Sharing Cultural Heritage: the Clavius on the Web Project

Matteo Abrate*, Angelo Mario Del Grosso◇, Emiliano Giovannetti◇, Angelica Lo Duca*, Damiana Luzzi◇, Lorenzo Mancini◇, Andrea Marchetti*, Irene Pedretti+, Silvia Piccini◇

*Institute of Informatics and Telematics,
◇Institute of Computational Linguistics "A. Zampolli"
National Research Council (CNR)
Via Moruzzi, 1 - Pisa, Italy
*name.surname@iit.cnr.it, ◇name.surname@ilc.cnr.it
+ Historical Archives of Pontifical Gregorian University,
Piazza della Pilotta, 4 00187 Roma, Italy
i.pedretti@unigre.it

Abstract

In the last few years the amount of manuscripts digitized and made available on the Web has been constantly increasing. However, there is still a considerable lack of results concerning both the explicitation of their content and the tools developed to make it available. The objective of the Clavius on the Web project is to develop a Web platform exposing a selection of Christophorus Clavius letters along with three different levels of analysis: linguistic, lexical and semantic. The multilayered annotation of the corpus involves a XML-TEI encoding followed by a tokenization step where each token is univocally identified through a CTS urn notation and then associated to a part-of-speech and a lemma. The text is lexically and semantically annotated on the basis of a lexicon and a domain ontology, the former structuring the most relevant terms occurring in the text and the latter representing the domain entities of interest (e.g. people, places, etc.). Moreover, each entity is connected to linked and non linked resources, including DBpedia and VIAF. Finally, the results of the three layers of analysis are gathered and shown through interactive visualization and storytelling techniques. A demo version of the integrated architecture was developed.

Keywords: language technologies for digital cultural heritage, lexica and ontologies, data visualization

1. Introduction

The Historical Archives of the Pontifical Gregorian University (APUG) contain and preserve more than 5,000 manuscripts, testifying to the intellectual works and teaching activities of the Jesuits of the Roman College (1551-1773), one of the main research places in modern Europe.
The Clavius on the Web project pays a special attention to the manuscripts related to Christophorus Clavius (1538-1612), a Jesuit mathematician and astronomer, one of the most respected and influential scholars of his time.
Clavius was one of the main authors of the calendar reform under pope Gregory XIII (Gregorian Calendar), and also strongly encouraged the introduction of the mathematic disciplines (astronomy, geometry, algebra) in the *ratio studiorum* of the Jesuit colleges. In this respect he influenced deeply the history of teaching of Modern Europe and, more generally, of the all Western World. He taught mathematics at the Roman College and wrote many books on mathematics and astronomy, on which scholars, as Descartes and Mersenne, were formed (Lattis, 1994).
The Clavius on the Web initiative is a pilot project which allows to deeply investigate morphological, lexical, and semantic content by up-to-date methods and technologies. This approach could be used also to study and analyze other authors' handwritten works.
However, the first stage of the Clavius on the Web project takes into account just the two manuscripts containing the letters[1], more than 300 papers which attest the correspondence between Clavius and some important characters of his time, such as Galileo Galilei, Tycho Brahe or Guido Ubaldo Del Monte. The project deals with these documents from a twofold perspective: linguistics and dissemination. From the linguistic perspective, it aims at processing the contents of the manuscripts, by developing a computational lexicon of mathematical and astronomical terminology, a statistical morphological tagging and lemmatization of Latin and a semantic annotation module. As to dissemination, the aim is to create a platform which would facilitate the browsing of manuscripts on the Web, by exploiting techniques of visualization and storytelling. More specifically, the visualization integrates the digitized version of the manuscripts, their transcription/translation with the extracted linguistic resources and other facilities from the Web of Data. The final purpose is, thus, to build a Web portal to tell stories about Clavius as well as a Linked Dataset, available as a Linked Data node. All these features will allow to use the resources both by scholars and by no-expert users.

The partners involved in the project are the Historical Archives of the Pontifical Gregorian University (APUG), the Institute of Computational Linguistics (ILC), and the Institute of Informatics and Telematics (IIT) of the Italian National Research Council (CNR).

The remainder of the paper is organized as follows: after a review of related works (section 2.), section 3. introduces the project workflow describing the different process phases. After the transcription in XML-TEI encoding and the manual annotation described in section 4., the linguistic

[1] APUG 529-530

and lexical annotation process are illustrated (section 5.). Section 6. is devoted to linked data and data visualization. Finally, conclusions and future works are discussed (section 7.).

2. Related Work

The Clavius on the Web Project is the result of an accurate study on the cutting-edge solutions for general purpose digital libraries or mono-thematic archives.

Most initiatives focus on Web access to repositories of texts and images available through collaborative and interconnected frameworks (Bozzi, 2013).

On one hand, Gallica[2], the Library of Congress[3], as well as international projects such as Internet Archive[4] or Europeana[5] are the best examples of general purpose initiatives. On the other hand, the Van Gogh[6], Wittgenstein[7] and Nietzsche[8] are projects related to the digital preservation of single authors' works.

Similar to the contribution's proposal are also platforms such as Knowledge Circulation in the 17th Century[9] and Darwin Correspondence Project[10], which explore and analyze corpus of letters in innovative and terrific ways. Google, on its side, is upgrading this digital field with the well-known Google Books[11], and increasingly popular projects like Google Cultural Institute[12] or Google Glass[13]. Perseus Project[14] and the Homer Multi-text Project[15] represent, among others, the state of the art concerning ancient, linked, open, and digital archives. The former is the largest archive for the Graeco-Roman world and the classical Greek and Latin texts. It provides data and tools for linguistic analysis, such as treebanks, and annotated entities in OAC[16] compliant with RDF format[17].

On the other side, the Homer Multi-text Project is a framework for digital philology concerning texts and manuscript images of the Iliad and the Odyssey. Within this project the CTS/CITE architecture[18] has been developed in order to cite textual passages and digital objects. A worth-considering transcription framework for unstudied manuscripts has been developed by the Transcribe Bentham project[19].

Other important initiatives are DARIAH[20], TextGrid[21], and Clarin[22]. The aim of the DARIAH project is to develop an European infrastructure and a data platform able to provide and to integrate services for digital arts and Humanities (Blanke et al., 2011). TextGrid provides integrated tools and a collaborative virtual research environment for analyzing texts, and gives computer support for digital editing purposes and for philological works (Neuroth et al., 2011). Finally, the Clarin infrastructure aims at making language resources and technologies available, in particular to the humanities and the social sciences research communities (Váradi et al., 2008).

The COST action Interedition[23], in conclusion, promoted the development of interoperable tools and shared methodologies in the digital scholarly editing field.

3. Project Workflow

Figure 1 shows the workflow for each research team involved in the project. Starting from the digitization of the APUG manuscripts, the workflow includes two other preliminary steps: the transcription of the letters written to or by Clavius and the translation of several texts from Latin into Italian and English.

Transcriptions and translations are marked up using TEI-XML P5[24] and then the texts are tokenized adopting a CTS[25] compliant structure. This operation is the first step of the linguistic analysis, which includes both lemmatization (with morphological analysis as well) and lexical annotation. The latter is functional to an automatic semantic annotation and to the construction of a specific lexicon related to the Clavius' domain.

A deeper semantic annotation is carried out manually in order to enrich the texts by using a specifically developed ontology schema. As well as a visual manuscript and a knowledge graph browsing tool, the final user interface also includes a storytelling platform, in order to provide a better educational source. Together with this, a Linked dataset related to Clavius has also been built and integrated.

4. From manuscripts to manual semantic annotation

The Clavius' manuscripts have been digitized in compliance with the up-to-date practices of preservation. While the letters were already transcribed by Ugo Baldini and Pier Daniele Napolitani (Clavius, 1992), the translations have been done from scratch to ease the reading by non-academic users.

In order to build a semantic infrastructure suitable to describe this documentation, an ontology schema was created. Some classes were inspired by already existing ontologies or conceptualization schemes: FOAF[26], CIDOC[27] CRM[28] and FRBRoo[29].

[2] http://gallica.bnf.fr
[3] http://catalog.loc.gov
[4] http://archive.org
[5] http://www.europeana.eu
[6] http://www.vangoghletters.org
[7] http://www.wittgensteinsource.org
[8] http://www.nietzschesource.org
[9] http://ckcc.huygens.knaw.nl
[10] https://www.darwinproject.ac.uk
[11] http://books.google.it/
[12] http://www.google.com/culturalinstitute/
[13] http://www.google.com/glass/start/
[14] http://www.perseus.tufts.edu
[15] http://www.homermultitext.org
[16] http://www.openannotation.org/
[17] http://www.w3.org/RDF/
[18] http://www.homermultitext.org/hmt-doc/cite/
[19] http://blogs.ucl.ac.uk/transcribe-bentham
[20] https://dariah.eu/
[21] https://www.textgrid.de/
[22] http://www.clarin.eu/
[23] http://www.interedition.eu/
[24] http://www.tei-c.org/Guidelines/P5/
[25] http://www.homermultitext.org/hmt-doc/
[26] http://www.foaf-project.org/
[27] http://www.cidoc-crm.org/
[28] http://www.cidoc-crm.org/
[29] http://www.cidoc-crm.org/frbr_inro.html

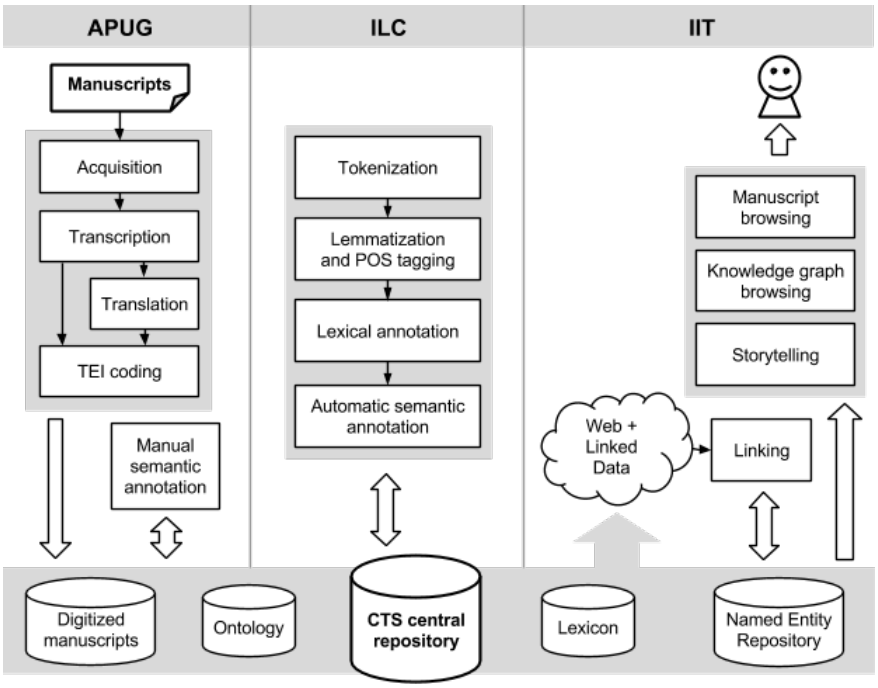

Figure 1: Project workflow.

This ontology schema provides the vocabulary to perform semantic annotation (Agosti et al., 2007). Such kind of annotation helps to bridge the ambiguity of natural language in expressing notions and their computational representation in a formal language, by associating instances of the ontology to text elements. Each annotated text element also acquires the property values describing an instance (i.e. biographical information). For example, in the text of a letter written in 1588, Galileo Galilei refers to Clavius using the acronym "VSRM" ("Vostra Signoria Molto Reverenda"). By linking the term "VSMR" to the "Christophorus Clavius" instance of the ontology PERSON class, a search looking for "Clavius" also returns the context in which he is referred to as "VSRM".

The semantic annotation is performed on the basis of a set of predefined classes: Person, Group/Institution, Location, Instrument (e.g. scientific or mathematical instruments), Astronomical entity (e.g. planets, stars), Work, Manuscript, Edition, Event (e.g. solar eclypses), Date. Scholars who perform the annotation have a deep knowledge of mathematical and astronomical theories of Clavius'age and of the domain entities mentioned in the letters. The annotators work focuses only on the original text of the letters, most of them written in Latin or ancient Italian.

5. Lexico-linguistic and automatic semantic annotation

5.1. TEI and CTS

The guidelines and the XML schema provided by the Text Encoding Initiative (TEI) have been used for digital editing purposes, to mark up the authoritative transcriptions (from the philological and linguistic points of view) and to provide an accurate physical description of Clavius letters (from the codicological and paleographic points of view). The adoption of the TEI standard allowed to achieve three main goals: i) to encode the logical divisions of the text (e.g. sentences, paragraphs, etc.), ii) to use a systematic notation for assigning standard URN identifiers to texts and their fragments, iii) to share data for the scientific community of reference. The TEI default encoding embeds every kind of information about a document directly in the text (inline annotation): in order to make the addition and retrieving of shared data easier, a citational approach (stand-off annotation) was adopted. The protocol provided by Canonical Text Services (CTS) architecture gives a valid solution for indexing textual passages by standard URN identifiers (Smith and Blackwell, 2012). CTS hierarchical notation derives explicitly from Functional Requirements for Bibliographic Records (FRBR) model[30] and from the Ordered Hierarchical Content Objects (OHCO) document structure (Renear, 2004). Through the proposed method it is possible to manage different textual units as strings of references (URIs) at different granularity (e.g. paragraph, sentence, etc). Therefore, without burdening the XML-TEI encoded file, the architecture handles annotations through the citational stand-off URN mechanism, spanning from the whole document exemplar to the single sequence of characters (token). Thanks to CTS notation, for example, some of the letters written in Latin were linked sentence by sentence to their translations into Italian and English.

5.2. Linguistic analysis of Latin

The linguistic module provides tokenization and lemmatization components for the analysis of the Latin sources by means of statistical part-of-speech tagging techniques. Tokenization and lemmatization of texts are the basic steps to

[30] http://www.homermultitext.org/hmt-doc/cite/cts-urn-overview.html

obtain linguistic-based indexing and search (Bamman and Crane, 2008). Indeed, a lemma can be regarded as a label of a set grouping all the relative word-forms. The output of Clavius tokenization provides, on one hand, the basic elements for language processing and, on the other hand, creates the proper URNs identifier according to the subreference notation of the CTS architecture[31]. In this way each token has its own URN in CTS notation, which is global and unique.

The aforementioned tokenization phase enables a canonical and shared approach, allowing a high degree of decoupling between the textual source entities and the related analyses, and thus providing a sound integration mechanism among the different processing tools.

Figure 2 shows the components developed to obtain (a) the sentences, referred to by the CTS identifier; (b) the morphological analysis of each CTS-token, gathered both by a statistical approach (Halácsy et al., 2007) and by a checking electronic lexicon (Bamman and Crane, 2008); and finally (c) the lemma, obtained by querying the Perseus repository with word-form and its part-of-speech.

The morphological analysis component (PoS Tagger) uses a statistical model that must be trained for the classification process. To obtain an accurate PoS tagging it is necessary to train the classifier using data as similar as possible to texts to be analyzed. For this purpose, the annotated corpora of the Perseus Latin Treebank[32] (Bamman et al., 2008; Bamman and Crane, 2011) have been used as the training set.

All the steps described above are automatic and, consequently, intrinsically prone to introduce errors. To obtain an error-free analysis the results have been therefore manually checked and proof-read using an *ad hoc* Java Web-based application which is still in development.

5.3. Automatic Semantic Annotation

In the perspective of digitizing a large amount of manuscripts, it would be useful to provide a tool for automatic semantic annotation. For this purpose a Named Entity Recognition (NER) stochastic component will be developed. It will be trained starting from the manually annotated corpus of letters. In this respect some critical issues will be faced: (a) the language: most of the project corpus is written in scientific Latin of renaissance age; since the linguistic annotation is a prerequisite to NER, the performance might be influenced by the morphological and lexical differences with respect to classical Latin; (b) the size of the training set: the accuracy of classification systems based on machine learning techniques depends on the size of the training corpus. Since in this project the number of available pre-annotated texts is quite modest, accuracy is likely to be not very good; (c) the semantic classes: the NER component will probably be more accurate in detecting certain entities than others: some difficulties are expected in detecting entities that describe Events or in distinguishing between Manuscripts and Editions.

[31] http://www.homermultitext.org/hmt-doc/standards/tokenization.html

[32] http://www.homermultitext.org/hmt-doc/standards/tokenization.html

5.4. Computational Lexicon

Within this project an electronic lexicon of the mathematical-astronomical terminology used by Clavius in his correspondence and, more generally, in his *Opera Mathematica*[33] is being built. It is well known that Clavius helped to build a rich and unambiguous mathematical vocabulary, thus providing the common language of European mathematics for the following centuries.

The architecture of the lexicon is based on a well-established model in the context of Computational Lexicography, SIMPLE (Lenci et al., 2000; Ruimy et al., 2003), which strongly inspired the ISO Lexical Markup Framework and has already been customized to represent Saussures terminology (Ruimy et al., 2012; Ruimy et al., 2013). Based on a revised version of the theory of Generative Lexicon (Pustejovsky, 1995), SIMPLE permits the definition and description of the internal structure of lexical units, even those characterized by a more complex semantic content, the componential and relational nature of word meaning being emphasized. An ontology consisting of mono- and multi-dimensional semantic types, a wide network of semantic relationships and a rich set of semantic features allow structuring the key concepts of the domain and defining the relationships between them. More specifically, the semantic content of each instance of an ontological class is defined in a lexical entry, which describes the lexical-semantic relations that the term entertains with others on both a paradigmatic level (relations of hypernymy, hyponymy, meronymy, holonymy) and the syntagmatic axis; it also specifies the semantic and morphological distinctive traits (PoS), and gives information on the domain of use and, whenever appropriate, on the type of event denoted. For the first time the main terms of mathematics (*punctum*, *quantitas*, etc.) and astronomy (*motus planetorum*, *revolutio coelestis*, etc.) of Clavius' age receive a rich and highly structured representation of their semantic content. Such a lexicon should provide a deeper knowledge of the overall domain terminology and may open up new paths of analysis that have not yet been explored.

6. Linked Data, Visualization and Storytelling

6.1. Linked Data

The Clavius Linked Dataset (CLD) contains three types of entities related to Clavius: person, location and letter. Their relationships are extracted from the manuscripts. The specific information about each entity is retrieved from Web open sources (e.g. Wikipedia) through the Linking process, and is checked manually.

The strength of CLD consists in external and internal links established among entities. External links connect entities to sources of the Web, available both as linked datasets and standard Web sites. They are both generic and domain-

[33] Clavius' Opera Mathematica, published in Mainz in 1611-1612, consists of five volumes. See: http://www.e-rara.ch/zut/content/titleinfo/1182315

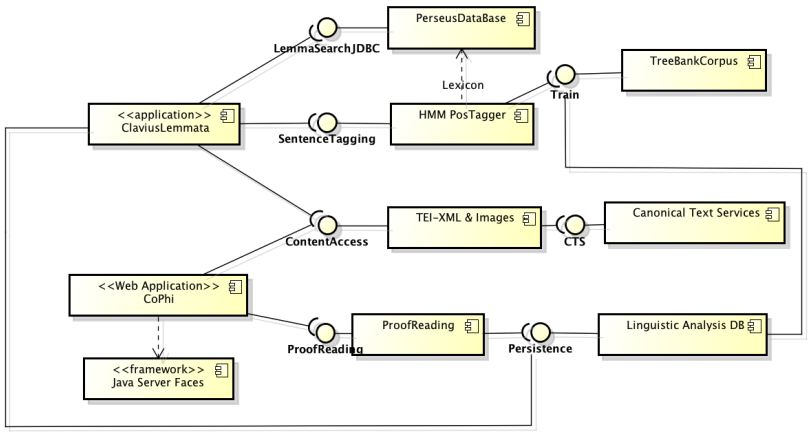

Figure 2: Linguistic components for semi-automatic annotation.

specific[34]: DBpedia[35], Wikipedia[36], Treccani[37] and VIAF[38] describe people. GeoNames[39], DBpedia and Wikipedia are used for locations. Internal links establish relationships among entities, within the same repository. Figure 3 shows the relationships among the three entities contained in the dataset. The figure shows the resources as ellipses and the objects (literals) as rectangles. The prefix used for the implemented dataset is `clavius:`. As for the other vocabularies, the used prefixes are `dcterms:` for the Dublin Core ontology, `foaf:` for FOAF, `bibo:` for Bibo and `gn:`for the GeoNames ontology. The figure describes the letter sent from Galilei to Christophorus Clavius on 1588, 8th January. Such a letter is identified by the ID `clavius:letter-13`, it was created by Galileo Galilei (identified by the ID `clavius:person-61`) and sent to Christophorus Clavius (ID `clavius:person-135`). Finally, the letter was written in Firenze (Florence), which is identified by the ID `clavius:location-19`.

The Clavius Linked Dataset also provides other useful information. For example, for each person, it provides a short biography and the links to the letters sent to Clavius.

Figure 4 shows a part of the RDF graph surrounding the resource `clavius:person-61`, which corresponds to the person "Galileo Galilei" (through the property `foaf:name`) connected to the linked data resource describing Galileo in DBpedia (through the property `owl:sameAs`) and to external non linked data resources (Wikipedia, Treccani and VIAF) through the property `dbpedia-owl:wikiPageExternalLink`. In addition, the Galileo node is linked to local resources: namely, the letters he wrote (the figure shows only one letter, `clavius:letter-13`) through the property `foaf:made` and people he knew (the figure shows Christophorus Clavius, identified by resource `clavius:person-135`), through the property `foaf:knows`.

No further information about locations is provided[40] since it can be easily retrieved by following the connected external links. Likewise the computational lexicon (Section 5.4.) and the ontology (Section 4.), the Clavius Linked Dataset is available as a Linked Data node [41] and released under the Creative Commons CC BY-SA license[42].

6.2. Visualization and storytelling

Three HTML5 Web user interfaces are provided to let users browse all the aforementioned data: the first is focused on the manuscript, the second on the visualization of the annotations and the third is devoted to storytelling.

In the manuscript interface (Figure 5), users can both see the digital image of the manuscript and zoom in to appreciate its details, and read its transcription and translations. This interface is based on the *Edition Visualisation Technology* (EVT) tool[43], an open-source software for the visualization of TEI-based digital editions (Rosselli Del Turco et al., 2008). EVT, originally developed for the visualization of other manuscripts, is currently being adapted to suit the needs of Clavius on the Web. A collaborative effort is also underway to help the authors with reengineering and technological update tasks, in order to develop an advanced, generic platform for the visualization of manuscripts on the Web.

The annotation visualization is conceived to reach and be appealing to a vast public, including users that have no technical nor domain knowledge. Like in an information graphic, typography, symbols, layout and choice of color play a central role. Moreover, the design follows the theoretical and practical advices from the field of *information visualization*, a discipline of computer science that studies how to graphically represent data to best exploit the computational power of human vision. Specifically, the visu-

[34]Belonging to the cultural field.
[35]http://dbpedia.org
[36]http://www.wikipedia.org/
[37]http://www.treccani.it/
[38]http://viaf.org/
[39]http://www.geonames.org/
[40]Geographical coordinates for locations.
[41]http://claviusontheweb.it/linked_data/index.html
[42]http://creativecommons.org
[43]Available from http://sourceforge.net/projects/evt-project/.

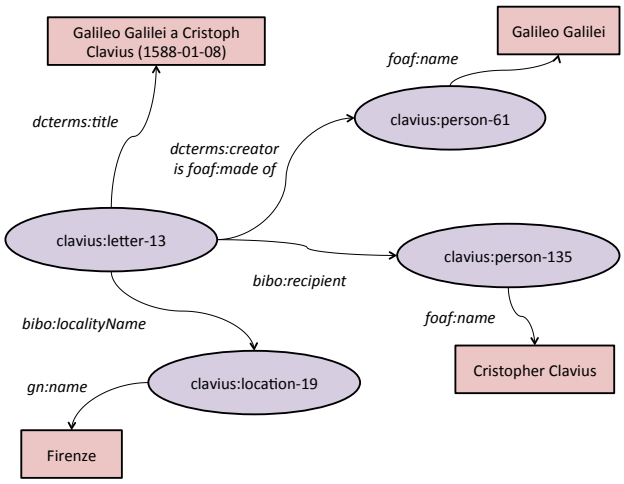

Figure 3: The relationships among people, letters and locations.

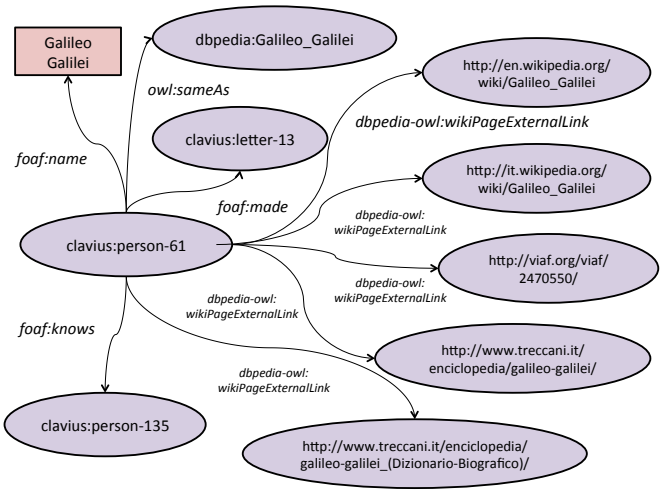

Figure 4: The RDF graph for resource clavius:person-61 (Galileo Galilei).

alization seeks to reduce the cognitive load of the user by representing data with the use of *preattentive variables*, visual depictions that are quickly and easily identified by our vision system without requiring a high use of active attention (Healey et al., 1996). An in-depth introduction to this topic can be found in (Ware, 2004).

Figure 6 represents the current visualization prototype, where three different aspects of the linguistic analysis are highlighted.

A third user interface takes a complementary approach to explore Clavius on the Web's data, by leveraging storytelling techniques to let users learn important stories about Clavius. A multimedial, Web-based framework is being developed by using state-of-the-art Web technologies, and is being implemented in collaboration with domain experts.

A first prototype (shown in Figure 7) tells the story of the life of Clavius, the main events concerning his career, and his travels. A map and a timeline are provided to complement textual narration, giving users the ability to grasp both chronological and geographical aspects of the life of Clavius.

7. Conclusions and Future Works

A new approach to the digitization was illustrated, based on multilayered annotation and visualization of a selection of Christophorus Clavius letters owned by the Historical Archives of the Pontifical Gregorian University.

Currently, the data model shared among the different annotation layers is basically a graph composed of XML nodes, where edges represent URI references between the different elements of analysis. However, work is now oriented to expose the obtained knowledge base as a Linked Open Data node accessible via SPARQL, by converting each assertion into an RDF statement.

As far as the linguistic annotation is concerned, the tools have been trained by using treebanks of classical Latin. Since most of the project corpus is written in Renaissance scientific Latin, future work will aim at the enhancement of the accuracy of the Latin lemmatizer. It will be realized through the implementation of a circular process in-

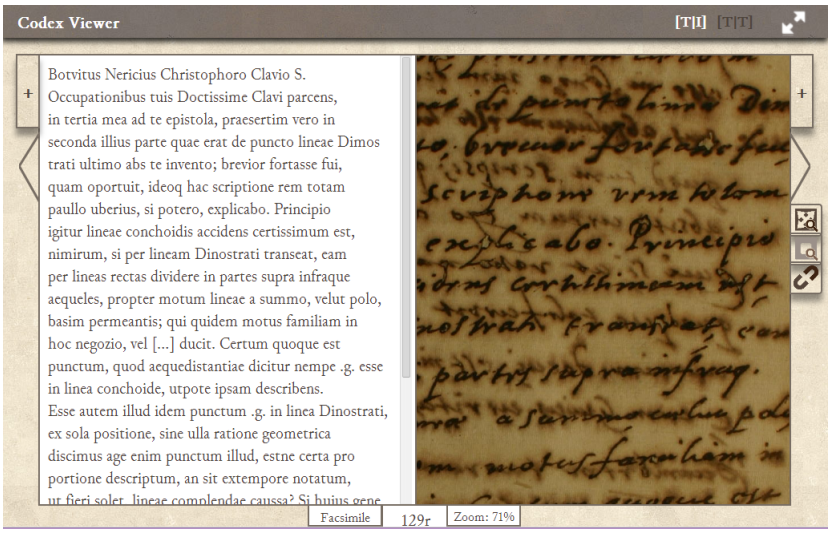

Figure 5: The Edition Visualization Technology manuscript interface (Rosselli Del Turco et al., 2008), showing a letter from Botwid Nericius to Christophorus Clavius.

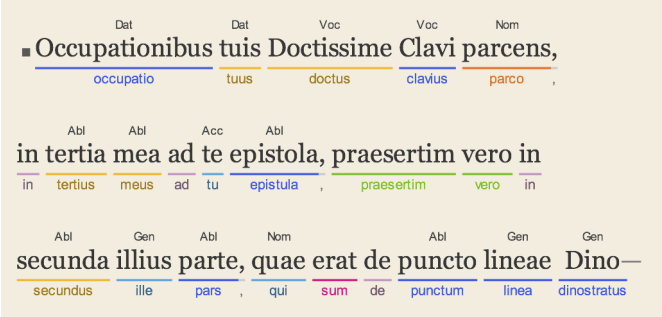

Figure 6: The current prototype of the annotation visualization. Sentence splitting is represented by means of a gray square introducing each new sentence. A token is indicated by an underline, and its lemma of reference appears underneath. The part-of-speech is indicated by the color of the underline and the lemma: the token can be a noun (blue), a verb (purple), an adjective (yellow), etc. The Latin grammatical case (such as nominative, genitive, etc.) is shown above the token. The original spacing, punctuation and line breaking of the text are preserved.

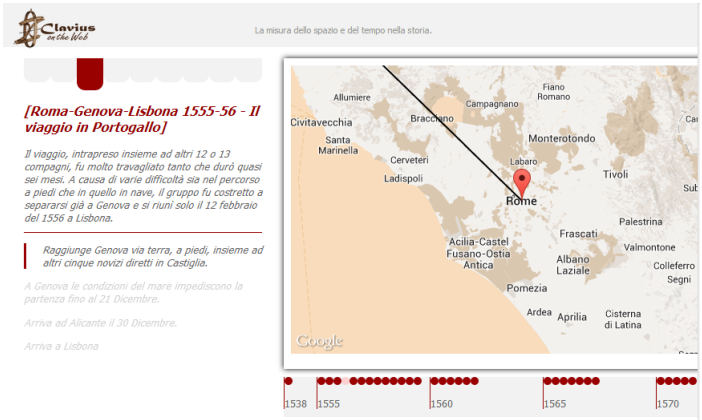

Figure 7: A first prototype of storytelling interface, telling the story of the life of Clavius by highlighting the main events of his career on a timeline and its travels on a map.

volving the re-training of the model on the basis of each proof-reading session conducted by the scholar.

In addition, as to dissemination purposes, an integrated Web platform will be released, allowing a user both to visu-

alize the manuscripts with their annotation and to perform himself the annotation.

The system architecture and the data structures have been designed following three basic principles: i) platform independence, realized through both stand-off annotations (CTS) and data formats described with general markup specifications, ii) component-based design, and iii) open source, since the software is going to be released under GPL license and the data under Creative Commons. On this basis, from the perspective of reusability, the overall procedure and most of the developed technologies and resources can be easily customized and applied to the processing of, ideally, any kind of textual resource.

8. Acknowledgements

This work has been carried out within the Clavius on the Web project.

9. References

Maristella Agosti, Giorgetta Bonfiglio-Dosio, and Nicola Ferro. 2007. A historical and contemporary study on annotations to derive key features for systems design. *Int. J. on Digital Libraries*, 8(1):1–19.

David Bamman and Gregory Crane. 2008. Building a dynamic lexicon from a digital library. In Ronald L. Larsen, Andreas Paepcke, Jos Luis Borbinha, and Mor Naaman, editors, *JCDL*, pages 11–20. ACM.

David Bamman and Gregory Crane, 2011. *Language Technology for Cultural Heritage*, chapter The Ancient Greek and Latin Dependency Treebanks. Springer, Berlin.

David Bamman, Marco Passarotti, Roberto Busa, and Gregory Crane. 2008. The Annotation Guidelines of the Latin Dependency Treebank and Index Thomisticus Treebank: the Treatment of some specific Syntactic Constructions in Latin. In *LREC*. European Language Resources Association.

T. Blanke, M. Bryant, M. Hedges, A. Aschenbrenner, and M. Priddy. 2011. Preparing dariah. In *E-Science (e-Science), 2011 IEEE 7th International Conference on*, pages 158–165, Dec.

Andrea Bozzi. 2013. G2A: a Web application to study, annotate and scholarly edit ancient texts and their aligned translations. *Studia graeco-arabica*, 3(1):159–171.

Cristophorus Clavius. 1992. *Corrispondenza, Baldini U. and Napolitani, P.D. (eds.)*. University of Pisa, Department of Mathemathics, Pisa.

Péter Halácsy, András Kornai, and Csaba Oravecz. 2007. HunPos: An open source trigram tagger. In *Proceedings of the 45th Annual Meeting of the ACL*, pages 209–212, Stroudsburg. Association for Computational Linguistics.

Christopher G Healey, Kellogg S Booth, and James T Enns. 1996. High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(2):107–135.

James M Lattis. 1994. *Between Copernicus and Galileo: Christoph Clavius and the Collapse of Ptolemaic Cosmology*. The University of Chicago Press, Chicago.

Alessandro Lenci, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, et al. 2000. Simple: A general framework for the development of multilingual lexicons. *International Journal of Lexicography*, 13(4):249–263.

Heike Neuroth, Felix Lohmeier, and Kathleen Marie Smith. 2011. Textgrid–virtual research environment for the humanities. *International Journal of Digital Curation*, 6(2):222–231.

James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press, Cambridge, MA.

Allen H. Renear, 2004. *A Companion to Digital Humanities*, chapter Text Encoding. Blackwell, Oxford.

Roberto Rosselli Del Turco, Raffaele Masotti, Julia Kenny, Chiara Leoni, and Jacopo Pugliese. 2008. Edition Visualisation Technology: a simple tool to visualize TEI-based digital editions. *The Linked TEI: Text Encoding in the Web*, page 208.

Nilda Ruimy, Monica Monachini, Elisabetta Gola, Nicoletta Calzolari, MC Del Fiorentino, Marisa Ulivieri, and Sergio Rossi. 2003. A computational semantic lexicon of italian: Simple. *Computational Linguistics in Pisa, Special Issue*, 18:821–864.

Nilda Ruimy, Silvia Piccini, and Emiliano Giovannetti. 2012. Les Outils Informatiques au Service de la Terminologie Saussurienne. In *SHS Web of Conferences*, volume 1, pages 1043–1056. EDP Sciences.

Nilda Ruimy, Silvia Piccini, Emiliano Giovannetti, and Andrea Bellandi, 2013. *Guida per un'edizione digitale dei manoscritti di Ferdinand de Saussure*, chapter Lessicografia computazionale e terminologia saussuriana, pages 161–179. Edizioni dell'Orso, Alessandria.

D. N. Smith and C.W. Blackwell. 2012. Four URLs, Limitless Apps: Separation of Concerns in the Homer Multitext Architecture. In *Donum natalicium digitaliter confectum Gregorio Nagy septuagenario a discipulis collegis familiaribus oblatum: A Virtual Birthday Gift Presented to Gregory Nagy on Turning Seventy by His Students, Colleagues, and Friends*, Cambridge MA. The Center of Hellenic Studies of Harvard University.

Tamás Váradi, Steven Krauwer, Peter Wittenburg, Martin Wynne, and Kimmo Koskenniemi. 2008. Clarin: Common language resources and technology infrastructure. In *LREC*.

Colin Ware. 2004. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.