Modelling and Publishing the “Lexicon der indogermanischen Verben” as Linked Open Data

Valeria Irene Boano[1], Francesco Mambrini[1], Marco Passarotti[1] and Riccardo Ginevra[1]

[1]*Università Cattolica del Sacro Cuore, Milan, Italy*

Abstract

This paper describes the modelling and publication of part of the etymological information in the *Lexicon der indogermanischen Verben*, an etymological dictionary of verbs attested in ancient Indo-European languages, as Linguistic Linked Open Data. The lexicon has been made interoperable with a set of lexical and textual linguistic resources for Latin in the Lila Knowledge Base.

Keywords

Linked Open Data, Ontolex Lemon, Etymology, Etymological dictionary

1. Introduction

Over the past decades, several linguistic resources for historical languages have become available in digital format. This has given scholars the chance to access and exploit them in a quicker and deeper way.

In particular, several different linguistic resources are available for Latin today. They consist of textual corpora, such as the LASLA corpus[1] for Classical Latin and the *Index Thomisticus* Treebank [1] for Medieval Latin, and lexical resources, like the Lewis and Short dictionary [2] and the Logeion metadictionary[2].

Due to the centuries-long lexicographic tradition for the Latin language, its lexical resources comprise a number of etymological dictionaries. A dictionary is defined as etymological when it contains information about the etymology of its entries, that is about their origin and historical development: for Indo-European (IE) languages, etymological dictionaries often put their entries in relation with reconstructed Proto-Indo-European (PIE) roots, minimal lexical units to which the dictionary's entry and further related formations may be traced back. It is also often explained by which morphological processes the attested word has been formed from the root.

Even if the available resources (for Latin and beyond) provide a huge amount of linguistic information, at the present day, their full exploitation is still hindered by their isolation. In fact, most resources can be accessed only individually, and cannot interact.

Isolation of resources is an issue because each resource can really reach its potential only when it is made interoperable with other (types of) resources. Today, interoperability between linguistic resources can be obtained by describing and publishing their data according to the principles of the Linked Open Data paradigm [3]. As a consequence, in recent years, the amount of linguistic resources published as Linked Open Data has been raised substantially, as witnessed by the growing size of the Linguistic Linked Open Data Cloud (LLOD Cloud)[3]. In particular, the LiLa Knowledge Base[4] represents a successful example of Linked Open Data (LOD) principles applied to linguistic resources for Latin.

Although LiLa currently makes quite a number of linguistic resources for Latin interoperable, there is still a large set of digitized materials to interlink in the Knowledge Base. Among them is an important etymological dictionary, the *Lexicon der indogermanischen Verben* (LIV) [4]. This paper describes the process of transforming information contained in this dictionary into a LLOD resource, linked to LiLa. Section 2 describes the LiLa architecture and the LIV structure. Section 3 details the modelling of the resource, and the linking process. Section 4 describes some possible examples of exploitation and interaction. Finally, Section 5 discusses conclusions and sketches the future work.

2. The LiLa Knowledge Base and the *Lexicon der indogermanischen Verben*

2.1. The LiLa Knowledge Base

The LiLa Knowledge Base (KB) [5] is a linguistic hub for Latin, containing FAIR [6] linguistic resources, published as LOD. As usual in LLOD, structural interoperability

CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 — Dec 02, 2023, Venice, Italy

valeriairene.boano01@icatt.it (V. I. Boano); francesco.mambrini@unicatt.it (F. Mambrini); marco.passarotti@unicatt.it (M. Passarotti); riccardo.ginevra@unicatt.it (R. Ginevra)

CEUR Workshop Proceedings (CEUR-WS.org)

[1] http://web.philo.ulg.ac.be/lasla/.

[2] https://logeion.uchicago.edu/.

[3] https://linguistic-lod.org.

[4] https://lila-erc.eu.

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

between resources is based on the Resource Description Framework (RDF) [7], which is the data model used for the Semantic Web [8]. Conceptual interoperability [9] is achieved by using common ontologies built and adopted by the LLOD community, such as the Ontolex Lemon model[5] and the OLiA ontology[6] [10].

LiLa is built around the so-called Lemma Bank, which contains a set of more than 200k Latin lemmas, taken from the database of the morphological analyser LEMLAT [11] and constantly extended. Each lemma of the Lemma Bank is a gateway between the different linguistic resources linked to the Knowledge Base, starting from the assumption that words (indexed by their lemmas) can be used as the point of contact between textual resources (which are made of occurrences of words), lexical resources (which describe words) and NLP tools (which process words).

Each entry in the LiLa Lemma Bank is an instance of `ontolex:Form`[7]. In particular, the `lila:Lemma`[8] is a form that can be linked to an `ontolex:LexicalEntry`[9] via the property `ontolex:canonicalForm`[10], which identifies the canonical form used to represent a lexical entry. Every other realization of a word is linked to the lexical entry via the property `ontolex:lexicalForm`[11]. Each lemma and each form may also be described with other properties, which give information, for example, about phonetic representation, Part of Speech (POS) tagging and other grammatical features.

Lexical resources are connected to LiLa by linking the `ontolex:LexicalEntry` of each resource to the `lila:Lemma` via the property `ontolex:canonicalForm`. Once a linguistic resource is linked to the KB via the Lemma Bank, all the interoperable resources can be queried together, using a SPARQL endpoint[12] also through a user-friendly interface[13].

The textual resources connected to LiLa so far include more than 3,5M words from Latin texts of different eras, such as the LASLA corpus[14], the *Index Thomisticus* Treebank [12], containing works of Thomas Aquinas, and *UDante*, a Universal Dependencies[15] treebank for Dante Alighieri's Latin works [13]. The lexical resources of LiLa include a derivational lexicon, *Word Formation Latin* [14], a manually checked subset of the Latin WordNet connected to a valency lexicon [15], the *Etymological dictionary of Latin and other Italic Languages* [16, 17] and a resource of principal parts of Latin words, *PrinParLat* [18].

2.2. The *Lexicon der indogermanischen Verben*

The *Lexicon der indogermanischen Verben*, also known as LIV, is an etymological dictionary of verbs attested in ancient Indo-European languages. After the first edition, curated by Helmut Rix [19] and published by Reichert Verlag in 1998, a second edition was published in 2001 with additions and corrections by Martin Kümmel and Helmut Rix [4].

The LIV is the main reference work for Proto-Indo-European verbal roots and contains three types of information:

- **Reconstructed Proto-Indo-European verbal roots**, which coincide with the entries of the dictionary and are provided with their presumed lexical meaning and their phonological structure. For each root, the corresponding index in the *Indogermanisches etymologisches Wörterbuch* [20] is specified as well.
- **Reconstructed Proto-Indo-European primary verbal stems**, which are either root formations or are formed by adding to the roots primary affixes that mainly express categories of aspect and actionality. The meaning of the stem is usually not specified.
- **Word forms that are historically attested in ancient IE languages**, which show how the Proto-Indo-European stems evolved in the various daughter languages. Each attested form is provided with its lexical meaning in the respective language. At the end of certain entries are sometimes listed innovative verbal stems that may ultimately be traced back to Proto-Indo-European roots, but are unlikely to directly reflect Proto-Indo-European verbal stems, having been created according to language-specific productive patterns.

The original data used during the linking process consists of a spreadsheet containing information from the LIV extracted and structured by Thomas Olander, with the collaboration of Simon Poulsen and Anders Richardt Jørgensen, and shared with us by the authors. For each of the 7,888 LIV entries the spreadsheet records its root, stem and attested forms.

The LIV is copyrighted by the Reichert Verlag, and LiLa is not authorized to reproduce the full content of the dictionary. The publisher has however agreed to allow us to model the basic etymological relations between

[5] https://www.w3.org/2016/05/ontolex/.
[6] https://acoli-repo.github.io/olia/.
[7] http://www.w3.org/ns/lemon/ontolex#Form.
[8] https://lila-erc.eu/lodview/ontologies/lila/Lemma.
[9] http://www.w3.org/ns/lemon/ontolex#LexicalEntry.
[10] http://www.w3.org/ns/lemon/ontolex#canonicalForm.
[11] http://www.w3.org/ns/lemon/ontolex#lexicalForm.
[12] https://lila-erc.eu/sparql/.
[13] https://lila-erc.eu/query/.
[14] https://www.lasla.uliege.be/cms/c_8508894/fr/lasla.
[15] https://universaldependencies.org/.

the PIE roots, the stems and the Latin words and stems, provided that explicit bibliographical attribution is given to the linguistic reconstruction.

This is the information that we modelled to be linked to the LiLa Knowledge Base, as described in the following section.

3. Modelling and linking the LIV

Making linguistic resources interoperable means using a shared set of vocabularies for knowledge description, as defined in specialized ontologies, to represent the information contained in them. The process of linking, on the other hand, aims to connect this information to a wider network of data, so that a meaningful context is provided[16]. Within the network of LiLa, this step means that all entries of a lexical resource must make reference to the canonical forms of the Lemma Bank, as described above.

This section details how we modelled our target information from the LIV, how we applied such modelling to the publication of these data as LOD and how we linked them to the LiLa collection.

3.1. Modelling

In the lexical resources linked to LiLa, etymological information has been expressed using the `lemonEty` extension of the Ontolex-Lemon model [21]. This ontology was used in LiLa to represent loanwords from Greek [22] and for the *Etymological Dictionary of Latin and the other Italic Languages* [16, 17].

The set of classes and properties of `lemonEty` are suitable to express the etymological information of the LIV too, but, compared to the aforementioned dictionaries in LiLa, a more complex modelling and a series of extensions are also required.

The `lemonEty` ontology establishes etymological relations between instances of the Ontolex's class `LexicalEntry`. In particular, a special subclass called `Etymon` is reserved for lexical items of the source language that are introduced in order to explain the history of the entries in the target language.

Two core classes of `lemonEty` that are particularly important are `Etymology`[17], and `EtyLink`[18]. The former "reifies the whole process of etymological reconstruction as scientific hypothesis" [17, p. 22]. Etymological links, on the other hand, connect "linguistic elements" from the source language to the corresponding elements of the target.

In applying this model to the LIV data, it is crucial to define what the "linguistic elements" connected via etymological links are. The previously mentioned lexical resources rested on a simple model where the etymological links involved only Latin lexical entries and etymons from a source language, so that e.g. the Latin word *abacus*[19] was the target of a link that had its source in the Greek etymon *ábax* 'reckoning board'.

In the LIV, on the other hand, relations are established between:

- **Inflected forms** of a historical language (e.g. Latin). In the case of Latin, those forms are used in the LIV as placeholders for all forms derived from the same **stem**; so, for instance, the Latin 1st-person perfect *fidi* stands for all forms from the perfect stem of the verb *findo* 'to cleave, split';
- **the PIE stems**, to which the inflected forms and stems of Latin (and other languages) must be traced back;
- **the PIE root** that underlies the PIE stems.

In the case of Latin, thus, the LIV documents etymological relations between a PIE and a Latin stem (the latter represented by a Latin inflected form). While the PIE root (e.g. *b^hei̯d-) and the Latin target lexical item (e.g. *findo*, inclusive of all its stems) can be conceptualised as lexical entries, the stems and the word forms must be described using concepts from other vocabularies.

For the Latin forms and stems we reused the individuals of the class `Stem`[20] provided by *PrinParLat*, a lexical resource listing all Latin "principal parts". Principal parts are sets of inflected wordforms from which the content of all the other paradigm cells can be inferred[21].

For the perfect stem, *PrinParLat* already includes all forms linked to their `Stem` therein, which could thus be immediately reused. As for the present stems, however, the related forms were not available, and had therefore to be generated and linked to their `Stem` via the property `ontolex:lexicalForm`[22].

Some specific information provided by the dictionary that could not be represented with any of the available

[16] This is the fifth and final step in Berners-Lee's five-star rating system: see https://www.w3.org/DesignIssues/LinkedData.html.

[17] http://lari-datasets.ilc.cnr.it/lemonEty#Etymology.

[18] http://lari-datasets.ilc.cnr.it/lemonEty#EtyLink.

[19] http://lila-erc.eu/data/lexicalResources/IGVLL/id/LexicalEntry/abacus.

[20] https://lila-erc.eu/lodview/ontologies/prinparlat/Stem. This class will be defined as a sub-class of `morph:Morph`, when the module will be released.

[21] https://github.com/CIRCSE/PrinParLat.

[22] Note that, following the conventional definition of a form in *PrinParLat*, which is more granular than the one used by LiLa, we chose to recreate the form for the same paradigmatic slot that is traditionally chosen as lemma. *PrinParLat* creates a unique form for each graphical or spelling variant of a form; lemmas in the Lemma Bank, on the other hand, unify them under a single form. In many cases, this choice results in a duplicated form for the 1st-person singular present indicative verb.

modules required the creation of *ad hoc* classes. In particular, some Latin stems that trace back to PIE roots, but are unlikely to directly reflect a PIE stem, are classified by the LIV as *Neubildungen*, that is 'innovations', since they have been created according to language-specific productive patterns. These innovations cannot be traced back to a PIE stem, so that no etymological link can be created. We therefore created a specific class `Innovation`, which contains all those innovative stems.

Moreover, some Latin entries are defined by the dictionary as 'remodelings' (*Umbildungen*): their stems may be traced back to PIE stems, but have been reshaped following language-specific productive patterns (e.g. Latin *fodio* has lost the first syllable of the PIE reduplicated stem $^*b^h\acute{e}\text{-}b^h od^h h_2/b^h d^h h_2\text{-}$). The remodeled Latin stems are now defined as instances of the new class `Remodeling`.

3.2. Linking

LIV provides etymological information for other IE languages in addition to Latin. Since, however, LiLa is limited to Latin resources, we restricted our attention only on entries where a connection to Latin forms was explicitly mentioned.

In total, we identified 550 Latin forms linked to PIE roots, 354 of which corresponded to the main lemma of a verb; the remaining 196 were instead analysed as inflected forms.

The forms were analysed with the UDPipe pipeline[23], in order to perform the POS tagging of all forms and lemmatization of 196 inflected forms. The results were manually checked, which confirmed a good accuracy of 97% for POS-tagging (only 11 cases were incorrectly tagged), but much lower performances for lemmatization (87 out of 196, i.e. 44%).

For each of the remaining lemmas in the manually corrected set, we created a lexical entry in our new etymological resource. The canonical forms of these entries were identified by matching the lemma strings with the written representations[24] of the lemmas in the LiLa Lemma Bank. In 143 cases, manual disambiguation was needed, as the query returned more than one possible match. In one case, it was not possible to link the form (*tātōd*) to any lemma in the Lemma Bank: we decided against adding the invariable form to the Lemma Bank and instead created a LIV `LexicalEntry` *tātōd*, without connecting it to any lemma.

Moreover, a set of 11 entries required a special treatment. For a series of entries, like for instance *$pleh_1$- [4, p. 482], in fact, the LIV does not point to a single Latin word form, but rather to a whole lexical root that is analogous to LiLa's lexical base [5, p. 191]. This morpheme represents a lexical element that is neither a prefix nor a suffix and is shared by all members of a derivational family. Comparably, for instance, the Latin hyphenated form *-pleo* in the entry *$pleh_1$- is used in the LIV as a placeholder for all the possible Latin verbs that can be formed adding different preverbs to the same base (e.g. *compleo* 'to fill up', *depleo* 'to empty', *expleo* 'to fill up'...).

In those cases where the LIV uses this notation, we chose to create one lexical entry for each verb connected to the corresponding lexical base in LiLa (e.g. the 'base of *pleo*'[25]).

Once the lexical entries had been created, the correct stems in the *PrinParLat* resource were easily identified by leveraging the advantages of the LOD model. In fact, each LiLa's lemma is linked to the appropriate stems via an instance of the *PrinParLat* class of `Flexeme`[26]; the stems for a lexical entry are therefore easily recoverable once the LiLa lemma is known.

3.2.1. A LIV lexical entry linked to LiLa

The Figure 1 is taken from the LodLive[27] visualization in LiLa. It shows an example of how a LIV lexical entry (*glubo* 'to peel') was modelled and linked to the Lemma Bank.

On the left side of the figure is the LIV `LexicalEntry` *glubo*, which is linked to LiLa's lemma *glubo* via the property `canonicalForm`: this simple but crucial link allows us to connect the LIV etymological relations with the other resources of the Knowledge Base.

Then, the LIV `LexicalEntry` is connected via the property `lexicalRel`[28] to two *PrinParLat* Latin stems, the present stem *glub-* and the perfect stem *glups-*. Each stem is connected with a Latin form: the present form *glubo* is part of the LIV resource, and is thus linked to the present stem via the property `lexicalForm`; on the other hand, the perfect form *glupsi* is part of the *PrinParLat* resource, and is thus linked to the perfect stem via the property `consistsOf`.

We then link the Latin stems to their Proto-Indo-European ancestors. Each of the two Latin stems is in fact the `etyTarget` of an `EtyLink` (*Etymology link: pres glubo* and *Etymology link: perf glupsi*), which connects them to their `etySource`, that is the corresponding PIE stem (for the present $^*g/\hat{g}l\acute{e}\underset{\frown}{u}b^h\text{-}/g/\hat{g}lub^h\text{-}$, and for the perfect $?^*g/\hat{g}l\acute{\bar{e}}\underset{\frown}{u}b^h/g/\hat{g}l\acute{e}\underset{\frown}{u}b^h\text{-}s\text{-}$). These etymological links

[23] https://lindat.mff.cuni.cz/services/udpipe/.

[24] In Ontolex Lemon, written representations are the different graphical variants of a form. See http://www.w3.org/ns/lemon/ontolex#writtenRep.

[25] http://lila-erc.eu/data/id/base/107. Note that we excluded those verbs that are only documented in lexicons of Medieval Latin, such as the Du Cange dictionary [23].

[26] http://lila-erc.eu/ontologies/prinparlat/Flexeme.

[27] "LodLive project provides a demonstration of the use of Linked Data standards (RDF, SPARQL) to browse RDF resources". http://en.lodlive.it/.

[28] http://www.w3.org/ns/lemon/vartrans#lexicalRel.

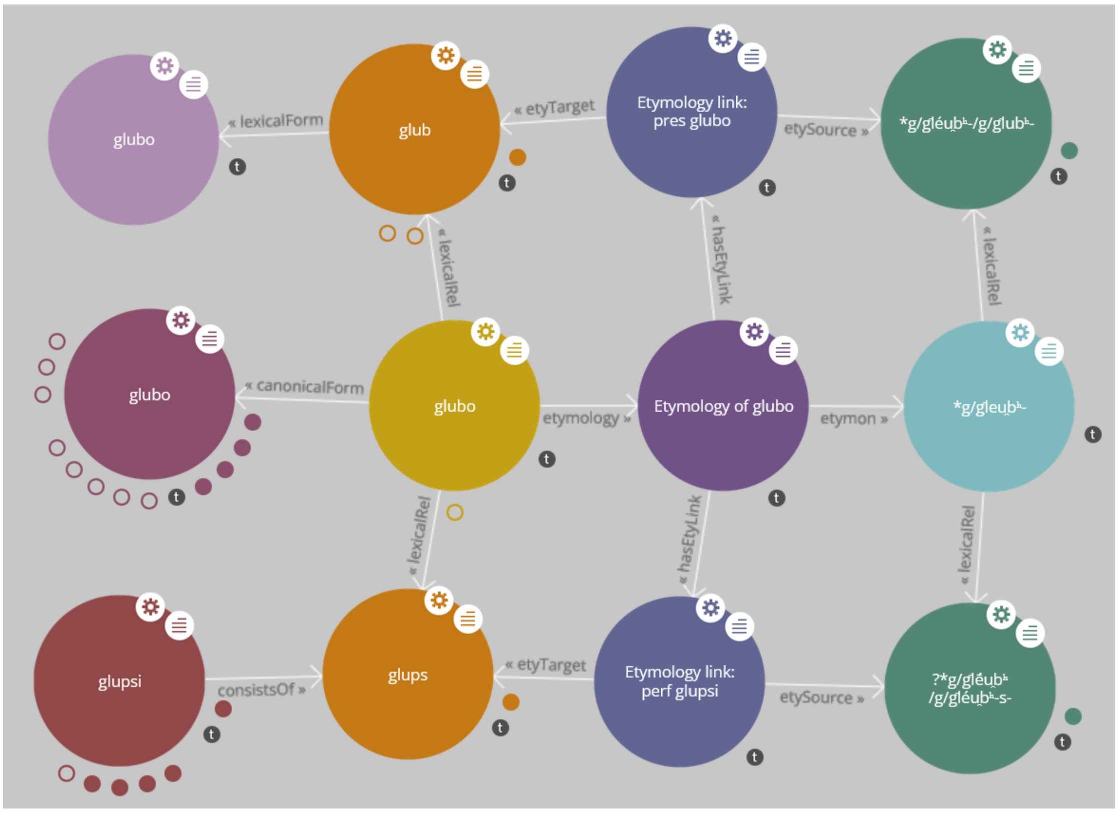

Figure 1: The linking of LIV etymological relations: the case of *glubo*.

reify the etymological relations that the LIV postulates between the stems, and constitute the bridge between Latin and PIE.

On the right side of the figure is the PIE symmetrical counterpart of the model. The PIE root **g/ĝléu̯b^h-* (which is an individual of the class `Etymon`, subclass of the class `LexicalEntry`) is linked to the two PIE stems via the property `lexicalRel`, in the same way as the `LexicalEntry` is linked to the Latin stems.

Finally, the generic etymological relation between the PIE root and the Latin lexical entry is reified by the `Etymology` class: this class establishes a link between them via the properties `etymon` and `etymology`, respectively. The `Etymology` is also connected with the two `EtyLink`, thanks to the property `hasEtyLink`, and thus constitutes a central crossroad between the LIV lexical items.

4. Querying the LIV data in LiLa

The modelling and linking work has, as shown above, benefited greatly from the advantages provided by the LOD paradigm. The re-use of the lemmas from the LiLa Lemma Bank as canonical forms for the LIV entries has allowed us to retrieve the stems from *PrinParLat*, as well as all words derived from a handful of selected lexical bases.

The exploration of the full set of words derived via the regular Latin word-formation rules from verbs of IE origin can be extended to all the entries in the LIV, beyond the 11 entries explicitly marked as bases in the dictionary. In fact, 342 regular entries of the LIV currently linked to a LiLa lemma are connected to a lexical base. Via this relation, we can access a set of 4,019 other verbs. Also, by leveraging the links in the LiLa network to textual resources, we can easily access the earliest occurrences in the corpora.

The full network of the resources linked to LiLa allows for even more advanced inquiries in historical linguistics

Entry URI	Lexical Base	Nr. of verbs
liv:5	Base of *facio*	256
liv:147	Base of *ago*	71
liv:157	Base of *fio*	67
liv:252	Base of *fero*	53
liv:167	Base of *capio*	51
liv:260	Base of *eo*	49

Table 1
LIV Entries connected to the most productive lexical bases in LiLa, with nr. of verbs (LIV entry excluded) connected to it.

and in the study of Latin lexicon. Table 1[29] reports the most productive bases connected to entries in the LIV, with the number of other verbs linked to each lexical base (note that the lemma of the LIV entry was excluded from the calculation).

As can be seen, the most productive words are some of the most common verbs belonging to the oldest IE substratum of Latin, like *facio* 'to do, make', *fero* 'to bring' or *capio* 'to seize, take'. Indeed, a joint query between the two dictionaries with IE etymologies, viz. LIV and [16], and the lexical bases in LiLa confirm that the bases that have at least one lemma that is traced back to PIE are considerably richer and more productive than those lexical families without any inherited lexemes. While the former have on average 23.70 members, the latter display only an average of 4.86 members. This fact can be easily explained considering that the latter group is mostly made up by loanwords, which are generally technical terms (especially from the Greek scientific or technical lexicon), and tend to be more specialised and less productive in terms of word formation.

The two dictionaries combined provide now information on PIE etymologies for 1,473 lemmas: 1,393 are connected to entries in the *Etymological Dictionary of Latin and the Other Italic Languages* [16, 17], 355 in the LIV. In particular, 275 lemmas are shared by the two resources: for these entries, it is therefore now possible to use LiLa to compare the approach to etymological reconstruction by the LIV and the dictionary by de Vaan.

5. Conclusion and Future Work

Linking a set of data from the LIV to LiLa enhances the Knowledge Base with etymological information about the processes that, starting from PIE roots, have led to the formation of the Latin word forms. Given the highly lexically-based nature of the architecture of the Knowledge Base, this makes the linking of the LIV an important achievement of the LiLa project.

The information provided by the LIV is now interoperable with that of the several other lexical resources currently interlinked through LiLa and can be queried together with the textual data provided by the Latin corpora published in the Knowledge Base.

In collecting and publishing as LOD the wealth of different digital resources for Latin built so far, an important challenge is to impact the scholarly community that has long been using the data provided today by these resources. The web-based interoperability among resources permitted by the LiLa Knowledge Base makes it possible to exploit such wealth of (meta)data like never before, in terms both of the quantity of the (meta)data under analysis and of the quality of the process leading to their retrieval. Interlinking through the Knowledge Base a set of data from the LIV, a reference lexical resource for the communities of Classicists and Historical Linguists, is expected to help overcome the challenge of making the use of LiLa a daily presence in the life of scholars who work in the fields of Classics and Historical Linguistics.

Finally, it is worth considering that the LIV provides etymological information not only about the Latin word forms, but, for each PIE root, it also reports a set of word forms which reflect the same root in several other IE daughter languages. The availability of this information allows for substantial research work to be performed in the near future. Indeed, by applying the principles of the Linked Data paradigm and reusing the same vocabularies adopted in LiLa to interlink the distributed linguistic resources for Latin, it is now possible to move one step further from the Latin language and aim to make interoperable word forms from several IE languages, by using the collection of PIE roots provided by the LIV as a kind of pivot resource to interlink them all.

6. 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. We thank the Reichert Verlag publisher for granting us permission to use some data from the LIV. We thank Matteo Pellegrini for his valuable insights about PrinParLat.

References

[1] M. Passarotti, The project of the index thomisticus treebank, Digital classical philology. Ancient Greek and Latin in the digital revolution 10 (2019) 299–319.

[2] C. Lewis, C. Short, A Latin Dictionary. Founded on Andrews' edition of Freund's Latin dictionary, Clarendon Press, Oxford, 1879.

[29] The namespace liv in Tab. 1 refers to the URL http://lila-erc.eu/data/lexicalResources/LIV/id/LexicalEntry/.

[3] C. Chiarcos, P. Cimiano, T. Declerck, J. P. McCrae, Linguistic linked open data (LLOD). introduction and overview, in: Proceedings of the 2nd Workshop on Linked Data in Linguistics (LDL-2013): Representing and linking lexicons, terminologies and other language data, Association for Computational Linguistics, Pisa, Italy, 2013, pp. i – xi. URL: https://aclanthology.org/W13-5501.

[4] H. Rix, ed., LIV. Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen, 2nd ed., Reichert Verlag, Wiesbaden, 2001.

[5] 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.

[6] M. Wilkinson et al., The fair guiding principles for scientific data management and stewardship., Scientific Data 3 (2016). doi:https://doi.org/10.1038/sdata.2016.18.

[7] O. Lassila, R. R. Swick, Resource Description Framework (RDF) Model and Syntax Specification, 1998. URL: https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/.

[8] T. Berners-Lee, The semantic web, Scientific American 284 (2001).

[9] N. Ide, J. Pustejovsky, 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, 2010.

[10] C. Chiarcos, M. Sukhareva, Olia – ontologies of linguistic annotation, Semantic Web 6 (2015) 379–386. doi:10.3233/SW-140167.

[11] 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.

[12] F. M. Cecchini, M. Passarotti, P. Marongiu, D. Zeman, Challenges in Converting the Index Thomisticus Treebank into Universal Dependencies, in: Proceedings of the Second Workshop on Universal Dependencies (UDW 2018), Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 27–36. URL: https://aclanthology.org/W18-6004. doi:10.18653/v1/W18-6004.

[13] F. 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 (CLiC-it 2020), Bologna, 2020.

[14] E. Litta, M. Passarotti, F. Mambrini, The Treatment of Word Formation in the LiLa Knowledge Base of Linguistic Resources for Latin, in: Proceedings of the Second International Workshop on Resources and Tools for Derivational Morphology (DeriMo 2019). 19-20 September 2019, Prague, Czechia, Institute of Formal and Applied Linguistics, Charles University in Prague, Prague, Czech Republic, 2019, pp. 35–43. URL: https://ufal.mff.cuni.cz/derimo2019/pdf-files/derimo2019.pdf.

[15] F. Mambrini, M. Passarotti, E. Litta, G. Moretti, Interlinking Valency Frames and WordNet Synsets in the LiLa Knowledge Base of Linguistic Resources for Latin, in: M. Alam, P. Groth, V. de Boer, T. Pellegrini, H. J. Pandit, E. Montiel, V. Rodríguez Doncel, B. McGillivray, A. Meroño-Peñuela (Eds.), Further with Knowledge Graphs. Studies on the Semantic Web 53, IOS Press, Amsterdam, 2021. URL: https://ebooks.iospress.nl/doi/10.3233/SSW210032. doi:10.3233/SSW210032.

[16] M. de Vaan, Etymological Dictionary of Latin and the Other Italic Languages, Brill, Leiden and Boston, 2008.

[17] F. Mambrini, M. Passarotti, Representing Etymology in the LiLa Knowledge Base of Linguistic Resources for Latin, in: Proceedings of the Globalex Workshop on Linked Lexicography. LREC 2020 Workshop, European Language Resources Association (ELRA), Paris, 2020, pp. 20–28. URL: https://lrec2020.lrec-conf.org/media/proceedings/Workshops/Books/GLOBALEX2020book.pdf. doi:10.5281/zenodo.3862156.

[18] M. Pellegrini, Flexemes in theory and in practice, Morphology (2023) 1–35.

[19] H. Rix, ed., LIV. Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen, Reichert Verlag, Wiesbaden, 1998.

[20] J. Pokorny, Indogermanisches etymologisches Wörterbuch (IEW), Francke Verlag, 1959.

[21] A. Khan, Towards the representation of etymological data on the semantic web, Information 9 (2018).

[22] G. Franzini, F. Zampedri, M. Passarotti, F. Mambrini, G. Moretti, Græcissâre: Ancient Greek Loanwords in the LiLa Knowledge Base of Linguistic Resources for Latin, in: Proceedings of the Seventh Italian Conference on Computational Linguistics. Bologna, Italy, March 1-3, 2021, CEUR-WS.org, Bologna, 2020, pp. 1–6. URL: http://ceur-ws.org/Vol-2769/paper_06.pdf.

[23] C. Du Cange, Bénédictins de Saint-Maur, P. Carpentier, L. Henschel, L. Favre, Glossarium Mediae et Infimae Latinitatis, Léopold Favre, Niort, 1883-1887.