Formal ontologies for narrative text analysis

Fabio Ciotti

Università di Roma Tor Vergata

AIUCD 2016 - Università Ca’ Foscari Venezia, Sept 7-9, 2016
Looking back

- In Digital Humanities world the attraction for novelty leads to the enthusiastic adoption of new methods and tools, without an adequate epistemological and methodological consideration.

- Sometimes, to make a step forward in scientific research we need to look back.

- My point is this: the evolution of semantic technologies gives us an apparatus of computational methods that can help to operationalize the semiotic/structuralist tradition and to adapt it to the context of the digital turn in literary studies.
Data mining algorithms in general are independent from the context (they can be applied indifferently to stock exchange transactions as to very large textual corpora). They individuate similarities and recurring patterns independently from the semantics of data. Humanities and literary data are heavily contextualized.

Text mining are agnostic toward the granularity of the data to which they are applied. Texts are only sequences of n-grams, and the probabilistic rules adopted to calculate the relevance of a given set of n-grams are completely independent from the fact that the units of analysis are individual coded characters, or linguistic tokens of greater extension.

If a very large textual set is composed of documents spread over a long period of time, diachronic variation of the form and usage of the language (both on the syntactic and semantic level) can invalidate purely quantitative and statistic measures.
Data in literary studies do not precede formal modeling; on the contrary, they are the product of modeling. It is very dubious to assume innocently a data set as the starting point of a meaningful analysis.

Meaning in literary texts is multi-layered, and some layers do not have direct lexicalization or they have a very complex and dispersed one (think to aspects of a narrative text at different abstraction level like anaphora, themes, plot and fabula, actants). Purely quantitative analysis apply only to textual “degré zéro”, on which the secondary modeling systems of literature builds their significance.

Texts are essentially intentional objects: the meaning of a word; the usage of a metaphor; the choice of a metric or rhythmic solution in a poetic text are determined by the attribution of sense and meaning by the author and by the reader. Intentional phenomena do not follow regular pattern and are hardly (if ever) detectable by statistical methods.
A methodology for digital literary studies

- One of the underlying assumptions of the distant reading approach (and in general of any purely quantitative approach) is quite analogous to the reductionist stance in cognitive sciences.

- Interesting literary phenomena can be reduced without residues to material linguistic phenomena, that in turn are completely accessible to purely quantitative and statistic/probabilistic methods.

- We can say that a purely quantitative approach to literary objects is eliminativist towards the intentional concepts of critical discourse.
A methodology for digital literary studies

- Texts are intentional objects
- Interpretation is based on the production and application of a set of intentional notions and terms to explain what the text means and how
- Semiotic and structural critics have tried to explain or reformulate them in more formal and abstract concepts that preserve the intentional nature of text and interpretation
A methodology for digital literary studies

- Franco Moretti has adapted the epistemological concept of operationalizing proposed by P.W. Bridgman: “Operationalizing means building a bridge from concepts to measurement, and then to the world. In our case: from the concepts of literary theory, through some form of quantification, to literary texts”

- This purely quantitative conception is too constrictive in my view. I think that
  - computational formal modeling is a suitable kind of operational translation of literary theoretical terms and concepts, though not directly producing quantitatively measurable “proxies” of those terms
  - the convergence between semiotic/structuralist theories and methods and contemporary ontologies and linked data oriented practices represents a big chance for the future development of Digital Literary Critics
  - this kind of Rich Data can also enhance the efficacy of text mining technologies, so we must not see it in contrast with those tools and methods
Creating formal models based on explicit conceptualization and logical foundation grants that all the discourses are firmly grounded to a common “setting” of the domain. We all (try) to speak of the same thing.

Formal ontologies permit the application of computational inferences and reasoning methods to express explanation and make predictions. Their grounding in description logic has made possible the development of efficient automatic reasoner and inference engines.

Semantic Web modeling provides methods to compare and eventually merge different ontologies; the Open World Assumption, ensures the functionality of the model even if it is incomplete or conceived as a work in progress.
A methodology for digital literary studies

- Making ontologies and linking them to texts is not only a way for ensuring already existent knowledge exchange, but also a way to build knowledge, since:
  - it asks for making explicit the tacit knowledge which is a major part of literary critical work
  - it asks for finding the textual correlatives to the abstract and theoretical notions that populate the literary theories once they are formalized as ontologies; this can limit the risk of having purely speculative concepts that do not cope with the reality of textual traditions.
An ontology for narrative

- A project to analyze some of the relevant concepts of narratology studies and model them using a formal ontology and develop a formal model in OWL 2 DL

- The object of our modelization are
  - the notion of character and of actor/actant
  - the notion of narrative/fictional world/space, following the idea of “narrative semiotic space”
Ontology of character

- Characters are one of the most essential parts of a literary text, since they are typically the agents of the plot, building the basic structural elements of narrative action.
- Readers of a story are interested in characters not only because of their functional role in the plot, but because they are meaningful entities.
- Literary characters are formed through characterization, the description of their appearance, acts and intentions.
Ontology of character

- Characters can be analyzed in two modalities:
  - by considering their structural function
  - by analyzing their characterization.

- Our model, based on the work carried out since the early 80s by Giuseppe Gigliozzi follows this second modality

- In this model, the character is conceived as a frame composed of a set of structured properties
Ontology of character

- Cappuccetto Rosso, nel quadro della storia che la costruisce, è solo il coagulo spaziotemporale di una serie di qualità fisiche e psichiche (semanticamente espresse come "proprietà"), tra cui anche le proprietà di essere in relazione con altri coaguli di proprietà, di compiere certe azioni e di patirne altre [Eco, Lector in fabula, 130]
Ontology of character

Il Personaggio è visto, quindi, come un’unità complessa che viene esplosa nei suoi costituenti.

Per ogni personaggio viene stilata una Lista di marche funzionali alla diegesizzazione del ruolo che quel personaggio, nell’universo del dicibile, rappresenta.

La Lista comprende marche che definiscono la qualità e marche che rappresentano il mondo delle azioni. Entrambe le marche possono essere: ipercodificate e “variabili” (o contestuali).

Un primo esempio chiarirà meglio la configurazione della Lista:

<table>
<thead>
<tr>
<th>personaggio / ruolo Principe</th>
</tr>
</thead>
<tbody>
<tr>
<td>marche qualitative ipercodificate: bello, ricco</td>
</tr>
<tr>
<td>marche variabili: coraggioso, generoso</td>
</tr>
<tr>
<td>marche diegetiche ipercodificate: va alla ricerca, combatte</td>
</tr>
<tr>
<td>marche diegetiche derivate:</td>
</tr>
<tr>
<td>da (coraggioso): difendere</td>
</tr>
<tr>
<td>da (generoso): donare, offrire</td>
</tr>
<tr>
<td>marche qualitative espanse derivate:</td>
</tr>
<tr>
<td>da (coraggioso): audace, temerario</td>
</tr>
<tr>
<td>da (generoso): disponibile, donare.</td>
</tr>
</tbody>
</table>

Date certe marche, la serie degli atti attribuibili all'individuo deve confermare la loro isotopia, intesa come criterio di compatibilità semica.

[Gigliozzi 91, pp. 57-58]
Ontology of character

- Another contribution to the structure of our ontology of fictional character is the re-factorization of the well-known notion of narrative function proposed by Greimas.

- If the actant is an abstraction of the concept of character/actor, in complex narratives one character can generate multiple personification or roles, with different set of semantic traits that in turn are actualized by different isotopies at the syntagmatic level.

La reinterpretazione linguistica delle dramatis personae che abbiamo proposto a partire dalla descrizione proppiana della fiaba russa ha cercato di stabilire anzitutto una distinzione fra gli attanti, che appartengono a una sintassi narrativa, e gli attori, riconoscibili nei discorsi particolari in cui si trovano manifestati.

[Greimas, Del senso 2, p. 45]
Ontology of character

Provisional OWL Classes

Class: Actant
  ◦ Class ActantType

Class: Character
  ◦ Class: Role

The relation between the class of character and the classes of actantial roles related semantic characterization is expressed by the LA property "semiotics:expresses" and OWL 2 class punning

Class: Intention

Class: Quality
  Class: PhisycalQ
  Class: MentalQ

Class: Action
  Class: HypercodA
  Class: DerivedA

Class: Object
  Class: RealOb
  Class: MentalOb

Class: Place

Class: Event
  Class: RealE
  Class: ImagE
Ontology of character

Provisional OWL Properties

ObjectProperty: hasQuality
- Domain: Actor
- Range: Quality

ObjectProperty: doAction
- Domain: Actor
- Range: Action

ObjectProperty: hasObject
- Domain: Actor
- Range: Object

ObjectProperty: inhabitPlace
- Domain: Actor
- Range: Place

ObjectProperty: participateEvent
- Domain: Actor
- Range: Events

ObjectProperty: hasIntention
- Domain: Actor
- Range: Intentions
There is a deep relation between the notion of character and that of narrative space.

According to Lotman (1972), the narrative space to which is bounded defines each character, and the hero is the only character that can move between different narrative spaces.

The concept of narrative space can be formalized using the notion of fictional possible world, whose definition is given by Eco (1979, pp 128-30):

- “Definiamo come mondo possibile uno stato di cose espresso da un insieme di proposizioni dove per ogni proposizione o p o o ~ p. Come tale un mondo consiste in un insieme di individui forniti di proprietà”. 
Conclusions and next steps

- Definition of the fictional space ontology and of its relation with that of characters
- Definition of a set of primitive abstract action typology (f.e. revisiting Roger Shank primitives), to derive all the actual actions of the characters... or not?
- Definition of an ontology for fictional events, so that character actions can relate to their collocation in the textual time, as much as fictional space can relate to characters places
Conclusions and next steps

- Exploration of possible convergence with other narrative ontological frameworks, namely the one developed by Bartalesi, Meghini and Metilli (2016) and Ontomedia (Jewell at al. 2005)

- Mapping with other relevant ontologies and encoding scheme in the cultural heritage domain for interoperability (e.g., Text Encoding Initiative and CIDOC-CRM)

- Explore the application of the ontology to storytelling modelling and design in social sciences, media and game studies, enterprise communication