Exploring the Betrothed Lovers*

Andrea Bolioli[1], Matteo Casu[2], Maurizio Lana[3], and Renato Roda[4]

1 **Cross Library Services srl**
    Trento, Italy
    `bolioli@cross-library.com`
2 **CELI srl**
    Torino, Italy
    `casu@celi.it`
3 **Dipartimento di Studi Umanistici**
    Università degli Studi del Piemonte Orientale, Vercelli, Italy
    `m.lana@lett.unipmn.it`
4 **Dipartimento di Culture, Politiche, Società**
    Università degli Studi di Torino, Torino, Italy
    `renato.roda@unito.it`

─── **Abstract** ───

We present the ongoing activities and the first results achieved in a research project concerning the understanding of narrative in the high school. Students and teachers experimented with new ways to learn linguistic and digital skills, by using a collaborative learning environment built around the novel *I Promessi Sposi*. We analyzed the literary text, extracting social networks of characters and other fundamental narrative elements (sequences, locations, etc.), in order to provide the students with appropriate tools and resources to conduct their own inquiries on the novel.

1 Introduction

In this paper we describe the ongoing activities and the first results achieved in a research project concerning the understanding of narrative in the high school. The project partners are four Italian high schools (in the province of Trento), the Human Language Technology research unit at FBK-irst, two small enterprises, and two public educational organizations (IPRASE and Centro Formazione Insegnanti). The research project *Sharing Educational Content* (Sèduco) aims to study, develop and test digital tools for the creation, management and sharing of educational resources in the Italian upper secondary school (age 14 to 18).

Two ongoing activities in the project are concerned with the topic of narrative understanding:

- the linguistic and narratological analysis, with computational tools, of the historical novel *I promessi sposi* (*The Betrothed*) by Alessandro Manzoni,
- the collaborative creation of an anthology of Italian literature, which is named Antologia 2.0.

This paper focuses on the first issue. Students and teachers experimented with new ways to learn linguistic and digital skills, by using a collaborative learning environment[1] built around the novel *I Promessi Sposi*. This educational application can be considered an example of "semantic mashup" for narrative, a discovering method which is gaining attention and it is considered attractive by general users (see for example `http://lotrproject.com`) as well as a good training camp by researchers. In agreement with Inderjeet Mani, we believe "that computational narratology has the potential to revolutionize the way we create and study literature" [6].

The remainder of the paper is organized as follows. Section 2 briefly describes the *Promessi Sposi* learning environment. Section 3 presents the basic ontology underlying the system. Section 4 describes the social networks of narrative interactions created and used in the environment.

2   The learning environment as a semantic mashup

I promessi sposi is a very important Italian novel written by Alessandro Manzoni and published in its final illustrated version in 1840. For many and various reasons it was recognized as a work representing the cultural and linguistic unity of Italy and rapidly became a book read and studied in the secondary high school. Today it constitutes a difficult matter of study for students because of the growing distance from its themes, matter, and linguistic flavor (nineteenth century vocabulary and syntax).

The work can be – in a typical postmodern way – rebuilt while keeping it intact. This has been done by making it the center of a universe of relation, tools and activities.

Relations: around the work relations are built first of all among schools whose students work on the text, among schools and SMEs, and among schools and territorial institutions.

Tools: we built a set of tools allowing the students to browse the text following characters in theirs wanderings, to pay attention to a specific place in order to look at passing by characters, to explore the text by means of social network graphs, and so on.

Activities: in addition to "exploration" activities, students can link self-made reports, photographs, images (such as the drawings by Francesco Gonin), other texts, bibliographic references to the appropriate passages.

The basis for this rebuilding of the work has been the annotation of the novel for entity mentions and narrative sequences. We annotated a version of Promessi sposi [7] for Entity mentions of type Person (characters) and Location using two NER tools (TextPro by FBK-HLT and Sophia Semantic Engine by CELI). The automatic annotation were manually corrected. The novel has been split into sequences, that is the basic narrative units studied in high schools, (chunks of text) concerning a specific action or place or character. Dialogs, being the object of a specific interest, constitute always distinct sequences. Given this criteria, this operation has been done by humans.[2]

What actually happens is that students can browse the novel following characters or searching for places, can find the sequences pertaining to their interest, can read them, can publish content related to a specific sequence or to a block of sequences linking it to them, can study graphs showing relations between characters based on their co-presence in the same place or on their being part of a same dialogue.

---

[1] `http://seduco.cross-library.com/promessisposi/`
[2] A portion of the resulting dataset will be available to the research community.

Even technical institute students, with a few literary interests, have enjoyed the novel and they succeeded in performing linguistic and narrative analysis that would not have been possible in a traditional frontal lesson based solely on the paper book.

3   Ontological modelling of *I Promessi Sposi*

An interactive application with such a rich set of contents has faced the task of finding a good balance between exploiting all the information (acquired via automated or manual annotation) and avoid its explosion in terms of conceptual complexity.

Our treatment of *I Promessi Sposi* contains different sets of entities, which are linked by means of relations: the novel itself, discourse elements (Chapters, Paragraphs, Textual chunks), narrative elements (Sequences, Agents and Locations).

As for the novel, we considered the FRBR (Functional Requirements for Bibliographic Records) model [5], which distinguishes between the abstract Work from its Expressions (e.g., different revisions by the author) and its Manifestations.

Discourse elements, such as chapters and paragraphs, are defined by the author, and – given a certain edition – they are fixed. Ontologically, in order to treat this subset of entities, we reused classes from DoCO, from the SPAR (Semantic Publishing And Referencing) set of ontologies [9].

Narrative elements are the most interesting in this learning environment. We defined our own subclasses for a sequence: signally, Narrative (including Dialogues and Monologues), Descriptive and Reflexive kinds of sequences. Sequences are in many-to-many relation with paragraphs. A narrative sequences is linked to a location (the place of the narrative unit), to characters which act in it, and possibly to speakers, which are not necessarily the same.

Characters are, for the most part, persons or group of persons. A relevant exception is played by the *Divine Providence*, which can be considered as an agent in Manzoni's work. For these reasons we adopt the Friend Of A Friend (FOAF) vocabulary [3], which provides the class `foaf:Agent` with subclasses `foaf:Person` and `foaf:Group`, with the property `foaf:member` to link a group (such as the *bravi*, in our case) to its members.

The notion of character is not a class in our model, because we interpret it as a role played by an agent in a work. For this reason we modelled *character* as an OWL object property between a `foaf:Agent` and a `frbr:Work`.

Places and locations play a particular role in the model, for they can be regions, municipalities (such as Milan, Lecco) as well as relative locations (e.g., Lucia's house). Another geographic information in the novel is given by journeys between locations. In order to structure places – e.g., declaring that Lucia's house is in Lecco – we reuse the `parentFeature` property from the GeoNames ontology [10]. Another aspect about the places deals with the changing of places in time: some historical places do not exist anymore, have changed name or have changed status (e.g., the home of Lucia was in Acquate, formerly a municipality, now a neighborhood in Lecco). Again, we use the GeoNames ontology when possible and custom properties when necessary. In the context of the application these aspects are related to two maps depicting the places of the novel: namely, a XVIIth century map of northern Italy, and the current Google map.

4   SNA in *I Promessi Sposi*

In narratology, SNA (Social Network Analysis) has been used mostly as a new instrument for the study of plot evolution. By the extraction of the interactional networks of characters

from narrative works and the subsequent synthesis of the obtained data in network graphs it is possible to open a whole new perspectives to better comprehend the dynamics and the structure of a narrative plot. Even excluding the numerous quantitative analysis options available, the mere rearrangement of the narration from the written context to a fairly clear and understandable display – where the whole plot is summarized in a single eidetic structure – represents a powerful explanatory enhancement [8]. This SNA analytical approach has already been employed with a fairly large selection of different literary text, ranging from Shakespeare's tragedies [8] to Lewis Carroll's *Alice in Wonderland*[1], including the whole Marvel comics universe [2].

However, the vast heuristic potential of quantitative network analysis methods has rarely been employed for studying a text as structured and as complex as *I Promessi Sposi* is. In addition, it is the first time it is used in an educational context (high schools), as far as we know. The composite narration of Alessandro Manzoni's most famous work comprises plentiful flashbacks, historical and ethical excursuses, changes in the main narration perspective.

Such narrative complexity challenges the heuristic potential of SNA techniques: the available methodologies must be carefully selected and polished to obtain the most effective analytical tool. While the identification of the network's "nodes" with the characters of the studied texts is a choice common to the absolute majorities of literary researches that employed network analysis methodologies, the definition of "edges" appears less immediate and more difficult. One option, in such sense, focuses on "conversational edges" [4], where an edge is formed between two characters/nodes every time the studied text features an explicit dialogue between aforementioned characters.

The graph in Figure 1 represents the complete conversational network of *I Promessi Sposi*, elaborated with the SNA visualization software Gephi. The dimension of each node is due to the number of conversational interactions ("degree" in SNA terminology) in which that node's character is involved: in fact the largest node, in the very middle of the graph, embodies Renzo Tramaglino, the me lead of the novel. The thickness of each "edge" is directly correlated to the number of interactions between the relative couple of nodes: if a dyad of characters shares many dialogues, the edge that tie their node will be more thick. Different colors distinguish each community of characters (that is a set of highly interconnected nodes).

Remarkably, the community optimization algorithm perfectly recognizes the various "narrative blocks" of characters interactions: for example, the group of green nodes at the left of Renzo, quite isolated from the rest of the network, identifies one of the novel's most famous flashbacks, centered on the origin of Gertrude's character.

Alongside the "conversational network", we have also extracted a second network from *I Promessi Sposi*: while the nodes still represent the main characters of the novel, this time an edge between two characters identify their co-occurrence in the same narrative sequence. The graph in Figure 2 illustrates this second action-oriented network. As it is possible to notice even at a first glance, while the general structure of the two networks is similar, there are few interesting differences: for example, in the second graph the node Lucia Mondella, the main female character, has a slightly larger degree than the one of Renzo, and achieve the status of larger hub of the graph. This is particularly interesting since the main plot of the book splits very soon into 2 separate narrative lines, one for each of the titular betrothed lover. According to our graphs, role of Lucia is more action-oriented while the one of Renzo more involved in conversational interactions.

The twofold SNA approach used, however, revealed not to be only a valuable visualization/summarizing education tool, but also a powerful instrument per a deeper exploration of the text's narrative structure and characters dynamics. With results that sometimes go

beyond the mere teaching/learning scope that is the main focus of our project. In fact, thanks to the quantitative metrics obtained from the two characters network, it was even possible to open new perspective of narrative analysis and to discover aspects unknown (and unexpected) even to our team's experts in the work of Alessandro Manzoni. For example, while investigating on the relational "importance" of each character in the novel through the evaluation of betweenness centrality (a measure of brokerage centrality, that is the value of a node as bridge between other nodes in a given network), we were surprised to find out that some characters considered moderately secondary have an higher value than some of the main actors of the plot. Figures like "il Griso" – the main antagonist's sidekick – and even more "Agnese" – Lucia's mother – play a way stronger bridging role in the relational balance than stronger narrative characters like the charismatic villain (then reformed helper) "Innominato" or the iconic "Don Abbondio". This is just one example of the many analytic exploration perspectives opened by the SNA approach, perspectives than can be investigated by both students and the researchers, taking advantage of the nature of versatile research tool – and not only passive visualization solution – of the network graph.

──── **References** ────

**1**   Apoorv Agarwal, Augusto Corvalan, Jacob Jensen, and Owen Rambow. Social network analysis of Alice in Wonderland. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, pages 88–96. Association for Computational Linguistics, 2012.

2   Ricardo Alberich, Joe Miro-Julia, and Francesc Rosselló. Marvel universe looks almost like a real social network. Preprint, 2002.

3   Dan Brickley and Libby Miller. The Friend Of A Friend (FOAF) vocabulary specification, 2007. `http://xmlns.com/foaf/spec/`.

4   David K. Elson, Nichoas Dames, and Kathleen McKeown. Extracting social networks from literary fiction. In Jan Hajič, Sandra Carberry, Stephen Clark, and Joakim Nivre, editors, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 138–147, Stroudsburg, PA, 2010. Association for Computational Linguistics.

5   IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records: final report, 1997. As amended and corrected through February 2009.

6   Inderjeet Mani. *Computational Modeling of Narrative.* Number 5 in Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

7   Angelo Marchese, editor. *Alessandro Manzoni: I Promessi Sposi.* Arnoldo Mondadori Editore, 1985.

8   Franco Moretti. Network theory, plot analysis. *New Left Review*, 68:80–102, 2011.

9   David Shotton and Silvio Peroni. DoCO, the document components ontology, 2011. `http://purl.org/spar/doco`.

10   Bernard Vatant. GeoNames Ontology – Semantic Web. `http://www.geonames.org/ontology/`.

A   Character Networks
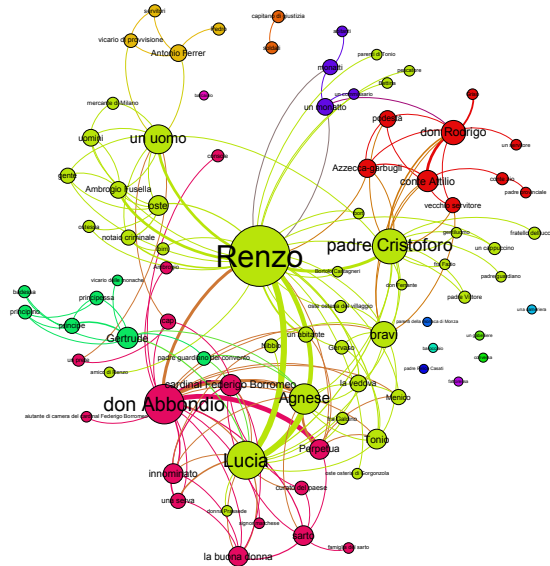
Figure 1 Conversational network.
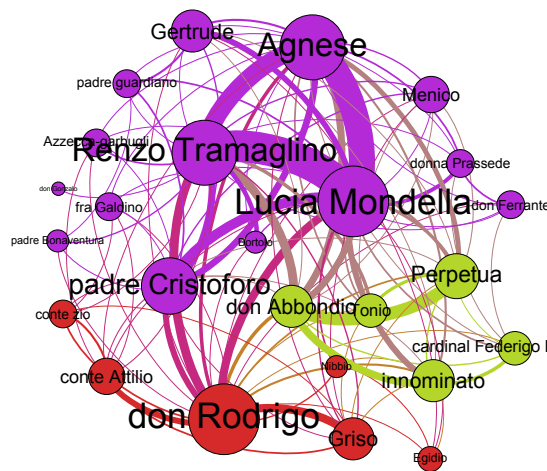
Figure 2 Co-occurrence in narrative sequences network.