PREMOVE In LiLa: Integrating Latin Preverbed Motion Verbs With WordNet And VerbNet

Andrea Farina[1], Marco Passarotti[2], Francesco Mambrini[2]
Matteo Pellegrini[3], Giovanni Moretti[2], Eleonora Litta[2]
[1]King's College London, United Kingdom
[2]Università Cattolica del Sacro Cuore, Milan, Italy
[3]University of Surrey, Guildford, United Kingdom
andrea.farina@kcl.ac.uk, marco.passarotti@unicatt.it, francesco.mambrini@unicatt.it,
matteo.pellegrini@surrey.ac.uk, giovanni.moretti@unicatt.it, eleonora.litta@unicatt.it

Abstract

PREMOVE is a diachronic dataset of Ancient Greek and Latin PREverbed MOtion VErbs, providing manually curated morphological, syntactic, and semantic annotations for almost three thousand verbal occurrences. This paper presents the integration of PREMOVE into the LiLa Knowledge Base of Latin, linking its semantic annotations to WordNet (WN) and VerbNet (VN). We describe the RDF conversion using OntoLex-Lemon and FrAC, enabling explicit modelling of token-level attestations and dataset-level provenance. The resulting linked resource achieves full FAIR compliance and supports complex SPARQL queries, allowing users to explore motion semantics across lexical, textual, and semantic layers. Example SPARQL queries demonstrate how researchers can retrieve attested forms for specific WN synsets or VN classes, supporting reproducible linguistic research and cross-resource exploration of motion semantics in ancient languages.

Keywords: PREMOVE, Lila Knowledge Base, OntoLex-Lemon

1. Introduction

Preverbs, i.e., prefixes attaching onto verbs like *under-* in Eng. *undergo*, play a central role in the study of language, contributing not only to the understanding of verb semantics and syntactic constructions, but also to broader questions in historical linguistics, diachronic syntax, and lexical typology (Booij and Van Kemenade, 2003; McGillivray, 2014). This holds especially for ancient Indo-European languages, among which Ancient Greek and Latin. Despite their importance and the great number of qualitative and quantitative studies on preverbs, until recently no systematic datasets documenting these phenomena across languages and periods were available, limiting opportunities for computational analysis and cross-resource comparison. To address this gap, a diachronic dataset of Ancient Greek and Latin PREverbed MOtion VErbs (PREMOVE) has now been created (Farina, 2025a).

In PREMOVE, verbal occurrences are annotated for morphology, syntax, and semantics (Farina, 2024) with token-level information including manual word-sense annotations linked to external resources such as WordNet (WN) (Fellbaum, 1998) and VerbNet (VN) (Kipper-Schuler, 2005). Originally released in CSV format and deposited in the Oxford Text Archive via CLARIN-UK[1], PREMOVE ensures long-term preservation and persistent citation. Depositing PREMOVE into CLARIN ensures compliance with many of the principles for FAIR data (Wilkinson et al., 2016), while also achieving three stars in Berners-Lee's 5-star rating system for Linked Open Data (LOD) (Berners-Lee, 2006-2009). In this paper, we describe how PREMOVE's semantic annotations – both WN- and VN-based – have been linked to the LiLa Knowledge Base (KB), where various language resources for Latin are made interoperable through their publication in accordance with the principles of the LOD paradigm (Passarotti et al., 2020). This integration has elevated the dataset's LOD compliance from three to five stars, also ensuring higher semantic interoperability thanks to the adoption of Semantic Web standards (Farina, 2025b). From a practical point of view, integration into LiLa allows PREMOVE's annotations to interact with a larger ecosystem of Latin lexical and textual resources, enhancing interoperability and enabling semantic querying across corpora and lexical resources.

The primary contributions of this paper are twofold: (i) the design and implementation of a FrAC- and OntoLex-Lemon-based RDF conversion of PREMOVE, capturing token-level attestations and dataset-level provenance, and (ii) the integration of the Latin data included in PREMOVE into LiLa, focusing on the alignment of WN synsets and VN classes with token- and lemma-level identifiers.

The paper is organized as follows. Section 2 introduces Linguistic LOD. Section 3 introduces the LiLa KB and the PREMOVE dataset. In Section 4, we describe the preprocessing steps required to

[1]http://hdl.handle.net/20.500.14106/2579

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 3672–3683
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

prepare PREMOVE for linking and RDF conversion. Section 5 presents the data modeling approach adopted for linking the PREMOVE annotations, relying on the OntoLex-Lemon model with FrAC attestations to represent token-level information. Sections 6 and 7 provide details of the resulting RDF triples and illustrate example SPARQL queries that can be performed now that PREMOVE is integrated into LiLa. Finally, Section 8 concludes the paper and outlines directions for future work.

2. Linguistic Linked Open Data

Linguistic Linked Open Data (LLOD) represent an integration of semantic web principles and linguistic resources, with the main goal of making linguistic (meta)data interoperable on a global scale. This approach leverages the principles of Linked Data (Berners-Lee et al., 2001), such as the use of unique URIs, standardized data models, such as RDF (Resource Description Framework) (Lassila et al., 1998), and linking between datasets, to create an open language knowledge network.

An important role in the world of LLOD is played by the scientific community, which has the responsibility to engage in constant dialogue in order to identify common practices for representation (through shared ontologies) and, consequently, for the publication of language resources. The recently concluded COST Action Nexus Linguarum[2] has been an excellent opportunity to bring together the LLOD community, achieving both the creation of new modules extending existing ontologies and the organization of training activities. Regarding the former, Nexus Linguarum extensively worked on OntoLex-Lemon (McCrae et al., 2017), the standard model for representing lexical and terminological resources as LOD, further developing related models such as FrAC for for frequency, attestation and corpus information[3] and Morph for morphology[4]. As for the latter, the COST Action organized a series of Datathons on LOD and developed a MOOC with two modules, one foundational and one advanced (Gracia et al., 2025).

There are many initiatives connected to the world of LLOD, among which two of the most general are particularly worthy of mention: the LLOD Cloud (Linguistic Linked Open Data Cloud) (Cimiano et al., 2020, 29-41)[5], a network of linguistic resources published as LOD, and ELEXIS[6], a European infrastructure for linking lexicographic resources and promoting the creation of digital lexicons as LOD.

[2] https://nexuslinguarum.eu

[3] https://acoli-repo.github.io/ontolex-frac/

[4] https://ontolex.github.io/morph/

[5] https://linguistic-lod.org/llod-cloud

[6] https://elex.is

DBnary (Sérasset, 2015) is a resource that converts Wiktionary data to LOD, offering multilingual lexical information. DBpedia (Auer et al., 2007) is a project aiming to extract LOD-structured content from Wikipedia, thus allowing users to semantically query relationships and properties of Wikipedia resources, including links to other related data sets. BabelNet (Navigli and Ponzetto, 2010) is an encyclopedic dictionary that provides concepts and named entities lexicalized in many languages. Words are connected with a large number of semantic relations and grouped into sets of synonyms.

In the LOD landscape, a sector that has seen a large growth over the last decade is the one dedicated to building, sharing, and exploiting LOD collections of various kinds of (meta)data related to the Ancient World. This has particularly concerned documents and materials related to Latin and Ancient Greek. Several different kinds of data are concerned by the various LOD initiatives and data sets related to the Ancient World. Geographical data on the Ancient World are published as LOD by initiatives such as Pleiades (Elliott, 2021), Pelagios (Simon et al., 2016), ArkeoGIS (Brovelli and Maurino, 2000), and Trismegistos (Depauw and Gheldof, 2014), which also publishes epigraphic documents, manuscripts, as well as information on ancient individuals and authors. The latter are also addressed by the Standard for Networking Ancient Prosopography (Bodard et al., 2017), while artifacts and archaeological data are handled by the Ariadne project[7].

Regarding the publication of linguistic resources for Classical languages as LOD, the LiLa KB of interoperable resources for Latin[8] represents the main initiative and the quintessential use case of LLOD.

3. LiLa and PREMOVE

3.1. The LiLa Knowledge Base

LiLa (Linking Latin) is a LOD Knowledge Base developed to promote interoperability across a broad spectrum of textual and lexical resources for Latin. Its conceptual model revolves around two primary components:

1. the Lemma Bank, a collection of approximately 200,000 Latin lemmas (canonical citation forms of lexical items) published as LOD and originating from the LEMLAT 3.0 morphological analyzer (Passarotti et al., 2017) ;
2. a set of linguistic resources for Latin published as LOD and interconnected through the Lemma

[7] https://ariadne-infrastructure.eu

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

Bank, including corpora, lexica, and dictionaries.[9]

As new resources are integrated, the Lemma Bank is continually expanded, while resources link back to the Lemma Bank by connecting their lexical entries in lexical resources and individual word occurrences (tokens) in textual resources to the corresponding lemma in the LiLa Lemma Bank.

The LiLa KB leverages several established ontologies to represent the (meta)data of interlinked linguistic resources. Chief among these are POWLA for corpus data (Chiarcos, 2012), OLiA for linguistic annotation (Chiarcos and Sukhareva, 2015), and Ontolex-Lemon for lexical data (McCrae et al., 2017). In addition, LiLa employs its own ontology to model lemmas in the Lemma Bank as instances of the class `lila:Lemma`, defined as a subclass of `ontolex:Form`. The class `lila:Lemma` has a specific subclass `lila:Hypolemma`, whose instances are citation forms that belong to a word's regular inflectional paradigm but receive a different PoS tag or degree of comparison than their 'most canonical' lemma, including participles, gerundives, deadjectival adverbs, and comparatives.

For lexical resources, each lexical entry, modeled using the class `ontolex:LexicalEntry`, is connected to its corresponding lemma in the Lemma Bank through the property `ontolex:canonicalForm`. With respect to textual resources, tokens are represented as instances of the class `Terminal` in the POWLA ontology and linked to their corresponding lemma in the Lemma Bank via the property `lila:hasLemma`.

3.2. PREMOVE: Scope and Annotation

The PREMOVE dataset (Farina, 2026) includes annotations of preverbed motion verbs in Ancient Greek and Latin, with the aim of supporting both data-driven historical cross-linguistic research (Farina et al., 2025b) and computational reuse (Farina et al., 2025a). In its current release, PREMOVE contains 2,834 verbal occurrences, encoded in a CSV format with 42 columns and more than 119,000 individual annotations. Each row corresponds to a uniquely identified verbal token, linked to its source text and enriched with multiple layers of linguistic and contextual information. Table 1 provides an overview of the Latin texts whose annotation is included in PREMOVE. The corpus has been designed to balance chronological distribution, literary genres, and textual registers, so as to capture a representative sample of preverbed motion verbs across different contexts of use. For a deeper explanation of textual choices, we refer to Farina (2026).

[9]The full list of resources currently interlinked in LiLa is available at `https://lila-erc.eu/data-page/`.

Table 1: Overview of Latin Texts in the Corpus.

Author	Text	Token Count
Ennius	Annales	1,194
Plautus	Amphitruo	9,988
	Mostellaria	9,780C
Caesar	De bello Gallico 1-4	20,498
Cicero	In Catilinam 1-3	11,625
	De amicitia	9,471
Sallust	Bellum Catilinae	10,655
Livy	Ab Urbe condita 1-2	39,913
Virgil	Aeneid	63,719
Propertius	Elegies 1.1-1.22	4,384
Horace	Satires	7,048
Seneca	De ira	22,614
	Medea	5,639
Tacitus	Historiae 1	11,852
Suetonius	Life of August	13,915
Apuleius	Metamorphoses 1–5	23,358

Although each token is annotated at different levels (morphology, syntax, semantics), for the purposes of this paper we focus on the semantic layer of PREMOVE. Each verbal occurrence is assigned both a semantic class from VN, grouping verbs by shared lexical-semantic properties (e.g., Eng. *go*, *advance*, *come*, or *arrive* are attributed class *ESCAPE-51.1-1*), and a sense from WN, providing a standardized semantic representation. In WN, word senses are given in the form of sets of synonyms (synsets), with a unique ID and a gloss explaining the sense (e.g., Eng. *go*, *locomote*, *travel*, and *move* share synset `01839438-v`, glossed as 'change location; move, travel, or proceed, also metaphorically'). This combination of class-based and sense-based annotation enables fine-grained analyses of motion verbs while ensuring linking with other lexical resources and tools. In addition, the dataset records whether a given use is literal or figurative, a distinction that is especially relevant for preverbed verbs in Ancient Greek and Latin, where processes of lexicalization and semantic extension are frequent (Cuzzolin, 1995; Bertocci, 2017; Farina, 2025c). Beyond verb-centered annotation, PREMOVE also encodes the semantics of participants in motion events, i.e. Figure and Ground (Talmy, 1975, 1983, 2000), where the Figure denotes the moving entity and the Ground the reference location or entity relative to which motion occurs. These are also semantically annotated through WN synsets, allowing systematic comparison of spatial roles across texts and languages.

A central concern in corpus-based linguistic research is the reliability of semantic annotations (Ide and Pustejovsky, 2017). As PREMOVE was developed as part of a doctoral project (Farina, 2022-

2026) all annotations were produced by a single author. To provide an indication of consistency, Inter-Annotator Agreement (IAA) was calculated across three annotators on a small subset of 20 sentences[10], on different annotation layers (Farina, 2026). Some surface-level features (such as the literal/compositional meaning of preverbed forms) achieved perfect agreement ($\kappa = 1$). For the WN-based categories, agreement ranged between fair and almost perfect, reflecting the challenges of assigning fine-grained synsets to language data (e.g., Palmer et al., 2004), especially in classical languages (Biagetti et al., 2021). Noun (i.e., Figures, Grounds, spatial relations) semantics yielded higher scores compared to verb semantics, reflecting the higher degree of polysemy of verbs compared to nouns (Jamrozik et al., 2013; Krennmayr, 2017; King and Gentner, 2022), as well as the higher granularity of English synsets, where multiple closely related senses often exist. The agreement on VN annotations was lower, although still fair. This is probably due to the difficulty of mapping (Ancient Greek and) Latin verb categories onto the English-centric VN classes. Moreover, subtle distinctions between classes can be difficult to resolve, and verbs may belong to more than one category.

4. PREMOVE: Dataset Preparation

The integration of PREMOVE into LiLa required a sequence of preparatory operations designed to (i) normalize token- and lemma-level information; (ii) align semantic identifiers across WN versions; and (iii) link PREMOVE's sense-level annotations to LiLa token URIs. Below we describe the annotation types targeted for linking (Section 4.1), the cleaning and remapping operations (Section 4.2), and the SPARQL-driven retrieval strategy (Section 4.3).

Before detailing these steps, it is important to note that not all Latin texts included in PREMOVE (Table 1) are currently represented in LiLa. Specifically, Ennius' *Annales*, Plautus' *Mostellaria*, Livy's *Ab Urbe condita*, Suetonius' *Life of Augustus*, and Apuleius' *Metamorphoses* are missing, as they were not present in LiLa's sources. As a result, 518 out of 1,484 tokens (35%) drawn from these works could not be linked to LiLa and were therefore excluded from the linking procedure described below. Nevertheless, their inclusion in the broader PREMOVE corpus remains essential, as they provide valuable chronological and stylistic breadth for the study of preverbed motion verbs (Farina, 2026).

[10]The IAA for the semantics of preverbs, central to PREMOVE, was calculated on 103 occurrences (Farina and Ciletti, 2026), yielding substantial agreement ($\kappa = 0.813$, $p < .001$).

4.1. Target Annotations and Scope

Although PREMOVE encodes 42 columns of linguistic and contextual information, the linking effort started from the semantic layers (Section 3.2): WN-based sense annotations and VN-based verb classes.

A key structural feature of the two resources shaped the linking process. As mentioned in Section 3.1, the fundamental component of LiLa is the Lemma Bank. Lexical resources are organized around lexical entries (instances of the class `ontolex:LexicalEntry`), to which lexical metadata are associated – including the URIs of WN synsets, modeled as instances of `ontolex:LexicalConcept`. Each lexical entry is linked to its `lila:Lemma` in the Lemma Bank via the `ontolex:canonicalForm` property. Equally, tokens from corpora (each assigned a distinct URI) are connected to their `lila:Lemma` via a dedicated property (`lila:hasLemma`). PREMOVE, in turn, provides linguistic descriptions directly at the token level, encoding sense annotations for individual occurrences – thus enriching the token component of the LiLa infrastructure and effectively performing word sense disambiguation in context (Jurafsky and Martin, 2018). The integration therefore focuses on identifying, for each PREMOVE token, the corresponding token URI in LiLa and aligning its contextual sense annotation with the WN synset URIs available in LiLa. VN information, which is not currently part of LiLa, was linked externally – that is, added as an additional semantic layer aligned with the LiLa structure but maintained outside the LiLa graph.

4.2. Data Cleaning, Normalization, and Remapping

Before linking PREMOVE to LiLa, we applied a systematic cleaning and normalization pipeline to remove inconsistencies and harmonize orthography and token metadata. Main steps included, for instance: (i) orthographic normalization (*v*/*u* variants); (ii) unification of capitalization variants; (iii) removal of spurious whitespace and tokenization artefacts; and (iv) flagging of divergent manuscript readings for manual inspection.

Automated retrieval (Section 4.3) produced the majority of token matches, but also overgeneration and mismatches due to orthographic and tokenization differences between the corpora. Frequent issues included multiple candidate URIs for identical forms, misaligned annotations across works, clitic splits (*Noreiamque* vs. *Noreiam* + *que*), and spacing variants (*eiusmodi* vs. *eius modi*).

All automatically retrieved matches were subjected to manual validation. Correction followed explicit criteria: preference for identical textual context (preceding/following tokens), consistency with the source text, and alignment of orthographic variants after normalization. When no unambiguous correspondence could be established, the match was discarded rather than forced. Manual intervention was required only for the subset of automatically retrieved candidates exhibiting the mismatch classes described above. After validation and correction, 960 token matches were retained as valid links. This post-editing procedure can therefore be interpreted as precision-oriented filtering of the automated output, prioritizing reliability over coverage.

PREMOVE's WN synset offsets derive from the new WNs for historical languages (Biagetti et al., 2021), which are based on WN 2.0, whereas LiLa adopts WN 3.1. A synset remapping step was therefore required. Given the limited number of distinct synsets represented in PREMOVE, offsets were remapped manually by systematically comparing glosses, definitions, and sense inventories across versions. Out of the tokens considered for linking, 48 synset annotations raised mapping issues. These cases fell into the following categories:

1. **Minor lexical or phrasing changes not affecting the underlying synset** – remapped straightforwardly (e.g., 2.0 `v#01420490` 'run away' → 3.1 `02079709-v` 'run away quickly').
2. **Slight definitional shifts but semantically equivalent senses** – remapped to the closest corresponding 3.1 synset (e.g., 2.0 `v#01790203` 'come into existence or develop' → 3.1 `02629812-v` 'come into existence, originate').
3. **No direct 3.1 offset but an acceptable contextual equivalent** – selected on the basis of gloss comparison and corpus usage (e.g., 2.0 `v#00543260` 'make a retreat' → 3.1 `02018017-v` 'leave').
4. **No suitable 3.1 synset available** – annotation conservatively left unmapped to avoid introducing spurious alignments (e.g., 2.0 `v#01382085` 'meet').
5. **Lemma absent from the LiLa WN layer** – could not be mapped; left for future enrichment (one case only: *transuolo*).

All remapping decisions were manually verified to ensure semantic consistency across versions.

4.3. SPARQL-based Token Retrieval

Token identification in LiLa was obtained through context-sensitive SPARQL queries that search for a specific token and its immediate context (i.e., the preceding and following tokens). These queries can be executed via the LiLa SPARQL access point at https://lila-erc.eu/sparql/, or programmatically through the API endpoint at https://lila-erc.eu/sparql/lila_knowledge_base/sparql . A representative query is provided in Listing 1 below.

Listing 1: SPARQL query to retrieve the token URI of a specific token (*uictor*) given its surrounding context (previous and next token, i.e., *hoc* and *abibat*).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax
    -ns#>
PREFIX powla: <http://purl.org/powla/powla.owl#>

select ?token_URI ?token WHERE {
  ?token_URI rdf:type powla:Terminal ;
       rdfs:label "uictor" ;
       powla:next ?next_tok ;
       powla:previous ?prev_tok .

  ?next_tok rdfs:label "abibat" .
  ?prev_tok rdfs:label "hoc" .
  }
```

After manually correcting mismatches (Section 4.2), we used a second query to retrieve the URI of the passage (parent container) containing a matched token; this step was necessary to anchor PREMOVE rows to the exact textual passage in LiLa.

5. Ontology and Modelling: FrAC attestation

OntoLex-Lemon is a standard vocabulary for the representation of lexical information, providing classes and properties to describe both formal and semantic aspects of lexical entries. It consequently plays a key role in the modelling of lexical resources linked to the LiLa KB, including the ones that provide information relevant to the content of PREMOVE – such as the Latin WN (LWN). OntoLex-Lemon classes and properties, such as `evokes/isEvokedBy`, `LexicalConcept`, `LexicalSense`, are used to code the semantic information provided in the resources, such as WN synsets, while the property `canonicalForm` is used for the linking to LiLa lemmas (Section 3). However, as we have seen in Section 4.1, PREMOVE gives annotations on those same semantic pieces of information but at the level of occurrences of words in context.

In addition to the core module, OntoLex-Lemon is characterized by a modular architecture, with separate modules designed to address more specific issues, such as syntax and semantics (synsem), decomposition (decomp), variation and translation (vartrans) and linguistic metadata (lime). Furthermore, besides the main four modules that extend the core, a series of models are currently being drafted by the community to provide OntoLex-Lemon with greater expressivity and extend beyond the original use case of lexicalizing ontologies towards other area of linguistic analysis and Natural Language Processing.

One of those extensions, currently still in draft status, is the Ontolex Module for Frequency, Attestation and Corpus Data (FrAC; Chiarcos et al., 2020). The model was designed to support two main use cases, namely extending the usability of OntoLex-Lemon in NLP applications in computational semantics, and supporting corpus-based lexicography. To this end, FrAC provides classes and properties to express *observations* derived from textual resources. That design choice makes it the ideal vocabulary to bridge information at the lexical and textual level in PREMOVE. The fact that FrAC is designed as an extension of OntoLex also ensures total interoperability with LiLa.

The main class of the module is the Observable, which encompasses all concepts from Ontolex that can be observed in corpora (lexical entries, forms, senses, lexical concepts)[11]. As said, what PREMOVE "observes" in texts is the occurrence of WN synsets and VN predicate frames, which in OntoLex-Lemon are both defined as lexical concepts. The observations that can be predicated on the observables are frequencies, collocations and attestations. For the modeling of PREMOVE we focus on the latter.

In FrAC, attestations are conceptualized as forms of citation that include a quotation from a resource and "provide evidence for the existence of a certain lexical phenomena". PREMOVE annotations on Latin texts align completely with this definition. In FrAC, attestations record the text of the quotation from the original source. While more formal ontologies for bibliographic citation, such as LRMoo[12], FaBiO[13], are supported by FrAC, in our implementation we opted to leverage the link with LiLa and make use of the identifiers for textual passages for referencing. LiLa's corpora represent the canonical citations of classical texts (Smith, 2009) as a layer of stratified citation units (Mambrini et al., 2022). Thus, the first paragraph in the second chapter of the first book of Caesar's *Gallic War* (canonically referred to in Classical Studies with a short citation like "Caes. *BG* 1.2.1") in any editions in LiLa is identified with its own URI[14]. This URI can be associated with attestation instances of FrAC using the `frac:observedIn` dedicated object property, which suffices to identify the passage univocally and canonically, while also allowing users to locate it precisely within a corpus. A further level of detailed identification is provided by the `frac:locus` property, which points to the URI of the exact token bearing the attestation.

6. Linking PREMOVE: RDF Encoding

To illustrate the data modelling adopted for PREMOVE, we provide below two representative excerpts of the RDF encoding used for linking token-level attestations to WN and VN semantic identifiers. All code used to generate these triples, including preprocessing scripts and SPARQL mapping queries, is available on GitHub[15] for full transparency and reproducibility.

In addition to token-level annotations, all attestations are explicitly linked to the PREMOVE dataset as a whole. Each attestation is associated with the dataset-level URI through `dcterms:isPartOf` to provide clear provenance and to facilitate cross-resource interoperability. Following automated retrieval, manual correction, and WN remapping, we obtained 960 token-level matches between PREMOVE and LiLa (Section 4.2).

6.1. WN Synsets

Listing 2 illustrates an example of how WN-based sense annotations from PREMOVE are modelled as FrAC attestations, linking each `ontolex:LexicalConcept` to its textual evidence through the property `frac:attestation`.

Listing 2: RDF encoding of WN attestations.

```
@prefix : <http://lila-erc.eu/data/dataset/PREMOVE/
     id/attestation/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-
     syntax-ns#> .
@prefix ontolex: <http://www.w3.org/ns/lemon/ontolex
     #> .
@prefix sn: <http://lila-erc.eu/data/
     lexicalResources/LatinWordNet/id/LexicalConcept
     /> .
@prefix frac: <http://www.w3.org/ns/lemon/frac#> .
@prefix dcterms: <http://purl.org/dc/terms/> .

sn:08683242-n a ontolex:LexicalConcept ;
    frac:attestation :oppidum_attestation .

:oppidum_attestation a frac:Attestation ;
    rdf:value "Qui cum se suaque omnia in oppidum
        Bratuspantium contulissent..." ;
```

[11]The latest draft of FrAC can be read at https://github.com/ontolex/frequency-attestation-corpus-information/blob/master/index.md.

[12]https://cidoc-crm.org/lrmoo.

[13]https://sparontologies.github.io/fabio/current/fabio.html.

[14]E.g., this section in one corpus (LASLA) is http://lila-erc.eu/data/corpora/Lasla/id/corpus/CaesarBellum%20Gallicum/CiteStructure/Liber_1/Capitulum_2/Paragraphus_1.

[15]https://github.com/farina-andrea/PREMOVE_LiLa.

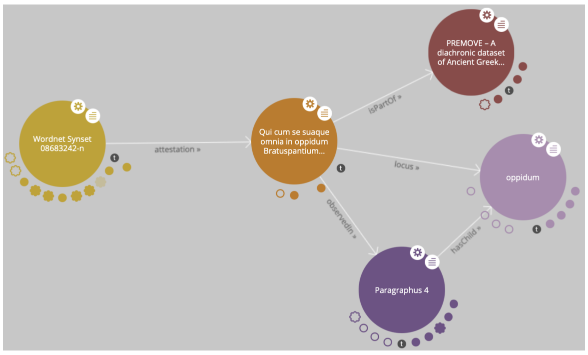

Figure 1: The FrAC attestation model: a WN `LexicalConcept` (synset) is linked to its textual attestation, connecting to dataset-level (`dcterms:isPartOf`) and corpus-structural (`frac:locus`, `frac:observedIn`) information.

```
    frac:locus <http://lila-erc.eu/data/corpora/
        Lasla/id/corpus/CaesarBellumGallicum/
        CaesBG2.BPN_t_0001666> ;
    frac:observedIn <http://lila-erc.eu/data/corpora
        /Lasla/id/corpus/CaesarBellumGallicum/
        CiteStructure/Liber_2/Capitulum_13/
        Paragraphus_2> ;
    dcterms:isPartOf <http://lila-erc.eu/data/
        dataset/PREMOVE/dataset> .
```

In this example, the LWN synset `sn:08683242-n` 'an urban area with a fixed boundary that is smaller than a city', lexicalized by *oppidum* 'town, fortified settlement', is declared as an `ontolex:LexicalConcept` and linked, via `frac:attestation`, to a specific instance of usage attested in Caesar's *Bellum Gallicum*. The `rdf:value` preserves the textual context, while `frac:locus` and `frac:observedIn` connect the attestation to the token-level and passage-level URIs in LiLa, respectively. Finally, `dcterms:isPartOf` ensures dataset-level provenance by declaring that each attestation belongs to the PREMOVE dataset. A schematic representation of this same *oppidum* attestation is shown in Figure 1, illustrating how the FrAC model connects WN concepts, corpus tokens, and textual evidence. The PREMOVE dataset is available as a LiLa dataset entity at `http://lila-erc.eu/data/dataset/PREMOVE/dataset`.

6.2. VN Classes

Listing 3 shows an RDF excerpt modelling VN classes from PREMOVE. Structurally, these triples closely mirror the approach used for WN: each VN class is instantiated as both an `ontolex:LexicalConcept` and a `pmo:SemanticClass`, and linked to token-level attestations via `frac:attestation`. For instance, VN class `vn33-escape-51.1` includes basic motion events (Section 3.2). In Listing 3, each attestation preserves the contextual form (`rdf:value`) and connects to the LiLa token (`frac:locus`) and the containing passage (`frac:observedIn`), exactly as in the WN model. Dataset provenance is recorded via `dcterms:isPartOf`.

Listing 3: RDF encoding of VN attestations.

```
@prefix : http://lila-erc.eu/data/dataset/PREMOVE/id
    /attestation/ .
@prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax
    -ns# .
@prefix ontolex: http://www.w3.org/ns/lemon/ontolex#
    .
@prefix vn: http://premon.fbk.eu/resource/ .
@prefix frac: http://www.w3.org/ns/lemon/frac# .
@prefix pmo: http://premon.fbk.eu/ontology/core# .
@prefix pmovn: http://premon.fbk.eu/ontology/vn#
    VerbClass .
@prefix dcterms: http://purl.org/dc/terms/ .

vn:vn33-escape-51.1 a ontolex:LexicalConcept , pmo:
    SemanticClass , pmovn:VerbClass ;
frac:attestation :exirent_attestation .

:exirent_attestation a frac:Attestation ;
rdf:value "Pisone consulibus ... exirent" ;
frac:locus http://lila-erc.eu/data/corpora/Lasla/id/
    corpus/CaesarBellumGallicum/CaesBG1.
    BPN_t_0000212
 ;
frac:observedIn http://lila-erc.eu/data/corpora/
    Lasla/id/corpus/CaesarBellumGallicum/
    CiteStructure/Liber_1/Capitulum_2/Paragraphus_1
 ;
dcterms:isPartOf http://lila-erc.eu/data/dataset/
    PREMOVE/dataset
 .
```

The adoption of this parallel triple structure enables PREMOVE to achieve uniform representation of lexical semantics, whether derived from WN or VN. The main difference is the origin of the identifiers: WN synsets are native to LiLa, whereas VN classes are external semantic resources reused for interoperability. Despite this, the consistent FrAC-based modelling enables seamless integration, cross-framework querying, and comparative analyses of lexical and event semantics across resources and languages.

7. Exploring PREMOVE Annotations with SPARQL

PREMOVE's RDF encoding of token-level attestations, linked to WN synsets and VN classes, enables flexible and fine-grained querying. This section presents representative SPARQL examples illustrating how semantic information can be retrieved at the token level, corpus attestations explored, and lexical items connected to external resources, demonstrating PREMOVE's interoperability within the LiLa KB and the benefits of its FrAC-based modelling.

Listing 4 presents a SPARQL query retrieving all token-level attestations of the WN synset `02013448-v` 'go away from a place'. As shown in Figure 2, this query returns the textual form of each attested token in PREMOVE, the corresponding token-level URI in LiLa, and the passage-level

URI. The result table is entirely downloadable. Using `frac:attestation` as the linking property allows researchers to trace each WN synset directly to the instances in which it occurs, enabling contextual analyses.

Listing 4: SPARQL query retrieving all PREMOVE attestations for WN synset `02013448-v`.

```
PREFIX frac: <http://www.w3.org/ns/lemon/frac#>
PREFIX ontolex: <http://www.w3.org/ns/lemon/ontolex
     #>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX lila: <http://lila-erc.eu/ontologies/lila/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax
     -ns#>

SELECT ?attestation ?token ?passage ?text
WHERE {
  VALUES ?syn {<http://lila-erc.eu/data/
       lexicalResources/LatinWordNet/id/
       LexicalConcept/02013448-v>}
    ?syn frac:attestation ?attestation .
    ?attestation rdf:value ?text ;
                 frac:locus ?token ;
                 frac:observedIn ?passage .
}
ORDER BY ?token
```

Building on this token-level retrieval, PREMOVE also supports corpus-sensitive analyses of sense distribution. Listing 5 illustrates this by querying the sense distribution of the Latin verb *abeo* across different works.

Listing 5: SPARQL query retrieving the distribution of WN synsets of *abeo* across works.

```
PREFIX frac: <http://www.w3.org/ns/lemon/frac#>
PREFIX ontolex: <http://www.w3.org/ns/lemon/ontolex
     #>
PREFIX lila: <http://lila-erc.eu/ontologies/lila/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax
     -ns#>
PREFIX powla: <http://purl.org/powla/powla.owl#>

SELECT ?syn ?docLayer (COUNT(?attestation) AS ?freq)
WHERE {
  ?lemma ontolex:writtenRep "abeo"@la .
  ?token lila:hasLemma ?lemma ;
         powla:hasLayer ?docLayer .
  ?docLayer rdf:type powla:DocumentLayer .
  ?attestation frac:locus ?token .
  ?syn frac:attestation ?attestation .
  FILTER(STRSTARTS(STR(?syn),
    "http://lila-erc.eu/data/lexicalResources/
         LatinWordNet/id/LexicalConcept/"))
}
GROUP BY ?syn ?docLayer
ORDER BY DESC(?freq)
```

The query selects all tokens corresponding to the lemma *abeo*, retrieves their associated attestations, and links each attestation to a WN synset via the `frac:attestation` property. The results are grouped by synset and text, producing a frequency table that shows how often each sense of *abeo* is instantiated in each work (Figure 3). This enables researchers to compare sense distribution across authors, genres, and periods, supporting corpus-driven and diachronic investigations (Farina, 2026). For example, the distribution of *abeo* illustrates how the verb predominantly expresses concrete physical departure in Early Latin, with frequent attestations of synsets such as `02013448-v` 'go away from a place' and `01839438-v` 'change location; move, travel, or proceed', especially in Plautus' *Amphitruo*. This distribution is partly influenced by genre: in archaic Latin comedy, motion verbs often function as stage-directional devices, whereas in later literary contexts – such as Seneca's tragedies, which were likely intended for recitation rather than performance – *abeo* more frequently develops abstract and metaphorical uses. In later texts, additional metaphorical senses become more prominent, including uses referring to abstract forms of departure such as relinquishing office or shifting social position.

To further illustrate the interoperability enabled by PREMOVE, Listing 6 presents a SPARQL query that connects WN synsets to VN classes via shared token-level attestations. The query again selects the WN synset `01788772-v` 'be brought to or as if to the point of death by an intense emotion such as embarrassment, amusement, or shame' and retrieves all VN classes that it instantiates. Using the `frac:attestation` property as a common linking mechanism, the query identifies which VN classes co-occur with a given WN concept at the token level (Figure 4). This allows researchers to explore how conceptual semantic distinctions encoded in WN align with structures represented in VN, thereby facilitating cross-resource semantic comparison within the LiLa KB.

Listing 6: SPARQL query retrieving all VN classes instantiated by WN synset `01788772-v`.

```
PREFIX frac: <http://www.w3.org/ns/lemon/frac#>
PREFIX ontolex: <http://www.w3.org/ns/lemon/ontolex
     #>
PREFIX pmo: <http://premon.fbk.eu/ontology/core#>
PREFIX pmovn: <http://premon.fbk.eu/ontology/vn#>

SELECT DISTINCT ?vnClass
WHERE {
  VALUES ?wnSyn {
    <http://lila-erc.eu/data/lexicalResources/
         LatinWordNet/id/LexicalConcept/01788772-v>
  }
  ?wnSyn frac:attestation ?attestation .
  ?vnClass frac:attestation ?attestation .
  FILTER(STRSTARTS(STR(?vnClass),
      "http://premon.fbk.eu/resource/"))
}
ORDER BY ?vnClass
```

Figures 5 and 6 visualize the RDF subgraphs defined by the SPARQL queries in Listings 4 and 5, respectively, illustrating two complementary access patterns to PREMOVE data. In both cases, nodes represent variables or resources, and directed edges correspond to RDF properties, making the underlying triple structure explicit. Figure 5 highlights the direct link between a WN synset (`?syn`) and its token-level attestations (`?attestation`) via `frac:attestation`, with attestations further connected to tokens (`?token`) through `frac:locus` and to textual passages via

attestation	token	passage	text
http://lila-erc.eu/da...	http://lila-erc.eu/data/corpora/Lasl...	http://lila-erc.eu/data/corpora/L...	Qui cum se suaque omnia in oppidu
http://lila-erc.eu/da...	http://lila-erc.eu/data/corpora/Lasl...	http://lila-erc.eu/data/corpora/L...	Dum ea conquiruntur et conferuntu
http://lila-erc.eu/da...	http://lila-erc.eu/data/corpora/Lasl...	http://lila-erc.eu/data/corpora/L...	Ea re constituta, secunda vigilia ma
http://lila-erc.eu/da...	http://lila-erc.eu/data/corpora/Lasl...	http://lila-erc.eu/data/corpora/L...	et calones, qui ab decumana porta

Figure 2: Sample results of the SPARQL query in Listing 4.

syn	docLayer	freq
http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalConcept/02013448-v	http://lila-erc.eu/data/corpora/Lasla/id/corpus/PlautusAmphitruo/D...	21
http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalConcept/02013448-v	http://lila-erc.eu/data/corpora/Lasla/id/corpus/VergiliusAeneis/Doc...	19
http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalConcept/01839438-v	http://lila-erc.eu/data/corpora/Lasla/id/corpus/VergiliusAeneis/Doc...	8
http://lila-erc.eu/data/lexicalResources/LatinWordNet/id/LexicalConcept/01839438-v	http://lila-erc.eu/data/corpora/Lasla/id/corpus/PlautusAmphitruo/D...	6

Figure 3: Sample results of the SPARQL query in Listing 5.

vnClass
http://premon.fbk.eu/resource/vn33-destroy-44
http://premon.fbk.eu/resource/vn33-die-42.4
http://premon.fbk.eu/resource/vn33-marvel-31.3

Figure 4: Results of the SPARQL query in Listing 6.

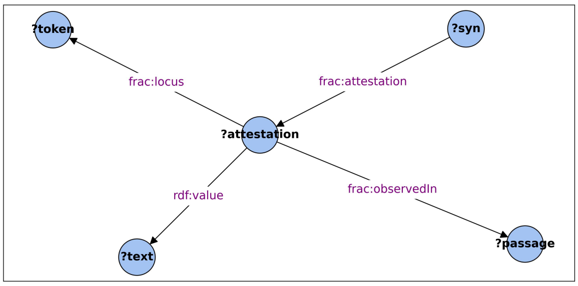

Figure 5: RDF subgraph of the SPARQL query in Listing 4.

frac:observedIn. Figure 6 extends this pattern by anchoring tokens to the lemma layer through lila:hasLemma and to document-level resources via powla:hasLayer, thereby enabling aggregation of attestations by synset and work.

8. Conclusion and Future Work

The publication of PREMOVE in LiLa represents an example of enhancing the efforts made to create a linguistic resource by integrating its (meta)data into a Linked Data KB and thus making them interoperable with those provided by other resources. The reuse of the tokens and synsets from the LWN published in LiLa and their linkage based on the annotations provided by PREMOVE enrich both LiLa, by extending its graph, and PREMOVE, by embedding it into the set of interoperable Latin resources. This also benefits the scientific community using LiLa, as users now have access to a (still small) portion of textual data disambiguated at the WN synset

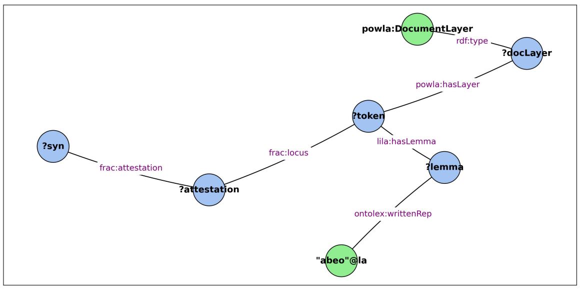

Figure 6: RDF subgraph of the SPARQL query in Listing 5.

level in LOD. As mentioned, PREMOVE includes additional annotation layers whose integration into LiLa and publication as LOD constitute the next natural step, posing new modeling challenges for further data types. Future expansions of the Latin component will follow the same linking approach described in this paper.

Recent research has also explored the use of Large Language Models (LLMs) for semantic annotation and Word Sense Disambiguation (WSD) in historical languages. In Farina and Ciletti (2026), LLMs were used to annotate preverb semantics, using a set of English prepositions/adverbs instead of WN synsets (Farina, 2026). We achieved an F1 of 0.596 for Latin and up to 0.8 after fine-tuning the model (Ciletti, 2026). In Farina et al. (2026), we employed LLM-based WSD using WN synsets to investigate diachronic lexical semantics in Latin and Ancient Greek, achieving an F1 score of 0.594 for Latin, with most discrepancies attributable to sense granularity rather than clear errors. These results indicate that LLMs can achieve good annotations which could support future extensions of PREMOVE by suggesting WN synsets or other semantic labels. Such integration would ideally follow a hybrid approach, where LLM-generated candidates are systematically validated within the curated linking framework presented in this paper.

9. Acknowledgments

This publication is based upon work from COST Action CA23147 GOBLIN - Global Network on Large-Scale, Cross-domain and Multilingual Open Knowledge Graphs, supported by COST (European Cooperation in Science and Technology, https://www.cost.eu. This work is also supported by UKRI under the Horizon Europe Guarantee (grant number UKRI947) for the project COALA (Computational Corpus Annotation for Quantitative Analysis of Latin Lexical Semantics) successfully evaluated by the ERC.

10. Author contributions

AF is responsible for conceptualisation, methodology, formal analysis, software implementation (including all code used for analysis), manual annotation of the dataset, and funding acquisition (COST Action CA23147 GOBLIN). He wrote the original draft for Sections 1, 3.2-4.2, 6-8, and edited the final manuscript. MPa, FM, MPe, GM, EL are responsible for the project supervision. They wrote the original draft for Sections 2-3.1 and 4.2-5, and edited the final manuscript.

11. Bibliographical References

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Karl Aberer, Key-Sun Choi, Natasha Noy, Dean Allemang, Kyung-Il Lee, Lyndon Nixon, Jennifer Golbeck, Peter Mika, Diana Maynard, Riichiro Mizoguchi, Guus Schreiber, and Philippe Cudré-Mauroux, editors, *The Semantic Web*, volume 4825, pages 722–735. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Lecture Notes in Computer Science.

Tim Berners-Lee. 2006-2009. Linked Data - Design Issues. https://www.w3.org/DesignIssues/LinkedData.html.

Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The semantic web. *Scientific american*, 284(5):34–43.

Davide Bertocci. 2017. "Intensive" verbal prefixes in Archaic Latin. In Rosanna Benacchio, Alessio Muro, and Svetlana Slavkova, editors, *Biblioteca di Studi Slavistici*, 1 edition, volume 39, pages 41–58. Firenze University Press, Florence.

Erica Biagetti, Chiara Zanchi, and William Michael Short. 2021. Toward the creation of WordNets for ancient Indo- European languages. *Proceedings of the 11th Global Wordnet Conference, University of South Africa (UNISA)*, Global WordNet Association:258–266.

Gabriel Bodard, Hugh Cayless, Mark Depauw, and Leif Isaksen. 2017. Standards for Networking Ancient Person-data: Digital approaches to problems in prosopographical space. *Digital Classics Online*, 3(2):28–43.

Geert Booij and Ans Van Kemenade. 2003. Preverbs: an introduction. In Geert Booij, Jaap Van Marle, Geert Booij, and Jaap Van Marle, editors, *Yearbook of Morphology 2003*, pages 1–11. Springer Netherlands, Dordrecht. Series Title: Yearbook of Morphology.

Maria Antonia Brovelli and Andrea Maurino. 2000. ArcheoGIS: An Interoperable Model For Archaeological Data. *International Archives of Photogrammetry and Remote Sensing*, 33:140–147.

Christian Chiarcos. 2012. 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*, number 7295 in Lecture Notes in Computer Science, pages 225–239, Berlin/Heidelberg, Germany. Springer.

Christian Chiarcos, Maxim Ionov, Jesse de Does, Katrien Depuydt, Anas Fahad Khan, Sander Stolk, Thierry Declerck, and John Philip McCrae. 2020. Modelling Frequency and Attestations for OntoLex-Lemon. In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, pages 1–9, Marseille, France. European Language Resources Association.

Christian Chiarcos and Maria Sukhareva. 2015. OLiA – Ontologies of Linguistic Annotation. *Semantic Web*, 6(4):379–386.

Michele Ciletti. 2026. Premove$_llama3.3-70b_float16(revisionc1ddc08)$.

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

Pierluigi Cuzzolin. 1995. A proposito di sub vos placo e della grammaticalizzazione delle preposizioni. *Archivio Glottologico Italiano*, 80(1-2):12–143.

Mark Depauw and Tom Gheldof. 2014. Trismegistos: An Interdisciplinary Platform for Ancient World Texts and Related Information. In Łukasz Bolikowski, Vittore Casarosa, Paula Goodale, Nikos Houssos, Paolo Manghi, and Jochen Schirrwagen, editors, *Theory and Practice of Digital Libraries – TPDL 2013 Selected Workshops*, volume 416, pages 40–52. Springer International Publishing, Cham. Series Title: Communications in Computer and Information Science.

Tom Elliott. 2021. The Pleiadic Gaze: Looking at Archaeology from the Perspective of a Digital Gazetteer. *Classical Archaeology in the Digital Age – The AIAC Presidential Panel*. Publisher: Propylaeum.

Andrea Farina. 2022-2026. *The differences in Ancient Greek and Latin motion verbs as a way to understand the conceptualisation of reality in the two cultures*. UK Research and Innovation (ref. number: 2749398).

Andrea Farina. 2024. Guidelines for a linguistic annotation of preverbed verbs of motion. Artwork Size: 12168715 Bytes Pages: 12168715 Bytes.

Andrea Farina. 2025a. PREMOVE – a diachronic dataset of ancient greek and latin annotated PREverbed MOtion Verbs. *Oxford Text Archive*.

Andrea Farina. 2025b. Prevnet: Towards enriching clarin with resources for preverbs in ancient greek and latin.

Andrea Farina. 2025c. Source preverbs and motion verbs in Latin. A quantitative approach to ab- and ex-. *Journal of Latin Linguistics*, 24(2):1–38.

Andrea Farina. 2026. *A comparative study on Ancient Greek and Latin preverbed motion verbs. The creation of PrevNet as an expandable online resource for preverbs and preverbation*. King's College London: PhD Thesis.

Andrea Farina, Andrea Ballatore, and Barbara McGillivray. 2025a. Mapping meaning in latin with large language models: A multi-task evaluation of preverbed motion verbs and spatial relation detection in llms. In *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025) Cagliari, Italy, September 24-26, 2025*.

Andrea Farina and Michele Ciletti. 2026. Probing preverbs: Evaluating large language models on Latin and Ancient Greek preverbed motion verbs. Forthcoming. In press.

Andrea Farina, Michele Ciletti, Barbara McGillivray, and Andrea Ballatore. 2026. Sense-Based Annotation of Geographical Nouns in Ancient Greek and Latin: A Diachronic Study with LLMs. *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics*.

Andrea Farina, Paola Marongiu, Mathilde Bru, and Daniele Borkowski. 2025b. When Data Meets the Past: Data Collection, Sharing, and Reuse in Ancient World Studies. *Open Information Science*, 9(1):1–20.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Cambridge (MA): Mit Press.

Jorge Gracia, Slavko Žitnik, Max Ionov, Christian Chiarcos, Dagmar Gromann, Francesco Mambrini, Marco Passarotti, Armando Stellato, John P. McCrae, Gilles Sérasset, Andon Tchechmedjiev, Sara Carvalho, Penny Labropoulou, and Rute Costa. 2025. MOOC on Linguistic Linked Data. In *The Semantic Web. 22nd European Semantic Web Conference, ESWC 2025*, pages 157–173, Cham. Springer.

Nancy Ide and James Pustejovsky, editors. 2017. *Handbook of Linguistic Annotation*. Springer Netherlands, Dordrecht.

Anja Jamrozik, Eyal Sagi, Micah Goldwater, and Dedre Gentner. 2013. Relational words have high metaphoric potential. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 21–26, Atlanta, Georgia. Association for Computational Linguistics.

Daniel Jurafsky and James H. Martin. 2018. Word relations, senses, and disambiguation. In *Speech and Language Processing*.

Daniel King and Dedre Gentner. 2022. Verb Metaphoric Extension Under Semantic Strain. *Cognitive Science*, 46(5):e13141.

Karin Kipper-Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania, PA: PhD Thesis.

Tina Krennmayr. 2017. Metaphor and part-of-speech. In Elena Semino and Zsófia Demjen, editors, *The Routledge Handbook of Metaphor and Language*, pages 165–177. Routledge, New York.

Ora Lassila, Ralph R. Swick, World Wide, and Web Consortium. 1998. Resource Description Framework (RDF) Model and Syntax Specification.

Francesco Mambrini, Marco Passarotti, Giovanni Moretti, and Matteo Pellegrini. 2022. The Index Thomisticus Treebank as Linked Data in the LiLa Knowledge Base. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4022–4029.

John P McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. The OntoLex-Lemon Model: Development and Applications. In *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, pages 587–597, Brno, Czech Republic. Lexical Computing CZ s.r.o.

Barbara McGillivray. 2014. *Methods in Latin Computational Linguistics*. Leiden: Brill.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden. Association for Computational Linguistics.

Martha Palmer, Olga Babko-Malaya, and Hoa Trang Dang. 2004. Different sense granularities for different applications. In *Proceedings of the 2nd International Workshop on Scalable Natural Language Understanding (ScaNaLU 2004) at HLT-NAACL 2004*, pages 49–56, Boston, Massachusetts, USA. Association for Computational Linguistics.

Marco Passarotti, Marco Budassi, Eleonora Litta, and Paolo Ruffolo. 2017. The Lemlat 3.0 Package for Morphological Analysis of Latin. In *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).

Rainer Simon, Leif Isaksen, Elton Barker, and Pau de Soto Cañamares. 2016. The Pleiades Gazetteer and the Pelagios Project. In *Placing Names: Enriching and Integrating Gazetteers*, merrick lex berman, ruth mostern, and humphrey southall (eds.) edition, pages 97–109. Bloomington: Indiana University Press.

Neel Smith. 2009. Citation in Classical Studies. *Digital Humanities Quarterly*, 3(1).

Gilles Sérasset. 2015. DBnary: Wiktionary as a Lemon-based multilingual lexical resource in RDF. *Semantic Web*, 6(4):355–361.

Leonard Talmy. 1975. Figure and Ground in Complex Sentences.

Leonard Talmy. 1983. How Language Structures Space. In Herbert L. Pick and Linda P. Acredolo, editors, *Spatial Orientation*, pages 225–282. Springer US, Boston, MA.

Leonard Talmy. 2000. *Toward a Cognitive Semantics. Vol. 1: Concept Structuring Systems*. Cambridge (MA): Mit Press.

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018.