Linking the Lewis & Short Dictionary to the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin

Francesco Mambrini, Eleonora Litta, Marco Passarotti, Paolo Ruffolo
CIRCSE Research Centre
Università Cattolica del Sacro Cuore
Largo Gemelli, 1 - 20123 Milan, Italy
francesco.mambrini@unicatt.it, eleonoramaria.litta@unicatt.it,
marco.passarotti@unicatt.it, paolo.ruffolo@posteo.eu

Abstract

This paper describes the steps taken to include data from the Lewis & Short bilingual Latin-English dictionary into the Knowledge Base of linguistic resources for Latin LiLa. First, data were extracted from the original XML and matched with entries in LiLa, overcoming ambiguities and structural inconsistencies in the source. Subsequently, senses were modelled using the Ontolex Lemon Lexicographic module (lexicog), so that they could be included in the LiLa Knowledge Base and thus made interoperable with the (meta)data of the linguistic resources for Latin therein interlinked.

1 Introduction

Since the pioneering times of 1949, when the Jesuit Roberto Busa persuaded Thomas Watson Sr., CEO of IBM, to fund his project aimed at processing the Latin texts of Thomas Aquinas with computers (Jones, 2016), scholars in the areas of Computational Linguistics, Literary Computing and Digital Humanities have built a plethora of linguistic resources for both modern and historical languages.

Particularly over the last two decades, many and diverse linguistic resources have been made available for Latin. These consist in corpora of texts spanning different eras and genres[1], dependency treebanks[2] and lexica[3]. These digital resources join the large set of textual and lexical resources that were created over the centuries for Latin: textual collections, thesauri, lexica, glossaries and mono/bilingual dictionaries. Among the latter, we could mention, for instance, the *Oxford Latin Dictionary* (Glare, 1968), the *Dictionary of medieval Latin from British sources* (Ashdowne et al., 1975), the Forcellini lexicon (Forcellini and Facciolati, 1871) and the still under construction *Thesaurus Linguae Latinae* (Ehlers, 1968), many of which are today accessible also in digital format.

However, the impact of these digital resources on the everyday work of classicists is still limited. On the one side, this is due to the still existing divisive dichotomy between "traditional" Humanities and computational approaches. On the other, it is a matter of fact that classicists are not yet put in the best condition to fully exploit all available resources for ancient languages, as these are currently scattered across the web in uncommunicative blocks, using different query languages, data formats, annotation criteria and tagsets. The last decade has seen a number of exploratory solutions to tackle the sparseness of linguistic resources. Among them, the European infrastructure CLARIN[4] represents a common hub where data and metadata of resources collected in single repositories (at national level) can be searched (through the so-called Virtual Language Observatory) and processed with different tools (through the CLARIN Language Resource Switchboard). As for Classical languages, *Logeion*[5] is a meta-

[1] See, for example, *Musisque deoque* for Classical Latin poetry (Manca et al., 2011), *CLaSSES*, containing epigraphic material (De Felice et al., 2015), the large corpus of Classical Latin prose and poetic texts by LASLA (Denooz, 2007) and *CroALa*, which brings together writings by Croatian authors produced between the 10th and 20th centuries (Jovanović, 2012).

[2] *Index Thomisticus* Treebank (Passarotti, 2019), Late Latin Charter Treebank (Cecchini et al., 2020a), UDante (Cecchini et al., 2020b), PROIEL (Eckhoff et al., 2018) and Latin Dependency Treebank (Bamman and Crane, 2011).

[3] Such as, for instance, valency and subcategorisation lexica (Passarotti et al., 2016; McGillivray and Vatri, 2015), the Latin WordNet (Minozzi, 2017) and word lists (Tombeur, 1998; Ramminger, 2008).

[4] https://www.clarin.eu.

[5] https://logeion.uchicago.edu/lexidium.

dictionary that allows to query together the lexical entries of several dictionaries for Ancient Greek and Latin, while *Corpus Corporum*[6] is a meta-collection that allows searches across more than twenty different corpora for Latin. However, what such initiatives still lack is to provide a real interoperability between distributed resources, which would result in interaction at both syntactic (structural) and semantic (conceptual) level.

Syntactic interoperability is defined as 'the ability of different systems to process (read) exchanged data either directly or via trivial conversion', using a common data model consisting of shared protocols and data formats. Semantic interoperability, on the other hand, is 'the ability to automatically interpret exchanged information meaningfully and accurately in order to produce useful results', by using a set of common linguistic data categories defined in *ad-hoc* ontologies (Ide and Pustejovsky, 2010).

Attaining syntactic and semantic interoperability between distributed linguistic resources is the objective of the Linguistic Linked Open Data (LLOD) community, which applies the principles of the Linked Data paradigm (Bizer et al., 2008) to the (meta)data contained in linguistic resources. As for Classical languages, the LiLa Knowledge Base (KB)[7] (Passarotti et al., 2020) makes textual and lexical resources for Latin interact through a commonly used data model, called the *Resource Description Framework* (RDF) (Lassila et al., 1998), and ontologies developed and shared by the LLOD community. In this way, the linked resources become interoperable with each other as well as with those for other languages described following the same structural and conceptual principles.

Based on a large collection of "canonical forms" (lemmas) - the so-called "Lemma Bank", LiLa achieves interoperability between resources by linking all those entries in lexical resources and tokens in corpora that point to the same lemma in the LiLa collection.

The lexical resources for Latin linked so far to LiLa include a word formation lexicon (Pellegrini et al., 2021), a polarity lexicon (Sprugnoli et al., 2020), an etymological dictionary (Mambrini and Passarotti, 2020) and a joint resource providing a manually checked subset of the Latin WordNet and a valency lexicon (Mambrini et al., 2021). The most recent among the LiLa connections is the bilingual Latin-English dictionary by Charlton Lewis and Charles Short (1879). The inclusion of this type of lexicon in LiLa was much needed, as no resource providing semantic information consisting of translations and definitions was available in the network of connected resources before. Since Lewis & Short is the first lexical resource of its kind included in LiLa, the process of its linking to the KB opened a number of LLOD-related challenges.

This paper describes how such challenges have been tackled and is organised as follows: Section 2 describes the Lewis & Short dictionary in its main characteristics. Section 3 discusses the ontologies involved in the modelling phase, the challenges that need to be overcome in the representation of the linguistic data as LLOD (3.1), and the strategies adopted to represent the dictionary entries using the chosen vocabularies (3.2). Finally, Section 4 discusses conclusions and highlights directions for future work.

2 The "Lewis & Short" Dictionary

2.1 The Printed and Digital Dictionary

The *Latin Dictionary*, curated by Ch. T. Lewis and Ch. Short and commonly referred to as the "Lewis & Short" (L&S), was published by Harper and Oxford University Press in 1879 (Lewis and Short, 1879). Though based on previous work by German scholars, it remained a standard in Latin lexicography in the English-speaking world until it was superseded by the *Oxford Latin Dictionary* (Glare, 1968).

In the digital age, its importance rests on two grounds. On the one hand, its relevance for the history of Classical Scholarship is undeniable. On the other hand, also on account of its copyright status, as the dictionary belongs now to the public domain, the L&S has quickly become one of the most used and best curated digital Latin dictionaries on the web. Following the same workflow used for the *Greek-English Lexicon* (Liddell et al., 1940), the Perseus Project has developed a widely used digital edition of the dictionary based on the standards of the Text Encoding Initiative (TEI) (Rydberg-Cox, 2002). The digital L&S has been incorporated in the word-search tools available on the Perseus website and in a series of other

[6] http://www.mlat.uzh.ch/MLS/.
[7] https://lila-erc.eu.

desktop and web applications.[8]

Perseus' TEI edition is the point of departure of our work.[9] Though its publication was a remarkable achievement, this electronic text is not exempt from occasional flaws and inconsistencies, which had to be taken into account.

In the digital edition, entries from the L&S are based on an XML encoding of the whole dictionary. The XML structure, albeit not always consistent, offers the following information about each word:

1. Entry: the headword. Entries are encoded within the TEI element `<entryFree>` and are 51,596 in total.[10]

2. Information about inflection, encoded as attributes in the XML and visualised in the output reproducing the customary descriptions for Latin dictionaries, e.g. a masculine noun of the second declension (e.g. *gallus* 'cock') is followed by the genitive singular ending of the word ('i'), and the abbreviation for gender 'm.' (e.g. *gallus, i*, m.).

3. Etymological or derivational information, encoded within the same element `<etym>`.

4. Sense(s): these act as containers where the meaning of the word is matched with a number of representative citations from Classical Latin sources. Each citation is accompanied by its canonical reference (e.g. "Cic. Sen. 8, 26" for a reference to Cicero, *De Senectute*, chapter 8, paragraph 26).

Entries can contain what we call "sub-entries", words that are not given a record of their own, but are discussed within another entry. Usually, these sub-entries consist of lexicalised present and past participles like, for example, *adolescens* 'young man' – sub-entry of *adolesco* 'to grow up'; another instance is the substantivised forms of adjectives, such as *verum* 'the truth' – sub-entry of *verus* 'true'. Sub-entries are encoded within the `<sense>` element and followed by the same type of inflectional information structured as the main entries.

[8]One example is the app *Diogenes* for querying corpora of Greek and Latin texts: `https://d.iogen.es/`.

[9]The digital edition is available from the repository of the Perseus DL and is distributed under a CC BY SA 4.0 license: `https://github.com/PerseusDL/lexica`.

[10]See `https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-entryFree.html`.

2.2 Linking the L&S to LiLa

The LiLa KB includes about 200,000 canonical forms, each of which is described by a series of properties that record the part of speech (PoS), the full morphological description and the inflectional category. Also, the data property "written representation", defined in the ontology Ontolex (see Section 3.1), registers all the attested spellings of any lemma. Publishing a lexical resource as LLOD within LiLa means to both represent its information using the appropriate standards and vocabularies (Section 3.1) and to link the dictionary entries to the right form in LiLa by matching the lemmas used to index the records to the appropriate form in the KB.

In order to achieve the latter goal, firstly we had to normalise the spelling of the L&S dictionary lemmas by removing upper case initials and substituting *j* with *i* and *v* with *u* in order to mirror LiLa's conventions. Then, after mapping part-of-speech and inflectional information between resources, we extracted 31,142 1:1 matches, 2,998 1:N matches and 4,553 1:0 matches, on the basis of the tuple written representation - PoS. The latter group was subsequently matched only on the basis of graphical representation, at which point we obtained 946 1:1 matches and 50 1:N matches. Of the remaining 3,557 unmatched entries, 1,289 were successfully analysed by the morphological analyser Lemlat (Passarotti et al., 2017), leaving 2,239 definitely unmatched entries. After resolving multi-word spellings and graphical variants, the unmatched entries were all added to the LiLa Lemma Bank, while 1:N matches were manually disambiguated and matched to the relevant lemmas.

3 Modelling Lexical Entries

3.1 LiLa, Ontolex and lexicog

As said, the LiLa KB for Latin resources is built around a collection of canonical forms that can be used both as head words of dictionaries or as "targets" for the lemmatisation of corpora (Passarotti et al., 2020). These lemmas are modelled using the Ontolex ontology, a now *de facto* standard of the LLOD community (Cimiano et al., 2020; McCrae et al., 2017). In particular, lemmas in the LiLa KB are defined as forms of words that are linked (or are ready to be linked) to lexical entries via the property "canonical form" of the Ontolex

ontology.[11]

Ontolex provides several classes and properties to describe the relationships that lexical entries have with, on the one hand, the grammatical forms attested in language and, on the other, the senses and the meanings of words. The core Ontolex module, however, imposes a series of restrictions that make its classes and properties ill-suited to represent the information in most standard dictionaries. The class Lexical Entry from the core Ontolex module, for instance, is inadequate to represent entries that license multiple syntactic interpretations, such as words that are registered in a dictionary as both adverb and conjunction. Subentries like the noun *verum* from the adjective *verus*, formed by a process of substantivisation from the word in the main entry, would also produce a mismatch between the dictionary and the lexical entry. Finally, the L&S, as most dictionaries, defines the senses of all but the most simple words by grouping them in sense clusters; those clusters are generally organized into hierarchies with multiple levels of nesting, from the most general to the most specific sense, a structure for which Ontolex has no suitable representation.

In order to overcome these issues, the Ontolex community has developed a specific extension of the ontology called the "OntoLex lexicography module" or `lexicog` (Bosque-Gil and Gracia, 2019).[12] The module is explicitly designed to capture the structural information expressed in a lexicographic resource and is primarily intended to support the conversion of lexicographic data that are not native to Ontolex. Retro-digitised dictionaries like the L&S are thus a perfect use case.

As said, `lexicog` focuses on the structural properties of dictionaries and does not attempt to convey any lexical, or indeed linguistic information, which are left to the classes and properties of Ontolex. The most important of these structural elements introduced in the vocabulary is that of the Lexicographic Entry. In `lexicog`, an entry is a container that represents a lexicographic article or record as it is arranged in the source (Bosque-Gil and Gracia, 2019). Thus, while a *lexical* entry (as defined in Ontolex) is an item in the lexicon of a given language, a *lexicographic* entry is a record in a linguistic resource that documents or discusses some properties of a given lexical item.

Lexicographic entries are a special subset of a larger class called Lexicographic Component. Apart from whole dictionary articles (the entries), components can be used to represent senses, sense groups or subentries (like the substantivised *verum*) within lexicographic entries.

It is important to stress once again that components represent only structural units; all linguistic information that is conveyed within these units must be expressed using Ontolex. The property `lexicog:describes` provides a link between the two dimensions, so that a lexicographic entry can be said to *describe* a lexical entry (as defined in Ontolex). In the same way, the lexicographic components that discuss a sense of a word or introduce a subentry, *describe* that specific lexical sense (as defined in Ontolex) or another lexical entry.

3.2 Lexicographic and Lexical Entries in the L&S

The LLOD version of the L&S linked to LiLa is now available online in the LiLa KB.[13] The entries can also be searched using LiLa's query interface and SPARQL endpoint.[14]

Figure 1 shows a visualisation of how the information from a sample entry, the adjective *hosticus* in the L&S dictionary, is represented in LiLa. In particular, the interplay between the linguistic and structural information is reflected in the complex relation between the lexical and lexicographic entries.

The L&S distinguishes two senses for the word: "belonging to an enemy, hostile" and "belonging to a stranger, foreign". Following the Ontolex approach, these meanings are represented by the two 'triangles' between the lexical entry (the light green node on the left), the concepts evoked by the word (gray-blue nodes), and the senses, labeled 0 ad 1, that mediate between them (greenish-yellow nodes).

The lexical entry is described by a *lexicographic entry*, identified by the id n21014 (inherited from the TEI XML file of the Perseus DL), while a specific lexicographic component describes each of the two senses (n21014_0 and n21014_1, respectively). What is particularly relevant is that the component n21014_0, which corresponds to the

[11] `http://www.w3.org/ns/lemon/ontolex#canonicalForm`.

[12] `https://www.w3.org/ns/lemon/lexicog#`.

[13] `http://lila-erc.eu/data/lexicalResources/LewisShort/Lexicon`.

[14] `https://lila-erc.eu/query/`, and `https://lila-erc.eu/sparql/`.

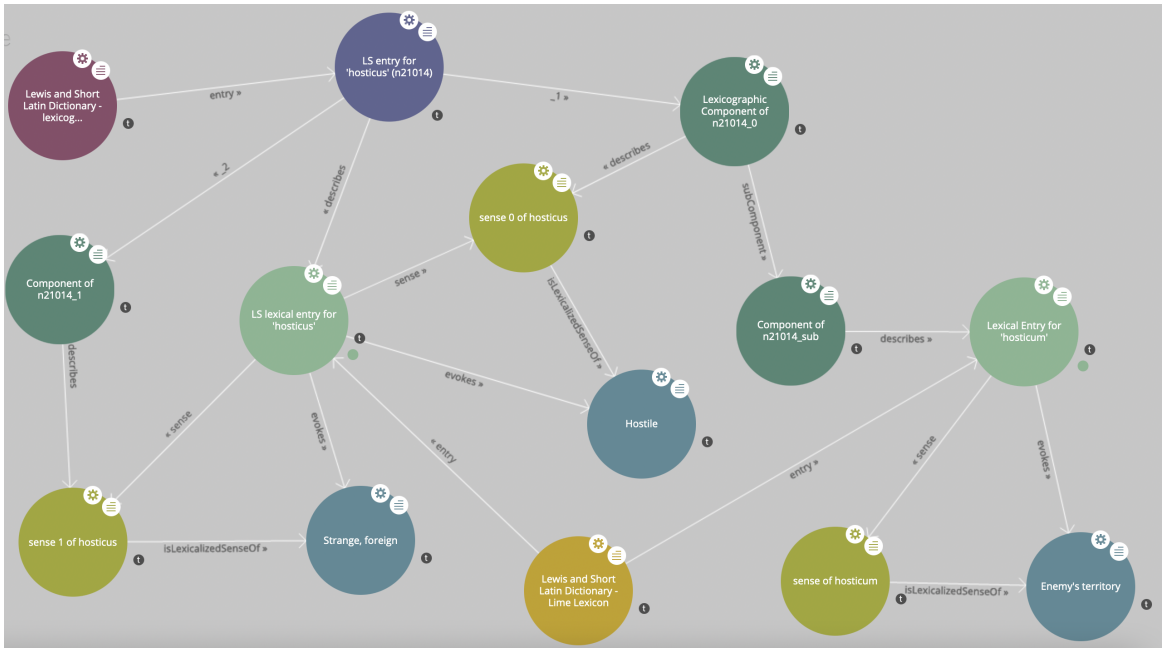

Figure 1: An entry in the LiLa's representation of the L&S.

sense "hostile", is linked to a sub-component that describes the lexical entry of the noun *hosticum*, a substantivised usage of the neuter adjective that means "the enemy's territory". That section of the entry that discusses the subentry "hosticum", which is itself a section of the paragraph dedicated to the first sense, is thus linked (via the "describes" property) to a different lexical entry.

4 Conclusions and Future Work

Perhaps even more than for any other modern language, a great number of lexical resources, either bi- or monolingual, is available for Latin, many of which have already been digitised and disseminated on the web. In this paper, we described a model of how this huge wealth of information can be published using the modern standards of the Semantic Web. The greatest advantage of this approach is that all the lexical resources published according to the same data model can be integrated in a wider network of linguistic information, along with the other digital resources that are connected to it. In the case of the L&S in LiLa, the Latin lexical entries of the bilingual dictionary can be queried together with the information about the same words provided by the other linguistic resources linked to the lemmas in the KB.

One example of the fruitful interactions between resources is the possibility to investigate the polysemy of words in relation to their derivation, as recorded in the Word Formation Latin resource, which is also linked to LiLa (Litta et al., 2020). The adjective *hosticus* of Figure 1, for instance, clearly inherits its two main senses ('hostile' and 'foreign') from the same polysemy of the noun *hostis* 'stranger' or 'enemy', from which it is derived. At the same time, while other resources in LiLa describe the senses of words, such as the Latin WordNet (Franzini et al., 2019; Mambrini et al., 2021), the complex relations between those senses (whether, for instance, one sense is interpreted as a specialised derivation from another) is generally available only in traditional lexical resources like the L&S.

The solutions we found to address the challenges raised by the representation of the L&S in LLOD will be reused when we will link further bilingual, as well as monolingual, dictionaries of Latin to the KB. Including such lexical resources in LiLa is an important achievement, as it makes it possible for the KB to interact with linguistic (meta)data for languages other than Latin. Undoubtedly, such an inter-linguistic (re)use of distributed resources is one of the objectives of the LLOD community, to which LiLa contributes by steadily providing it also with new (kinds of) linguistic resources represented in LLOD.

Acknowledgments

This 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.

References

Richard Ashdowne, David R Howlett, and Ronald Edward Latham. 1975. *Dictionary of medieval Latin from British sources*. Oxford University Press.

David Bamman and Gregory Crane. 2011. The ancient greek and latin dependency treebanks. In *Language technology for cultural heritage*, pages 79–98. Springer.

Christian Bizer, Tom Heath, Kingsley Idehen, and Tim Berners-Lee. 2008. Linked data on the web (ldow2008). In *Proceedings of the 17th international conference on World Wide Web*, pages 1265–1266.

Julia Bosque-Gil and Jorge Gracia. 2019. The OntoLex lemon lexicography module. `https://ontolex.github.io/lexicog/`.

Flavio Massimiliano Cecchini, Timo Korkiakangas, and Marco Passarotti. 2020a. A new latin treebank for universal dependencies: Charters between ancient latin and romance languages. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 933–942.

Flavio Massimiliano Cecchini, Rachele Sprugnoli, Giovanni Moretti, and Marco Passarotti. 2020b. Udante: First steps towards the universal dependencies treebank of dante's latin works. In *CLiC-it*.

Philipp Cimiano, Christian Chiarcos, John P. McCrae, and Jorge Gracia. 2020. *Linguistic Linked Data: Representation, Generation and Applications*. Springer, Cham.

Irene De Felice, Giovanna Marotta, and Margherita Donati. 2015. Classes: A new digital resource for latin epigraphy. *IJCoL. Italian Journal of Computational Linguistics*, 1(1-1):125–136.

Joseph Denooz. 2007. Opera latina: le nouveau site internet du lasla. *Journal of Latin Linguistics*, 9(3):21–34.

Hanne Eckhoff, Kristin Bech, Gerlof Bouma, Kristine Eide, Dag Haug, Odd Einar Haugen, and Marius Jøhndal. 2018. The proiel treebank family: a standard for early attestations of indo-european languages. *Language Resources and Evaluation*, 52(1):29–65.

Wilhelm Ehlers. 1968. Der thesaurus linguae latinae. prinzipien und erfahrungen. *Antike und Abendland*, 14(1):172–184.

Egidio Forcellini and Jacobo Facciolati. 1871. *Lexicon totius latinitatis*, volume 3. Typis seminarii.

Greta Franzini, Andrea Peverelli, Paolo Ruffolo, Marco Passarotti, Helena Sanna, Edoardo Signoroni, Viviana Ventura, and Federica Zampedri. 2019. Nunc Est Aestimandum. Towards an evaluation of the Latin WordNet. In Raffaella Bernardi, Roberto Navigli, and Giovanni Semeraro, editors, *Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)*, pages 1–8, Bari, Italy. CEUR-WS.org.

Peter GW Glare. 1968. *Oxford latin dictionary*. Clarendon Press, Oxford.

Nancy Ide and James Pustejovsky. 2010. What does interoperability mean, anyway? toward an operational definition of interoperability for language technology. In *Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong, China*.

Steven E Jones. 2016. *Roberto Busa, SJ, and the emergence of humanities computing: the priest and the punched cards*. Routledge.

Neven Jovanović. 2012. Croala. enhancing a tei-encoded text collection. *Journal of the Text Encoding Initiative*, (2).

Ora Lassila, Ralph R. Swick, World Wide, and Web Consortium. 1998. Resource description framework (rdf) model and syntax specification.

Charlton T. Lewis and Charles Short. 1879. *A Latin Dictionary. Founded on Andrews' edition of Freund's Latin dictionary*. Clarendon Press, Oxford.

Henry Liddell, Robert Scott, and Henry Stuart Jones. 1940. *A Greek-English Lexicon*. Clarendon Press, Oxford, 9 edition.

Eleonora Litta, Marco Passarotti, and Francesco Mambrini. 2020. Derivations and Connections: Word Formation in the LiLa Knowledge Base of Linguistic Resources for Latin. *The Prague Bulletin Of Mathematical Linguistics*, 115:163–186.

Francesco Mambrini and Marco Passarotti. 2020. Representing etymology in the lila knowledge base of linguistic resources for latin. In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, pages 20–28.

Francesco Mambrini, Marco Passarotti, Eleonora Litta, and Giovanni Moretti. 2021. Interlinking valency frames and wordnet synsets in the lila knowledge base of linguistic resources for latin. In *Further with Knowledge Graphs*, pages 16–28. IOS Press.

Massimo Manca, Linda Spinazzè, Paolo Mastandrea, Luigi Tessarolo, and Federico Boschetti. 2011. Musisque deoque: Text retrieval on critical editionse. *J. Lang. Technol. Comput. Linguistics*, 26(2):127–138.

John P. McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. The OntoLex-Lemon Model: development and applications. In *Proceedings of eLex 2017*, pages 587–597.

Barbara McGillivray and Alessandro Vatri. 2015. Computational valency lexica for latin and greek in use: a case study of syntactic ambiguity. *Journal of Latin Linguistics*, 14(1):101–126.

Stefano Minozzi. 2017. Latin wordnet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'information retrieval. *Strumenti digitali e collaborativi per le Scienze dell'Antichità*, (14):123–134.

Marco Passarotti, Berta González Saavedra, and Christophe Onambele. 2016. Latin vallex. a treebank-based semantic valency lexicon for latin. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2599–2606.

Marco Passarotti, Marco Budassi, Eleonora Litta, and Paolo Ruffolo. 2017. The Lemlat 3.0 Package for Morphological Analysis of Latin. In Gerlof Bouma and Yvonne Adesam, editors, *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, volume 133, pages 24–31, Gothenburg. Linköping University Electronic Press.

Marco Passarotti, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, and Rachele Sprugnoli. 2020. Interlinking through lemmas. the lexical collection of the lila knowledge base of linguistic resources for latin. *Studi e Saggi Linguistici*, 58(1):177–212.

Marco Passarotti. 2019. The project of the index thomisticus treebank. In *Digital Classical Philology*, pages 299–320. De Gruyter Saur.

Matteo Pellegrini, Eleonora Litta, Marco Passarotti, Francesco Mambrini, and Giovanni Moretti. 2021. The two approaches to word formation in the lila knowledge base of latin resources. In *Proceedings of the Third International Workshop on Resources and Tools for Derivational Morphology (DeriMo 2021)*, pages 101–109.

Johann Ramminger. 2008. *Neulateinische Wortliste. Ein Wörterbuch der Lateinischen von Petrarca bis 1700*. Thesaurus Linguae Latinae.

Jeffrey A Rydberg-Cox. 2002. Mining Data from an Electronic Greek Lexicon. *Classical Journal*, 98(2):183–188.

Rachele Sprugnoli, Francesco Mambrini, Giovanni Moretti, and Marco Passarotti. 2020. Towards the modeling of polarity in a latin knowledge base. In *WHiSe@ ESWC*, pages 59–70.

Paul Tombeur. 1998. *Thesaurus formarum totius Latinitatis: a Plauto usque ad saeculum XXum; TF.[2]. CETEDOC Index of Latin forms: database for the study of the vocabulary of the entire Latin world; base de données pour l'étude du vocabulaire de toute la latinité*. Brepols.