The Hypermedia Dante Network Project

Gaia Tomazzoli[1], Leyla M. G. Livraghi[2], Daniele Metilli[3], Nicolò Pratelli[4], Valentina Bartalesi[5]

[1] Università di Pisa – gaia.tomazzoli [at] fileli.unipi.it
[2] Università di Pisa – leyla.livraghi [at] fileli.unipi.it
[3] ISTI-CNR e Università di Pisa – daniele.metilli [at] isti.cnr.it
[4] ISTI-CNRUniversità di Pisa – nicolo.pratelli [at] isti.cnr.it
[5] ISTI-CNRUniversità di Pisa – valentina.bartalesi [at] isti.cnr.it

ABSTRACT

In this paper, we present the Hypermedia Dante Network (HDN) project. First, we briefly introduce the relevant state of the art on Dante's commentaries and their digital representation, and we outline the project goals. In the main section, we present the core features of the HDN ontology, an evolution of the DanteSources ontology that aims at representing knowledge about Dante's primary sources as they are identified by a vast range of commentaries. Then, we describe the tool that has been developed to process Dante's commentaries and populate the HDN ontology. Finally, we comment on the project's usability and possible outcomes for both scholars and common users.

KEYWORDS

Dante Alighieri, digital libraries, intertextuality, semantic web, commentaries

TALK

1   INTRODUCTION

Dante's *Commedia* has elicited detailed and systematic exegesis since its very first reception. Given the poem's encyclopedic features, one of the main tasks of commentaries has always been to identify Dante's vast array of sources. This incredibly rich body of knowledge is, therefore dispersed in dozens of commentaries; in the past twenty years, Dante scholars have worked to make such knowledge easily available through an impressive enterprise of digitization. One of the main outcomes of said enterprise is the Dartmouth Dante Project (DDP) [5], an open-access database where readers and scholars can search and browse the full text of more than 70 commentaries on Dante's *Commedia*, written in Italian, Latin, and English from 14th century to the present day; our work will initially rely on 50 of said commentaries. However, the DDP is only searchable through textual queries, making it difficult for users to retrieve meaningful information.

In recent years, ISTI-CNR and the University of Pisa have joined forces in the first project to ever apply Semantic Web standards to Dante studies. Said project, called DanteSources [1], developed a formal ontology for representing knowledge about primary sources in Dante's so-called minor works. The users of DanteSources can perform structured queries on those sources of Dante's *Vita nova*, *De vulgari eloquentia*, *Convivio*, *Monarchia*, and *Rime* that have been identified by recent commentaries. DanteSources distinguishes between explicit citations, strict citations, and generic citations. Users can interrogate the knowledge base by primary source, by source author, by thematic area, or by type of reference.

Hypermedia Dante Network (HDN) [6] aims to complement and expand on DanteSources, which is limited to the so-called minor works, by gathering and representing the substantial amount of additional information about Dante's primary sources contained in commentaries on Dante's *Commedia*. Confronted with a significantly more complex text and an incomparably wider tradition of commentaries, HDN team members could not just adapt the existing framework, but rather had to rethink the entire DanteSources ontology.

2   THE HDN ONTOLOGY

The HDN ontology is an evolution of the one developed in DanteSources [1], integrated with the Narrative Ontology [2], and having as reference ontologies the CIDOC CRM [3] and its extension FRBRoo [4] (including its in-progress reformulation, LRMoo [7]). The fundamental concept described by the ontology is the reference. A reference is an exegetic relation between a text and another entity that is identified by a commentary. A reference can be expressed by the following statement:

A says that knowledge about B can be enriched through C

where A, the *source* of the reference, is a fragment of text that asserts the reference, and usually comes from a commentary; B, the *subject* of the reference, is a fragment of text that is clarified by the reference – and in our case study it occurs within a work of Dante's; C is the *object* of the reference, i.e., the entity that is being referred to.

Within the class of references, we identify three distinct subclasses: external supports, *loci paralleli*, and citations. *External supports* are references in which the source identifies an object-text that proves useful to support its interpretation of the subject-text. *Loci paralleli* are references in which the source identifies, within an object-text, an idea, a character, or a stylistic feature that, for being used in a similar way as in the subject-text, helps to clarify it, without necessarily implying a direct derivation. *Citations* are references in which the source establishes that the subject-text directly derives from the object-text. These are clearly the most specific references, and they lie at the centre of our ontology.

We identified three characteristics of a reference: the citation *type*, its *content*, and the *relationship* between subject and object. These characteristics always apply to citations, but may not apply to external supports and *loci paralleli*. We also represent internal references, i.e., quotations to works by the same author of the subject-text. These kinds of references are simply modelled as citations in which the author(s) of the subject-text is the same as the author(s) of the object-text.

Characteristics of the reference

The ontology describes three types of citations, which were already identified in the DanteSources ontology: *explicit citations* (references explicitly made by Dante in his work), *strict citations* (references to a specific work identified by a scholar), *generic citations* (references to a concept put forward by a scholar).

The *content* of the citation is an explicit or implicit reference to a textual place. We have identified 3 types and 7 sub-types of content:

a   Textual correspondences: they are references to textual fragments identified by linguistic and stylistic coincidence. These are further divided into: 1. *image* (the source identifies within the object an image as a precedent for a description or conceptualization made by Dante in the subject), and 2. *stylistic feature* (the source identifies a linguistic and/or rhetorical precedent for an expression included in the subject-text).

b   Thematic correspondences: they are references to textual fragments identified by thematic coincidence. These are further divided into: 3. *character* (the source relates a character mentioned by Dante to a specific object-text), 4. *episode* (the source identifies a literary episode as an antecedent for Dante's subject-text), and 5. *topography* (the source believes that a literary memory acts behind Dante's mention of a place).

c   Conceptual correspondences: they are references to textual fragments identified by conceptual coincidence. These are further subdivided into: 6. *motif* (the source identifies within a specific object-text the precedent for the development of one of Dante's motifs), and 7. *theory* (the source identifies within the object-text a theory upon which Dante's text relies).

The relationship between the subject and the object of the citation is classified according to three categories: *correction/contradiction*, *extension/re-elaboration*, and *confirmation/homology*. These three types of relation are inspired by the corresponding properties that are defined in the CiTO ontology [8].

Implementation of the ontology

The conceptualization has been formalized in an ontology that has been developed as an extension of the CIDOC CRM and FRBRoo (and its in-progress reformulation LRMoo).

The commentaries to Dante's text are instances of the Commentary class, which is a subclass of F2 Expression. The subparts of the commentaries, including notes, are instances of E90 Symbolic Object. To highlight Dante's works, we introduced the Dante's Work class as a subclass of F2.

The ontology defines a class for each reference category, that is: Reference, External Support, Loci Paralleli, and Citation. Loci Paralleli and External Support are subclasses of Reference. Citation is a subclass of Loci Paralleli. Reference is defined as a subclass of E73 Information Object.

The taxonomy of the classes seen so far is shown in Fig. 1 (the ontology classes in light blue and the subclass relations in blue).
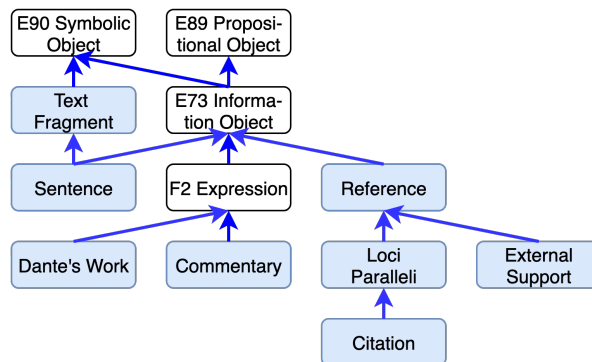

Figure 1: the main classes of the HDN ontology

The citation types and subject-to-object relationships are individuals, that is, instances of the E55 Type class. To distinguish them, the ontology includes two subclasses of E55: Citation Type, having as instances: genericCitation, strictCitation and explicitCitation; Subject-Object Relationship, having as instances correction, extension, and confirmation.

The subject of the Citation can vary significantly, therefore we consider it as an instance of E1 CRM Entity, the most general class of CRM.

An instance of Text Fragment is linked to the part of the work to which it belongs by the property R15i is Fragment Of. A part of the work (e.g., a *canto*) is linked to the part that includes it (e.g., a *cantica*) by property P148 is component of.

The references have as elements the source, the subject, and the object. The links between these elements and the references they belong to are represented by the following properties: hasRefSource links a reference to its source, and it is a subproperty of P129i is subject of; hasRefSubject links a reference to its subject, and it is a subproperty of P106 is composed of; hasRefObject links a reference to its object, and it is a subproperty of P129 is about.

Ontology Population Tool

In order to extract the knowledge contained in the commentaries on Dante's Commedia and represent it according to our ontological model, we are developing a tool that will be distributed to scholars to populate the ontology.

The commentaries that will be processed in the HDN project are a selection of the corpus digitized by the Dartmouth Dante Project [5]. The text of the commentaries was pre-processed to identify the notes and, for each note, the following elements: the body of the note, the fragment of Dante's text, the verse, the canto and the cantica.

Once the pre-processing has been performed, scholars will use the tool to populate the ontology with the information that cannot be extracted automatically, that is, the references found in each note and all their characteristics (type, function, content, etc.).

3 CONCLUSIONS AND OUTLOOK

The outcomes of Hypermedia Dante Network will be manifold.

First of all, our project will make meaningful information quickly and easily available to users despite the vastness and complexity of both Dante's *Commedia* and its commentaries. The knowledge base will be made available through a user-friendly web application, where different queries can be performed. For example, a scholar could be interested in retrieving information about the main sources for Dante's topography, or, on the other hand, about the most significant uses of a particular source. Not only will HDN provide the user with information about which authors and works are quoted by Dante and in what forms, but it will also distinguish the different ways in which commentaries have put forward sources to explain Dante's poem, thereby giving a more detailed account of what features of a given author or source were more important to the poet at a given phase of his reception. By doing so, it will help us deepen our understanding of both Dante's poem and the almost self-sufficient tradition of its commentaries.

Although such web application is mainly intended for scholarly use, it will also prove useful in familiarizing everyday readers with research on Dante's primary sources, and it will prompt new and unexpected explorations in the masterpiece of Italian literature. We intend to design a web application that is flexible, accessible, and usable by a wide public.

In the future, we plan to progressively enrich our knowledge base with new commentaries and new languages, in order to make knowledge on Dante's primary sources available to everyone. Moreover, we hope that our endeavour will inspire the development of intertextuality-based ontologies for other literary works, thereby extending the scope of our multifaceted approach to stylistic, linguistic, and thematic features of literary texts.

REFERENCES

1. Bartalesi, Valentina, Carlo Meghini, and Daniele Metilli. 'A Conceptualisation of Narratives and Its Expression in the CRM'. *International Journal of Metadata, Semantics and Ontologies* 12, no. 1 (2017): 35–46.

2. Bartalesi, Valentina, Carlo Meghini, Daniele Metilli, Mirko Tavoni, and Paola Andriani. 'A Web Application for Exploring Primary Sources: The DanteSources Case Study'. *Digital Scholarship in the Humanities* 33, no. 4 (2018): 705–723.

3. Doerr, Martin. 'The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata'. *AI Magazine* 24, no. 3 (2003): 75–75.

4. Doerr, Martin, Chryssoula Bekiari, Patrick LeBoeuf, and Bibliothèque nationale de France. 'FRBRoo, a Conceptual Model for Performing Arts'. In *2008 Annual Conference of CIDOC, Athens*, 15–18, 2008.

5. Hollander, Robert. 'The Princeton Dante Project'. *Humanist Studies & the Digital Age* 3, no. 1 (2013): 53–59.

6. Meghini, Carlo, Mirko Tavoni, and Michelangelo Zaccarello. 'Mapping the Knowledge of Dante Commentaries in the Digital Context: A Web Ontology Approach'. *Forthcoming* Romanic Review 111, no. 2 (2020).

7. Riva, Pat, and Maja Žumer. 'FRBRoo, the IFLA Library Reference Model, and Now LRMoo: A Circle of Development', 2017.

8. Shotton, David. 'CiTO, the Citation Typing Ontology'. *Journal of Biomedical Semantics* 1, no. S6 (2010).