Using Graph Databases for Historical Language Data: Challenges and Opportunities

Barbara McGillivray[1], Pierluigi Cassotti[2,*], Pierpaolo Basile[2], Davide Di Pierro[2] and Stefano Ferilli[2]

[1] *King's College London, Strand Campus, Strand, London, WC2R 2LS, United Kingdom*

[2] *University of Bari Aldo Moro, Department of Computer Science, Via E. Orabona 4, Bari, 70125, Italy*

Abstract

The integration of semantic information into language resources has the potential to open up new avenues of enquiry into the mechanisms of language change. We present the first experiments in integrating data from Latin textual corpora and language resources into a graph database via the GraphBRAIN Schema and show the potential of this model for research into the mechanisms of semantic change in Latin.

Keywords

Knowledge Graphs, Latin Corpora, Semantic Change Detection, Graph Data Model

1. Introduction

Research in Historical Linguistics often requires the analysis and support of heterogeneous data and tools, such as lexical resources, encyclopaedias, and large corpora. Nevertheless, these resources are often siloed. Graph Databases present an ideal opportunity to combine the advantages of DataBase Management Systems (DBMSs) for handling individuals (scalability, storage optimization, efficient handling, mining and browsing of the data, etc.) with the high-level functionalities available in Knowledge Bases (KBs). Graph DBMS are intrinsically designed to store schemaless data. Differently from traditional DBMSs like the relational [1] or object-oriented [2] ones, they lack predefined structures. Following this approach, Neo4j [1], one of the most common graph DBMSs, does not provide support for introducing ontology definitions based on labels and/or arcs. The absence of a schema may lead to ambiguity when reading and managing data in downstream applications due to the inherent ambiguity of the words used for expressing concepts. Hence, the semantics becomes blurred.

To address these issues, we propose the use of GraphBRAIN [3] as a solution. GraphBRAIN consists of a graph database which follows the Labelled Property Graph (LPG) [4] structure.

19th IRCDL (The Conference on Information and Research science Connecting to Digital and Library science), February 23–24, 2023, Bari, Italy

*Corresponding author.

barbara.mcgillivray@kcl.ac.uk (B. McGillivray); pierluigi.cassotti@uniba.it (P. Cassotti); pierpaolo.basile@uniba.it (P. Basile); davide.dipierro@uniba.it (D. Di Pierro); stefano.ferilli@uniba.it (S. Ferilli)
0000-0003-3426-8200 (B. McGillivray); 0000-0001-7824-7167 (P. Cassotti); 0000-0002-0545-1105 (P. Basile); 0000-0002-8081-3292 (D. Di Pierro); 0000-0003-1118-0601 (S. Ferilli)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

[1] https://neo4j.com/

CEUR-WS.org/Vol-3365/short7.pdf

This structure stores nodes with specific labels, arcs which represent relationships among nodes and properties on both nodes and arcs. Properties are stored in the format of key/value pairs. GraphBRAIN requires KB designers to define a data schema which operates also as an ontology. GraphBRAIN provides a mapping mechanism for exporting schemes into an SW-compliant language, the Web Ontology Language (OWL). These schemes guide access through all the CRUD operations on the database but also ensure interpretability and interoperability among different applications. Following the schemes, applications become compliant with each other. Neo4j, in its Enterprise Edition, does not provide any constraint definition process. In other versions, it supports a few constraints like *unique node property constraints*, *node property existence constraints*, *relationship property existence constraints* and *node key constraints*. Evidently, these tools are not as expressive as ontology definitions.

In [5], we adopted GraphBRAIN technology to model time-sensitive linguistic knowledge in a graph database, describing a time-sensitive model of linguistic knowledge that can be used for graph databases. In this paper, we show an application of this model to the lexical semantic analysis of Latin data, i.e. the analysis of the meanings of Latin words. Differently from previous approaches, such as Basile et al. [6], Hamilton et al. [7], and Carlo et al. [8], we exploit graph database potentialities to detect semantic changes in specific concepts.

Latin is in a particularly favourable position among historical languages for the large-scale analysis of semantic change processes, thanks to a number of factors. First, Latin researchers now enjoy unprecedented access to digital data covering over two thousand years of history. Thanks to the ERC-funded LiLa project [2], seven Latin language resources and six corpora have been linked at the level of word lemmas so far, making Latin a unique case among historical languages. Second, we have access to extensive computational language resources for Latin, Latin WordNet [9], and digitised dictionaries of Latin, which provide rich information about words' semantics and examples of usage. Finally, focussing on Latin allows us to investigate semantic change processes over long time spans. Latin has one of the longest recorded histories of any human language, making it naturally suitable for quantitative studies [10]. The first inscriptional records date from the sixth century BCE, and Latin continues to be used to the current day by the Catholic Church and some academic and legal institutions around the world. Written Latin diverged from the spoken vernaculars in the second half of the first millennium of the Christian era, but it remained in use as one of the principal channels of communication across most of Europe for the next thousand years. The humanists' conscious effort to reproduce Classical Latin led to a range of interesting developments, particularly affecting the neo-Latin lexicon to enable the expression of new concepts [11]. This extensive chronological span has raised the question of the extent to which Latin is seen as a dead or fossilised language (e.g. Herman [12], Butterfield [13]). However, it remains an open question to what extent this fossilisation affected the semantics of words, as we know that the Latin lexicon, in this respect, has remained dynamic (over 4,500 words have acquired new meanings since the Renaissance; Demo 2022). The extent to which post-classical Latin can really be considered as a "fixed" language (Leonhardt [14], Roelli [15], Langslow [16]) from the point of view of its ability to generate new meanings of words is still largely unknown beyond anecdotal evidence.

In Section 2 we present the Linguistic Knowledge Graph, in Section 3 we describe the Latin

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

data that we worked on, and in Section 4 we show how we loaded the Latin data into the Linguistic Knowledge Graph. Finally, in Section 5 we draw some conclusions and outline future directions of work.

2. The Linguistic Knowledge Graph

The Linguistic Knowledge Graph (LKG) aims to capture different aspects of lexical resources, such as relations between words and concepts, morphological, and syntactical information. Moreover, LKG covers diachronic aspects of language, such as the date of publication of a document, and the birth and death of an author. The schema we designed takes inspiration from the ontological lexicon model LeMON [17]. For space constraints, we report in Table 1 node types and in Table 2 the relationships adopted for diachronic analysis. The lexical unit is represented as node of type *InflectedWord* or *Lemma*, which are subclass of *Word*, i.e. *Lemma IS_A Word* and *InflectedWord IS_A Word*. The *Lemma* can be a multi-word expression (mwe), in this case, the flag mwe is set to True. The respective lemma of an *InflectedWord* can be retrieved exploiting the relationship *HAS_LEMMA* between *InflectedWord* and *Lemma*. The *LexiconConcept* is used to represent the word's meanings, and each instance of *LexiconConcept* represents a different meaning. For example, the *LexiconConcept* can represent the senses reported on a sense inventory, e.g. synsets in WordNet [18]. The relationship between a word and its meaning is expressed using the relationship *HAS_CONCEPT* among instances of *Word* and instance of *LexiconConcept*. Multiple relationships can be defined over couples of *LexiconConcept* using the reflexive relationship *SEM_RELATION*. At the same time, reflexive relationships over the Word instances can be described by the *LEX_RELATION* relationship.

The document structure from which words are extracted can be represented at different levels of granularity: *Sentence*,*Text*, *Document*, and *Corpus*. In particular, each excerpt can be represented as *Text* or *Sentence*, which is a subclass of *Text*. A *Text* may belong to (*BELONG_TO*) a *Document* and a *Document* can be part of (*BELONG_TO*) a *Corpus*. The occurrences of a word in a particular *Text* are traced by the relationship *HAS_OCCURRENCE* among *Word* and *Text*. In the case of sense-annotated corpora, such as SemCor, is possible to specify the occurrences of senses using the relationship *HAS_EXAMPLE* among *LexiconConcept* and *Text*. Currently, the LKG takes into account two types of metadata: author and language. The relationship *HAS_AUTHOR* among nodes of type *Text* and nodes of type *Person* determines the author of a *Text*. The relationship *HAS_LANGUAGE* among nodes of type *Text*, *Document*, *Corpus*, and *Word* to nodes of type *Language* specifies the respective language.

The time is modelled using two classes of nodes: *TimeInterval*, and *TimePoint*, both subclasses of *TemporalSpecification*. The *TimeInterval* type is used when the date is not precisely stated, while the *TimePoint* is used in cases where the date is fixed. The start and end extremes of the *TimeInterval* nodes can be specified using the respective relationships *startTime* and *endTime*. In the current version of the LKG, time specification is supported for *Person* and *Text*. More specifically, the date of birth and death of authors is specified using the relationship *BORN* and *DIED* between *Person* and *TemporalSpecification*. The publishing date of a text is specified by the relationship *PUBLISHED_IN* among *Text* nodes and *TemporalSpecification* nodes.

Table 1
LKG classes with their respective superclasses and attributes.

Class	Superclass	Attributes
Word		value:String
Lemma	Word	value:String
		posTag:String
		mwe:Boolean
InflectedWord	Word	value:String
Stem		value:String
LexiconConcept	Concept	id:String
		resource:String
Text		value:String
Sentence	Text	
Document		title:String
Corpus		name:String
TemporalSpecification		name:String
		description:String
TimePoint	TemporalSpecification	Year:Integer
		Month:Integer
		day:Integer
TimeInterval	TemporalSpecification	
Person		name:String
		lastname:String
Language		iso639-1:String
		iso639-2:String
		enName:String
Category		id:String

3. Latin data

The data we loaded into the graph consists of a portion of the LatinISE corpus [19] annotated at the level of dictionary senses. LatinISE is a Latin corpus covering the period from the fifth century BCE to the twenty-first century and contains 10 million word tokens, semi-automatically lemmatised and part-of-speech tagged. The metadata fields in LatinISE indicate text identifier, author, title, dates, century, genre, url of source, and optionally book title/number and character names (for plays). The annotated dataset was produced as part of the SemEval shared task on Unsupervised Lexical Semantic Change Detection [20]. 40 Latin lemmas ("target words") are selected, of which 20 are known to have changed their meaning with the advent of Christianity (for example, *beatus*, which shifted its meaning from 'fortunate' to 'blessed') and 20 are known to not have changed their meaning between the BCE era and the CE era. For each of the 40 lemmas, 60 sentences are randomly extracted from LatinISE, 30 of them are from texts dated in the BCE era, and 30 from texts dated in the CE era. Each sentence was annotated by at least one expert annotator, according to the DuReL framework [21]. The annotators were asked to judge the semantic relatedness of an instance of usage of a target word with respect to the list of its dictionary definitions using a four-point scale (Unrelated, Distantly Related, Closely Related, and Identical). The definitions were taken from the Latin portion of the Logeion online dictionary (https://logeion.uchicago.edu/) containing Lewis and Short's *Latin-English Lexicon* (1879) [22], Lewis' *Elementary Latin Dictionary* (1890) [23], and Du Fresne Du Cange et al. [24]. See McGillivray et al. [25] for further details about the dataset and its annotation framework.

Table 2
LKG relationships with their respective subject, object and attributes.

Relationship	Subject	Object	Attributes
IS_A	Sentence	Text	id:Integer
	Lemma $\cup InflectedWord$	Word	id:Integer
BELONG_TO	Text	Document	id:Integer
	Document	Corpus	id:Integer
	Text	Category	

HAS_OCCURRENCE	Word	Text	begin:Integer
			end:Integer

{LEX_RELATION}	Word	Word	
HAS_LEMMA	Word	Lemma	
HAS_CONCEPT	Word	LexiconConcept	grade:Float
HAS_EXAMPLE	LexiconConcept	Text	
HAS_DEFINITION	LexiconConcept	Text	
REFER_TO	LexiconConcept	Concept	
{SEM_RELATION}	LexiconConcept	LexiconConcept	
PUBLISHED_IN	Text $\cup Document \cup Corpus$	TemporalSpecification	
HAS_AUTHOR	Text $\cup Document \cup Corpus$	Person	
BORN	Person	TemporalSpecification	
DIED	Person	TemporalSpecification	
startTime	TimeInterval	TimePoint	
endTime	TimeInterval	TimePoint	
HAS_LANGUAGE	Text $\cup Document \cup Corpus \cup Word$	Language	

```
MATCH
(centuryNode:TimeInterval)-[:startTime]->(startCentury:TimePoint),
(centuryNode:TimeInterval)-[:endTime]->(endCentury:TimePoint),
(pubNode:TimeInterval)-[:startTime]->(startPub:TimePoint),
(pubNode:TimeInterval)-[:endTime]->(endPub:TimePoint),
(text:Text)-[:PUBLISHED_IN]->(pubNode)
WHERE
centuryNode.description="century"
WITH text,
centuryNode,
CASE WHEN endPub.Year > endCentury.Year THEN endCentury.Year ELSE endPub.Year END as minEnd,
CASE WHEN startPub.Year > startCentury.Year THEN startPub.Year ELSE startCentury.Year END as maxStart
WITH *,
CASE WHEN minEnd-maxStart+1 > 0 THEN minEnd-maxStart+1 ELSE 0 END as time_overlap
ORDER BY time_overlap DESC
WITH text,
collect({century:centuryNode})[0] AS max
WITH *,
max.century as century
CREATE (text)-[r:CLUSTER]->(century)
RETURN text,century
UNION ALL
MATCH
(centuryNode:TimeInterval)-[:startTime]->(startCentury:TimePoint),
(centuryNode:TimeInterval)-[:endTime]->(endCentury:TimePoint),
(text:Text)-[:PUBLISHED_IN]->(point:TimePoint)
WHERE
centuryNode.description="century" and
point.Year>=startCentury.Year and
point.Year<=endCentury.Year
WITH text, centuryNode as century
CREATE (text)-[r:CLUSTER]->(century)
RETURN text, century;
```

Listing 1: Clustering publishing date by centuries

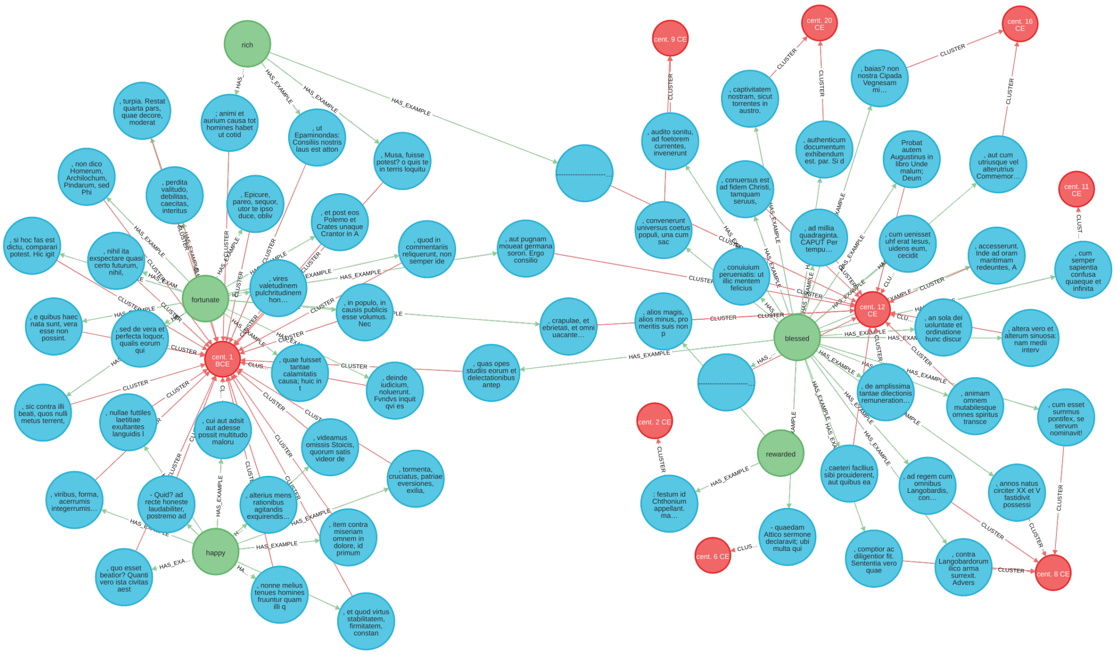
rich
fortunate
happy
blessed
rewarded
cent. 1 BCE
cent. 2 CE
cent. 6 CE
cent. 8 CE
cent. 9 CE
cent. 11 CE
cent. 12 CE
cent. 16 CE
cent. 20 CE
HAS_EXAMPLE
CLUSTER
, turpia. Restat quarta pars, quae decore, moderat
; animi et aurium causa tot homines habet ut cotid
, ut Epaminondas: Consiliis nostris laus est atton
, Musa, fuisse potest? o quis te in terris loquitu
, non dico Homerum, Archilochum, Pindarum, sed Phi
, perdita valitudo, debilitas, caecitas, interitus
, Epicure, pareo, sequor, utor te ipso duce, obliv
, et post eos Polemo et Crates unaque Crantor in A
, si hoc fas est dictu, comparari potest. Hic igit
, nihil ita exspectare quasi certo futurum, nihil,
, vires valetudinem pulchritudinem hon…
, quod in commentariis reliquerunt, non semper ide
, aut pugnam moueat germana sorori. Ergo consilio
, e quibus haec nata sunt, vera esse non possint.
, sed de vera et perfecta loquor, qualis eorum qui
, in populo, in causis publicis esse volumus. Nec
, crapulae, et ebrietati, et omni uacante…
, quae fuisset tantae calamitatis causa; huic in t
, deinde iudicium, noluerunt. Fvndvs inquit qvi es
, quas opes studiis eorum et delectationibus antep
, sic contra illi beati, quos nulli metus terrent,
, nullae futtiles laetitiae exultantes languidis l
, cui aut adsit aut adesse possit multitudo maloru
, videamus omissis Stoicis, quorum satis videor de
, tormenta, cruciatus, patriae eversiones, exilia,
, viribus, forma, acerrumis integerrumis…
- Quid? ad recte honeste laudabiliter, postremo ad
, alterius mens rationibus agitandis exquirendis…
, item contra miseriam omnem in dolore, id primum
, quo esset beatior? Quanti vero ista civitas aest
, nonne melius tenues homines fruuntur quam illi q
, et quod virtus stabilitatem, firmitatem, constan
, audito sonitu, ad foetorem currentes, invenerunt
, captivitatem nostram, sicut torrentes in austro.
, authenticum documentum exhibendum est. par. Si d
, baias? non nostra Cipada Vegnesam mi…
Probat autem Augustinus in libro Unde malum; Deum
, aut cum utriusque vel alterutrius Commemor…
, conuersus est ad fidem Christi, tamquam seruus,
, conuenerunt universus coetus populi, una cum sac
, ad millia quadraginta. CAPUT Per tempu…
, cum uenisset uhf erat Iesus, uidens eum, cecidit
, accesserunt. Inde ad oram maritimam redeuntes, A
, cum semper sapientia confusa quaeque et infinita
, conuiuium perueniatis: ut illic mentem felicius
, alios magis, alios minus, pro meritis suis non p
, an sola dei uoluntate et ordinatione hunc discur
, altera vero et alterum sinuosa: nam medii interv
, de amplissima tantae dilectionis remuneration…
, animam omnem mutabilesque omnes spiritus transce
, cum esset summus pontifex, se servum nominavit!
: festum id Chthonium appellant. ma…
- quaedam Attico sermone declaravit; ubi multa qui
, caeteri facllius sibi prouiderent, aut quibus ea
, ad regem cum omnibus Langobardis, con…
, annos natus circiter XX et V fastidivit possessi
, comptior ac diligentior fit. Sententia vero quae
, contra Langobardorum ilico arma surrexit. Advers

Figure 1: Graph for the Latin word *beatus*.

4. Loading the Latin data in the Linguistic Knowledge Graph

For each instance of the target words in the Latin corpus we encode:

- the author as *Person*,
- the manuscript as *Document*,
- the year as *TimePoint* if the date is certain, *TimeInterval* otherwise,
- the sentence (left context, target word and right context) as *Text*,
- the definitions of the Lewis and Short Dictionary as *LexiconConcept*,
- the word lemma as *Lemma*,
- the inflected forms of the target words as *InflectedWord*,
- the scores associated with each *LexiconConcept* as properties of the *HAS_EXAMPLE* and *HAS_OCCURRENCE* relationships.

In order to simplify and make the visualisation more effective, we created the *HAS_EXAMPLE* relationship only in cases where the annotation reported a score of 4. In addition, to make more evident the distribution of senses with respect to centuries, we associate each date of publication of the texts with the reference century. We do this via the query given in Listing 1. In case a *Text* is not associated with a specific *TimePoint*, it will be linked with the century having the greatest overlap with the *TimeInterval* of the text itself. On the other hand, for texts for which a precise date is specified, the query associates the *Text* with the respective century of its year. The centuries are represented as *TimeInterval*, and the description attribute is validated with “century”. A new relationship, called *CLUSTER*, is so created among nodes of type *Text* and nodes of type *TimeInterval* to indicate the century.

A subgraph for the word *beatus* is shown in Figure 1. The graph shows the nodes representing the texts from which the word *beatus* is extracted, the centuries and the senses given in the Lewis and Short Dictionary. The relationships among these nodes are *CLUSTER* and *HAS_EXAMPLE*. The former connects nodes of type *TimeInterval* and nodes of type *Text*, see 1. The latter links *LexiconConcept*s and *Text*s. Most occurrences of the word *beatus* in the reference corpus are dated 1st century BCE and 11th century CE. One can immediately notice a difference in the distribution of the senses: “happy” and “fortunate” on the one hand are associated with the time period BCE (see the cluster of nodes on the left of Figure 1), and “blessed”, on the other hand, is associated with the time period CE (see the cluster of nodes on the right of Figure 1). In fact, only one sentence in the dataset displays the sense “blessed” in the first century BCE. Similarly, only two sentences dated CE contain the word *beatus* with the meaning of “fortunate”, the latter, on the other hand, is dated 1079-1142 CE and is an excerpt from the Sermones of Petrus Abaelardus.

5. Conclusions

In this work, we introduced an application of LKG for Latin data. It appears to be an interesting and novel approach to tackling the analysis of diachronic corpora. Furthermore, differently from previous approaches, it gives rise to explainable results since we take advantage of explicit relationships modelled as graphs. The LKG seems to lead to promising results, and it is ready forfurther investigations into Lexical Semantic Change Detection (LSCD). Future developments include a better visualization of resources, machine-learning-based techniques for automatic LSCD and an interface for querying and analysing the LKG data.

Acknowledgement

This work fulfils the research objectives of the PNRR project FAIR - Future AI Research, spoke 6 - Symbiotic AI, CUP H97G22000210007, as well as the CHANGES - Cultural Heritage Active Innovation for Next-Gen Sustainable Society, CUP H53C22000860006.

References

[1] H.-P. Kriegel, M. Pfeifle, M. Pötke, T. Seidl, The paradigm of relational indexing: A survey, in: BTW 2003–Datenbanksysteme für Business, Technologie und Web, Tagungsband der 10. BTW Konferenz, Gesellschaft für Informatik eV, 2003.

[2] E. Bertino, L. Martino, Object-oriented database management systems: concepts and issues, Computer 24 (1991) 33–47.

[3] S. Ferilli, Integration strategy and tool between formal ontology and graph database technology, Electronics 10 (2021). URL: https://www.mdpi.com/2079-9292/10/21/2616. doi:10.3390/electronics10212616.

[4] C. Sharma, R. Sinha, A schema-first formalism for labeled property graph databases: Enabling structured data loading and analytics, in: Proceedings of the 6th ieee/acm international conference on big data computing, applications and technologies, 2019, pp. 71–80.

[5] P. Basile, P. Cassotti, S. Ferilli, B. McGillivray, A New Time-sensitive Model of Linguistic Knowledge for Graph Databases, CEUR Workshop Proceedings, 2022, p. 69.

[6] P. Basile, A. Caputo, G. Semeraro, Temporal random indexing: a tool for analysing word meaning variations in news, in: M. Martinez-Alvarez, U. Kruschwitz, G. Kazai, F. Hopfgartner, D. P. A. Corney, R. Campos, D. Albakour (Eds.), Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), volume 1568 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2016, pp. 39–41. URL: http://ceur-ws.org/Vol-1568/paper7.pdf.

[7] W. L. Hamilton, J. Leskovec, D. Jurafsky, Diachronic word embeddings reveal statistical laws of semantic change, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, Volume 1: Long Papers, The Association for Computer Linguistics, 2016. URL: https://doi.org/10.18653/v1/p16-1141. doi:10.18653/v1/p16-1141.

[8] V. D. Carlo, F. Bianchi, M. Palmonari, Training temporal word embeddings with a compass, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI, AAAI Press, 2019, pp. 6326–6334. URL: https://doi.org/10.1609/aaai.v33i01.33016326. doi:10.1609/aaai.v33i01.33016326.

[9] S. Minozzi, 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'Antichita (2017) 123–134.

[10] H. Pinkster, Sintassi e semantica latina, Rosenberg & Sellier, 1991.

[11] J. Ramminger, Latin and the early modern world: linguistic identity and the polity from petrarch to the habsburg novelists, 2016.

[12] J. Herman, Vulgar Latin. Translated by Roger Wright, The Pennsylvania State University, 2000.

[13] D. Butterfield, A companion to the latin language, 2011.

[14] J. Leonhardt, Latin: Story of a World Language, The Belknap Press of Harvard University Press, 2013.

[15] P. Roelli, Latin as the Language of Science and Learning, De Gruyter, 2021.

[16] D. R. Langslow, Bilingualism in ancient society, 2002.

[17] T. Declerck, P. Buitelaar, T. Wunner, J. McCrae, E. Montiel-Ponsoda, G. Aguado de Cea, Lemon: An ontology-lexicon model for the multilingual semantic web. (2010).

[18] G. A. Miller, WORDNET: a lexical database for english, in: Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, USA, February 23-26, 1992, Morgan Kaufmann, 1992. URL: https://aclanthology.org/H92-1116/.

[19] B. McGillivray, A. Kilgarriff, Tools for historical corpus research, and a corpus of Latin, in: P. Bennett, M. Durrell, S. Scheible, R. J. Whitt (Eds.), New Methods in Historical Corpus

Linguistics, Narr, Tübingen, 2013, pp. 247–257.

[20] D. Schlechtweg, B. McGillivray, S. Hengchen, H. Dubossarsky, N. Tahmasebi, Semeval-2020 task 1: Unsupervised lexical semantic change detection, in: A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May, E. Shutova (Eds.), Proceedings of the Fourteenth Workshop on Semantic Evaluation, SemEval@COLING 2020, Barcelona (online), December 12-13, 2020, International Committee for Computational Linguistics, 2020, pp. 1–23. URL: https://doi.org/10.18653/v1/2020.semeval-1.1. doi:10.18653/v1/2020.semeval-1.1.

[21] D. Schlechtweg, S. Schulte im Walde, S. Eckmann, Diachronic Usage Relatedness (DURel): A framework for the annotation of lexical semantic change, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, Louisiana, 2018, pp. 169–174. URL: https://www.aclweb.org/anthology/N18-2027/.

[22] C. T. Lewis, C. Short, A Latin Dictionary, Founded on Andrews' edition of Freund's Latin dictionary revised, enlarged, and in great part rewritten by Charlton T. Lewis, Ph.D. and Charles Short, Clarendon Press, Oxford, 1879.

[23] C. T. Lewis, An Elementary Latin Dictionary, American Book Company, New York, Cincinnati, and Chicago, 1890.

[24] C. Du Fresne Du Cange, G. A. L. Henschel, P. Carpentier, J. C. Adelung, L. Favre, Glossarium mediæet infimælatinitatis, L. Favre, Niort, 1883-1887.

[25] B. McGillivray, D. Kondakova, A. Burman, F. Dell'Oro, H. Bermúdez Sabel, P. Marongiu, M. Márquez Cruz, A new corpus annotation framework for latin diachronic lexical semantics, Journal of Latin Linguistics 21 (2022) 47–105. doi:https://doi.org/10.1515/joll-2022-2007.