolution)”

Lösungsansatz

- bewusste Informationsverluste auf Ebene der formalen Aussagen (mit Identifikatoren verknüpft)
- Zusätzlich: Abbildung als Strings

Tripel-Ebene (2): Unvollständigkeit, Unbestimmtheit, Ungewissheit

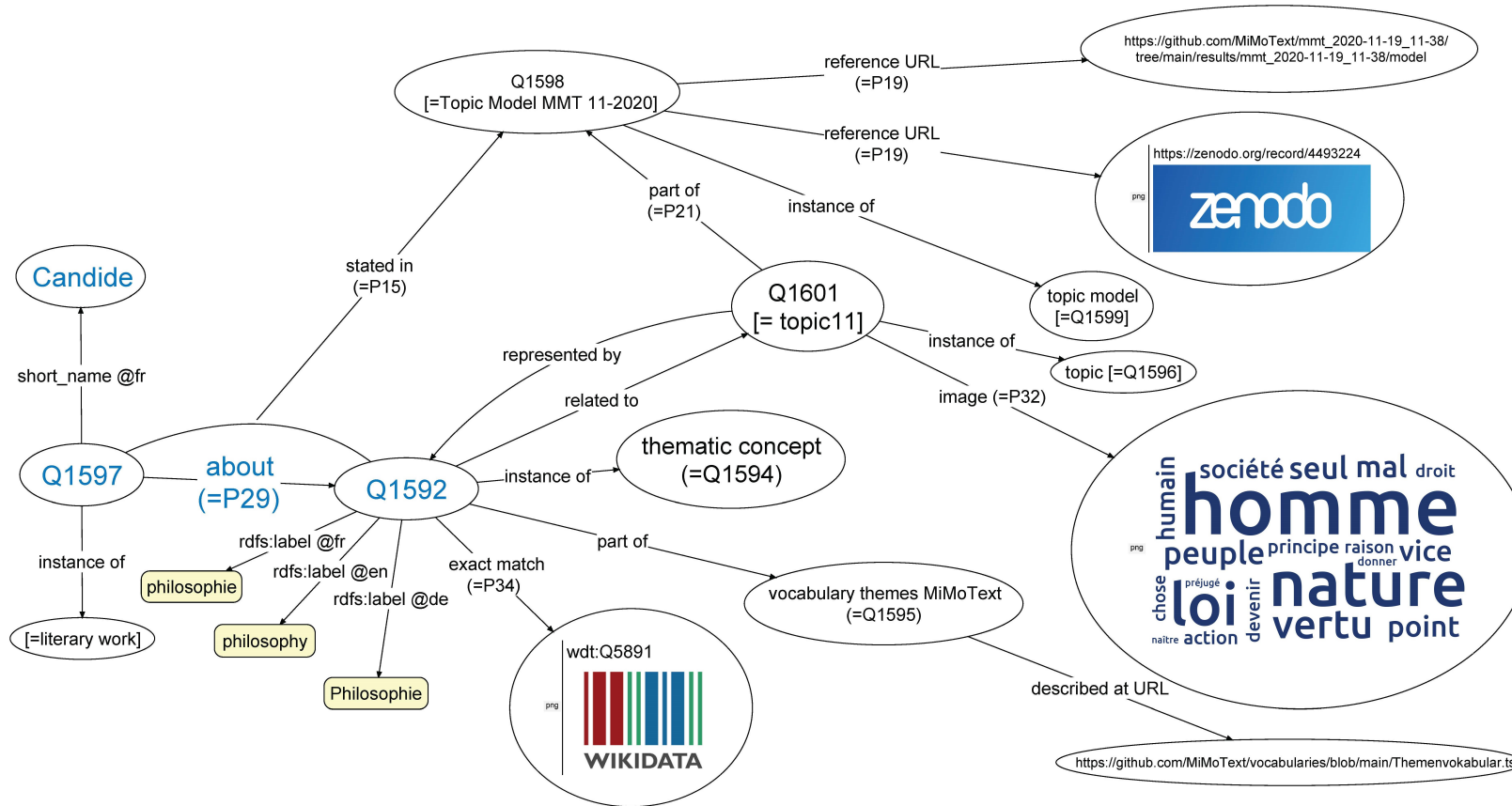
Bibliographie-Daten

- fehlende Angaben zur Erzählform: Ursache unklar
- fehlende Publikationsdaten: Versehen (Identifier repräsentiert Erstpublikationsdatum)

Lösungsansatz

- Ergänzung, soweit ermittelbar

Tripel-Ebene (3): Probabilität



Problem

- Jedes Topic hat eine Wahrscheinlichkeit in jedem Text
- Top-5 werden genannt

Lösung (geplant)

- Wahrscheinlichkeit der Topics über Qualifier repräsentieren

Meta-Aussagen (reification) als Schlüssel im LOD-Paradigma

Widersprüchliche Aussagen können

- nebeneinander in einem Wissensnetzwerk stehen
- durch Quellenangabe, Qualifizierung markiert werden

Unterscheiden lassen sich:

a) konfligierende Aussagen

Werk X → wurde publiziert → 1789 | 1791

b) komplementäre Aussagen

Werk X → handelt von → Philosophie | Reise | Trauer | Nahrung

c) Aussagen von unterschiedlicher Granularität

Werk X → spielt in → Buenos Aires | Konstantinopel | Europa | Amerika

6. Fazit

Konzept-Ebene

- überwiegend linguistische Mehrdeutigkeit, die mit verschiedenen Tools und Ansätzen adressiert wird

Tripel-Ebene

- Multiperspektivität und Subjektivität können im Linked Open Data-Paradigma abgebildet werden
- Mehrwert ergibt sich nicht aus Einzeldeutungen eines literarischen Textes, sondern aus Zusammenführung vieler Aussagen
- => datenbasierte Literaturgeschichte

Chancen datenbasierter Literaturgeschichts-Schreibung und LOD als Modellierungsansatz

- Widersprüchliche Aussagen können nebeneinander stehen
- Sie können in einem Wissensnetzwerk zusammengeführt werden
- Die Transparenz durch Referenzierung der Quellen ist möglich
- Ermöglicht auch gezieltere Suchoptionen
 - Aussagen aus bestimmten Quellentypen
 - Aussagen auf bestimmten Methoden basierend

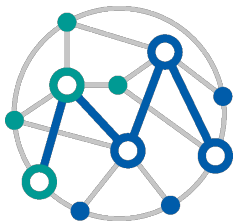
=> Pluralität von Perspektiven, Heterogenität der Quellen: Mehr-Deutigkeit

=> Mehrwert Differenzierung von Phänomenbereichen und Ebenen (im Hinblick auf verschiedene Informationsqualitäten)

- Bauer, Matthias, Joachim Knappe, Peter Koch, and Susanne Winkler. 2010. "Dimensionen der Ambiguität." Zeitschrift für Literaturwissenschaft und Linguistik 40 (2): 7–75. <https://doi.org/10.1007/BF03379835>.
- Gius, Evelyn, and Janina Jacke. 2016. "Zur Annotation Narratologischer Kategorien Der Zeit Guidelines Zur Nutzung Des CATMA-Tagsets (Version 2.0)." <https://heureclea.de/wp-content/uploads/2016/11/guidelinesV2.pdf>.
- Kuczera, Andreas, Thorsten Wübbena, and Thomas Kollatz. 2019. "Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Zur Einführung in diesen Band." Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera / Thorsten Wübbena / Thomas Kollatz. Wolfenbüttel 2019. (= Zeitschrift für digitale Geisteswissenschaften - ZfdG / Sonderbände, January, 4). https://doi.org/10.17175/SB004_013.
- Mariani, Fabio. 2022. "'Probably Sold to Paalen, Possibly by Exchange': Vagueness, Incompleteness, Subjectivity, and Uncertainty in Digital Art Provenance." In . https://wp.unil.ch/lit/files/2022/06/COMHUM_2022_paper_5.pdf.
- Motro, Amihai. 1997. "Sources of Uncertainty, Imprecision, and Inconsistency in Information Systems." In Uncertainty Management in Information Systems, edited by Amihai Motro and Philippe Smets, 9–34. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-6245-0_2.
- Piotrowski, Michael. 2019. "Accepting and Modeling Uncertainty." Die Modellierung Des Zweifels – Schlüsselideen Und -Konzepte Zur Graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera / Thorsten Wübbena / Thomas Kollatz. Wolfenbüttel 2019. (= Zeitschrift für digitale Geisteswissenschaften / Sonderbände: 4). https://doi.org/10.17175/SB004_006A.
- Wagner, Andreas. 2019. "Ambiguität und Unsicherheit: drei Ebenen eines Datenmodells." Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera / Thorsten Wübbena / Thomas Kollatz. Wolfenbüttel 2019. (= Zeitschrift für digitale Geisteswissenschaften / Sonderbände: 4). https://doi.org/10.17175/SB004_007.
- Wagner, Wiltrud. 2020. "5 TInCAP: A Corpus for Annotating Ambiguity." In Idioms and Ambiguity in Context, by Wiltrud Wagner, 92–119. De Gruyter. <https://doi.org/10.1515/9783110685459-008>.



Vielen Dank!



Projekt: <https://mimotext.uni-trier.de>

Folien: <https://zenodo.org/communities/mimotext-tcdh>