Using Formal Models in the Humanities: challenges and misdirections

Michele Pasin

Centre for Computing in the Humanities

Kings College, London

michele.pasin@ kcl.ac.uk

Summary

1. The Digital Humanities
   - where we're coming from / what we're aiming at
   - examples of problems we have
2. Modeling humanities subjects #1: philosophical ideas
   - approach
   - modeling patterns for navigation
3. Modeling humanities subjects: citations
   - problem description
   - first results
4. Conclusions

Digital Humanities: (still) looking for a definition

I think of "Digital Humanities" is a vague and ambiguous term; in fact, I think it causes a fair amount of confusion both inside and outside of the academy. DH is an umbrella term that, depending on who you are talking to, covers a huge territory: everything from applied text analysis and corpus stylistics to the more esoteric and theoretical realms of video game criticism.

Digital Humanities is the use of digital technologies for the pursuit of a greater understanding of the humanities

A field of study that looks at the application of digital technology to the research areas of the humanities. It is not necessarily undertaking that humanities research, but examining the methodology and provision of possibilities for enabling that research or new conceptions of related research. The majority of work in this area, however, gets characterised by those outside the field as IT professionals just doing a bit of programming work for them while uselessly banging on about long-term preservation formats, open data, and not really doing real research.

Univ. of Alberta, Tapor Wiki: http://bit.ly/1oSvX

Digital Humanities: key facts

[...] with the advent of the World Wide Web, digital humanities has broadened its reach, yet it has remained in touch with the goals that have animated it from the outset: **using information technology to illuminate the human record, and bringing an understanding of the human record to bear on the development and use of information technology.**

http://www.digitalhumanities.org/companion/

- essentially, two broad ways to look at it:

- **from humanities**, to computing... eg. 'new media studies'

- **from computing**, to the humanities... eg 'applied computing'

Digital Humanities: key facts

- varying relationship to computer science
  - using vs building?

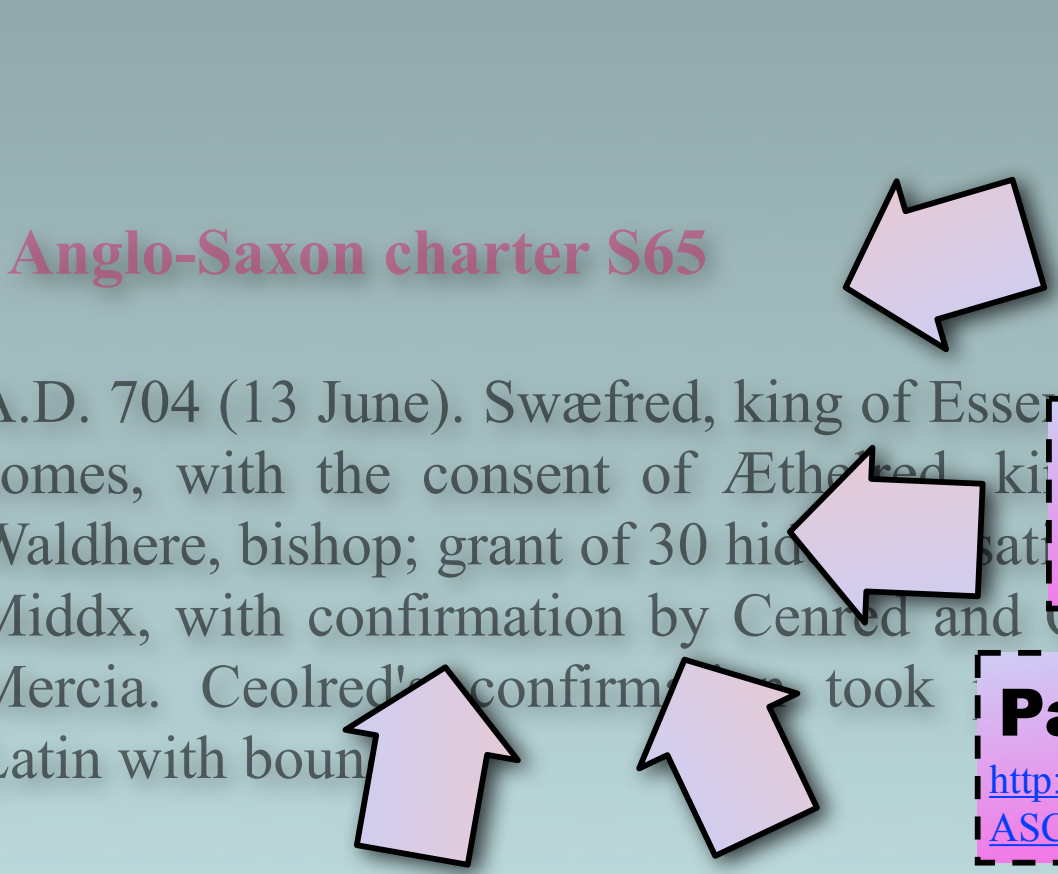
- not one, but many DH!
  - archeology
  - history
  - classical studies
  - literary studies, linguistics, and lexicography
  - musicology
  - philosophy and religion
  - art history

Digital Humanities: increasing need for data integration solutions

Anglo-Saxon charter S65

A.D. 704 (13 June). Swæfred, king of Essex, and Pæogthath, comes, with the consent of Æthelred, king of Mercia, to Waldhere, bishop; grant of 30 hides (cassati) at Twickenham, Middx, with confirmation by Cenred and Ceolred, kings of Mercia. Ceolred's confirmation took place at Arcencale. Latin with bounds.

Digital Humanities: increasing need for data integration solutions



Kemble
http://www.trin.cam.ac.uk/kemble/singlesheets/4-24.html

Langscape
http://www.langscape.org.uk/descriptions/editorial/L_65_000.html

Pase
http://www.pase.ac.uk/pase/apps/ASC/persons.jsp?sourceKey=341

ESawyer
http://www.esawyer.org.uk/content/charter/65.html

ASchart
http://www.aschart.kcl.ac.uk/content/charters/text/s0065.html

Digital Humanities: increasing need for data integration solutions

Transaction Factoid (PASE)

Title Factoid (KEMBLE)

Anglo- harter S65

A.D. 704 (13 Ju Swæfred, king of Essex, and Pæogthath, comes, with the c sent of Æthelred, king of Mercia, to Waldhere, bishop; grant of 30 hides (cassati) at Twickenham, Middx, with confirmation by Cenred and Ceolred, kings of Mercia. Ceolred's confirmation took place at Arcencale. Latin with bo nds.

Confirmation Factoid (PASE)

Place Factoid (LANGSCAPE)

Data integration in DH: desiderata

Project/Data	ASChart	E-Sawyer	Kemble	LangScape	PASE
				x	x
				x	x
				x	x
				x	x
					x
				x	
Language	x	x			x
Language of part of text				x	
					x
					x
Editorial notes			x	x	x

- **data sharing**
  - maintain provenance and integrity
  - eliminate redundancy
  - allow for comparative perspective (e.g. visualise conflicts of interpretations)

- **models exposure**
  - what is an event (e.g. Anglo-Saxon project: what is a transaction in PASE?), a person, a place?
  - Can a certain consensus be reached? Necessity to establish community of practices around modeling exercises, clusters of consensus around knowledge domains or specific disciplines

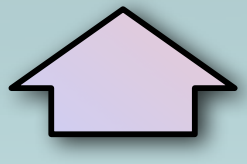
A cross-discipline KR research agenda in DH?

Viewpoints, ideas

People and ph. objects

Information objects

Historical events & time



- history
- performing arts
- visual arts
- law
- religion
- art history
- literary studies
- linguistics
- philosophy
- archeology
- classical studies
- musicology
- lexicography

A cross-discipline KR research agenda in DH?

Contradictory information

Interpretation events

Debate, argumentation

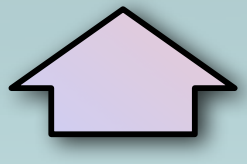
Uncertainty, partial knowledge

Viewpoints, ideas

People and ph. objects

Information objects

Historical events & time



- history
- performing arts
- visual arts
- law
- religion
- art history
- literary studies
- linguistics
- philosophy
- archeology
- classical studies
- musicology
- lexicography

example #1..

modeling philosophy

PhiloSurfical: background and rationale

- PhiloSURFical (2007): learning through semantic navigation
  - prototyped with Wittgenstein's **Tractatus** Logico-Philosophicus (1921)
  - **Annotation of learning materials** *by means of* a domain ontology
  - Reasoning on annotated resources
  - Dynamic reorganization according to different perspectives
  - Mechanisms for **contextual navigation**
  - Tools for providing not answers, but documents!

- Other notable projects:
  - InPhilo Project (USA, 2007)
    Ontological backbone for the Stanford Encyclopedia of Philosophy
    Funded by the *National Endowment for the Humanities Digital Humanities Initiative*
  - Discovery Project (Europe, 2006)
    Generic framework for collaborative annotation/navigation in the *philo*-SW
    Funded by the EU EcontentPlus Grant

What we'd like: a vision...

Interpretative

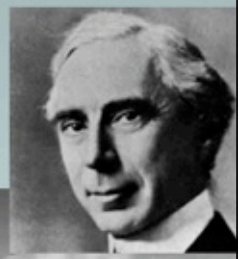
Max Black 1909-1988,
Commentary on the Tractatus

7
What we cannot speak about we must pass over in silence.

Ludwig Wittgenstein 1889-1951,
Tractatus Logico-Philosophicus

Comparative

B. Russell, 1872-1970



6.2
Mathematics is a logical method. The propositions of mathematics are equations, and therefore pseudo-propositions.

Cambridge, 1920-1930

Textual

empiricism platonism

logicism

realism formalism

Theoretical

epistemology

ph.of logic

ph.of mathematics

ph.of language

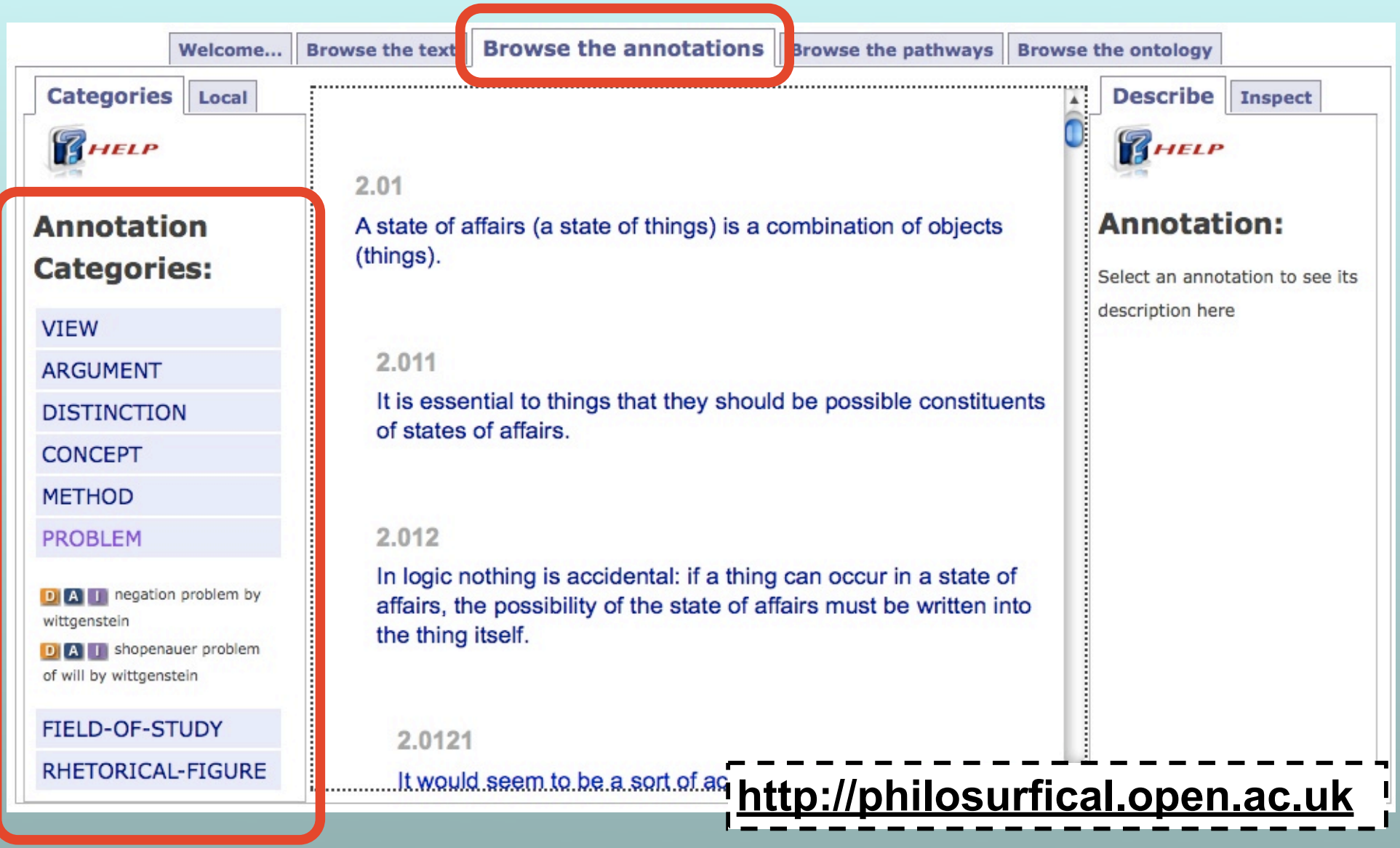
Theoretical

F. Ramsey, 1903-1930

G.E.Moore, 1873-1948

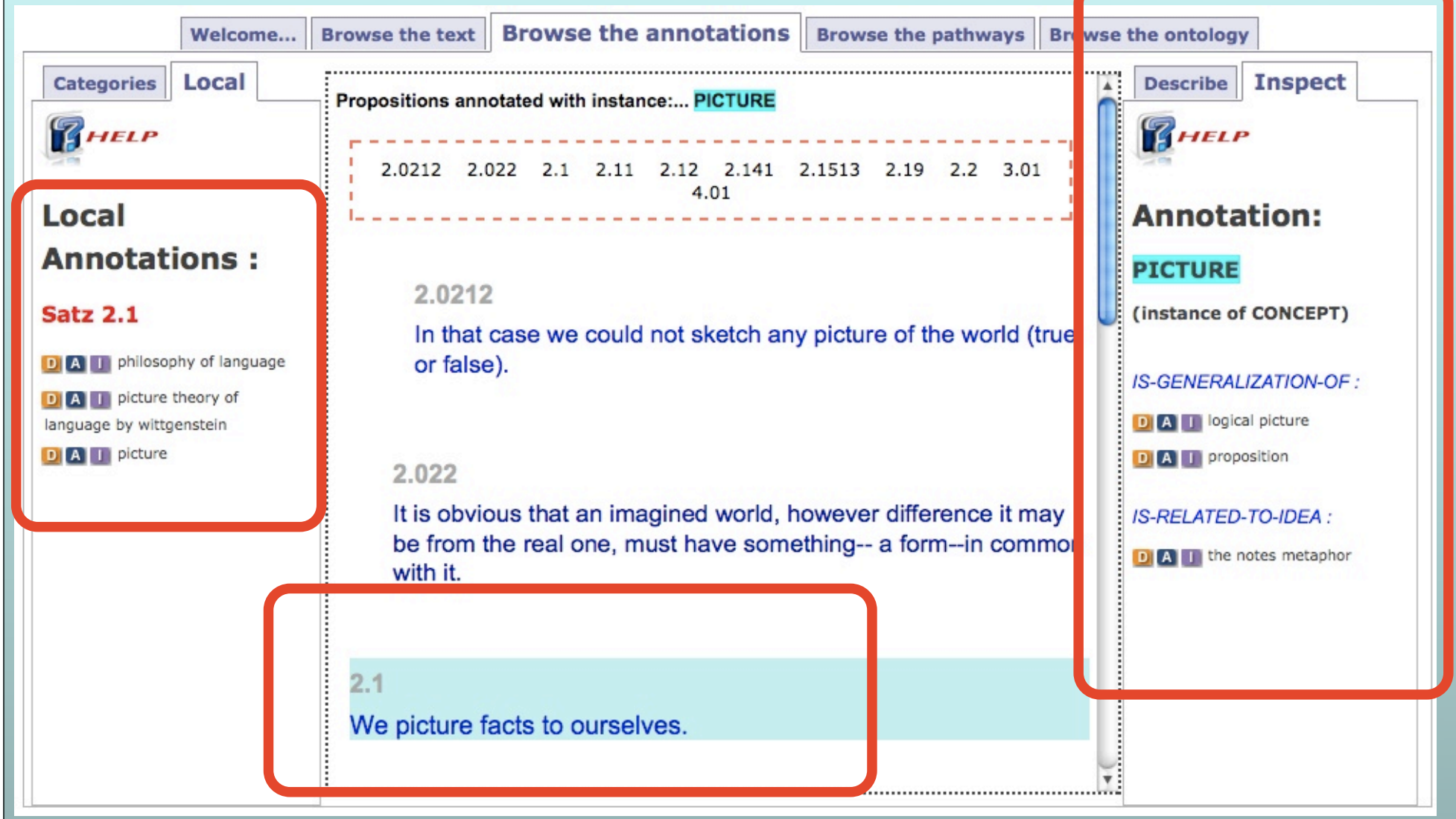
PhiloSURFical: main page

PhiloSURFical: annotation categories

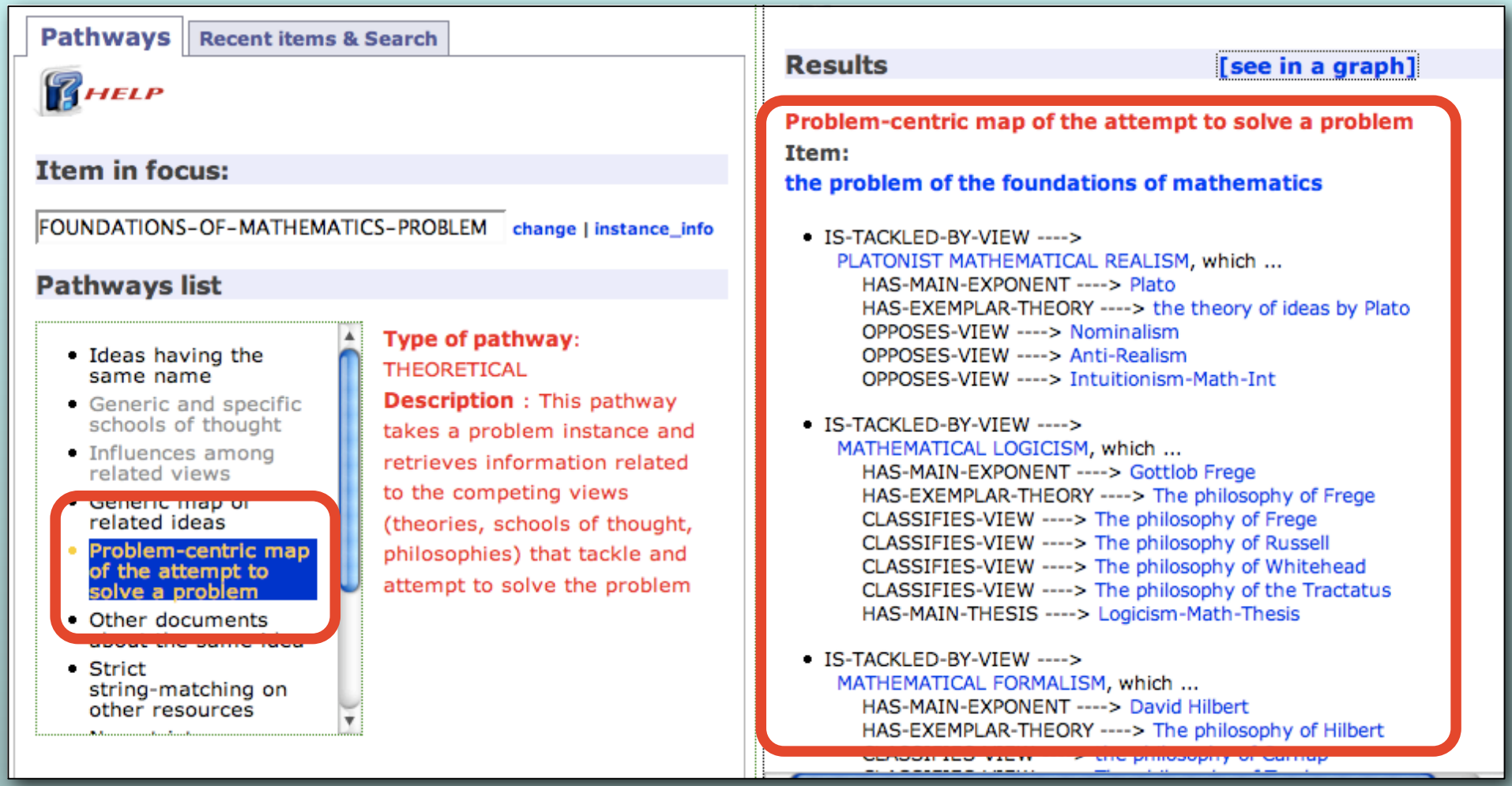


http://philosurfical.open.ac.uk

PhiloSURFical: local annotations



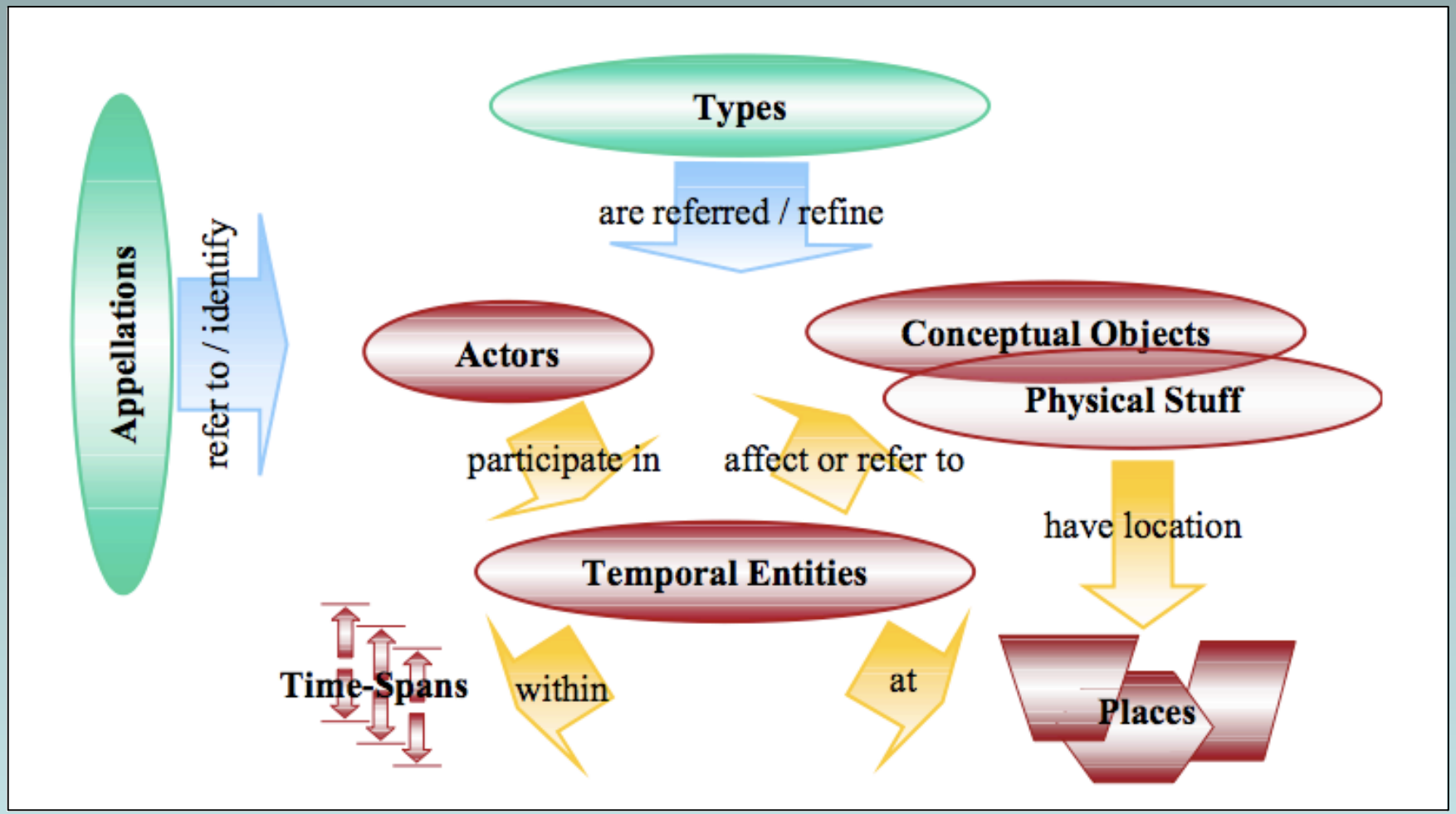
Pathway: problem-centric map



http://philosurfical.open.ac.uk

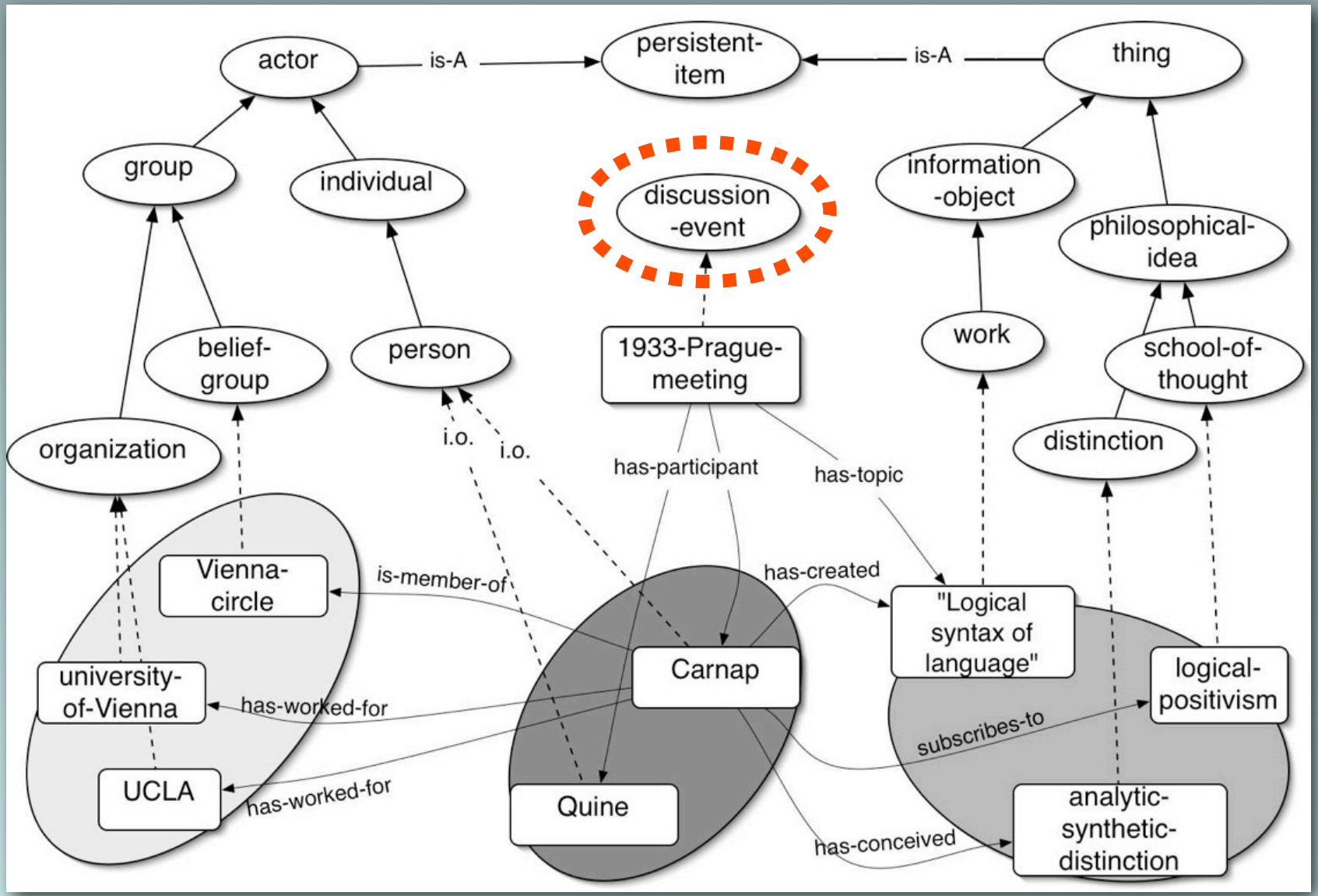
CIDOC Conceptual Reference Model

1996 ICOM initiative, 2006 **ISO standard** (version 4.2)

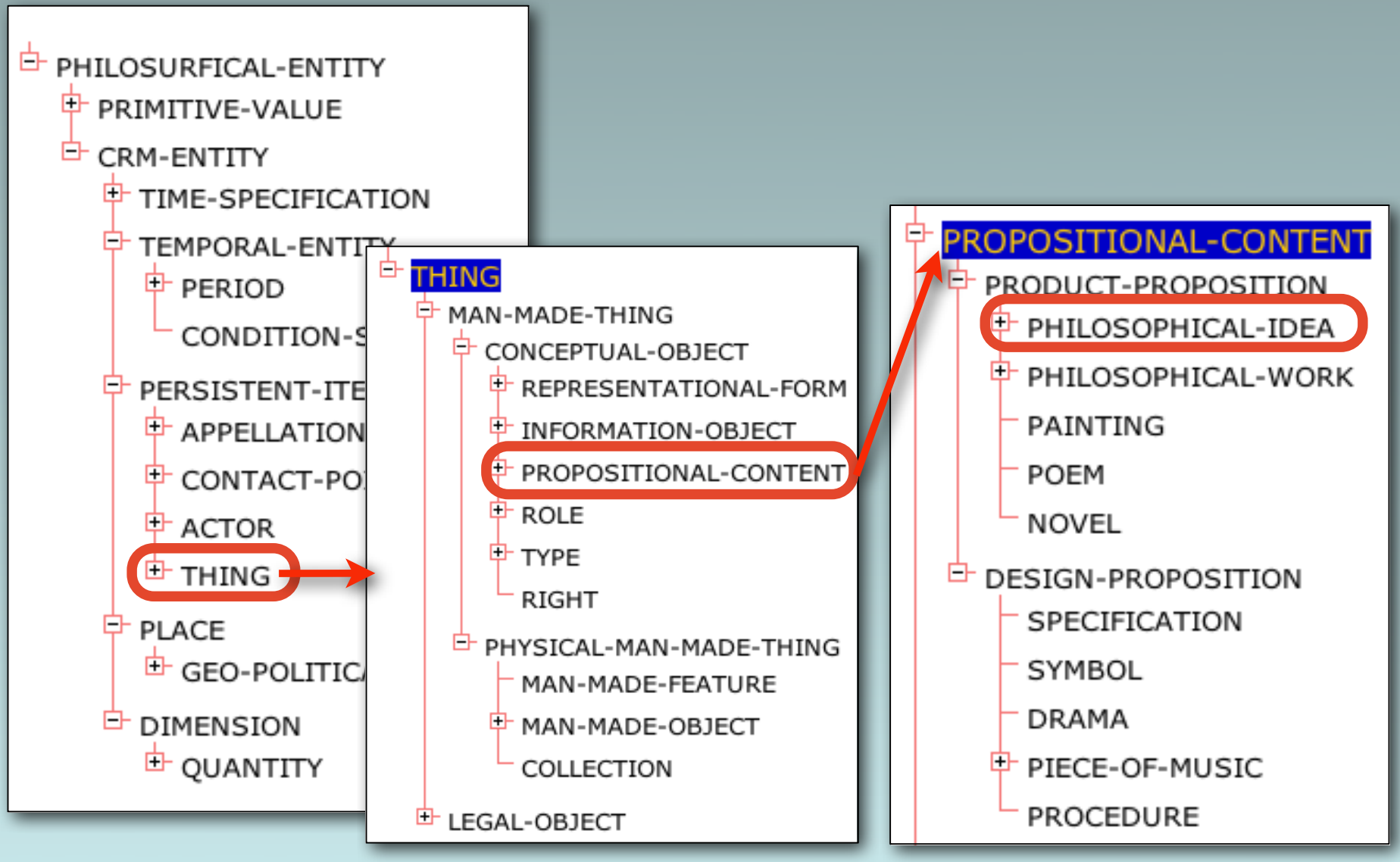


Adapted from CIDOC specification, 2005

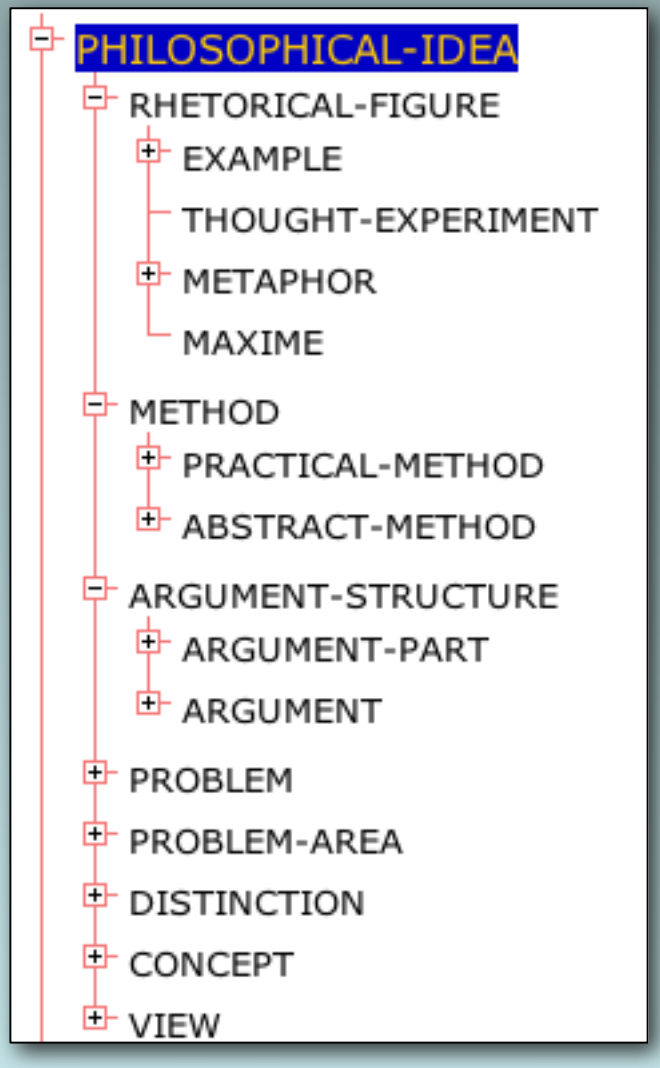
Example: a philosophical event



Onto Walkthrough: conceptual objects



Onto Walkthrough: philosophical ideas

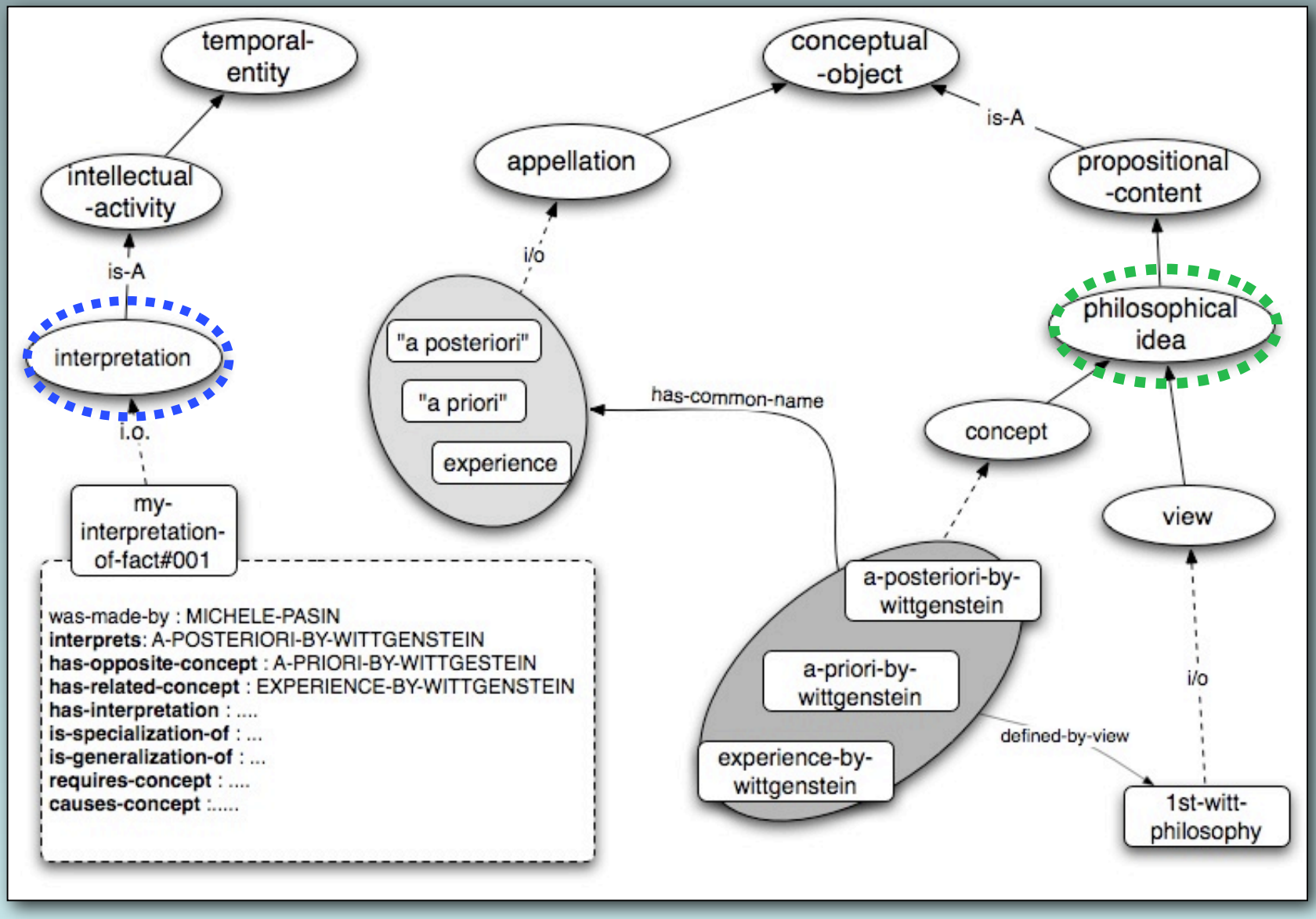


- **Constructivist** approach: 'pragmatic minimalism'

  ..."stone" can be a concept, if there's a view defining it!

- **Goal**: individuate the types of non-physical objs which play a role in the construction of viewpoints!

Important: interpretations vs ideas

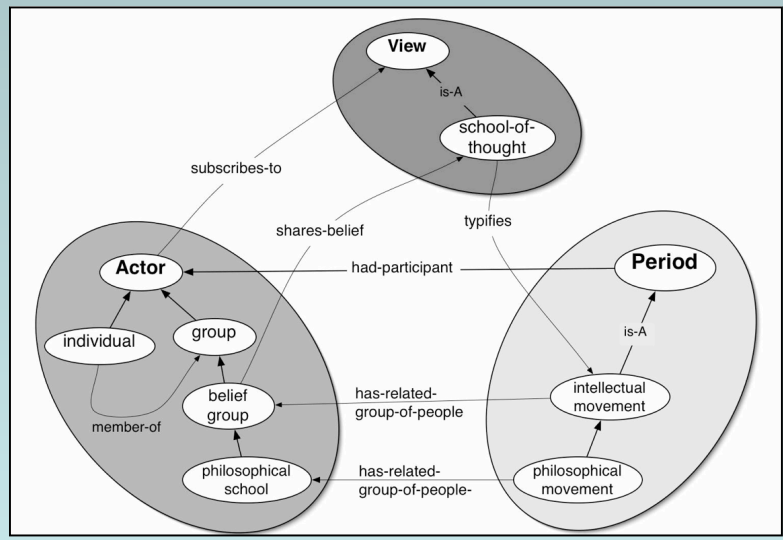


Modeling pattern I: how many rationalisms?

- Problem: natural language often hides type-differences

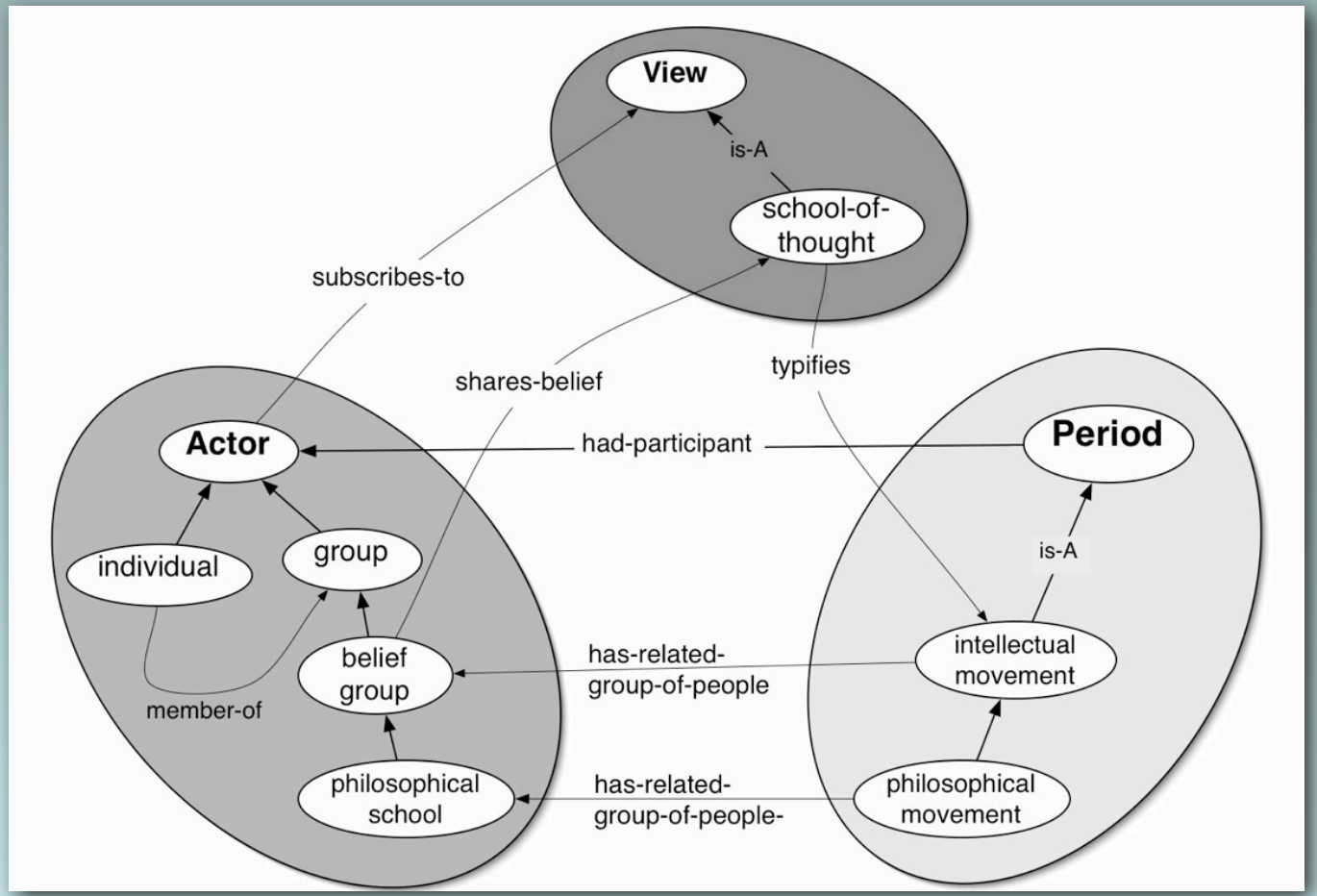
- Tip: taking advantage of natural language ambiguities, so to present resources which are potentially explicative

- Advantage: it allows navigations of ontologically distant entities (belief-groups, views, events)



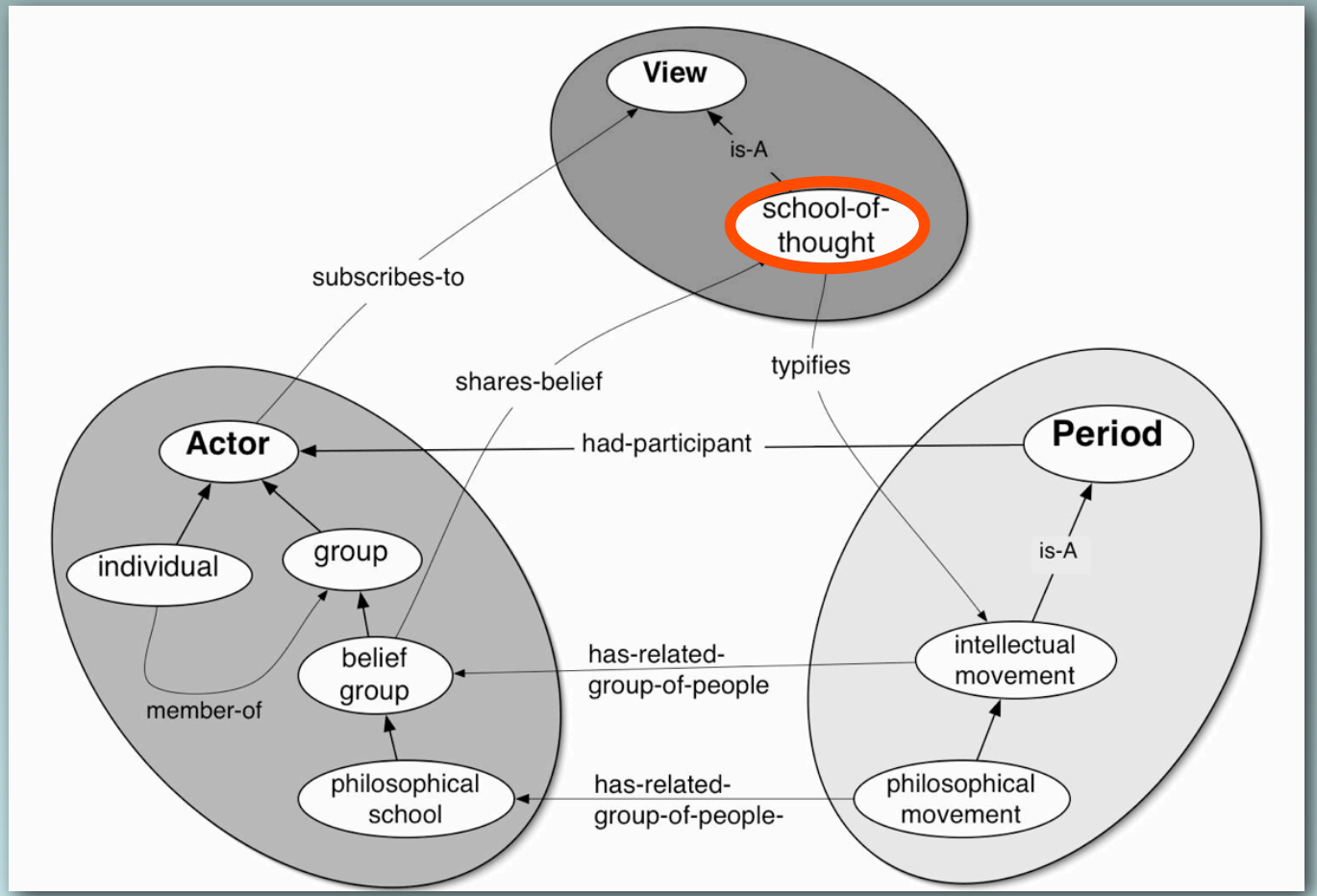
Ex. I: how many senses of rationalism?

*"This theory is clearly a new and re-shaped **rationalism**"*
*"Descartes was one of the founders of modern **rationalism**"*
*"Throughout history, the attacks of **rationalism** against empiricism has diminished"*



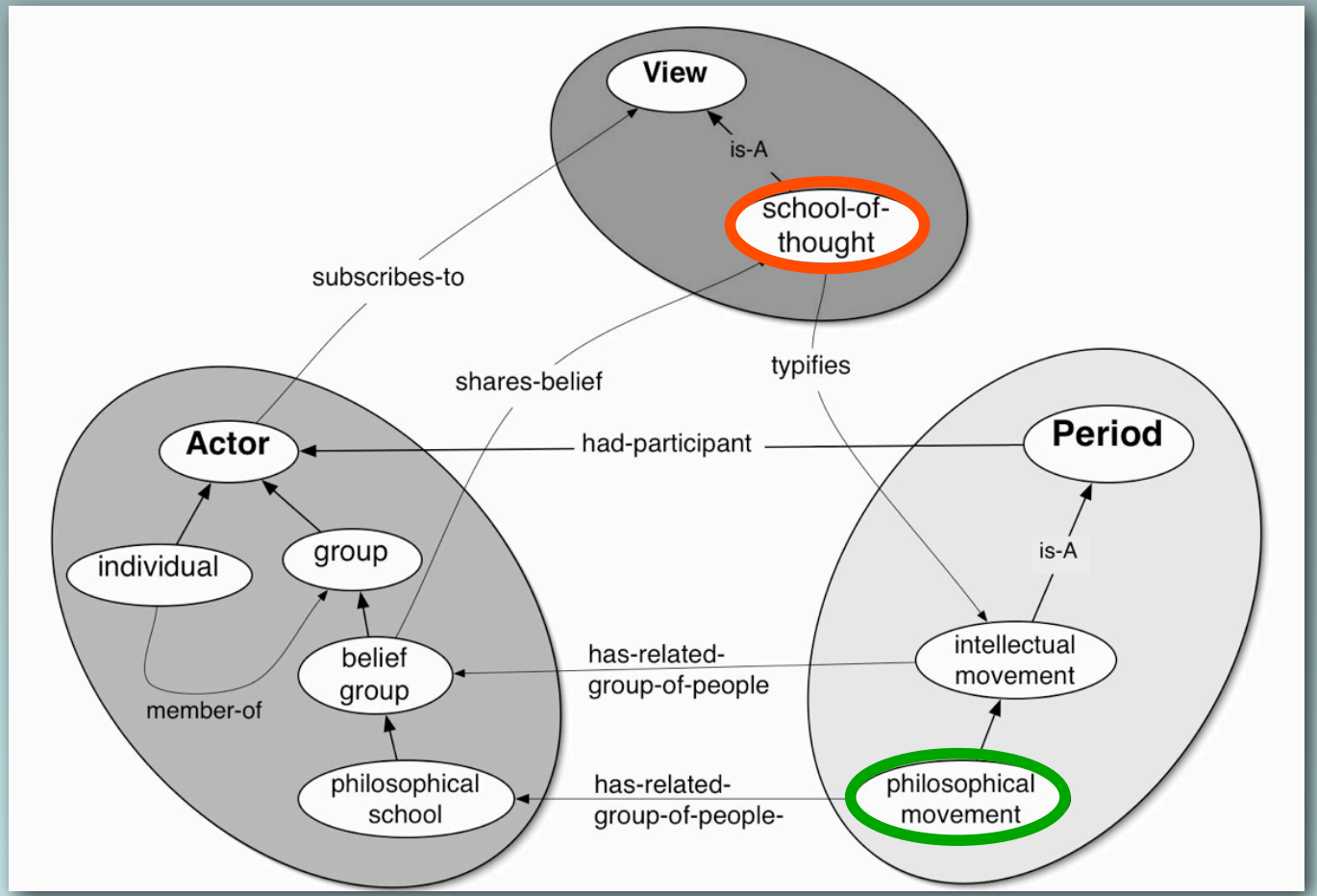
Ex. I: how many senses of rationalism?

*"This theory is clearly a new and re-shaped **rationalism**"*
"Descartes was one of the founders of modern rationalism"
"Throughout history, the attacks of rationalism against empiricism has diminished"



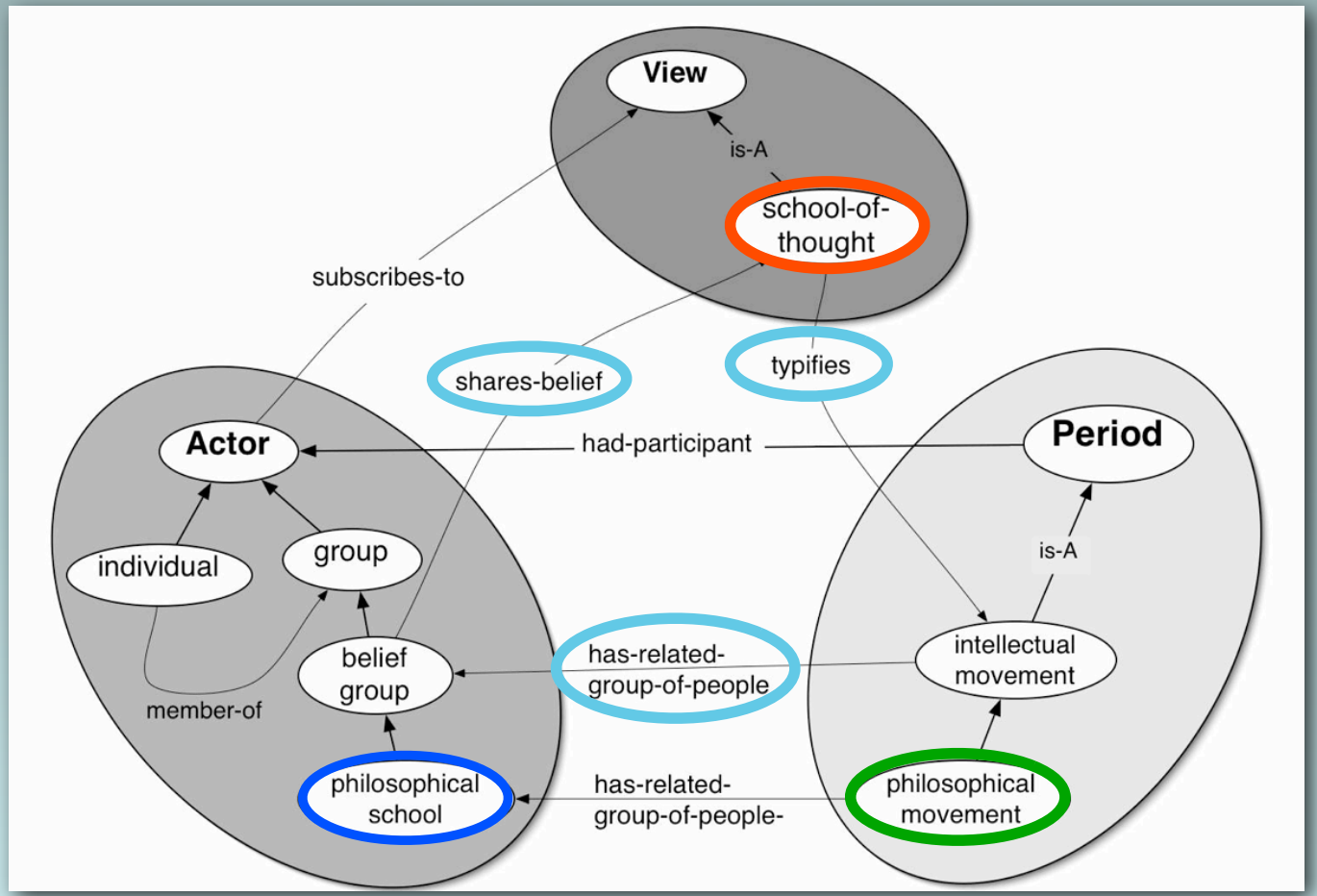
Ex. I: how many senses of rationalism?

*"This theory is clearly a new and re-shaped **rationalism**"*
*"Descartes was one of the founders of modern **rationalism**"*
"Throughout history, the attacks of rationalism against empiricism has diminished"



Ex. I: how many senses of rationalism?

*"This theory is clearly a new and re-shaped **rationalism**"*
*"Descartes was one of the founders of modern **rationalism**"*
*"Throughout history, the attacks of **rationalism** against empiricism has diminished"*

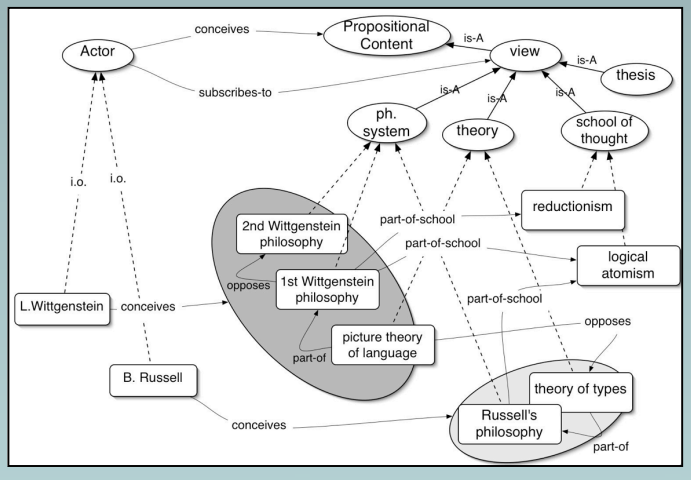


Pattern II : various granularities for theories

- Concerns only the “philosophical idea” branch of the ontology

-Taxonomies of viewpoints?

- Dolce: not suited for philosophical ideas
- Cyc: very convoluted, unusable
- Wordnet: flat hierarchy

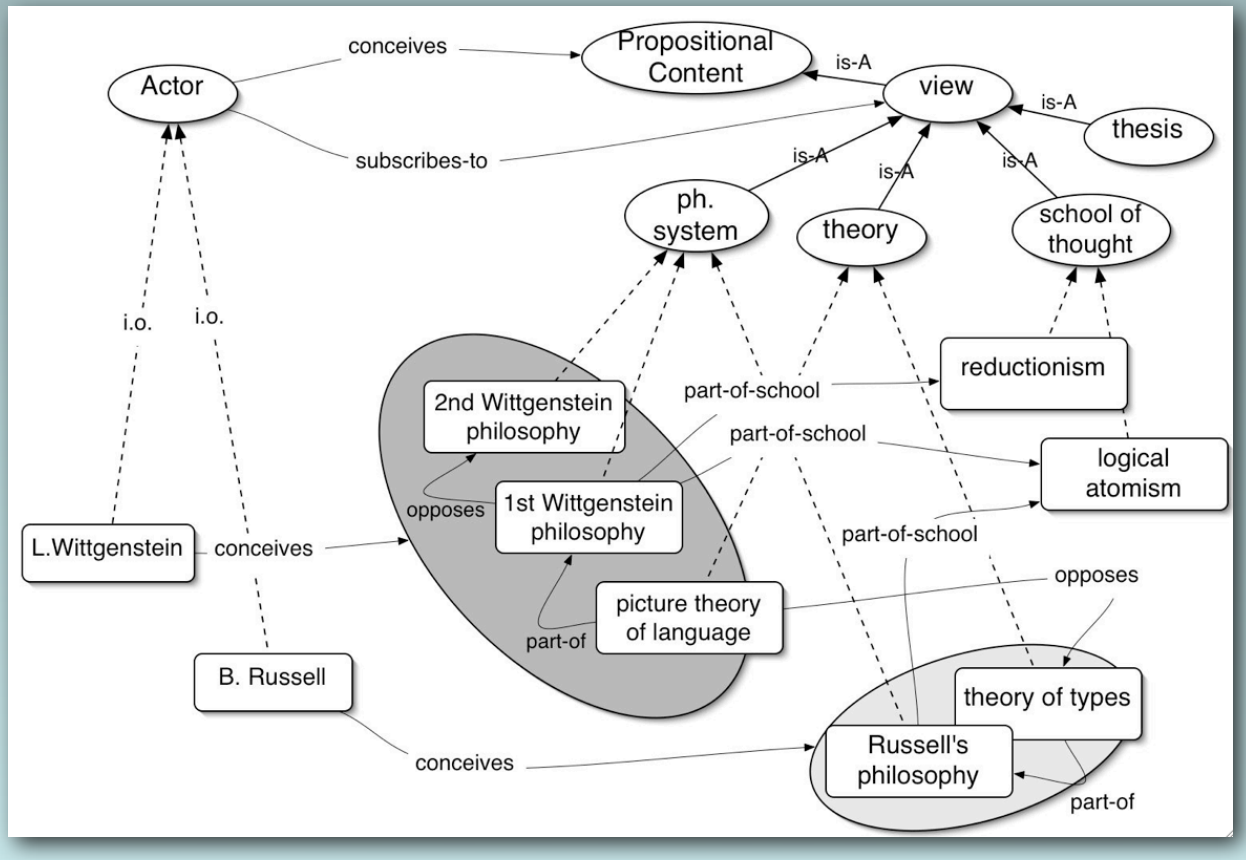


- Allows navigations which give viewpoints a “theoretical” context

  - e.g. views an author had, within a problem area, consistent with a school of thought etc..

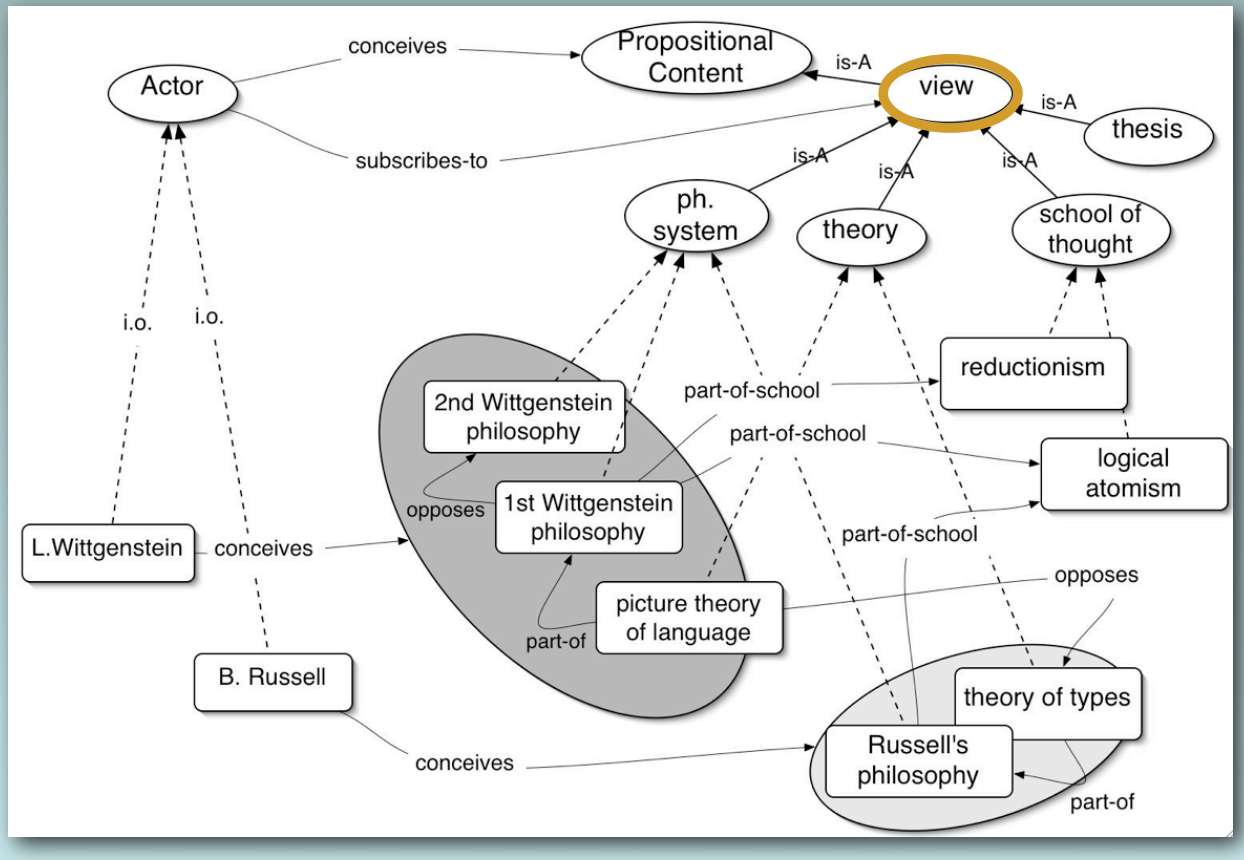
Ex. II: not all views are theories!

"Wittgenstein's philosophy, differently from Frege's one, deals also with problems typical of aesthetics"
"The 2nd Wittgenstein philosophy is much inspired from a kantianism, than from a logical positivism"
"Within the pictorial theory of language, Wittgenstein demonstrated that we can derive complex sentences from atomic ones"



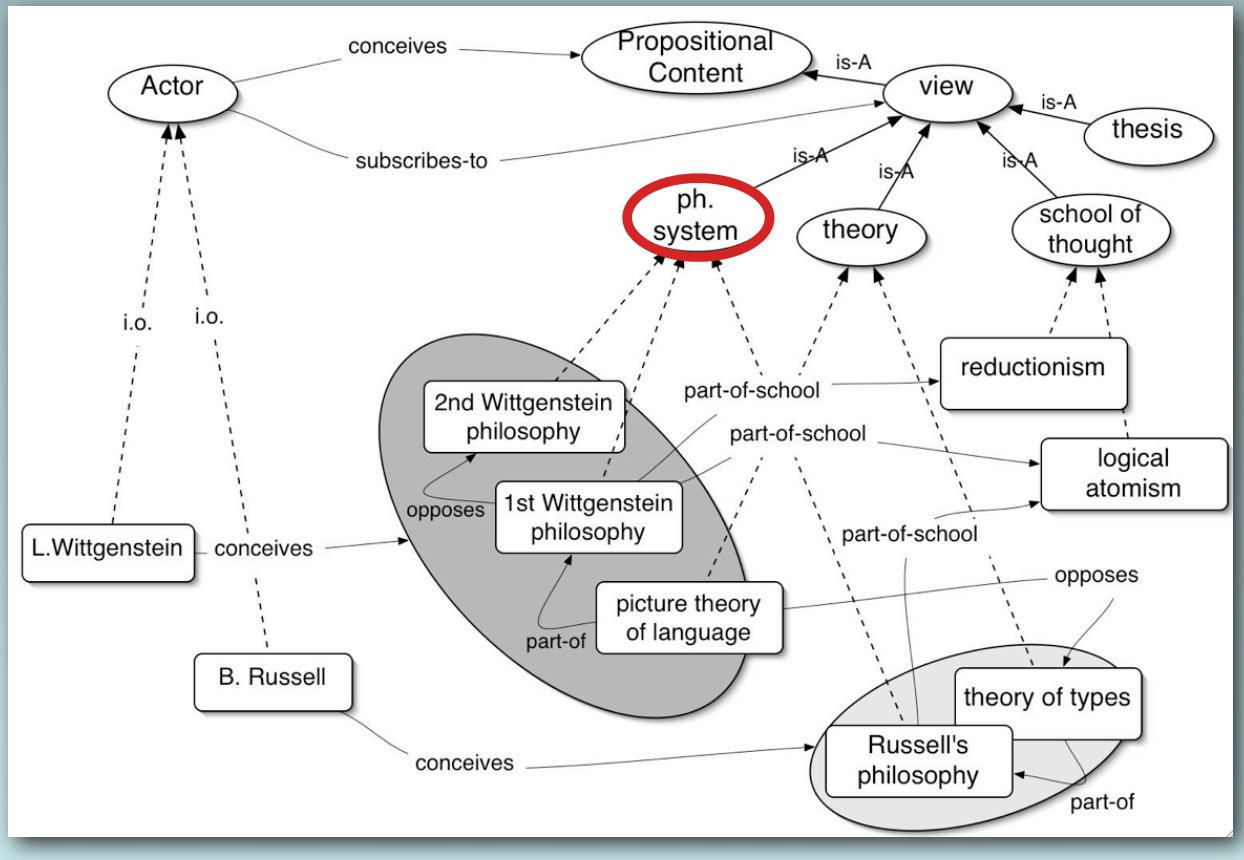
Ex II: not all views are theories!

*"**Wittgenstein's philosophy**, differently from **Frege's one**, deals also with problems typical of aesthetics"*
*"The **2nd Wittgenstein philosophy** is much inspired from a **kantianism**, than from a **logical positivism**"*
*"Within the **pictorial theory of language**, Wittgenstein demonstrated that **we can derive complex sentences from atomic ones**"*



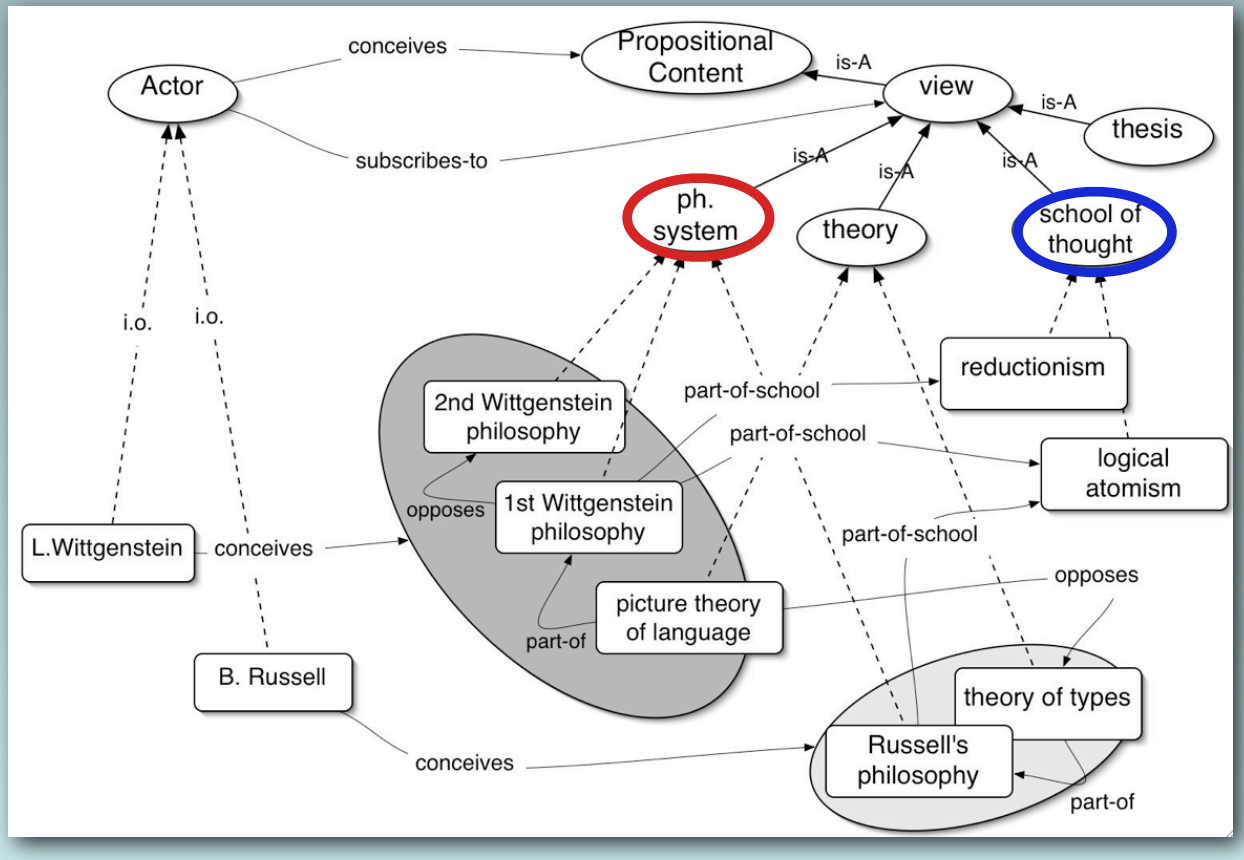
Ex. II: not all views are theories !

*"**Wittgenstein's philosophy**, differently from **Frege's one**, deals also with problems typical of aesthetics"*
*"The **2nd Wittgenstein philosophy** is much inspired from a **kantianism**, than from a **logical positivism**"*
*"Within the **pictorial theory of language**, Wittgenstein demonstrated that **we can derive complex sentences from atomic ones**"*



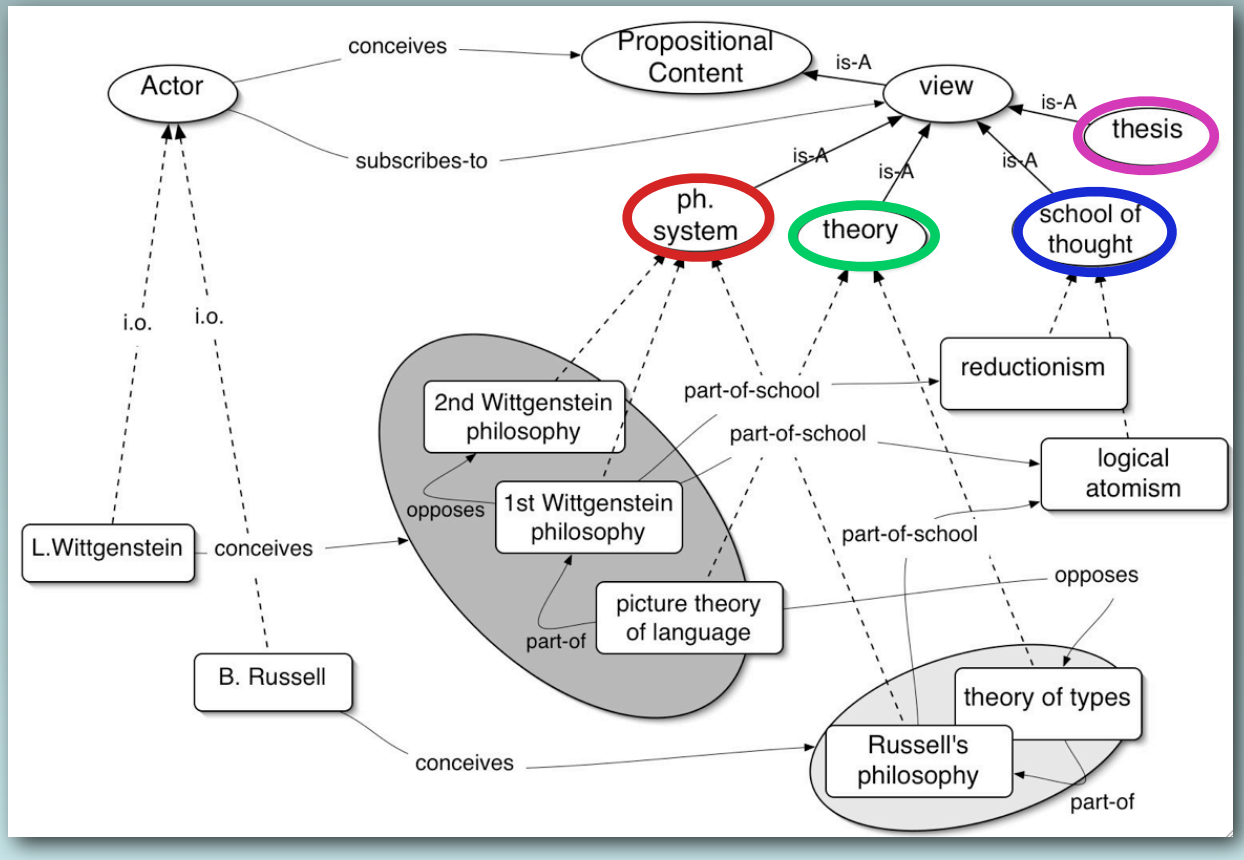
Ex. II: not all views are theories !

*"**Wittgenstein's philosophy**, differently from **Frege's one**, deals also with problems typical of aesthetics"*
*"The **2nd Wittgenstein philosophy** is much inspired from a **kantianism**, than from a **logical positivism**"*
*"Within the **pictorial theory of language**, Wittgenstein demonstrated that **we can derive complex sentences from atomic ones**"*



Ex. II: not all views are theories

*"**Wittgenstein's philosophy**, differently from **Frege's one**, deals also with problems typical of aesthetics"*
*"The **2nd Wittgenstein philosophy** is much inspired from a **kantianism**, than from a **logical positivism**"*
*"Within the **pictorial theory of language**, Wittgenstein demonstrated that **we can derive complex sentences from atomic ones**"*



Lesson learned: modeling patterns for navigation

- **Disclaimer**: different from “normal” ontology modeling patterns!
  - not focused on architectural issues
  - not involved in the ontology creation process
  - they are not prescriptive!

- Purpose: interpreting a (philosophical) concept/text, so to create ***applicable*** formal models for navigation
  - they open up new senses which can be used for exploring a subject domain

- **Strategy**: taking advantage of natural language ambiguities, overlapping word senses, hidden categories in language

example #2 ..

modeling citations

Modeling citations: background

- It’s very much a work in progress...

  - work started by Matteo Romanello @ Kings College in 2010
  - purpose: create a framework for text mining of citations in classics / ‘ancient’ references management
  - ontology as the backbone of the system

- Other projects looking at citations

  Shotton. **CiTO**, the Citation Typing Ontology, (2009)

  D'Arcus and Giasson. **BIBO**: Bibliographic Ontology (2009)

  Kruk et al. **MarcOnt**: integration ontology for bibliographic description formats. (2005)

  Gruber. The Bibliographic Data Ontology (1992)

The classic way to model citations

```
<http://example1.com/citingwork>
  cito:cites <http://example2.com/citedwork> ;
  cito:inTextCitationFrequency [
    a cito:InTextCitationCount ;
    cito:inTextCountValue "10"^^xsd:integer ;
    cito:inTextCitationTarget
          <http://example2.com/citedwork> ;
    ] ; .
<http://example2.com/citedwork>
   cito:isCitedBy <http://example1.com/citingwork> ;
   cito:globalCitationFrequency [
     a cito:GlobalCitationCount ;
     cito:globalCountValue "206"^^xsd:integer ;
    cito:globalCountSource
          <http://scholar.google.com>;
    cito:globalCountDate "2009-03-11"^^xsd:date ;
    ] ; .
```

Shotton. CiTO, the Citation Typing Ontology, (2009)

Textual Transmission History

Smith, J. Edward Sherburne's *Letters to Dudley Carleton*, 1950, XII

[...] The Prentizes on shrove Tewsday last, to the nomber of 3. or .4000 comitted extreame insolencies; part of this nomber, taking their course for Wapping, did there pull downe to the grownd 4 houses, spoiled all the goods therein, defaced many others, & a Justice of Peace coming to appease them, while he was reading a Proclamacion, had his head broken with a brick batt. Th' other part, making for Drury Lane, where lately a newe Playhouse is erected, they besett the house round, broke in, wounded divers of the players, broke open their trunckes, & whatt apparrell, bookes, or other things they found, they burnt & cutt in peeces; [...]

Gerald Eades Bentley (1968), *The Jacobean and Caroline Stage*, vol 6 Oxford:Oxford University Press, 54-6

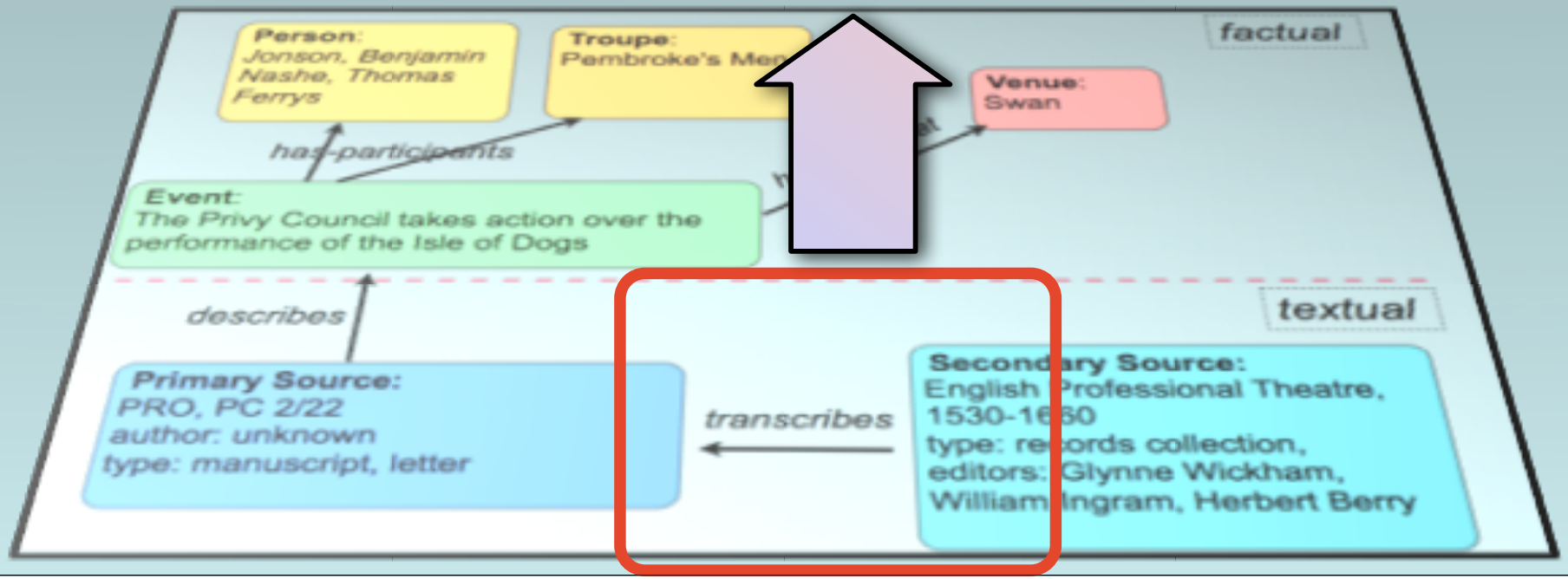
[...] The most detailed account was transcribed by Halliwell-Phillipps from a letter among the State Papers in the Public Record Office. The letter was written by Edward Sherburne and is dated '8th Marche, 161[illegible]' (Sherburne, 1616, VI):

The Prentizes on shrove Tewsday last, to the number of 3. or .4000 comitted extreme insolencies; part of this nomber, taking their course for Wapping, did there pull downe to the ground four houses, spoiled all the goods therein, defaced many others, and a Justice of Peace coming to appease them [...]

Textual Transmission History in EMLoT

EMLoT Transcription Record:

Authored by: Tanya Hagen
Primary Source Transcribed: PRO, PC 2/22
Location of Transcription (in Sec.Source): pag. 120, vol. 2
Primary Source as Cited in Sec.Source: PRO, PC 2/22/m.724d.
Transcription Notes: excerpted; parenthetical date(s); parenthetical note(s); spelling and punctuation standardized



Further evidence: Canonical Citations

Classics: three levels of information objects

Primary sources: inscriptions, papiry, manuscripts

Secondary sources: modern editions of works by ancient authors (text established by editors on the basis of primary sources)

Tertiary sources: works by modern authors (i.e. monographs, commentaries, journal papers, etc.)

Citation of resources

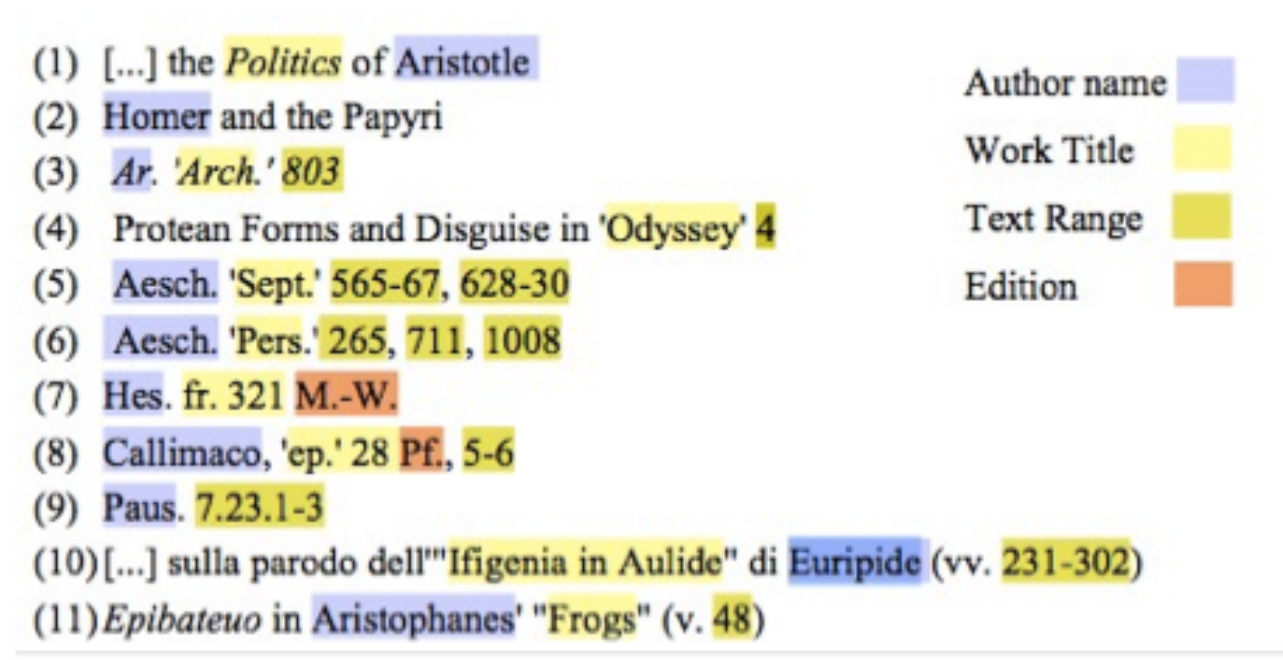
Different citation STYLES at the level of scholarly communities (i.e. journals, schools, personal choices, etc.)

Different styles need to be modeled on CANONICAL forms (inspired by secondary sources)

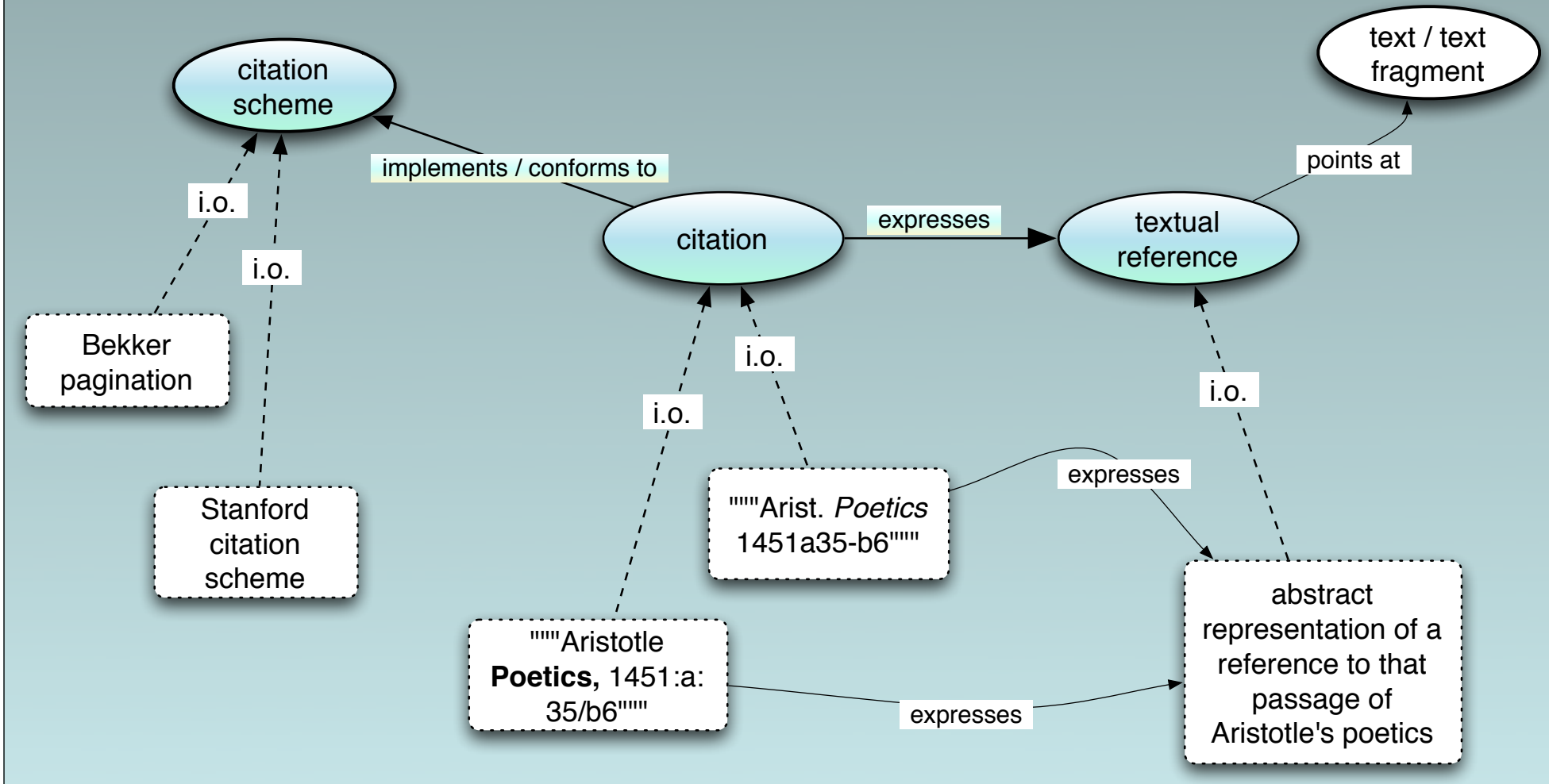
Canonical Citations: coordinate systems for ancient works

- what is a canonical reference?

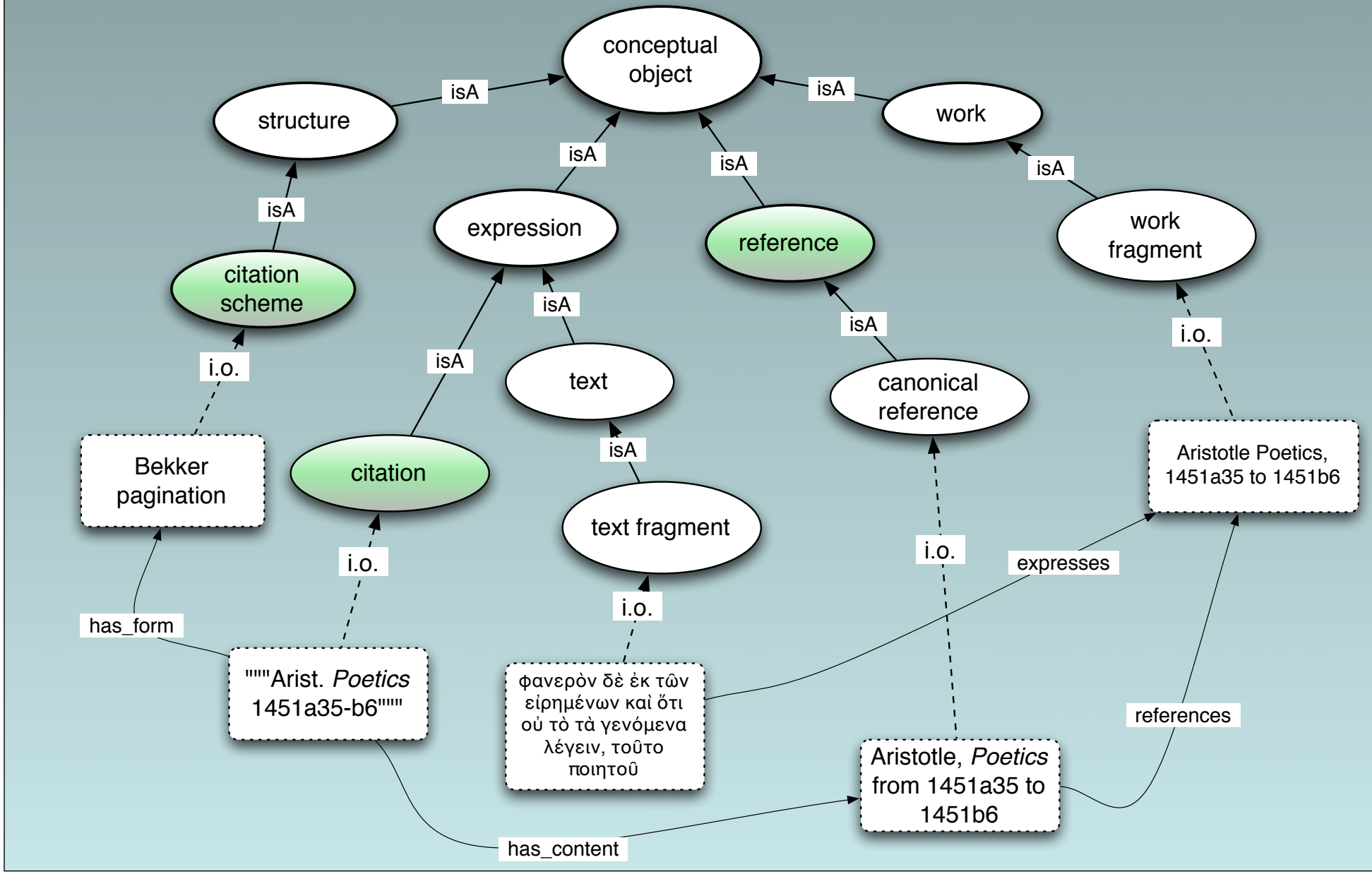
- encodes the ‘way to find a passage’, using a logical scheme (vs a physical one) agreed upon by the community



Citations as first class objects?



Early attempts to model Canonical R.



Wrapping up

The Digital Humanities

- rapidly evolving field(s)
- data integration solutions are strongly needed

Modeling humanities subjects can't always be reduced to similar models in scientific ones

- eg modeling philosophical ideas
- eg modeling citations *in* the humanities

There's more: modeling the 'discourse' of humanities subjects

- solutions already exist? Please let me know!
- future work?

. . . thank you . . .

Pasin and Motta. Ontological Requirements for Annotation and Navigation of Philosophical Resources. Synthese (2009)

Pasin and Ciula. Laying the conceptual foundations for data integration in the humanities. Digital Humanities 2009 (2009)

M. Romanello, M. Pasin. An Ontological View of Canonical Citations. Digital Humanities 2011 , Stanford, USA, Jun 2011. [forthcoming]

Romanello. Rethinking Critical Editions of Fragmentary Texts. ELPUB 2009

http://staff.cch.kcl.ac.uk/~mpasin/