Ontology-Based Observation Modeling: Focus on Literary Scholarship

Emilio M. Sanfilippo*
emilio.sanfilippo@cnr.it

*CNR ISTC Laboratory for Applied Ontology, Trento

14 October 2026
7th SCOR online seminar

Introduction

To represent the **source dependency** of information in ontologies.

Examples:

- John weights 70kg at t *according to balance* b
- Pottery p was manufactured in the 3rd century BC *according to device* d

But also, e.g.:

- John weights 68kg at t *according to balance* b'
- Pottery p was *not* manufactured in the 3rd century BC *according to device* d'

$\Rightarrow$ Representation of the *epistemic* dimension of information

Application contexts: Track and reason about data from different sources (perspectives), from industry and science to the humanities.

Application context

"KGs are central infrastructures for semantic interoperability, data integration, and structured knowledge representation across scientific, cultural, and industrial domains. However, many KG formalisms and engineering practices still treat statements as if their validity were independent of the conditions under which they are produced, interpreted, accepted, revised, or queried. *This assumption becomes limiting in domains where knowledge is contextual, evolving, contested, or perspectival, and where statements may depend on provenance, temporal scope, epistemic status, methodological assumptions, community viewpoints, or changing conceptual frameworks*" (emphasis is ours).

From the presentation of the Multi-dimensional Knowledge Graphs (MKG), 25 October, co-located with ISWC 2026 – the International Semantic Web Conference, Bari, Italy.

Computational Literary Studies (CLS)

A research field within the Digital Humanities:

- Application of computational methods – especially machine learning and natural language processing – to analyze literary corpora to support scholarship research

Focus on (at least):

1. Explicit intended meanings of the adopted concepts
2. Consensus-driven knowledge models
3. Transparency of methods for analysis
4. Replicability of results

See: Simone Rebora, *Computational Literary Studies: Theory and Methods*, Routledge, 2026

CLS meets Applied Ontology (AO)

Ontology-based approach to literary knowledge and data:

1. To support the *analysis* and *formal representation* of literary concepts for the sake of clarity and understanding;
2. To support the use of *consensus-driven*, *transparent* and *replicable* methods for literary scholarship (e.g., dataset, vocabularies, inference mechanisms, etc.)

Contribution

Ontology (in first-order logic and OWL):

⇒ To represent *interpretative data*, i.e., what scholars claim (data extracted from literary scholars' texts)

⇒ To reason over such data (e.g., for comparing claims)

Some requirements for the ontology:

- *What* is claimed
- *How* it is claimed (assertion, rejection, etc.)
- *Chains* of claims

Need to take into account:

- Claims that are *incompatible*
- Claims at *different levels* of detail

Publications

This talk is mainly based on:

- Sanfilippo, E.M., Masolo, C., Mosca, A., Tomazzoli, G. (2026). Modeling and Reasoning over Observations: An Ontology for Literary Criticism. Proceedings of CCLS 2026, Potsdam, May 28-29. Conference preprint (journal paper in publication)
- Masolo, C., Sanfilippo, E.M., Bottazzi, E., Ferrario, R., Mosca, A., Vilardo, M.M. (2025). An observational approach to representing interpretation. Applied Ontology, vol. 20, n. 4, 2025

Further work, aligned to the **DOLCE** foundational ontology, is presented in:

- Masolo, C., Porello, D., Borgo, S., Ferrario, R., Mosca, A., Sanfilippo, E.M., Stufano, R., Terkaj, W. (2026). Adding an Epistemological Layer to DOLCE. Proceedings of the 16th International Conference (FOIS 2026), vol. 429, IOS Press.

Observations

Classification of domain entities through properties – using different sorts of procedures (empirical, computational, discoursive, etc).

Examples (non-exhaustive):

- **Scholarly observation**: the similarity between two literary characters; attribution of genre; attribution of authorship, etc., through scholarly argumentation (but also computational analysis)
- **Empirical observation**: an artefact dated to the 3rd cent. b.C. through carbon-14 measurement

Main taxonomy

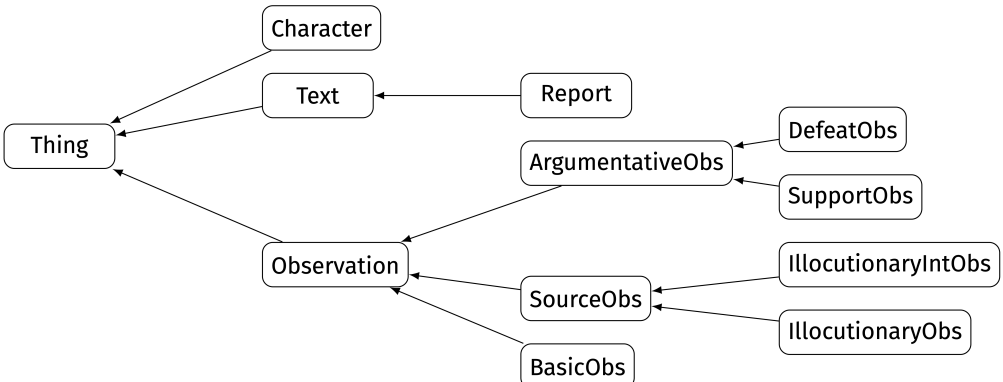


Figure: From Sanfilippo et al. (2026); boxes stand for classes; arrows for subsumption relations between classes

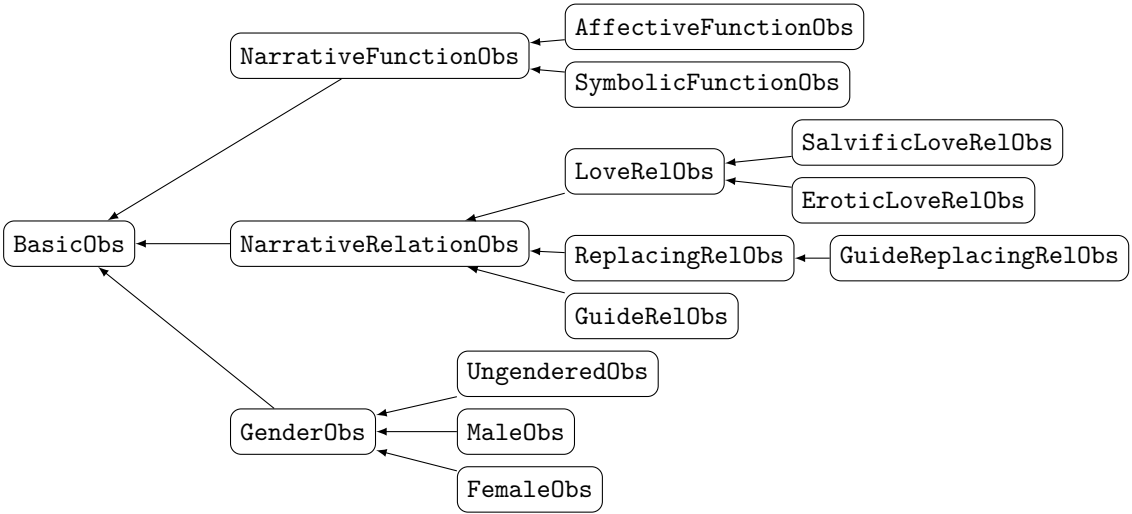
Observation types

Three types of observations:

1. **Basic** observations: domain specific, i.e., the core of *what* is claimed;
2. **Source** observations: make explicit the (asserting, rejecting) source of an observation:
   - 2.1 Illocutionary observations
   - 2.2 *Interpretative* illocutionary observations
3. **Argumentative** observations: subjective forms of inferences:
   - 3.1 Support
   - 3.2 Defeat

$\Rightarrow$ Basic and Argumentative observations have to fall in the scope of Source observations: the latter provide explicit information on their sources

Basic observations (example)



Basic Observation

Development in **three steps**:

- Step 1 Identify and taxonomically structure a *finite* set $\mathcal{P}$ of *unary* predicates, standing for different kinds of observations;
- Step 2 For each observation kind $P \in \mathcal{P}$, establish the number and type of entities involved in P's instances, i.e., the *arguments* of P;
- Step 3 Represent, when possible, relations between the $\mathcal{P}$s, e.g., *incoherence* and *correlation*.

$\Rightarrow$ Strong collaboration with domain experts

Example (Step 1 and Step 2)

Step 1 (taxonomy):

1. $\texttt{NarrativeRelationObs} \sqcup \texttt{NarrativeFunctionObs} \sqsubseteq \texttt{BasicObs}$
2. $\texttt{NarrativeRelationObs} \sqcap \texttt{NarrativeFunctionObs} \sqsubseteq \bot$

Step 2 (fix the arguments):

3. $\texttt{LoveRelObs} \sqsubseteq \exists^{=3}\, \texttt{hasArgument}$
4. $\texttt{LoveRelObs} \sqsubseteq \exists^{=1}\, \texttt{hasLover.Character}$
5. $\texttt{LoveRelObs} \sqsubseteq \exists^{=1}\, \texttt{hasBeloved.Character}$
6. $\texttt{LoveRelObs} \sqsubseteq \exists^{=1}\, \texttt{inText.Text}$

with `hasLover`, `hasBeloved`, and `inText` as subproperties of `hasArgument`.

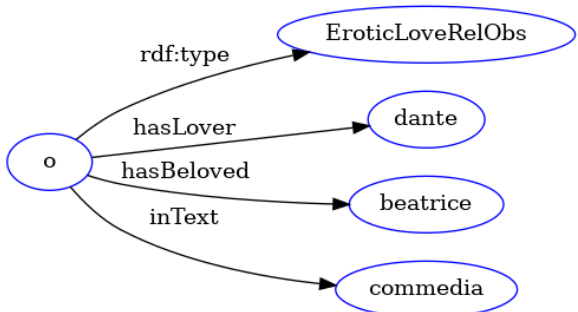
Example (cont'd)

With $\texttt{EroticLoveRelObs} \sqsubseteq \texttt{LoveRelObs}$:

- $\texttt{EroticLoveRelObs}(o) \sqcap \texttt{hasLover}(o, dante) \sqcap \texttt{hasBeloved}(o, beatrice) \sqcap \texttt{inText}(o, commedia)$

i.e., the (literary) character of Dante has (is observed as having) a form of erotic love for the character of Beatrice in the context of the *Commedia*.

Example (in RDF)



Namespaces:
https://www.loa.istc.cnr.it/cs#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#

Figure: RDF graph for the observation `EroticLoveRelObs`(o) (previous slide)

Example (identity rule)

Identity rule (in SWRL; `SameAs` element of the language):

7. $\texttt{EroticLoveRelObs}(?x), \texttt{hasLover}(?x, ?z), \texttt{hasBeloved}(?x, ?v), \texttt{inText}(?x, ?t),$
   $\texttt{EroticLoveRelObs}(?y), \texttt{hasLover}(?y, ?z), \texttt{hasBeloved}(?y, ?v), \texttt{inText}(?y, ?t)$
   $\rightarrow \texttt{SameAs}(?x, ?y)$

$\Rightarrow$ Similar identity rules apply to the leaf-nodes of the taxonomy in slide n.11.

Example (Step 3)

Step 3 (relations):

8. $\texttt{MaleObs}(?x), \texttt{hasArgument}(?x, ?z), \texttt{UngenderedObs}(?y), \texttt{hasArgument}(?y, ?z) \rightarrow \texttt{INC}(?x, ?y)$

Incoherence (INC) means that the observations are conflictual in their claims:

- $\texttt{MaleObs}(o_1) \sqcap \texttt{hasArgument}(o_1, dante)$
- $\texttt{UngenderedObs}(o_2) \sqcap \texttt{hasArgument}(o_2, dante)$
- $\texttt{INC}(o_1, o_2)$, reasoning through the rule above

Notice the difference with the use of "non-observational predicates", e.g.:

- $\texttt{Male} \sqcap \texttt{Ungendered} \sqsubseteq \bot$
- $\texttt{Male}(dante)$, $\texttt{Ungendered}(dante)$

  $\Rightarrow$ it generates an inconsistency in the knowledge base

Source Observations

Make explicit the **source** of observations:

- **Illocutionary observations**: make explicit what an interpreter beliefs with respect to a text ($\rightarrow$ without necessary access to the text);
- **Interpretative illocutionary observations**: make explicit how an interpreter *interprets* a text ($\rightarrow$ access to texts is necessary).

They can be either **assert** or **deny** observations.

Interpretative Illocutionary Observations

Source observations that express agents' *interpretation* of texts:

9. $\texttt{IllocutionaryIntObs} \sqsubseteq \exists^{=3}\, \texttt{hasArgument}$
10. $\texttt{IllocutionaryIntObs} \sqsubseteq \exists^{=1}\, \texttt{hasSource.Text}$
11. $\texttt{IllocutionaryIntObs} \sqsubseteq \exists^{=1}\, \texttt{hasInterpretedText.Text}$
12. $\texttt{IllocutionaryIntObs} \sqsubseteq \exists^{=1}\, \texttt{hasObservation.Observation}$

$\Rightarrow$ `IntAssertObs` and `IntRejectObs` are disjoint subclasses of `IllocutionaryIntObs`

$\Rightarrow$ **Intended meaning**: (the author of) the source asserts or reject an observation by interpreting a text

Interpretative Illocutionary Observations (example)

Example:

- $\texttt{IntAssertObs}(o_3) \sqcap \texttt{hasSource}(o_3, r_1) \sqcap \texttt{hasInterpretedText}(o_3, prt) \sqcap \texttt{hasObservation}(o_3, o) \sqcap \texttt{EroticLoveRelObs}(o) \sqcap \texttt{hasLover}(o, dante) \sqcap \texttt{hasBeloved}(o, beatrice) \sqcap \texttt{inText}(o, commedia)$

$\Rightarrow$ by interpreting the text prt, (the author of) report r_1 asserts that the character of Dante ...

Interpretative Illocutionary Observations (in RDF)

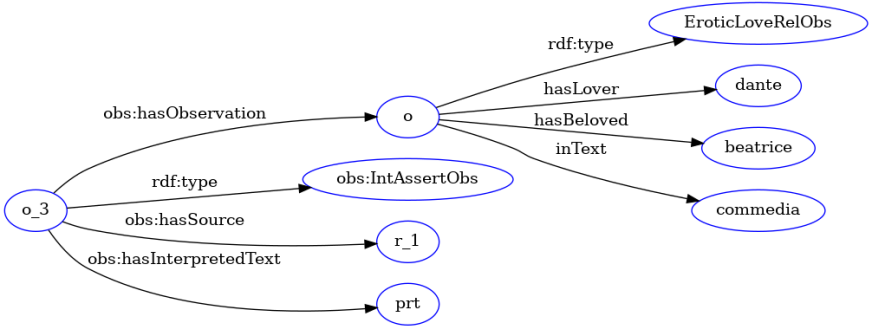


Figure: RDF graph for `IntAssertObs`(o_3) (previous slide)

Argumentative Observations

Observations that (partially) represent the argument put forward by an agent to support or defeat another observation.

Defeat:

13. $\texttt{DefeatObs} \sqsubseteq \exists^{=2}\texttt{hasArgument}$
14. $\texttt{DefeatObs} \sqsubseteq \exists^{=1}\texttt{hasDefeatingObs.Observation}$
15. $\texttt{DefeatObs} \sqsubseteq \exists^{=1}\texttt{hasDefeatedObs.Observation}$

$\Rightarrow$ Argumentative observations can be *asserted*, *denied*, etc.

- $\mathbf{iasr}(prt, par,$
  $\mathbf{def}(\mathbf{eroticLove}(dante, beatrice, par), \mathbf{salvificLove}(dante, beatrice, par)))$
  [syntactic shortcut; see Sanfilippo et al. (2026)]

Criteria for Analysis

Disputability of observations (syntactic shortcut):

– $\mathbf{iasr}(?r, ?t_1^c, \mathbf{iasr}(?t_1^c, ?t^l, ?o_1)), \mathbf{iasr}(?r, ?t_2^c, \mathbf{iasr}(?t_2^c, ?t^l, ?o_2)),$
$\mathbf{iasr}(?r, ?t_2^c, \mathbf{iasr}(?t_2^c, ?t^l, \mathbf{def}(?o_2, ?o_1))) \rightarrow \texttt{Disputable}(?o_1)$

Example (**simplified**; in shortcuts):

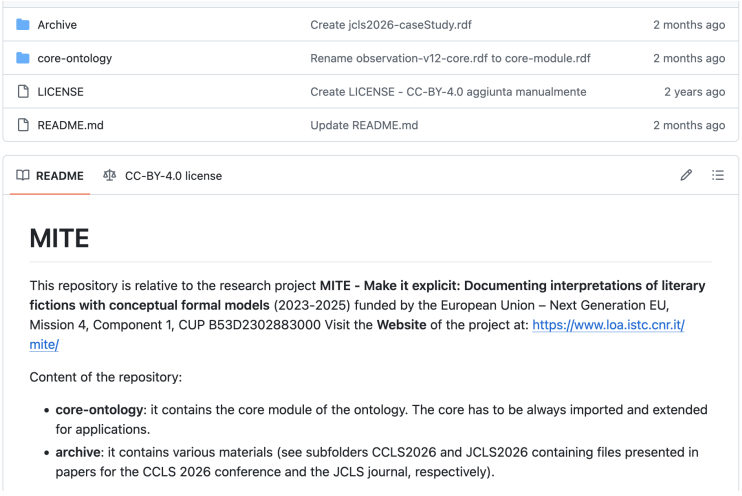
- $\mathbf{iasr}(psk, par, \mathbf{salvific}(dante, beatrice, par))$
- $\mathbf{iasr}(prt, par, \mathbf{erotic}(dante, beatrice, par))$
- $\mathbf{iasr}(prt, par,$
$\mathbf{def}(\mathbf{eroticLove}(dante, beatrice, par), \mathbf{salvificLove}(dante, beatrice, par)))$

$\Rightarrow$ by reasoning, $\mathbf{salvific}(dante, beatrice, par)$ is classified as `Disputable`

(see Sanfilippo et al. (2026))

Semantic Web Ontology

Available online: https://github.com/appliedontolab/MITE



Archive	Create jcls2026-caseStudy.rdf	2 months ago
core-ontology	Rename observation-v12-core.rdf to core-module.rdf	2 months ago
LICENSE	Create LICENSE - CC-BY-4.0 aggiunta manualmente	2 years ago
README.md	Update README.md	2 months ago

README CC-BY-4.0 license

MITE

This repository is relative to the research project **MITE - Make it explicit: Documenting interpretations of literary fictions with conceptual formal models** (2023-2025) funded by the European Union – Next Generation EU, Mission 4, Component 1, CUP B53D2302883000 Visit the **Website** of the project at: https://www.loa.istc.cnr.it/mite/

Content of the repository:

- **core-ontology**: it contains the core module of the ontology. The core has to be always imported and extended for applications.
- **archive**: it contains various materials (see subfolders CCLS2026 and JCLS2026 containing files presented in papers for the CCLS 2026 conference and the JCLS journal, respectively).

Conclusions

In a **nutshell**, ontology-based approach:

- To represent and reason over source-dependent data (general context)
- Tuned to interpretative data in literary criticism (specific context)

Core benefits:

1. Capable of dealing with multiple, perspectival claims (pluralism)
2. Capable of representing and reasoning over (chains of) incompatible claims
3. Capable of capturing shared as well as subjective forms of inferences

Conclusions (CLS specific)

An ontology-based approach provides a fundamental support for, e.g.,:

- Reducing ambiguities in the arguments of interpreters
- Making explicit the argumentative assumptions of the various positions on a given subject
- Qualifying the proximity or distance between the advocated theses
- Comparing the contents of the various positions with each other
- Identifying the discursive elements that generate convergence or divergence between them
- ...

Grazie!

Acknowledgements:
Group work with (alphabetic): Stefano Borgo, Roberta Ferrario, Claudio Masolo, Alessandro Mosca, Daniele Porello, Gaia Tomazzoli

Some this research was conducted in the scope of the project: MITE – Make it explicit: Documenting interpretations of literary fictions with conceptual formal models (2023-2025) funded by the European Union – Next Generation EU, Mission 4, Component 1, CUP B53D23028830001

https://www.loa.istc.cnr.it/mite/ (see Dissemination page)

Contact: emilio.sanfilippo@cnr.it