Editors C. Boutoura, A. Tsorlini, E. Livieratos
AUTH CartoGeoLab, 2019, ISSN 2459-3893

Günther Görz[1,2], Chiara Seidl[1,2], and Martin Thiering[1,3]

Linked Biondo: Modelling Geographical Features in Renaissance Texts and Maps

Keywords: spatial cognition, text annotation/analysis, map annotation, Linked Open Data

Summary
Bibliotheca Hertziana's project "Historical spaces in texts and maps"[4] aims at a cognitive-semantic analysis of Flavio Biondo's "Italia Illustrata" (1474) linking with contemporary maps (Goerz et al., 2018). At focus are relations between historical maps and texts aiming to explore the historical understanding of space and the knowledge associated with it. Our research combines cognitive-semantic parameters such as toponyms, landmarks, spatial frames of reference, geometric relations, gestalt principles and different perspectives with computational linguistic analysis. Contributing to Spatial Humanities, we are convinced that all maps are cognitive maps, depicting culture-specific spatial knowledge and practices. Recogito is being used as the main tool for static annotations of places and persons/peoples in text and maps. These are complemented by cognitive-linguistic spatial role markups by means of the *brat* tool. To achieve a deeper and more generic semantic level of linguistic and map-related annotations, we pursue the transition to an ontology-based representation. For this purpose, we defined a domain ontology based on the event-based CIDOC Conceptual Reference Model (CRM) and its spatio-temporal extension CRMgeo in OWL-DL, and appropriate mappings to be applied to the annotations exported in CSV, RDF and JSON formats. Using the CRM opens up a wide spectrum of interoperability and linking to many web resources, such as the gazetteers being used with Recogito. For ontological enrichment and processing of the primary data, we use the Virtual Research Environment WissKI with CRM as the top conceptual model. This allows for a semantic interpretation of annotations such that, e.g., for each place we generate an instantiated CRM description in triple format, ready for storage and publication as Linked Data. In the same fashion, mappings are applied to the results of spatial role labeling: These triples encode cognitive parameters, primarily "figure - spatial_relation – ground" constructions.

Research question and methodological approach

This paper focuses on the generation of research data in a long-term research project about historical understanding of social space and its change in the so-called Renaissance. The study of the relations between historical maps and texts aims to explore the historical understanding of space and the knowledge associated with it by taking up approaches from cognitive science and in particular cognitive linguistics. Cognitive maps depict culture-specific spatial knowledge and practices. This knowledge is represented in different ways, which change historically through different processes and practices. The epistemological focus is therefore framed by questions such as which forms of knowledge represent spatial relations, how can spatial transformation processes be represented and analyzed, and what is the connection between culture-specific practices and cognitive representations? In particular, the investigation of the relation of text and maps is of specific interest. In order to approach this complex of questions, our project combines cognitive-semantic parameters such as toponyms, landmarks, spatial frames of reference, geometric relations, gestalt principles and different perspectives with computational linguistic analysis methods

[1] Bibliotheca Hertziana/Max Planck Institute for the History of Art, Rome
[2] FAU Erlangen-Nürnberg, Department of Computer Science, Digital Humanities
[3] Technical University Berlin, Department of Linguistics
[4] http://biblhertz.it/en/research/research-projects-of-the-institute/historical-spaces-in-texts-and-maps-biondo-project/

according to our "Common Sense Geography"[5] approach. Using new text and map mark-ups and corpus-specific quantitative methods, historical texts are processed and reinterpreted. The very general categorization pattern of spatial relations in visual perception is based on the gestalt theoretic figure-ground asymmetry adapted from the founders of cognitive linguistics, Langacker 2008 and Talmy 2003. This asymmetry is not only constitutive in visual perception, but also in linguistic meaning components of a sentence; hence, instead of clausal phrases such as subject and object the theory uses figure and ground. The figure is a moving or conceptually movable entity whose site, path, or orientation is conceived as a variable the particular value of which is the relevant issue. The ground is a reference entity, one that has a stationary setting relative to a reference frame, with respect to which the figure's site, path, or orientation is characterized. (Talmy 2003: 184).

Flavio Biondo's work "Italia Illustrata", published posthumously (1392-1463), serves as the first case study. Flavio Biondo is rightly regarded as the real founder of archaeological science and antiquarian topography. Particularly in his Italia Illustrata, he draws on famous authors of Roman antiquity such as Livius, Vergil and Pliny. Greek authors such as Strabon and Ptolemaios, on the other hand, are rarely considered – despite their geographical and topographical content and despite the intentions of Biondo's work. An exception is the Latium book (Regio Latina), in which Biondo intensively uses a hitherto undiscovered Latin translation of Strabon. Strabon's work is not only used here as a data basis, but also as a structural principle: He uses Strabon's hodological description technique to locate all the Latin cities relative to each other and transform them into a narrative structure on the west coast of Italy and three Roman roads (Via Appia, Via Latina, Via Valeria). This raises the question of whether the Italia Illustrata was actually written using "many maps", as current Biondo research (Clavuot 1990 and others) generally assumes? Since only a few regional maps from the early 15th century have been preserved and even literary references to cartographic knowledge can hardly be found, this problem has to be put into a larger context and possible alternatives have to be discussed in an interdisciplinary approach. Our main hypothesis based in a cognitive linguistic background is that Biondo's narrative is based on cognitive maps or mental models enabling the reader to mentally triangulate different spatial references. In particular, it is necessary to discuss which strategies Biondo has used to collect, filter and process its heterogeneous material of historical, geographical, archaeological and art-historical information and to translate it into a text that can be read by a contemporary audience. Because the Latium book plays a key role within "Italia Illustrata", the interdisciplinary project begins with the analysis and commentary of Regio V: Latina. In contrast to previous publications on Biondo, which are largely exhausted in textual criticism, biographical, literary and art-historical references as well as in his "afterlife", particular attention is paid to the identification of toponyms and the geographical vocabulary, the reception of Strabon, contemporary cartography, and mental maps.

Sources, their preparation and conditioning

First of all, there are some fundamental issues regarding the available text sources. The textual transmission of Flavio Biondo's Italia Illustrata is rather complex due to different available editions. The reason is that modern editions are based either on the extant manuscripts (Biondo/Pontari 2011-14) or the first printed edition (Johannes Philippus de Lignamine, Rome 1474), supervised by Gaspare Biondo (Biondo/White 2005, 2016), or the "best-known, most-cited early printed edition" (Froben, Basel 1559: see Castner 2005, 2010). The manuscripts, written and reworked over a long-time span, are not uniform in orthography, punctuation, or style, thus making it difficult to use the text for digital and linguistic

[5] Geus and Thiering, 2014.

analyses. According to his own confession, Gaspare changed Flavio's original for stylistic reasons before printing; and the Froben edition clearly deviates in many instances from the earlier ones. The editio princeps has been transmitted in two versions, which are, contrary to the common belief, not identical, and textual deviations from White's and Castner's texts exist which already amount to more than one hundred just in the case of the Latium chapter. In the long run, the complete text shall be investigated. We decided to work with the White edition because it is still the closest one to the editio princeps; unfortunately, his English translation is not error free and is being inspected only for heuristic reasons and used for some comparative studies.

Preprocessing steps, Word Lists, Concordance and Part-of-Speech Tagging

As a prerequisite for the analysis and interpretation of Biondo's text in the framework of cognitive semantics, we designed a workflow for text analysis which contains automatic and semi-automatic processing steps. Its ultimate goal is to annotate, that is, to mark-up individual words or word sequences such as proper names of places (and also of persons and events) in different text sources and also spatial relations in texts in the framework of cognitive semantics. Moreover, these annotations will be related to similar annotations in (historical) geographical maps.

Basically, we work with two UTF-8[6] encoded text representations, plain text and TEI/XML.[7] All texts have been generated digitally, mostly based on spelling-corrected OCR (Optical Character Recognition).

To get a first overview of the language register used in our selected text passage, we ran some simple analysis procedures on the Latin text and its English translation as a whole. As for the tools, we used some command line scripts of our own, and we applied a KWIC Concordance program[8] and the Tree Tagger with the Stuttgart-Tuebingen Tag Set (STTS)[9]. For word lists and statistics, concordances, n-grams and collocations – which are de facto word co-occurrences – also the interactive AntConc[10] tool as well as the browser-based on-line Voyant tools[11] are very useful. All results for Latin and English which we got in batch processing mode have been put up on a project web page[12] for quick lookup. They include[13]:

- alphabetic word (form) list with frequencies; endings sorted word (form) list with frequencies; ascending and descending word (form) lists with frequencies,
- concordance (KWIC), and word index for concordance,
- word list, Tree-tagged (STTS), with lemmata; word list, Tree-tagged (STTS), with frequencies, sorted by tags.

[6] UTF-8 is a Unicode character encoding; http://unicode.org/. All URLs have been checked on Mar. 9, 2019

[7] TEI: Text Encoding Initiative, providing a modular set of tags for various text sorts; http://www.tei-c.org/. Although many software tools already accept TEI encoding as a text input format, there are still some which require plain text.

[8] KWIC for Windows Version 4.7 and 5.3 by Satoru Tsukamoto; http://downloads.informer.com/kwic-concordance/download/.

[9] http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html.

[10] AntConc is an easy to use word statistics and concordance program. It has several options to support semantic analysis, e.g., to annotate clusters, to mark single lexical items embedded in their textual context etc. See: http://www.laurenceanthony.net/software/antconc/.

[11] https://voyant-tools.org.

[12] http://www.biblhertz.it/forschung/forschungsprojekte-des-instituts/historische-raeume-in-texten-und-karten-biondo-projekt/ will link to the freely available results.

[13] For English, we also used Lancaster University's wmatrix3 toolbox (http://ucrel.lancs.ac.uk/wmatrix/.), which provides a semantic tagger, too. These tools are flexible in a variety of output formats, including XML.

For the annotation of Latin still only a few computational resources and tools are available. Besides a few scanned classical lexica, there is an online version of WordNet for Latin as a lexical resource.[14] A powerful freely available lemmatizer and morphological analyser for Latin texts is Collatinus[15] in web-based and standalone versions. The Perseus Digital Library provides access to its "word study tool", an online version of a Latin morphological analyzer.[16] For the well-known Tree Tagger[17] – a tool for annotating text word by word with part-of-speech, lemma, word class, and inflection information – many language models are provided, among them Latin and English. Therefore, we decided for some steps to work in parallel language mode, that is, with Latin and English as far as appropriate tools are available, synchronized sentence by sentence. Ideally, we could synchronize glossing word by word using linguistic terminology as known from field linguistics research (Levinson and Wilkins 2006; Thiering 2015).

Our goal in the first phase of the analysis and interpretation of Biondo's text in the framework of cognitive semantics is the identification of places by toponyms or definite descriptions, and of spatial relations between them. Hence, beyond the morphosyntactic and semantic information on words as provided by the mentioned tools we would also need information about grammatical structures. Parsers are computational tools for grammatical analysis beyond the word level which provide structural data about constructions and collocations and reveal the constituent and dependency structures of sentences. In particular the latter grammatical structures are particularly important for semantic representation (cf. Fischer and Ágel 2010). For Latin, only recently a promising dependency parser has become available (Straka 2017) which we have not yet applied to the full text. For heuristic reasons we had used the Stanford Parser[18] on the English translation, a lexicalized stochastic parser whose English language model generates dependency trees of reliable and sufficient quality.

Finally, we aim for annotated logical forms which express the spatial relations of different geographical objects described in the text. Because up to now no general tools for spatial role labeling have available, we decided for a semi-automatic annotation procedure.

Toponyms and spatial relations: Text and map annotation

For semi-automatic markup of places – and also of persons and named events – in the Latin and English **texts**, we use the Recogito 2 annotation tool[19], which contains a good Named Entity Recognition component. The plain text annotation provides an integrated geographical verification mode with several gazetteers, which provide the necessary information, where for historical texts we prefer Pleiades[20], but also the Digital Atlas of the Roman Empire (DARE)[21], and GeoNames. The annotation results can be exported in several formats, e.g. as Open Annotation/RDF[22] data, as tables (CSV, "comma-separated values") – which are particularly useful for comparison and further processing –, GeoJSON[23], and also as

[14] http://multiwordnet.fbk.eu/online/multiwordnet.php.

[15] http://outils.biblissima.fr/en/collatinus/.

[16] http://www.perseus.tufts.edu/hopper/morph?redirect=true&lang=la.

[17] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/.

[18] https://nlp.stanford.edu/software/lex-parser.shtml.

[19] cf. Simon et al. (2015), http://recogito.pelagios.org/.

[20] https://pleiades.stoa.org/.

[21] http://dare.ht.lu.se/.

[22] Open Annotation Data Model, http://www.openannotation.org/spec/core/.

[23] JSON (JavaScript Object Notation) format for encoding geographic data, http://geojson.org/.

simple TEI/XML[24] files with appropriate tags that are well suited as a basis for further tagging. Furthermore, Recogito (see below and fig. 1, 3) does also allow to annotate map images and to display (annotated) places on different types of maps like OpenStreetMap[25] or DARE.

The most important step of text processing is Spatial Role Labeling, i.e., the markup of spatial object descriptors and the relations between them. For this purpose, international standard proposals have been developed (cf. Mani 2010; also, SpaceEval Annotation Guidelines[26]) and at several international computational linguistics conferences, e.g. LREC (Linguistic Resources and Evaluation Conference)[27], competitions on automatic Spatial Role Labeling by means of machine learning techniques were organized. In this case, labeled texts have to be provided as training data. Independent of our decision to apply machine learning in the long run, we would have to manually label our actual text anyway to provide a labeled training corpus of considerable size. Therefore, we decided to use the interactive *brat* rapid annotation tool[28] and provided a configuration file defining all entities, relations, and events to be annotated according to the requirements for spatial construals – that is, the ascriptions of meaning components such as figure-ground asymmetries – and its parameters. To be compatible with the subsequent step of semantic modelling, the entity types of the Pleiades vocabulary were chosen.

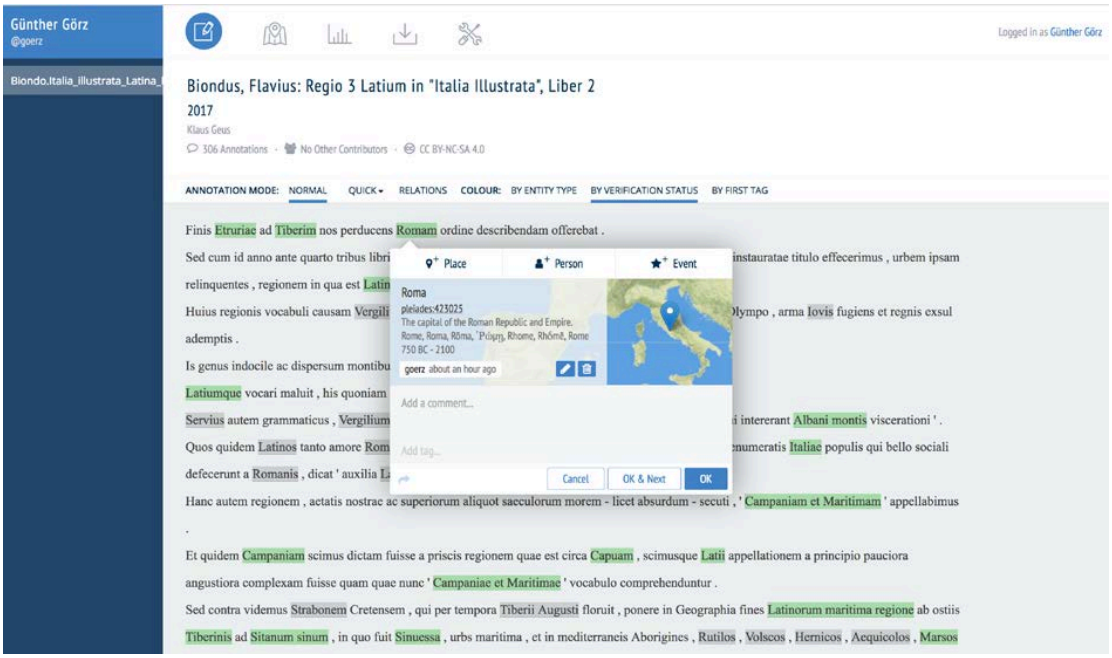

Figure 1: Text annotation with Recogito 2

The Latin text and its English translation were split into individual, aligned sentences. In *brat*, the marks and lines are entered graphically using the mouse-track function – basically one can drag from point A, a landmark, toponym or other spatial encoding, to a point B and thus related these two points to form a unit

[24] TEI: Text Encoding Initiative, https://tei-c.org/.

[25] https://www.openstreetmap.org/.

[26] http://jamespusto.com/wp-content/uploads/2014/07/SpaceEval-guidelines.pdf.

[27] https://www.cs.york.ac.uk/semeval-2013/task3/.

[28] http://brat.nlplab.org/index.html.

or construction. The results are stored in a purely text-based standoff format of which an XML version can be exported. *brat* also allows for a parallel display of aligned sentences, such that the already labeled English sentence can be shown statically together with the Latin to be labeled.

Besides the survey of text, it is also important to annotate **maps** in their relation to each other. Biondo mentions his use of (not identifiable) maps, but more discussion about the role they played for him is essential. In any case, it seems worthwhile to study maps of Italy of the fourteenth and fifteenth century in detail for a comparison of toponyms mentioned in the text and displayed on maps. We are convinced that there is a priority of texts over maps, i.e., the production of maps is in general based on texts (cf. Ptolemy's Geography, or portolan texts vs. portolan charts). In most cases, limitations of space are more restrictive for map images than for descriptions. On the other hand, visualization adds a new dimension to the understanding of geographical texts (cf. MacEachren 1995).

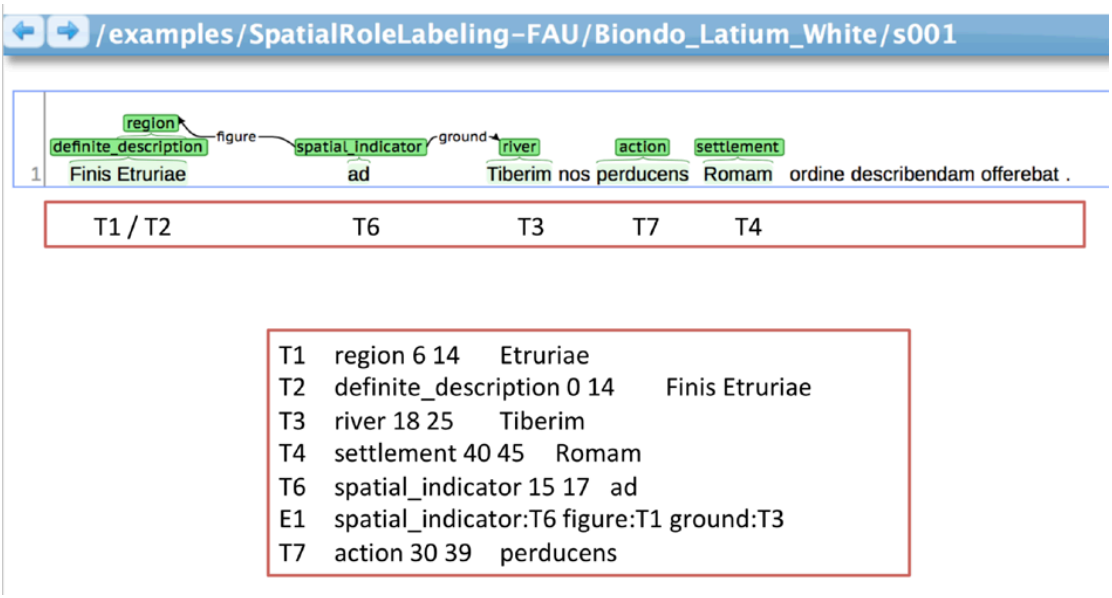

Figure 2: Example annotation with brat and its representation in standoff format

Our selection of maps for annotation comprises the earliest single maps of Italy by Paulinus Minorita (14th century), six further maps of Italy from the 15th century according to the excellent selection by Milanesi 2007/8, two relevant sections of the Tabula Peutingeriana – which show Roman streets –, some Portolan charts before 1465 for coast cities, and more than 25 Ptolemaic maps from the 15th century, traditional ones as well as *Tabulae Novae* from the redactions of Donus Nicolaus Germanus, after 1466.

In order to annotate toponyms and ethnonyms in maps the on-line tool Recogito 2 with its geographic verification mode is being used as well (fig. 3). Up to now a sufficient number of maps – Ptolemaic "traditional" and "novae" as well as others – have been annotated and their corresponding tables been analyzed, comparing the occurrences of toponyms and their spellings, and in relation to the text. Whereas toponyms on the "traditional" Ptolemaic maps correspond very well with the listing in the text (Geography, Book 3, Tabula 6), including Ptolemaic coordinates, there are no coordinate lists for the *Tabulae Novae* in editions of Ptolemy´s Geography. With the results of geographic verification, spatial relations and distances between the places can be calculated for maps and the text as well. Furthermore, we plan to visualize Biondo's imaginary routes in historical and modern maps. It is still a matter of debate

whether further investigations such as cartometric measurements would provide useful information for the interpretation. At this point we are performing some experiments with MapAnalyst[29], an image registration tool for the analysis of ancient maps.

Case Study: Ptolemaic Tabulae Novae

In a case study, we completely annotated and analyzed three *Tabulae Novae* of Italy, from the second and third redaction of Donnus Nicolaus Germanus[30]. In the course of several centuries many ancient places have been renamed or have disappeared, and many new places have originated. Therefore, it was an obvious task to "update" the maps by way of juxtaposing more recent maps to the traditional ones. Which more recent sources the *Tabulae Novae* were based upon is not known, but probably it was the same kind of sources which were also available for Biondo.

These three maps out of the *Tabulae Novae* are L20-nova[31] from the year 1466, L23-nova from 1467, and L26-nova from 1468, all ascribed to Donnus Nicolaus Germanus. Three different redactions are known today which were prepared by himself or under his supervision; whereas L20-nova and L23-nova are part of the second redaction, L26-nova is subject to the third one.

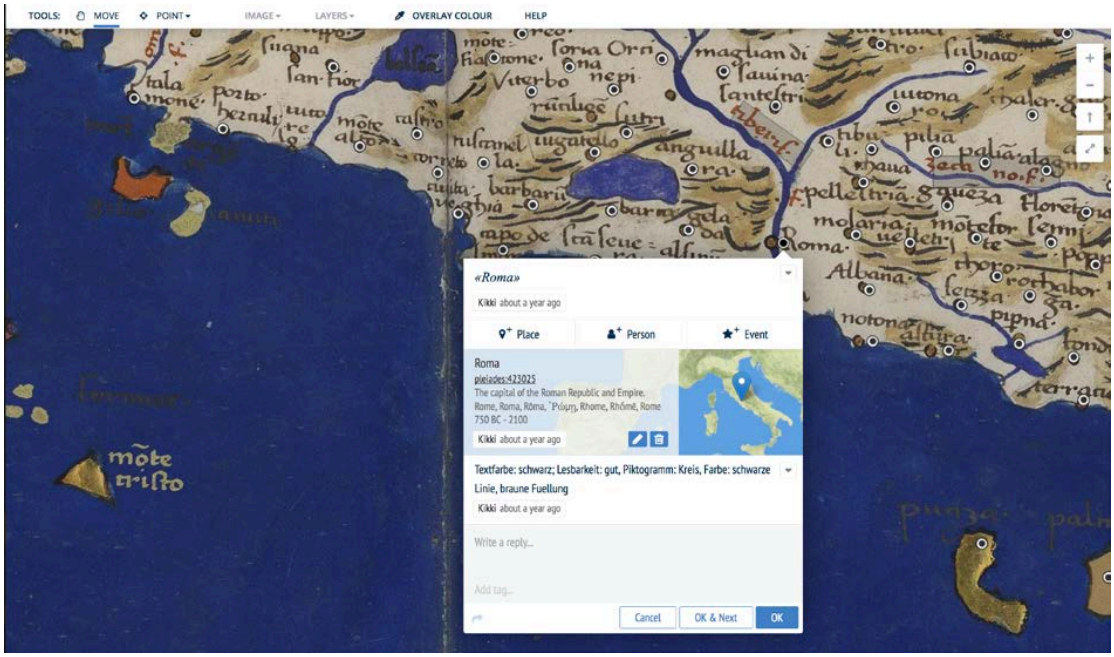

Figure 3: Map annotation with Recogito 2. A Ptolemaic Tabula Nova by Donnus Nicolaus Germanus (Fischer L23-nova), 1467

L23-nova is also a map from the second redaction, dated just one year later than L20-nova and is part of *Codex Urbinas Latinus 275* (*Bibliotheca Vaticana)*. From its external features, this map is nearly identical to the one before. Minor differences are to be found in the inscriptions. The names of the islands are

[29] http://mapanalyst.org/.

[30] cf. Gautier Dalché 2007.

[31] These identification labels are according to Joseph Fischer as quoted in: Klaus Geus, Der lateinische Ptolemaios, in: Ptolemaios et al. 2009, p. 356-364.

mostly set on the blue ground of the surrounding sea, not on the area of the islands. Due to the aging of the colors these inscriptions are more difficult to read. In general, however, both maps feature the same style of writing, so one may assign them to the same scribe. This conjecture is further confirmed by the identical use of abbreviations and diacritics. For instance, suffixes such as *–us* and *–um* are replaced by *ũ* (alsinũ = alsinum), and the tilde is set within a word to replace a following *n* (pallestrĩa = pallestrina). Both maps (L20-nova and L23-nova) don't write the usual *s*, but use the ſ instead (oſtia or aſtura).

L26-nova is a map out of Donnus Nicolaus Germanus' third redaction from 1486 and covers a double page of the *Codex Vaticanus Latinus 3811*. The information inscribed on the map is similar so far as the names of places are set horizontally in dark brown writing and combined with the same pictograms. But a distinction is made between major cities, where the names are put in majuscules, and smaller ones with their names in minuscules. The names of rivers are also set in parallel to the curvature of the river, and the names of islands are mostly placed in the blue water zone, like on map L23-nova. A golden inscription in majuscules, ITALIA, is put on the violet ground in a black framed box, the only indication of the name of the whole country in the three maps. In addition to these differences, close inspection of all the writing makes it evident that it has been done by another hand. This scribe partially uses the long *s (ſ)* (oſtia), but also the standard s when it stands as an initial (subiaco). Furthermore, suffixes are partially marked by diacritics like in the older maps (alsinũ = alsinum), partially they are written out (paliam). The tilde is used only rarely to replace the *n* (belmõte = belmonte).

Using *Recogito*, it was possible to identify further differences with reference to the names of places. L20-nova was indicated with 626 annotations, L23-nova with 647 and L26-nova with 615 annotations. For places, point annotation was used, whereas mountains, rivers or also lakes were annotated with rectangular boxes. A short description of the annotation (text color, readability, pictogram) was added and completed by geographic verification. If possible, the respective location was verified with the *Pleiades* gazetteer, in other cases also with *DARE* or, as the last option, with *Geonames*. If a location could not be verified[32], the annotation was marked with *flag this place* and a tagged with *settlemen*t, *river*, etc.

A problem in verification with a gazetteer is encountered when the available places do not refer to a city as such but to theatres, spas or other physical objects which are named by the city. In such cases it was tried, if possible, not to work with *Pleiades* but to use *DARE* or *Geonames* (as far as the city was contained there).[33] In general, by way of the annotations and verifications, it could be shown that there are several differences or a kind of development between the maps of the second and the third redaction, especially with regard to shape and handwriting. For some places, naming or spelling have changed, or they have disappeared.

a) Singular letter:

L20-nova /L23-nova: *lexi*; L26-nova: *Lesi* (pleiades: 442642)

b) *ſ* and s:

L20-nova: *chiverſ*; L23-nova: *chiaverſ*; L26-nova: *chiavers* (pleiades: 383555)

c) Tilde:

L20-nova /L23-nova: *levãto*; L26-nova: *levanto* (geonames: 3174793)

[32] Which applied to 25% in L20-nova, to 26% in L23-nova and to 24% in L26-nova.

[33] L20-nova: 63% annotated with *Pleiades*, 2% annotated with *DARE*, 35% annotated with *Geonames;*
L23-nova: 65% annotated with *Pleiades*, 2% annotated with *DARE*, 33% annotated with *Geonames;*
L26-nova: 63% annotated with *Pleiades*, 2% annotated with *DARE*, 35% annotated with *Geonames*.

d) Upper and lower case of initial letters:

   L20-nova /L23-nova: *lamatrice*; L26-nova: *Lamatrice* (geonames: 3183121)

e) Shift of lines:

   L20-nova /L23-nova: *alagno* (in one line);

   L26-nova: *alag-no* (over two lines) (pleiades: 422833)

f) Places not indicated on L26-nova:

   Sora (pleiades: 433126)*, Sausa* (pleiades: 167919)*, Fumione* (geonames: 3176486)*, Asti* (pleiades: 383669)*, Ostiglia* (pleiades: 393438)*, Frusolone* (pleiades: 432851)

g) Places not indicated on L23-nova:

   Panego (geonames: 3170937), the indication of the *Anguillare* (pleiades: 413015)

The observation that the scribe must have changed between the second and third redaction is also confirmed by the *Recogito* analysis of these three maps. In the course of reproduction of the maps during the three years considered here (1466–1468) specific changes can be recognized. Single letters were exchanged, or the ſ was replaced by the regular *s*, the use of tildes was reduced, and names of places were written out more and more. Conspicuously, some names are not indicated any more in the later version, because perhaps they could no longer be identified. In two cases where places are indicated in L20-nova and L26-nova but not in L23-nova a scribal error might be the cause, but the frequency of places which are missing in L26-nova speaks against such a general assumption.

Regarding our research question for the relation of historical spaces in text and maps we consider the use of the toponyms in both. Although the space for placing toponyms on maps is restricted it seems reasonable to compare the geographical annotations in the CSV tables of the three maps with the table of Biondo's Latium book in the *Italia Illustrata* (Biondo/White 2005/2016) that were drawn from *Recogito*. It was found that from 206 place names in the text the maps indicated 61 place names. Sixteen of these places are even identically spelled in the text and on the maps: *Astura, Alba, Civitella,* and *Ostia*[34]. The other places show minor differences in single letters or the endings, such as, e.g., *Palestrina/Pallestrina, Roccagorga/Rochaborga, Sermoneta/Sermona*. Substantial differences were also found; for four places, *Campania, S. Felice, Collona* and *Nettuno,* the names on the maps were given as *Patria, Gravina, Carrara (L20-nova)/Charrara /L23-nova/L26-nova)* and *Tareto.*

According to our present analysis, it can be stated that a considerable change in geographical knowledge occurred in the fifteenth century, which is reflected in the update of the traditional maps to "tabule novae". Also, some development has taken place in the course of the singular redactions. Places were spelled differently or disappeared in a short time. Apart from that, some coincidences were found by comparing the indication of places in the text and on the maps. Thus, it may be assumed that the redactions by Donnus Nicolaus Germanus were based on similar sources as were used by Biondo.

Generated research data and their semantic enhancement through ontology-based data modelling

As already mentioned, Recogito offers several formats for exporting text and image annotation results. Annotations of toponyms, but also persons – in our case also geographically important, because names of

[34] Astura, Alba, Ascoli, Benevento, Bologna, Civitella, Capua, Fondi, Fratta (only on L23-nova and L26-nova), Gaeta, Ostia, Ravenna, Roma, Subiaco, Tagliacozzo, Velletri (on L23-nova Velleti).

peoples represent regions – and named events can grosso modo be exported on two different levels of information:

- In RDF Open Annotation format and equivalently in JSON-LD[35], which identify the described object, the annotator, the data of annotation, the comment, and, if identified in a gazetteer, the URI of the identified place.
- As a CSV table or, equivalently, in GeoJSON, which presents the annotation content more in detail, including geographical information from the gazetteer(s).

In the latter format, in particular CSV tables, the following information is contained:

- UUID, a unique identifier for the particular annotation,
- QUOTE_TRANSCRIPTION, the textual transcription of the inscription,
- ANCHOR, the position of the annotated item in the text (character position) or image (pixel coordinates),
- TYPE, either PLACE or PERSON or EVENT
- URI, the unique web identifier of the place in a gazetteer, if the place can be geographically verified,
- VOCAB_LABEL, the spelling of the name in the gazetteer, including variants,
- LAT, geographical latitude of the place,
- LNG, geographical longitude of the place,
- PLACE_TYPE, a descriptor from a standardized vocabulary such as the Pleiades vocabulary, e.g. “settlement”, “river”, “mountain”, etc.
- VERIFICATION_STATUS, either geographically VERIFIED, or NOT_IDENTIFIABLE, if the place cannot be found in one of the available gazetteers, or UNVERIFIED, if the verification step has not been completed,
- TAGS, a free field which can be used for place types in the case of not identifiable places.
- COMMENTS.

Recently, a possibility for annotating relations between annotated entities in text has been introduced in Recogito 2. Relations can be exported as CSV tables – nodes and edges lists – ready for processing with the Gephi[36] graph visualization platform. Generally, the relations can be chosen arbitrarily, but it would be useful to have a predefined set such as in *brat* configurations and the constraints that come with them, therefore we use *brat* instead.

The cognitive linguistic annotations constitute the second kind of annotations. Based on the cognitive theoretical groundwork[37] of our project we identified a set of basic abstract spatial parameters:

- Toponyms (placenames, buildings, streets, squares, regions, etc.) and landmarks, natural vs. man-made (mountains, rivers, forests, etc.),
- Gestalt principles of figure–ground asymmetries as figure–trajectory/path [=spatial_relation]–ground triples,

[35] JSON (JavaScript Object Notation) for Linking Data, https://json-ld.org/

[36] https://gephi.org/

[37] Thiering 2015.

- Spatial frames of reference: relative/deictic, intrinsic/geometrical, absolute/allocentric,
- Topology and geometry
- Perspective: bird's/frog-eye perspective, hodological perspective, vectorial perspective,
- Distances: scale, scope, size; linguitically encoded in adjectives, adverbs, verbs, but mostly in adpositions and case systems,
- Metrical systems; encoded in verbal systems such as posture verbs and case systems,
- Motion events: source –trajectory–goal, and
- References to common sense knowledge such as itineraries or travel reports.

The annotations described so far are bound to the linguistic level, i.e., directly related to the text and map image "surface". To achieve a deeper and more generic semantic level, we pursue a transition to the methodological level of general knowledge representation.[38] So, the toponyms and other place descriptions in the cognitive-linguistic spatial role annotations – primarily "figure–spatial_relation–ground" constructions – can be identified, be enriched with general geographic information and linked to a variety of (online) resources. The semantic and epistemic level in which these representations will be anchored is given by domain models, so-called "formal ontologies",[39] which may be regarded as the conceptual kernels of appropriate domain theories. Hence, their underlying abstractions are integrated providing much more content about the conceptualization of space[40] and the geographic domain.[41]

For the conceptual framework we build upon the CIDOC Conceptual Reference Model (CRM), a fairly generic ("reference") ontology,[42] originally defined for the cultural heritage sector, and acknowledged as ISO[43] standard 21127 since 2006.[44] A decisive reason for choosing the CRM and its spatio-temporal extension CRMgeo[45] was, that being a standard it opens up a wide spectrum of interoperability and linking to many web resources. Ontological enrichment with the CRM as top conceptual model, which in its basic design is *event-based*, provides for example a generic "assignment event" which has open positions to be filled or linked with the semantic roles, resp., for agent, (material and immaterial) constituents, time-span, and place.[46]

[38] Cf., e.g., Allemang and Hendler 2011.

[39] A formal ontology defines the conceptual system of a domain of discourse; cf. Noy 2003.

[40] The concept of space is related to the question of the spatial orientation of the ancients, and more specifically to Biondo's adaption of ancient spatial thinking. Traditionally ancient geographical literature uses natural points of orientation (coast lines, rivers, mountains, winds, etc.), with which the observer may locate different directions or geographical objects. All geographical points of orientation depend on the perspective of an imaginary observer. See, e.g. papers on "Common Sense Geography" in Geus and Thiering 2014; also Dan et al. 2016.

[41] Cf., e.g., Guarino 1998 and Menzel 2002.

[42] http://www.cidoc-crm.org/.

[43] International Organization for Standardization.

[44] We implemented CRM in a Description Logic Language, the Semantic Web Ontology Language OWL-DL: Görz et al. (2008); http://erlangen-crm.org/. For OWL-DL, cf. https://www.w3.org/TR/owl2-primer/.

[45] Cf. Hiebel et al. (2016); see also http://new.cidoc-crm.org/crmgeo/sites/default/files/CRMgeo1_2.pdf.

[46] For a similar approach, see the ontological framework developed by Grossner et al. (2016), and in particular their Ontological Design Pattern for "Setting" which comes close to the definition of "Spacetime Volume" in CRM.

A remarkable feature of Recogito is that – if defined by the gazetteer – the places are further tagged with controlled terms from a thesaurus, e.g. the Pleiades vocabulary.[47] It allows for a more detailed characterization of the named place, e.g., as a settlement, a river, or a mountain, which can immediately be integrated with the ontology-based representation. The same controlled vocabulary is used for classifying the linguistically annotated entites in spatial role labeling, defined in a *brat* configuration.

With formal ontologies, we provide an answer to the question: What is the meaning of annotations? And, at the same time, in particular with the use of a standardized formal ontology like the CRM, we can directly transform the annotations into semantic representations ready for publication as Linked Open Data.

First of all, we defined a domain ontology for the description of historical maps and their content ("hmp:"), connected to the generic CRM/CRMgeo ("ecrm:" for Erlangen CRM), based on an extension of an ontology we had developed for a database of medieval maps several years ago (Görz 2007). The ontology offers a framework for the general metadata of maps and geographical texts as well as for descriptions of their content as provided by the mentioned annotations.[48] The meaning of each metadata component (property) is defined by a so-called "ontology path", i.e., a sequence of triples built from entities and properties of the ontology. As an example, in a map production event (*hmp:M9_Map_Production*) there is an actor, the Creator, defined by "*hmp:M28_Map –> hmp:A3i_was_produced_by –> hmp:M9_Map_Production –> hmp:A4_carried_out_by_map_author –> hmp:M1_Map_Author –> ecrm:P131_is_identified_by –> ecrm:E82_Actor_Appellation*". For illustration, we just present a small selection of metadata components of a historical map (*hmp:E28_Map*), in a simplified manner:

- ID: an *ecrm:E42_Identifier*
- Title: a *hmp:M2_Title*
- Creator: an *ecrm:E82_Actor_Appellation*
- Material: an *ecrm:E75_Conceptual_Object_Appellation*
- Production Place: an *ecrm:E44_Place_Appellation*
- Production Date: an *hmp:M10_Production_Date*
- Scale: an *ecrm:E75_Conceptual_Object_Appellation*

For each map we may have several images, in which depicted objects are annotated; so we have an analogous data model for images (*hmp:M34_Image*). For each annotated place (*hmp:M43_Annotated_Place* is a subclass of *crmgeo:SP6_Declarative_Place*) where the (geographical) contents of the annotations are encoded in the columns of the CSV tables, each column is transformed into a component for which similar ontology paths are defined. The annotated place is linked to the image by "*hmp:M43_Annotated_Place –> hmp:A43i_is_depicted by –> E36_Visual_Item –> P65i_is_shown_by –> hmp:M34_Image*". So, e.g., for QUOTE_TRANSCRIPTION, the path is *hmp:M43_Annotated_Place –> P48_has_preferred_identifier –> E42_Identifier*". Each annotation, represented as a line in the table, has a unique ID (UUID) and refers to a visual item (*E36*) which

[47] https://pleiades.stoa.org/docs/partners/pleiades-rdf-vocabulary. In principle, other controlled vocabularies could be considered as well, such as the ones by Getty which are available as Linked Open Data: http://www.getty.edu/research/tools/vocabularies/.

[48] For very similar work cf. Gkadolou and Stefanakis (2013) and Chalkias et al. (2017) with their Historical Map Ontology Design Pattern: http://ontologydesignpatterns.org/wiki/Submissions:HistoricalMap.

represents at least an inscription (and may in some cases also consist of an image like a wall or tower, etc.), e.g.:

- QUOTE_TRANSCRIPTION: an *ecrm:E42_Identifier*
- Type: an *ecrm:E55_Type* (here: PLACE, PERSON, or EVENT)
- URL: an *ecrm:E51_Contact_Point*
- VOCAB_LABEL: an *E44_Place_Appellation*
- LAT / LNG: *crmgeo:SP5_Geometric_Place_Expression*

There are also further data models for image series and works like map collections or atlases. Recogito offers also an export option for annotated texts in TEI/XML format. To define the semantics of TEI annotations[49] of named entities, i.e., place, person and event names in texts, mapping – defining the meaning of these tags in terms of CRM – is applied onto the respective TEI tags, as, e.g., outlined in Ore and Eide 2009.[50] Of course, in terms of Recogito annotations there is a strict equivalence between TEI tags and CSV table entries.

Linked Open Data with the Virtual Research Environment WissKI

In recent years semantic technologies have become increasingly popular to represent, manage and publish data in the humanities; therefore, Virtual Research Environments with semantic backends are used to build complex knowledge networks. Data is exposed as triples using RDF, and important vocabularies and thesauri are available as linked data. Ontologies like the CIDOC Conceptual Reference Model (CRM) are the semantic backbone of this approach and provide interoperability and data exchange beyond pure linking.

WissKI[51] is a ready-to-be-used web-based Virtual Research Environment and publishing framework that in its core relies on Semantic Web technologies to represent curated knowledge. The system enables digital humanists to produce high-quality linked data without having to cope with technical issues of the Semantic Web and ontologies and of the CIDOC CRM in particular. This is achieved by defining a mapping between traditional index card or tabular styles on the one

hand and graph-based linked data on the other. The mapping may be opaque to the users and only

be managed by a (modelling) administrator.

By default, data may be input and displayed either as structured data via forms or as free text. Free text may be input through a graphical editor and is semantically indexed in terms of named entity recognition results, calendar date specifications, mentioned events, and also technical terms as far as appropriate authority files such as gazetteers are available.

Form input provides mechanisms for error reduction by showing autocompletion hints that are backed by available authorities. From the textual annotations, RDF triples may be generated and be reused as structured data. Furthermore, the system allows the upload, derivation, display and annotation of images.

[49] For the verbal, albeit not formal definition of tags see the TEI Guidelines in http://www.tei-c.org/.

[50] As suggested by the late Sebastian Rahtz, the mappings are expressed by transformation rules, cf. Rahtz, TEI to CRM, http://tei.it.ox.ac.uk/Talks/2010-11-12-CRM/talk.pdf.

[51] http://wiss-ki.eu/ has been developed by our Digital Humanities Research Group at FAU Erlangen-Nuremberg in cooperation with the Germanic National Museum, Nuremberg and is actually used in more than 100 projects, cf. Görz and Scholz 2012.

From a technical perspective, WissKI is based on Drupal[52] ver. 8, a widely used Web Content Management System with a big and active user and developer community. It has a modular architecture and there exist a vast variety of third party extensions. Being such an extension, WissKI profits from a stable core system and also from these community contributions, providing all sorts of functionality.

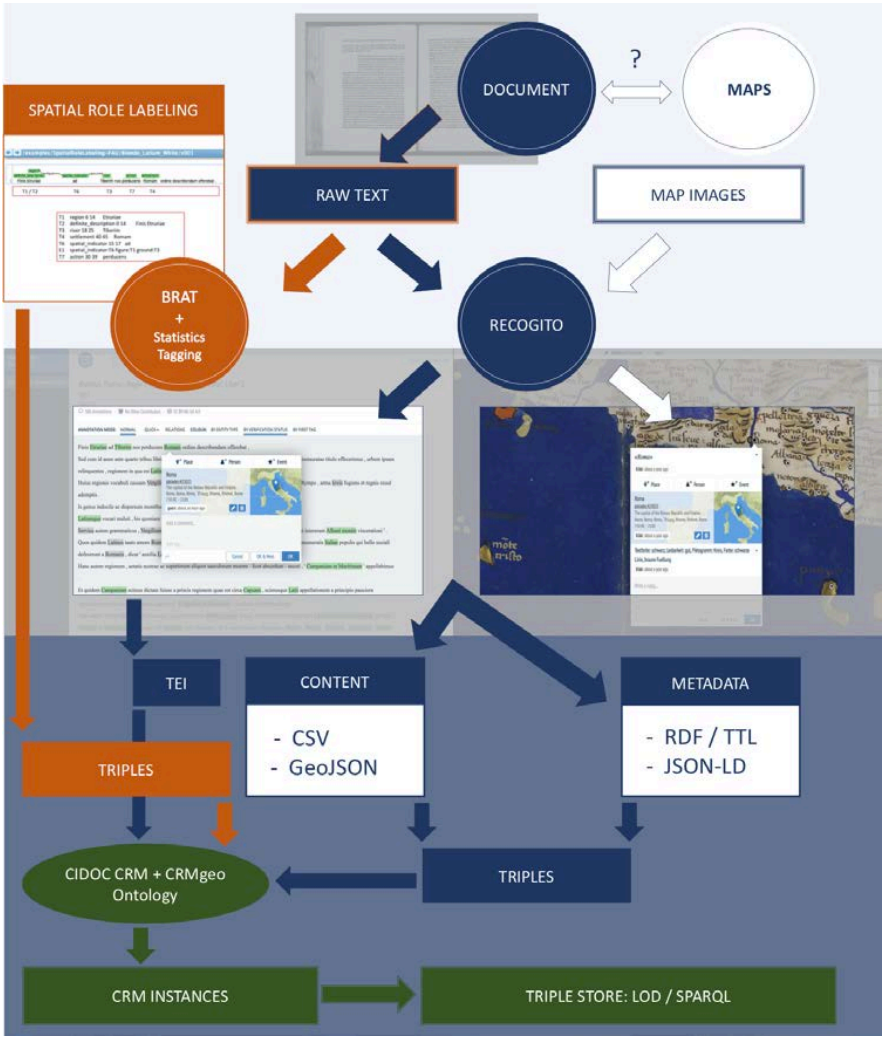

Figure 4: Generation, processing and ontological upgrading of annotation data

We decided to use WissKI for the representation, storage and access of all research data which are produced in our project. First of all, the system has to be configured with the ontology introduced above, ECRM with CRMgeo extended by hmp. Then, with its "pathbuilder" tool, all ontology paths for the metadata of images, image series, maps and works (collections) have been defined. In addition, for images the paths for the annotated content are defined in the same way; a URL provides access to the digital image. For each object type, WissKI generates an input form – and also an equivalent search form – based on the paths. Whenever value of a metadata component is entered, the underlying ontology path is instantiated and broken down into triples, which are stored in a triple database. For the transfer of the

[52] http://drupal.org/

annotations in CSV tables, WissKI provides a table input mode. Images and raw data of different types can be stored in WissKI as well, using Drupal pages for the latter ones. Hence, everything represented in the lower part of fig. 4 is subsumed by WissKI.

Using the semantically enriched geo-information from text (and map) annotations as CRM instances, the spatial entities ("figure", "ground") and relations obtained by spatial role labeling as "figure–spatial_relation–ground" triples can now be upgraded to this rich semantic level by linking data. Due to the fundamental underlying triple structure for all kinds of annotations the data are immediately ready for publication as standardized Linked (Open) Data. For this purpose, WissKI provides a SPARQL query interface. These triple data constitute a huge knowledge graph; they are the "raw material" for further research steps, i.e. the exploration of the historical understanding of spaces and the associated knowledge.

Conclusion and an outlook on cognitive maps

To conclude, we come back to our main hypothesis: Biondo's narrative is based on cognitive maps or mental models. These cognitive maps enable the reader to mentally reconstruct different spatial references or rather a spatial grid. Following the general idea that all maps are cognitive maps (Blakemore and Harley 1980), in addition to the analytic perspective described above, this idea provides us also with a synthetic view in the sense that we will use the data provided by the analytic steps to reconstruct plausible sketch maps in the near future.

We built a semi-automatic environment designed to facilitate annotating and analyzing historical texts and maps with linguistic and geographical content and outlined a new methodology based on the generated data which may help to understand and compare cognitive maps. Biondo presents a number of different spaces in his (re)construction of Italia. Our main aim is to analyze historical constructions of spaces, but also the spatial encodings from a cognitive semantic point of view. In Thiering et al. 2019, referring to Biondo's text, we present the different forms of knowledge represented in spatial relations and spatial perception as being elaborated from these representations. Biondo, in particular in many quotations of classical authors, makes frequent references to their naming of places, landmarks, toponyms and spatial relations. Facing the problem that many place names and ancient places were no longer extant or at least no longer identifiable at his time, he could not easily refer to a number of places to be used as a reference system. The text also refers to a number of different semiotic encodings such as other texts and maps. These intertextual traces are one of the major tasks to tackle spatial conceptions of the Renaissance with respect to the ancient world as the guiding spatial matrix.

Literature

Allemang, Dean and James Hendler (2011). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, Second Edition. Waltham, MA: Morgan Kaufmann/Elsevier.

Biondo/White, Jeffrey A. (edited and translated) (2005/2016). Biondo, Flavio *Italy illuminated*. Volume I, Books I–IV: 2005; Volume II, Books V–VIII, 2016. Cambridge, MA and London, England: Harvard University Press.

Biondo/Castner, Catherine J. (2005/2010). *Biondo Flavio's Italia Illustrata. Text, Translation, and Commentary, Volume I Northern Italy*. 2005. *Volume II Central and Southern Italy*. 2010. Binghamton, New York: Global Academic Publishing, Binghamton University.

Biondo/Pontari (2011–14). *Blondus Flavius: Italia Illustrata.* Pontari, Paolo (ed.). Roma: Istituto Storico Italiano per il Medio Evo.

Blakemore, Michael and Brian J. Harley (1980). Concepts in the History of Cartography - A Review and Perspective. *Cartographica*. *International Publications on Cartography,* 17/4, Monograph 26. University of Toronto Press: Toronto.

Chalkias, Christos, Christophoros Vradis and Margarita Kokla (2017). Managing geospatial information and ontologies of historical maps: Empirical evidences from the analysis of Kitchener's survey of Cyprus. In: Proceedings of the AGILE Pre-conference workshop "*Bridging space, time, and semantics in GIScience*". Wageningen, The Netherlands, 9 May 2017, 6 pp. (http://geosem.ntua.gr/images/Managing-geospatial-information-and-ontologies-of-historical-maps.pdf)

Clavuot, Ottavio (1990). *Biondos "Italia Illustrata" – Summa oder Neuschöpfung? Über die Arbeitsmethoden eines Humanisten*. Tübingen: Niemeyer.

Dan, Anca, Wolfgang Crom, Klaus Geus, Günther Görz, Kurt Guckelsberger, Viola König, Thomas Poiss and Martin Thiering (2016). *Common Sense Geography and Ancient Geographical Texts*. Berlin: eTOPOI 6, 571-597.

Fischer, K. and V. Ágel (2010). Dependency grammar and valency theory. In: *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 223-255.

Gautier Dalché, Patrick (2007). The reception of Ptolemy's Geography (end of the fourteenth to beginning of the sixteenth century). In: Harley, J.B. and David Woodward (ed.): *The History of Cartography, vol. 3: Cartography in the European Renaissance.* Chicago: Chicago University Press, 2007, Part 1, p. 285–364

Geus, Klaus and Martin Thiering (eds.) (2014). *Features of common sense geography: implicit knowledge structures in ancient geographical texts*. Zürich: Lit-Verlag.

Gkadolou, Eleni and Emmanuel Stefanakis (2013). A formal ontology for historical maps. In: Buchroithner, M. et al. (ed.), *Proceedings of the 26th International Cartographic Conference*. Dresden, 813 (16 pages online).

Görz, Günther (2007). Kognitive Karten des Mittelalters. Digitale Erschließung mittelalterlicher Weltkarten. In: Burckhardt, Daniel et al. (eds.), Geschichte im Netz: Praxis, Chancen, Visionen. Beiträge der Tagung .hist 2006, Historisches Forum: Bd. 10, I. Berlin: Clio-online und Humboldt-Universität zu Berlin, 539-572.

Görz, Günther, Martin Oischinger and Bernhard Schiemann (2008). An implementation of the CIDOC Conceptual Reference Model (4.2.4) in OWL-DL. In: *Proceedings CIDOC 2008 - The Digital Curation of Cultural Heritage.* Athens, Benaki Museum, 15.-18.09.2008. Athens: ICOM CIDOC, 1-14.

Görz, Günther and Martin Scholz (2012). WissKI: A Virtual Research Environment for Cultural Heritage. In: De Raedt, Luc et al. (Ed.): *20th European Conference on Artificial Intelligence, ECAI 2012, Proceedings*. Amsterdam: IOS Press (http://www2.lirmm.fr/ecai2012/), 2 pp.

Görz, Günther, Klaus Geus, Tanja Michalsky, and Martin Thiering (2018). Spatial Cognition in Historical Geographical Texts and Maps: Towards a cognitive-semantic analysis of Flavio Biondo's "Italia Illustrata". In: Boutoura, Ch., Tsorlini, A. (Ed.): *Digital Approaches to Cartographic Heritage.* 13th Conference of the Inter-national Cartographic Association Commission on Cartographic Heritage into the Digital, Madrid, 29-44.

Grossner, Karl, Krysztof Janowicz, and Karsten Keßler (2016). Place, Period, and Setting for Linked Data Gazetteers. In: Berman et al. *Placing Names: Enriching and Integrating Gazetteers*, 80–96.

Guarino, Nicola (1998). Formal ontology and information systems. In: Nicola Guarino (ed.). *Formal Ontology in Information Systems*. Proceedings of FOIS-98, Trento, Italy, 6-8 June 1998. IOS Press, Amsterdam, 3-15.

Hiebel, Gerald, Martin Doerr and Øyvind Eide (2016). CRMgeo: A spatiotemporal extension of CIDOC-CRM. *International Journal of Digital Libraries*, 1-9.

Langacker, Ronald. W. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Levinson, Stephen C. and David Wilkins (eds.) (2006). *Grammars of Space*. Cambridge University Press, Cambridge.

MacEachren, Alan M. (1995). *How Maps Work: Representation, Visualization, and Design*. New York and London: Guildford Press.

Mani, Inderjeet et al. (2010). SpatialML: annotation scheme, resources, and evaluation. *Language Resources and Evaluation*, 44/3, 263-280.

Menzel, Christopher (2002). Ontology theory. In: Jerome Euzenat et al. (eds.). *Ontologies and Semantic Interoperability*, Proc. ECAI-02 Workshop. CEUR-WS, 64. ECCAI, Lyon, 61-67.

Milanesi, Marica (2007/8). Antico e moderno nella cartografia umanistica: le grandi carte d'Italia nel Quattrocento. *Geographia Antiqua*, 16-17, 153-176.

Noy, Natalya (2003). Ontologies. In: Ali Farghaly (ed.). *Handbook for Language Engieers*. Stanford, CA: CSLI Publications, 181-211.

Ptolemaios, Klaudios, Alfred Stückelberger, and Florian Mittenhuber (2009). *Klaudios Ptolemaios Handbuch der Geographie: Ergänzungsband mit einer Edition des Kanons bedeutender Städte*. Basel: Schwabe Verlag.

Simon, Rainer, Elton Barker, Leif Isaksen, and Paul de Soto Canamares (2015). Linking Early Geospatial Documents, One Place at a Time: Annotation of Geographic Documents with Recogito. *e-Perimetron*, 10(2), 49–59.

Straka, Milan and Jana Straková. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In: *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies.* Vancouver, Canada: Association for Computational Linguistics, 88—99.

Talmy, Leonard (2003). *Towards a Cognitive Semantics, Vol. I+II.* Cambridge, MA: MIT Press.

Thiering, Martin (2015). *Spatial Semiotics and Spatial Mental Models: Figure-Ground Asymmetries in Language*. Berlin: De Gruyter Mouton.

Thiering, Martin, Günther Görz, Klaus Geus, and Chiara Seidl (2019). Spatial Cognition in Historical Geographical Texts and Maps: Towards a Cognitive-Semantic Analysis of Flavio Biondo's Italia Illustrata. In: Barker, Elton and Leif Isaksen (eds.). *Linking Places: Classification, Representation, and the Epistemology of Historical Geography*. Bloomington: Indiana University Press. In print.