Linking the Dictionary of Medieval Latin in the Czech Lands to the LiLa Knowledge Base

Federica Gamba[1], Marco C. Passarotti[2] and Paolo Ruffolo[2]

[1]Charles University, Faculty of Mathematics and Physics, Malostranské náměstí 25, 118 00 Prague, Czechia
[2]Università Cattolica del Sacro Cuore, largo A. Gemelli 1, 20123 Milan, Italy

Abstract

The paper presents the process of linking the *Dictionary of Medieval Latin in the Czech Lands* to the LiLa Knowledge Base, which adopts the Linked Data paradigm to make linguistic resources for Latin interoperable. An overview of the Dictionary and of the architecture of the LiLa Knowledge Base is first provided; then, the stages of the process of linking the *Dictionary* to LiLa's collection of lemmas are described. In conclusion, a query illustrates how interoperability allows for full exploitation of Latin resources.

Keywords

Linked Data, LiLa Knowledge Base, dictionary, Medieval Latin

1. Introduction

Many resources are available for Latin, making it a particularly privileged language among the historical ones. However, most often those resources are scattered, with their sparsity representing a substantial hindrance to the full exploitation of the information they contain. To overcome the sparsity of resources, stored in separate silos, the CIRCSE Research Center in Milan, Italy, started the LiLa - Linking Latin project[1] (2018-2023), which built a Knowledge Base to make all existing textual and lexical resources for Latin interoperable by adopting the four principles of the Linked Open Data (LOD) paradigm [1]: 1) use URIs as names for things; 2) use HTTP URIs so that people can look up those names; 3) when someone looks up a URI, provide useful information; 4) include links to other URIs. so that they can discover more things.[2]

The LiLa Knowledge Base has already a wide coverage in terms of interlinked resources. Classical Latin is naturally well-represented, as proved by the LASLA corpus, which includes 130 Classical Latin texts [2], and by the Lewis and Short dictionary [3], whose primary focus is on Classical Latin. Later stages of Latin are found as well in the Knowledge Base; for instance, the *Index Thomisticus* Treebank [4] comprises texts by Thomas Aquinas (1225–1274), the UDante treebank [5] encompasses Medieval Latin works written by Dante Alighieri, and the Computational Historical Semantics Corpus [6] includes e.g. the *Decretum Gratiani*, a collection of canon law compiled in the XII century.

However, while the LiLa Knowledge Base already extends over a large temporal range, its spatial coverage is not as wide. So far, no resource from the Eastern Europe areas where Latin was spoken has been linked. For this reason, we decided to link to LiLa the *Dictionary of Medieval Latin in the Czech Lands*, a lexical resource that aims at collecting the Latin vocabulary as it emerged in that area during the Middle Ages. The resource encompasses a late variety of Latin (1000-1500 CE), strongly tied to a specific geographical area. These two levels of variability, along the temporal and spatial axes, make it extremely interesting to link such a resource to the Knowledge Base, as we expect it to contribute to enlarge the amount of lemmas stored in the large collection of Latin lemmas that represents the core part of the whole architecture of LiLa.

The paper is organised as follows. Section 2 introduces the LiLa Knowledge Base. Section 3 describes the *Dictionary*. Section 4 outlines the process of linking the *Dictionary* to LiLa. Section 5 shows the added value of interoperability of Latin resources in LiLa by presenting a query on the *Dictionary* interlinked.

2. The LiLa Knowledge Base

The LiLa Knowledge Base [7] achieves interoperability between linguistic resources for Latin, by adopting a set of ontologies widely used to model linguistic information, as well as Semantic Web and Linked Data standards. Among the former, OLiA is used to model linguistic annotation [8], Ontolex-Lemon for lexical data [9, 10] and POWLA for corpus data [11]. As for the latter, the Resource Description Framework (RDF) [12] is a data model

CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 — Dec 02, 2023, Venice, Italy
gamba@ufal.mff.cuni.cz (F. Gamba); marco.passarotti@unicatt.it (M. C. Passarotti); paolo.ruffolo@unicatt.it (P. Ruffolo)
0000-0003-3632-0594 (F. Gamba); 0000-0002-9806-7187 (M. C. Passarotti); 0000-0002-9120-0846 (P. Ruffolo)

[1]https://lila-erc.eu/.
[2]https://www.w3.org/wiki/LinkedData.

used to describe information in terms of triples, consisting of: (1) a predicate-property that connects (2) a subject (i.e. a resource) with (3) its object (another resource or a literal). Data recorded in the form of RDF triples are queried via the SPARQL query language [13].

The architecture of the LiLa Knowledge Base is highly lexically-based, as it exploits the lemma as the most productive interface between resources and tools. Indeed, its core is the so-called Lemma Bank, a collection of around 200,000 lemmas taken from the database of the morphological analyser LEMLAT [14] and constantly extended. A `lila:Lemma`[3] is a subclass of `ontolex:Form`[4], whose individuals are the inflected forms of a lexical item. In particular, the lemma is a form that can be linked to a `ontolex:lexicalEntry`[5] via the property `ontolex:canonicalForm`[6], which identifies the form that is canonically used to represent a lexical entry.

To overcome divergent lemmatisation criteria that may possibly be adopted in resources, LiLa exploits three key properties. The symmetric property `lila:lemmaVariant`[7] connects different forms of the same lexical item that can be used as lemmas for that item, like for verbs with an active and a deponent inflection (e.g., *sequo* and *sequor* 'to follow'). The property `ontolex:writtenRep`[8] registers different spellings or graphical variants of one lemma, like for instance *conditio* and *condicio* 'condition'. For forms that can be reduced to multiple lemmas like participles – that can be considered either part of the verbal inflectional paradigm or as independent lemmas – a special sub-class of `lila:Lemma` called `lila:hypolemma`[9] is defined.

3. The Dictionary of Medieval Latin in the Czech Lands

The *Dictionary of Medieval Latin in the Czech Lands*[10] is a lexical resource developed at the Department of Medieval Lexicography of the Institute of Philosophy of the Czech Academy of Sciences. It aims to collect the vocabulary of Medieval Latin as it was used in the Czech lands from about 1000 CE, when Latin writing began in the area, to 1500 CE. In light of this aim, the *Dictionary* features three types of entries:

- Vocabulary taken from Classical Latin without any semantic change in the Middle Ages. Only source citations with a translation illustrate its meaning. E.g., *labellum* 'small lip'.
- Vocabulary taken from Classical Latin with changes. This type of entry is composed of two parts: first, ancient meanings are listed; then, the + sign introduces Medieval developments (syntactical alternations, new phrases, meanings of the word coined in Medieval times). E.g., *falcatus* 'curved' + 'shod'.
- Vocabulary that emerged during the Middle Ages. Such entries are marked with an asterisk (∗). Square brackets [ ] with etymology and references to other dictionaries including the word follow the heading of the entry. E.g., *emicamen* 'splendour, clarity'.

Moreover, the *Dictionary* relies on a differential method to capture all divergences – at several linguistic layers – of Medieval Latin vocabulary inherited from the ancient era as compared with the Classical norms. Indeed, it records language phenomena not attested in the 8th edition (and later unchanged editions) of Georges' Latin-German Lexicon [15].

The material the *Dictionary* is built upon amounts today to ca. 800,000 excerpt sheets, assembled from various sources of Czech provenance (diplomatical, official, belles-lettres, scientific literature, etc.). What is particularly valuable is that not only edited texts served as a source to build the *Dictionary*, but also several manuscripts and old prints from Czech and foreign libraries were used. The excerption of sources has been carried out from 1934, when the project of the *Dictionary* started, until the 1970s. In 1977 the first fascicle was published, illustrating editorial principles and lists of sources and abbreviations. Overall, the electronic database [16] is built upon, and comprises, the three volumes prepared by Silagiová and colleagues ([17], [18], [19]).

So far, letters A-M are covered, for a total of 48,452 entries. 24,943 out of these are full entries (provided with meanings, definitions, grammatical information, examples), whereas 23,509 are references that point to full entries (see 3.1). Fascicle 24, encompassing entries beginning with N, is currently under preparation.

The *Dictionary* is accessible through a dedicated website[11] and can be downloaded from the LINDAT/CLARIAH-CZ research infrastructure[12] as a compressed set of XML files.

3.1. XML Files

We provide a brief overview of the structure of the XML files of the *Dictionary*, as those data are relevant for the process of modeling information and linking the entries

[3] https://lila-erc.eu/lodview/ontologies/lila/Lemma.
[4] http://www.w3.org/ns/lemon/ontolex#Form.
[5] http://www.w3.org/ns/lemon/ontolex#LexicalEntry.
[6] http://www.w3.org/ns/lemon/ontolex#canonicalForm.
[7] http://lila-erc.eu/ontologies/lila/lemmaVariant.
[8] http://www.w3.org/ns/lemon/ontolex#writtenRep.
[9] https://lila-erc.eu/lodview/ontologies/lila/Hypolemma.
[10] The Czech title is *Slovník středověké latiny v českých zemích*; the Latin one *Latinitatis medii aevi lexicon Bohemorum*.
[11] http://lb.ics.cas.cz.
[12] http://hdl.handle.net/11234/1-4792.

to the LiLa Knowledge Base. The lexical entry for the adjective *exquisitus* 'exquisite' (Figure 1) will serve as an example of the XML files of the resource.

The whole entry is encoded as the value of an `entryFree` element, which contains a single unstructured entry in any kind of lexical resource, such as a dictionary or lexicon. Core information about the entry is provided through attributes: the lemma is given, together with a numerical unique identifier assigned to it; `georges='True'` or `'False'` specifies whether an entry for the same lemma is found or not in Georges' dictionary. Optionally, `hom_nr` distinguishes homographs, and `type='reference'` denotes that the entry is just a reference to a different one; for instance, the dummy entry for *geniculor* 'to bend the knee' is just a reference to its active counterpart *geniculo*, which, in light of that, is the only full entry of the two (with meanings, grammatical information, etc.). Then, in the `orth` element the lemma is stated once again as a value; `orth` includes the attribute `type` either with value 'lemma', if it is a full entry, or with value 'ref_all', if it is a reference.

Following the lemma, the `gramGrp` element encodes grammatical information about the lexical item, roughly corresponding to its Part of Speech (POS) and (possibly) its inflectional category. In the case of *exquisitus*, the value `<gramGrp> 3. </gramGrp>` indicates that it is an adjective of the first class, i.e. with three distinct endings for the three genders (*exquisitus, -a, -um*, respectively for the forms of masculine, feminine and neuter singular nominative).

The `sense` elements (possibly more than one for a same entry) capture the different meanings of a lexical item. For each `sense`, a definition `def` is provided both in Latin and in Czech, with the Czech one corresponding to a translation of the Latin counterpart. Some examples are listed as well, together with their source. The label *script. et form.* is used to record orthographic and morphological variants (e.g., *exequisitus* for *exquisitus*), while the label *metr.* for metrical ones.

4. Linking the Dictionary to LiLa

This section describes the process of linking the *Dictionary of Medieval Latin in the Czech Lands* to the LiLa Knowledge Base. The coverage of the linking task is not yet complete, as, so far, we have been working only with full entries (i.e., excluding those with `type='reference'`).

As mentioned in Section 2, in LiLa the lemma works as interface between interlinked resources. In light of the pivotal use of lemmas, the core operation at the base of the linking process is to perform a string match between the tuples *(lemma, POS)* in the resource to be linked and the lemmas and their POS in the LiLa Lemma Bank. The goal is to retrieve the correct lemma in the Lemma Bank corresponding to the lemma/POS used in the entry of the *Dictionary*.

The string match results in three possible outcomes: a) only one matching lemma/POS is found in the Lemma Bank; b) more than one matching lemma/POS is found, resulting in an ambiguity due to homography; c) no matching lemma/POS is found, as the couple is not present in the Lemma Bank.

The first outcome is overall straightforward and does not raise particular issues. The second one, i.e. multiple matches found, requires disambiguation to be performed. To this aim, grammatical information (inflectional classes) can be exploited, although they do not always guarantee a full resolution of the ambiguity; Subsection 4.1 elaborates on this. The third possibility, i.e. missing matches, represents the most interesting outcome; firstly, because it entails enlarging the Lemma Bank with new canonical forms of citation, and secondly because it allows to reflect about the peculiar aspects of the variety of the Latin vocabulary represented in the *Dictionary*, by focusing on those lexical items provided by the *Dictionary* that result as out-of-vocabulary with respect to the current Lemma Bank of LiLa.

4.1. Aligning Grammatical Information

In order to automatically disambiguate multiple matches, we exploit the grammatical information provided by the *Dictionary* in the `gramGrp` element. However, this information is not encoded in a fully standardised way, thus requiring an alignment to be performed. Indeed, we need to define a set of heuristics to align grammatical categories as they are encoded in the *Dictionary* and the set of tags employed in LiLa, which is based on the Universal POS tagset [20] and expanded with inflectional categories. As an illustration, the word *acus* 'needle' has *-us, f.* as `gramGrp`, i.e. the genitive ending and the gender. From that we can generalise and establish a correspondence between the genitive ending in *-us* together with the gender, as found in the *Dictionary*, and a NOUN with inflectional class `n4`[13] in LiLa.

In most cases, grammatical information provided by the *Dictionary* is sufficiently fine-grained to provide all elements needed to disambiguate the multiple linking to the Lemma Bank, as it roughly consists of POS and inflection class, like in the case of *acus*. Yet, sometimes only information corresponding to POS is available. Several substantives are marked just as *subst.* (e.g., *deptar*, type of medicinal plant), which makes it non-trivial, if possible at all, to infer an inflectional category.

[13] `n4` corresponds to fourth declension nouns.

```
<?xml version='1.0' encoding='utf8'?>
<entryFree   georges='True' lemma='exquisitus' n='263390'>
<orth type='lemma'>exquisitus</orth>
<gramGrp><norm/>3.<norm_end/></gramGrp>

<form><norm/>exequ- <bibl type='source' index_as='source'>LupCus 44</bibl><norm_end/></form>

<sense georges='false'>
<sense type='hier' n='a'>
<sense type='expl'>
<def lang='lat' index_as='definition-lat'><i/>electus, egregius <i_end/></def>
<def lang='cs' index_as='definition-cze'><i/> - vybraný, vynikající<i_end/></def>

</sense>
</sense>
<sense type='hier' n='b'>
<sense type='expl'>
<def lang='lat' index_as='definition-lat'><i/>quaesitus, singularis, insolitus <i_end/></def>
<def lang='cs' index_as='definition-cze'><i/> - hledaný, zvláštní, neobvyklý:<i_end/></def>

<div type='examples'>
<cit type='example' index_as='example'><norm/>deliciosi cibi e-i wymysleny <bibl type='source' index_as='source'>HusBethl II 75</bibl>.<norm_end/></cit>

</div>

</sense>
</sense></sense>

<ab type='formated'><b/>exquisitus<b_end/><norm/>  3.   <norm_end/><i/>script. et form.:<i_end/><norm/> exequ- |LupCus 44| <norm_end/><b/>  a<b_end/><no

</entryFree>
```

Figure 1: XML file of the *Dictionary* entry *exquisitus* 'exquisite'.

4.2. Linking to the Lemma Bank

After aligning the two tagsets, we proceed to link the *Dictionary* entries to the Lemma Bank. The one-to-one matches, i.e. lemmas in the *Dictionary* that match with just one lemma in the LiLa Lemma Bank with respect to both lemma and POS, have been considered validated. The following subsections discuss the two other scenarios, namely one-to-many and one-to-zero matches.

4.2.1. One-to-Many

The string match on lemma and POS results in 827 ambiguous matches. Therefore, we add inflectional class as a further constraint; as a result, 303 lemmas are disambiguated automatically, while 445 still remain ambiguous and need to be inspected manually. For instance, for *lacertus* a correspondence in the Lemma Bank is found with *lacertus* 'upper arm' and *lacertus* 'lizard; a seafish', both NOUNs of the second declension (inflectional class n2). Only the manual checking of the meaning can thus allow to retrieve the correct match.

4.2.2. One-to-Zero

After performing the string match on lemma and POS, no match is found in the LiLa Lemma Bank for 10,278 lemmas. Among those, we automatically handle adverbs, verbs and *pluralia tantum* to find out whether they could be linked to the Lemma Bank respectively as hypolemmas of an adjective, lemma variants of a corresponding verb with opposite voice (active if deponent and vice versa) or lemma variant of a noun in singular form. By defining a set of heuristics applied automatically, we find that: (a) 92 adverbs can be linked to the adjective they are derived from (e.g., *homagialiter* - *homagialis* 'of homage'); (b) 18 verbs can be linked to their counterpart with opposite voice (e.g., *attaedio* - *attaedior* 'to bore'); (c) 80 plural forms can be linked to their singular equivalent (*moscilli* - *moscillus* 'little habit').

A closer look at lemmas that remain unmatched (10,088) raises interesting insights, allowing for some linguistic considerations. First, clear evidence of areal contact is provided by forms like *bosako, -onis* and *kamennikko, -onis*. As the spelling reveals, these forms are the result of a contact with the language that was spoken in the area at that time, namely Old Czech. Indeed, *bosako* comes from the Czech form *bosák*, denoting a monk that by virtue of the rule has to walk barefoot, while *kamenniko* 'stonemason' derives from *kameník*. Additionally, several lemmas pertain to very specific domains. Consider e.g. *ascoa*, a sea animal, *igenecha*, a type of quadruped[14], or *cinapus*, a species of fish, as an example of vocabulary of fauna. Flora is found as well: e.g., *elipurgis*, corresponding to *Cynoglossum officinale*, *bulboquilon*, 'mandrake', and *atomana*, a herb. Similar forms evidently display the specificity of some domains covered by the *Dictionary of Medieval Latin in the Czech Lands*.

4.3. Results

The string match on lemmas and POS tags results in 55.5% one-to-one mappings; for 3.3% of entries more than one possible match was found, while for 41.2% no match was retrieved. The amount of lemmas that are not found in the Lemma Bank reflects the nature of the *Dictionary*, and especially its temporal, geographical and domain specificity. For comparison purposes, consider, for instance, that the process of linking the bilingual Latin-English

[14]Possibly the common genet.

dictionary by Lewis and Short, which is focused on Classical Latin, resulted in only 9% of unmatched lemmas [21]. The percentage of no-match entries increases to 70% in the case of the *Neulateinische Wortliste* by Ramminger [22], which covers a time range spanning between 1300 and 1700 and features entries mirroring contemporary changes in the society, e.g. *typographus* 'typographer'.

Figure 2 shows an example of an entry of the *Dictionary* (*exquisitus*) linked to the LiLa Knowledge Base. The (yellow) node in the center of Figure 2 is the `ontolex:lexicalEntry` for *exquisitus*, which is linked via the property `lime:entry`[15] to the node that represents the entire *Dictionary* (an individual of the class `lime:lexicon`[16]) and to the corresponding lemma in the Lemma Bank via the property `ontolex:canonicalForm`. The lexical entry works as gateway to all information associated to it in the resource. For instance, Figure 2 shows how the two meanings associated to *exquisitus* in its entry in the *Dictionary* are modeled. The two definitions provided by the resource (in Latin and in Czech) are linked to the lexical entry as individuals of the class `ontolex:lexicalSense`[17] via the property `ontolex:sense`[18]. Each sense is the specific lexicalisation of a more general `ontolex:lexicalConcept`[19] to which the sense is linked via the property `ontolex:isLexicalizedSenseOf`[20].

Although not visible in Figure 2, the lemma *exquisitus* in the Lemma Bank is linked via `ontolex:canonicalForm` to the entries for *exquisitus* in several other lexical resources and to its occurrences (tokens) in the textual resources interlinked in LiLa[21] .

5. Querying the Dictionary in LiLa

This Section presents a query to exemplify the added value of interoperability between the resources linked to LiLa[22]. The query, available within a set of preocompiled queries in the SPARQL endpoint of LiLa, retrieves all those lemmas whose entries in the *Dictionary* include the word *natura* 'nature' in their definition(s) and do not occur also in the Lewis and Short dictionary, and returns the number of their occurrences in the textual corpora linked to LiLa.

The 11 retrieved lemmas[23] occur in 5 corpora, for a total of 132 occurrences, 5 out of which are found in the Computational Historical Semantics corpus[24] , 104 in the *Index Thomisticus* Treebank,[25] 4 in UDante,[26] 1 in the CIRCSE Latin Library[27] (specifically, in Augustine's *Confessiones*) and 18 in the LASLA corpus[28] [2]. The results of the query confirm once again the specificity of the *Dictionary of Medieval Latin in the Czech Lands*. Having excluded Classical lemmas that can also be found in the Lewis and Short dictionary, what remains are mostly lemmas that occur in corpora featuring texts of later stages of Latin: for instance, the texts from the *Index Thomisticus* Treebank and UDante date back respectively to XIII and XIV centuries. The only exception is represented by the LASLA corpus, which includes Classical Latin. Yet, occurrences in LASLA are limited to the lemma *mollitia* 'softness, weakness', which is therefore attested in Classical times as well, while all the other lemmas appear to have originated later.

6. Conclusions

Linking the *Dictionary* to the LiLa Knowledge Base not only was a further step towards the full exploitation of linguistic resources for Latin, thanks to their interoperability, but also contributed to improve the degree of linguistic diversity represented in LiLa as for three aspects, that are particularly relevant for Latin as a language that was used for centuries all over Europe: (a) diachronic diversity: the *Dictionary* collects a portion of the Latin vocabulary that emerged in Medieval times; (b) diatopic diversity: the lexical resource includes items from a specific area, namely the Czech lands; (c) domain-based diversity: quite frequently the entries of the *Dictionary* belong to very specific domains (e.g., flora and fauna; see Section 4.2.2). The contribution of the lemmas from the *Dictionary* in enlarging the LiLa Lemma Bank is thus considerable both in terms of quantity and in terms of quality, and highlights the importance of linking to the Knowledge Base also resources that feature non-standard varieties of Latin.

In the near future, we intend to finalise the linking, by disambiguating ambiguous matches and adding missing lemmas to the Lemma Bank, as well as by including referencing lemmas besides full entries (see Section 3.1). We also intend to model citations of attestations, i.e. refer-

[15] http://www.w3.org/ns/lemon/lime#entry.
[16] http://www.w3.org/ns/lemon/lime#Lexicon.
[17] http://www.w3.org/ns/lemon/ontolex#LexicalSense.
[18] http://www.w3.org/ns/lemon/ontolex#sense.
[19] http://www.w3.org/ns/lemon/ontolex#LexicalConcept.
[20] http://www.w3.org/ns/lemon/ontolex#isLexicalizedSenseOf.
[21] For the full list of the resources currently made interoperable through LiLa, see https://lila-erc.eu/data-page/.
[22] The linguistic resources for Latin linked in LiLa can be queried either via a query graphical interface (https://lila-erc.eu/query/) or through a SPARQL endpoint (https://lila-erc.eu/sparql/).
[23] *Accidentalis*, *bestialitas*, *connaturalis*, *connaturalitas*, *contingentia*, *eligibilis*, *finitas*, *fumositas*, *leuiathan*, *materialitas*, *mollitia*.
[24] http://lila-erc.eu/data/corpora/CompHistSem/id/corpus.
[25] http://lila-erc.eu/data/corpora/ITTB/id/corpus.
[26] http://lila-erc.eu/data/corpora/UDante/id/corpus.
[27] http://lila-erc.eu/data/corpora/CIRCSELatinLibrary/id/corpus. Collection of Latin texts enhanced with different layers of linguistic annotation.
[28] http://lila-erc.eu/data/corpora/Lasla/id/corpus.

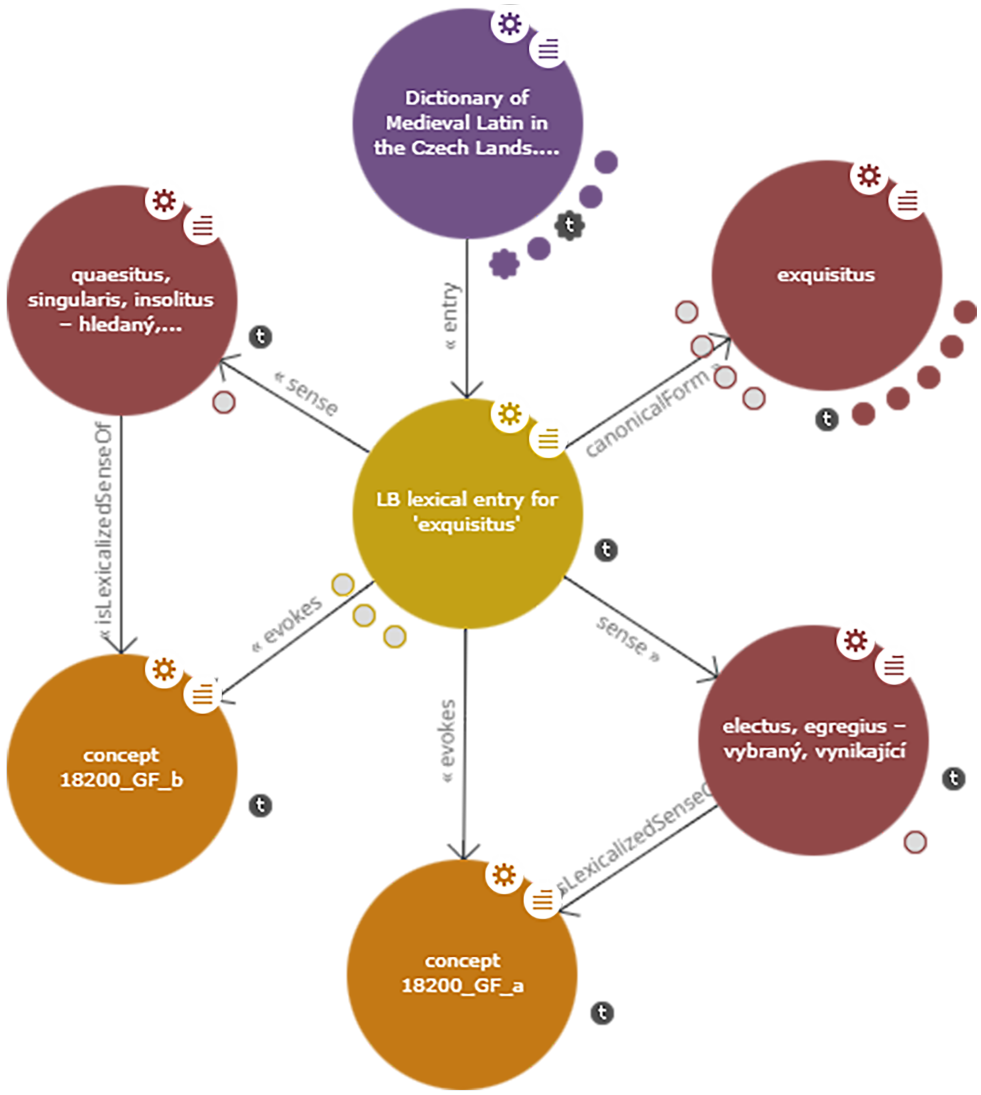

Figure 2: The entry for *exquisitus* after being linked to LiLa.

ences to other dictionaries where an entry is found and to sources of examples. Moreover, we plan to link to the Knowledge Base some documents from the same area and period as the *Dictionary*, such as the Czech Medieval sources from the AHISTO project[29]. However, these documents are currently available only as raw texts, and would need to be lemmatised before the linking. Given the peculiar nature of their Latin variety, conditioned by the Czech language and rich of local proper names, lemmatisation with the currently available trained models will probably provide low accuracy rates. Once again, this proves the importance of collecting non-standard Latin data (and resources) and investigating to what ex-

[29] https://nlp.fi.muni.cz/projekty/ahisto/portal.

tent Latin varieties differ.

Acknowledgments

The "LiLa - Linking Latin" project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme – Grant Agreement No. 769994. This work was partially supported by the Grant No. 20-16819X (LUSyD) of the Czech Science Foundation (GACR).
We want to thank Pavel Nývlt for his collaboration in providing the database "The Dictionary of Medieval Latin in Czech Lands", available via the LINDAT/CLARIAH-CZ Research Infrastructure, supported by the Ministry of Education, Youth, and Sports of the Czech Republic (Project No. LM2018101).

References

[1] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific american 284 (2001) 34–43.

[2] M. Fantoli, M. Passarotti, F. Mambrini, G. Moretti, P. Ruffolo, Linking the LASLA Corpus in the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin, in: Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 26–34. URL: https://aclanthology.org/2022.ldl-1.4.

[3] C. T. Lewis, C. Short, A Latin Dictionary, Clarendon Press, Oxford, 1879.

[4] M. Passarotti, The Project of the Index Thomisticus Treebank, in: M. Berti (Ed.), Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution, De Gruyter, Berlin, 2019, pp. 299–319.

[5] F. M. Cecchini, R. Sprugnoli, G. Moretti, M. Passarotti, UDante: First Steps Towards the Universal Dependencies Treebank of Dante's Latin works, in: Seventh Italian Conference on Computational Linguistics, CEUR-WS.org, Bologna, 2020, pp. 1–7.

[6] T. Geelhaar, A. Mehler, B. Jussen, A. Henlein, G. Abrami, D. Baumartz, T. Uslu, et al., The Frankfurt Latin Lexicon from Morphological Expansion and Word Embeddings to Semiographs, Studi e saggi linguistici 58 (2020) 45–81. doi:10.4454/ssl.v58i1.276.

[7] M. Passarotti, F. Mambrini, G. Franzini, F. M. Cecchini, E. Litta, G. Moretti, P. Ruffolo, R. Sprugnoli, Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin, Studi e Saggi Linguistici 58 (2020) 177–212.

[8] C. Chiarcos, M. Sukhareva, Olia–ontologies of linguistic annotation, Semantic Web 6 (2015) 379–386.

[9] P. Buitelaar, P. Cimiano, J. McCrae, E. Montiel Ponsoda, T. Declerck, Ontology lexicalisation: The lemon perspective, in: Proceedings of theWorkshops 9th International Conference on Terminology and Artificial Intelligence, 2011, pp. 33–36. URL: http://tia2011.crim.fr/Workshop-Proceedings/TIAW-2011.pdf, ontology Engineering Group - OEG.

[10] J. P. McCrae, J. Bosque-Gil, J. Gracia, P. Buitelaar, P. Cimiano, The Ontolex-Lemon model: development and applications, in: Proceedings of eLex 2017 conference, 2017, pp. 19–21.

[11] C. Chiarcos, POWLA: Modeling linguistic corpora in OWL/DL, in: The Semantic Web: Research and Applications: 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012. Proceedings 9, Springer, 2012, pp. 225–239.

[12] O. Lassila, R. Swick, Resource Description Framework (RDF) model and syntax specification, 1998. Available online at https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/.

[13] E. Prud'Hommeaux, A. Seaborne, SPARQL query language for RDF, W3C working draft 4 (2008).

[14] M. Passarotti, M. Budassi, E. Litta, P. Ruffolo, The Lemlat 3.0 package for morphological analysis of Latin, in: Proceedings of the NoDaLiDa 2017 workshop on processing historical language, 2017, pp. 24–31.

[15] K. E. Georges, H. Georges, Ausführliches lateinisch-deutsches Handwörterbuch, 2 vols, Hannovre/Leipzig: Hahnsche Buchhandlung (1913).

[16] J. Ctibor, P. Nývlt, On-line Dictionary of medieval Latin in the Czech lands, 2021. URL: http://hdl.handle.net/11234/1-4792, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

[17] Z. Silagiová, J. Černá, H. Florianová, P. Nývlt, H. Šedinová, K. Vršecká, Latinitatis medii aevi lexicon Bohemorum – The Dictionary of Medieval Latin in Czech Lands, Volume I (A-C), second, revised edition, 2018.

[18] Z. Silagiová, P. Nývlt, J. Černá, H. Florianová, B. Kocánová, H. Šedinová, K. Vršecká, Latinitatis medii aevi lexicon Bohemorum – The Dictionary of Medieval Latin in Czech Lands, Volume II (D-H), second, revised edition, 2019.

[19] Z. Silagiová, J. Černá, D. Martínková, B. Kocánová, M. Koronthályová, K. Vršecká, R. Mašek, J. Matl, H. Miškovská, P. Nývlt, H. Šedinová, I. Zachová, Latinitatis medii aevi lexicon Bohemorum – The Dictionary of Medieval Latin in Czech Lands, Vol-

ume III (I-M), 1995 to 2016. The electronic version has been created by Jan Ctibor.

[20] S. Petrov, D. Das, R. McDonald, A Universal Part-of-Speech Tagset, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 2089–2096. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/274_Paper.pdf.

[21] F. Mambrini, E. Litta, M. Passarotti, P. Ruffolo, Linking the Lewis & Short Dictionary to the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin, in: Proceedings of the Eighth Italian Conference on Computational Linguistics (CLiC-it 2021). Milan, Italy, January 26-28, 2022, 2021, pp. 214–220.

[22] F. Iurescia, E. Litta, M. Passarotti, M. Pellegrini, G. Moretti, P. Ruffolo, Linking the Neulateinische Wortliste to the LiLa Knowledge Base of Interoperable Resources for Latin, in: Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, Association for Computational Linguistics, Dubrovnik, Croatia, 2023, pp. 82–87. URL: https://aclanthology.org/2023.latechclfl-1.9.