

Nunc Est Aestimandum

Towards an Evaluation of the Latin WordNet

Greta Franzini*, Andrea Peverelli*, Paolo Ruffolo*, Marco Passarotti*,
Helena Sanna°, Edoardo Signoroni°, Viviana Ventura°, Federica Zampedri°

*CIRCSE Research Centre, Università Cattolica del Sacro Cuore, Milan, Italy

°Università degli Studi di Pavia, Pavia, Italy

greta.franzini@unicatt.it

Abstract

English. This paper describes a preliminary expansion and assessment of the Latin WordNet for the purposes of the *LiLa: Linking Latin* project. The objective of this study is to better understand the implications of expanding and evaluating the sense coverage of the Latin WordNet, with a view to identifying the most effective method for its refinement and inclusion in the LiLa Knowledge Base of Latin resources. Our test empirically demonstrates the inadequacy for Latin of a common semi-automated approach of expansion and informs potential lines of improvement for the resource.¹

1 Introduction

WordNets are among the most used lexico-semantic resources in Natural Language Processing (NLP). Indeed, their value is such as to warrant the annual *Global WordNet Conference*, which is now in its tenth edition.² In the words of Fellbaum (1998, p. 52):

WordNet [...] is perhaps the most widely used electronic dictionary of English and serves as the lexicon for a variety [sic] of different NLP applications including Information Retrieval (IR), Word Sense Disambiguation (WSD), and Machine Translation (MT).

Since the release of the *Princeton WordNet* (hereafter PWN) in the mid 1980s (Miller et al., 1990), interest in providing WordNets for modern

languages has far exceeded that for historical languages. With the exception of the *Historical Thesaurus of English*, whose purpose is not dissimilar to that of a WordNet but whose distinct structure sets it apart from this type of resource,³ the only two historical language WordNets in existence today are the Latin (Minozzi, 2017) and the Ancient Greek WordNets (Bizzoni et al., 2014): both have limited lexical coverage and the Latin WordNet (hereafter LWN) is particularly noisy (see Section 3). Their incompleteness poses significant challenges to a number of computational analyses, thus restricting the scope for lexico-semantic research.⁴

The study described here falls within the scope of the *LiLa: Linking Latin* project (Passarotti et al., 2019).⁵ In its wider effort to connect linguistic resources and NLP tools for Latin in a Linked Data Knowledge Base, LiLa is conducting a first assessment of the LWN. Besides being structurally compatible with LiLa, a refined LWN is essential to the Knowledge Base as a connector between Latin and resources in other languages, thus meeting a growing need in the field of Linguistic Linked Open Data (Chiarcos et al., 2013).

This paper describes a preliminary assessment of the LWN with a view to better understanding how to approach its expansion and evaluation: Sections 2 and 3 briefly outline existing research in WordNet evaluation and the structure of the LWN, respectively; Section 4 details our evaluation method; Section 5 discusses our preliminary results; finally, Section 6 summarises our contribution and focusses on directions for future research.

2 Related Work

Evaluation. To evaluate a WordNet is to evaluate its *coverage* of a specific linguistic domain or of

¹Copyright 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

²<http://globalwordnet.org/>

³<https://ht.ac.uk/>

⁴Most recently Franzini et al. (2018).

⁵<https://lila-erc.eu> (2018-2023).

an entire language (period), be that qualitative (accuracy) or quantitative (inclusivity). Among others, Bodenreider et al. (2003) conducted a quantitative evaluation of the bio-genetic domain in the PWN by mapping a list of relevant terms against manually-established semantic classes of nominal synsets, and proved PWN’s coverage to be satisfactory. A study by Hajič et al. (2004) sought to manually evaluate and improve the Czech WordNet using the lexico-semantic annotation of the Prague Dependency Treebank. In spite of achieving poor inter-annotator agreement, their outcome can inform future improvements of the resource.

The first automated, qualitative evaluation of a WordNet was performed by Nadig et al. (2008) on the PWN. Using dictionary definitions, the authors applied different extraction and matching algorithms to automatically validate 38,840 nominal synsets (corresponding to 103,620 lemmas) and 56,203 hypernym-hyponym noun pairs, reaching accuracy rates of 70% and 70.88%, respectively. These high rates are hardly surprising, given that the PWN is a handmade resource; nevertheless, they give us an indication as to what might be expected from a similar evaluation performed on automatically-generated WordNets.

Extension. Researchers looking to extend WordNets in languages other than English typically do so by semi-automatically comparing lemmas and synsets in their target language against the contents of the PWN with the help of bilingual dictionaries and linguistic resources. This is the case of the Arabic WordNet (AWN), extended through semi-automated comparison with a lexicon of modern standard Arabic and the PWN (Abouenour et al., 2013). As far as Latin is concerned, a parallel evaluation effort to the one described here is being conducted by the University of Exeter.⁶ In Exeter, the lexical coverage of the LWN has been automatically extended to 70,000 lemmas using *Freedict.com* as well as the Lewis and Short (1879) and Whitaker’s Words Latin dictionaries (hereafter L&S and WW) as sources, and synsets assigned through a ranking system of glosses.⁷

⁶<https://latinwordnet.exeter.ac.uk/>

⁷L&S: https://github.com/PerseusDL/lexica/tree/master/CTS_XML_TEI/perseus/pdllex/lat/ls; WW: <https://github.com/mk270/whitakers-words>; Freedict.com: <https://www.freedict.com/onldict/lat.html>

3 The Latin WordNet

The LWN was first created in 2004 following the *Expand Method* (Vossen, 2002, p. 52), that is, by automatically translating portions of the aligned Italian and English (PWN) data contained in MultiWordNet (hereafter MWN_I and MWN_E) into Latin with the help of bilingual dictionaries (Latin to English mostly from Glare (1982) via WW; Latin to Italian mostly from Pianezzola et al. (2001)). The LWN comprises 9,378 lemmas distributed across 8,973 synsets (Minozzi, 2017): 5,621 synsets are nominal (denoted by the initial n# in the ID), 2,283 verbal (v#), 775 adjectival (a#) and 294 adverbial (r#). Additionally, it provides two files of synset relations: one containing 13,771 language-dependent lemma-to-lemma relations, the other 4,588 synset-to-synset relations common to MWN (see Table 1).

<i>latin_relation.sql (lemma-to-lemma)</i>		
type	n	%
Antonymy	4,538	32.95%
Pertainymy/Derivation*	9,233	67.04%
<i>common_relation.sql (synset-to-synset)</i>		
type	n	%
Hyper/hyponymy	3,900	85.00%
Meronymy, part of	292	6.36%
Entailment (v)	90	1.96%
Attribute (n)	80	1.74%
Value of (a)	80	1.74%
Similar to (a)	54	1.17%
Cause (v)	34	0.74%
Meronymy, substance of	32	0.69%
Meronymy, member of	26	0.56%

Table 1: The distribution of lemma and synset relations across the LWN. *The *Pertainymy/Derivation* relation between lemmas is not well defined in the LWN documentation.

The criteria behind the selection of LWN lemmas remain unclear, and there are some noticeable gaps, both lexical (*amo, amare* ‘to love’) and relational (the adjectives *inaequabilis* ‘unequal’ and *aequabilis* ‘equal’ are placed in a relation of derivation only but could also count as antonyms). Examples of erroneous, modern senses inherited by the LWN from MWN_E are shown in Table 2. In point of fact, in his most recent publication, the creator of LWN states that the lexical coverage and the results of his automatic assignments need further evaluation and verification (Minozzi, 2017, p. 130).

lemma	synset_id	definition
ager	n#W0021124	in un database, ogni area in cui vengono registrate le singole informazioni che compongono il record [...]
capitolium	n#06188340	the federal government of the United States
genetrix	n#W0021113	titolo e appellativo che si dà alle suore professe o a quelle che hanno cariche particolari; sono venuta a fare atto d’obbedienza alla madre badessa di questo convento
voco	v#00720710	send a message or attempt to reach someone by radio, phone, etc; make a signal to in order to transmit a message [...]

Table 2: Synsets to be removed from LWN.

4 Evaluation method

For a close understanding of the implications of evaluating a WordNet, we formulated a first experiment combining a small, automated extension of the sense coverage of the LWN with a follow-up manual revision of their corresponding synsets. The purpose of this experiment was to measure the reliability and feasibility of these two approaches in order to identify the most effective compromise for LiLa.

Data. Firstly, we formatted LWN and all necessary Machine Readable Dictionaries for the task as relational SQL tables: these included WW, L&S, MWN_E and MWN_I .

Machine-recommended senses. Next, inspired by the work of Abouenour et al. (2013), we formalised a rudimentary algorithm in bash script to automatically extend the sense coverage of the LWN by proposing new synsets taken from the MWN_E . While aware that this method would introduce some noise, the neither exact nor approximate amounts could not be quantified *a priori*. Figure 1 exemplifies the algorithmic process: for the LWN adverb *velociter* ‘swiftly, quickly’, the algorithm 1) searched for joint lemma and PoS overlaps between LWN and WW; 2) where there was a match, it then looked for overlaps between the single-word WW glosses and MWN_E lemmas; 3) where these also matched, it checked the lemma’s corresponding synset(s) in MWN_E for that PoS against existing LWN synsets to 4) label machine recommendations as *NEW* (machine-suggested and not already present in LWN) or *COM*

(for “common”, i.e., machine-suggested but already present in LWN). Table 3 lists the results of the recommender system for *velociter*.

synset_id	definition	label
r#00051957	in a swift manner; she moved swiftly	NEW
r#00082992	with rapid movements; he works quickly	COM
r#00102338	with little or no delay; [...]	COM
r#00285860	without taking pains; [...]	COM

Table 3: Synset assignments for the adverb *velociter* to be evaluated by human raters.

The recommender system produced 121,098 lemma-synset entries for the whole LWN: 93,479 synset assignments (77.19%) were classified as *NEW*, 25,613 (21.15%) as *COM* and 2,006 (1.65%) as *OLD* (synsets present in the LWN only). Given the algorithm’s optimisation on recall, we expected these large numbers to include many false positives and homography, e.g., the verbs *edo*, *edere* ‘to eat’ (3rd conjugation) and *edo*, *edare* ‘to publish’ (1st conjugation) or *volo*, *velle* ‘to want’ (irregular conjugation) and *volo*, *volare* ‘to fly’ (1st conjugation).

Lemma selection. Next, for our test evaluation, we randomly selected 100 LWN-WW matched lemmas, 25 per PoS, featuring both *NEW* and *COM* synset assignments. This selection resulted in 3,746 lemma-synset entries to be evaluated.⁸

Manual evaluation. Of the five raters recruited for the task, four were in possession of intermediate Latin proficiency and one had expert (including spoken) knowledge of the language.⁹ Using a custom web annotation environment designed to facilitate the task and with Latin dictionaries at hand (Campanini and Carboni, 1993; Castiglioni and Mariotti, 1966 1979 1996 2007; Bianchi et al., 1972), raters were instructed to approve or reject synset assignments.

Unsurprisingly, our synset recommender generated irrelevant assignments, as shown in Table 4.

The evaluation was performed over a period of approximately two months and informed the formulation of guidelines to enforce consistency. Among other directives, the guidelines demanded that raters accept an assignment even if specific

⁸Of the 100 selected lemmas, 36 had multiple homographic entries with the same PoS.

⁹Those with intermediate Latin knowledge were pursuing a Master’s degree in Theoretical and Applied Linguistics, while the expert rater completed a Master’s in Modern Philology (“Lettere” and Semantics).

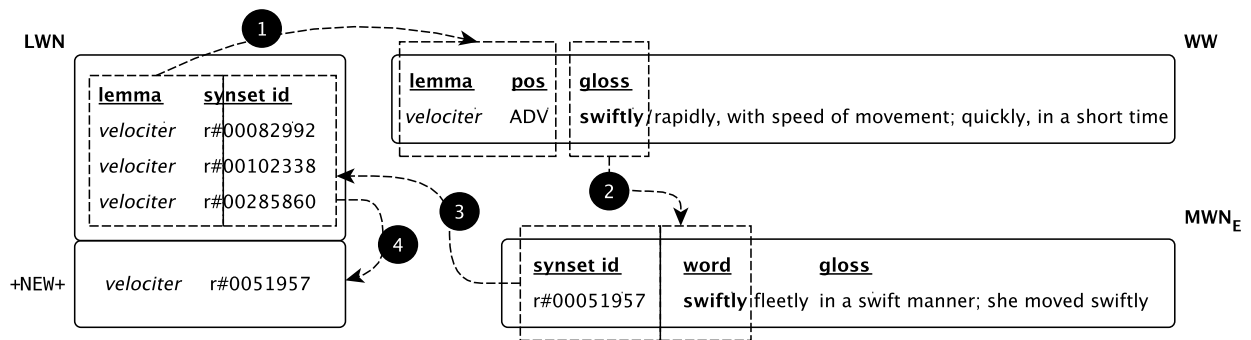


Figure 1: The algorithmic process of synset assignment. Here, a new MWN_E synset is added to *velociter*.

lemma	synset_id	definition
albus	a#01549077	used to signify the Confederate forces in the Civil War (who wore gray uniforms); a stalwart gray figure
caput	n#02805750	a toilet on board a boat of ship
contentus	a#00760259	slang for ‘drunk’
deprehendo	v#00733757	be the catcher, in baseball; Who is catching?
tonus	n#00319371	an all-fours game in which the first card led is a trump

Table 4: Machine-proposed synsets to be discarded from LWN.

to an idiomatic use of the lemma (e.g., *edo*, *edere* ‘to eat/consume/devour’ but *edere voces* ‘utter’); accept an assignment even if its specificity is not mirrored in the reference dictionary (e.g., while the specific sense ‘to sodomize’ for *caedo* is not explicitly mentioned in Castiglioni and Mariotti (2007), the verb is said to have sexual connotations as well);¹⁰ reject an assignment if the corresponding sense is not included in their reference dictionary; and reject an assignment should there be any other strong uncertainty not covered by the guidelines. The assessment of the relations, if any, *between* OLD synset assignments in our evaluation set was ignored at this stage.

Missing senses. Where applicable, raters were also instructed to make a note of missing senses, be those from the Classical, Medieval or Late periods of Latin.¹¹ Inclusion of these missing senses in the LWN is not described here but is planned future work (see Section 6). Examples are:

¹⁰IV ed., s.v., “caedo,” Def. fig. “in senso osceno, *sbatere*, Catull. 56, 7 e a.”

¹¹We do not consider contemporary Latin (19th and 20th centuries).

prudenter (r): skillfully;
puto, *putare* (v): to clean; to prune, trim
radix (n): radish; liquorice
tener (a): erotic, amorous; adaptable (style); soft (soil)

Inter-rater reliability agreement. Next, we measured inter-rater reliability (IRR) using percentage agreement without chance correction (McHugh, 2012). Percentage agreement was chosen over Fleiss Kappa (Fleiss, 1971) because the evaluation was performed in a controlled setting with low chances of guessing on a binary yes/no rating. We thus applied the following formula:

$$A_o(r) = \frac{abs(N_A(r) - N_R(r))}{N_V(r)}$$

where the observed agreement A_o on each lemma-synset relation (r) is calculated by dividing the absolute difference of accepted N_A and rejected N_R assignments by the total number of evaluations N_V . Agreement values range between 0.0 and 1.0, where 0.0 means no inter-rater agreement and 1.0 means perfect inter-rater agreement.

5 Results and discussion

In this section we assess IRR agreement rates against the table proposed by McHugh (2012, p. 279). As previously observed in related studies, lower agreements are not a reflection of raters’ inability to distinguish word meanings but, rather, of their difficulty in selecting the synsets that best fit their subjective opinion (Hajič et al., 2004, p. 28). Table 5 provides minimum (m_v), maximum (M_v) and average values of agreement (A_v) per type of synset assignment as well as standard deviations (S_v). The A_v values all fall within the *strong* tier of McHugh agreement (64-81%, corresponding to a square k agreement of .80-.90), but reveal that almost $\frac{1}{3}$ of all synsets was not reliably

rated.

type	n	m_v	M_v	A_v	S_v
OLD	35	0.200	1.000	0.691	0.345
COM	876	0.200	1.000	0.654	0.320
NEW	2,835	0.200	1.000	0.702	0.329

Table 5: Inter-rater agreement values grouped by type of synset assignment.

IRR agreement is a measure of both actual agreement but also of disagreement among raters. So, for a better understanding of the *quality* of both Minozzi’s and our own synset assignment, we calculated the acceptance rates of OLD, COM and NEW assignments. As Table 6 shows, the acceptance rates on all three types of assignment is very low, with an average 77% of all assignments being rejected by all raters and a tenuous average of 0.02% of unanimous acceptance. These results are particularly worrying for OLD and COM assignments, as they give us a first indication of the quality, and hence usability, of LWN.

type	n	Acceptance in %					
		0_r	1_r	2_r	3_r	4_r	5_r
OLD	35	65.7	14.2	5.7	2.8	11.4	0.0
COM	876	79.4	8.6	5.0	3.4	2.2	0.6
NEW	2,835	87.2	6.7	3.4	1.5	1.0	0.0

Table 6: Acceptance rates of synset types per number of raters (N_r).

As far as Part of Speech (PoS) is concerned, the most prolific syntactic category in terms of machine-proposed synset assignments were verbs, followed by nouns, adjectives and adverbs. Table 7 shows IRR agreement rates per PoS relative to the number of synset assignments; \bar{x} indicates the average or arithmetic mean of synsets per lemma (25 in total) per category. Nouns and verbs fared the best, with *strong* agreement on large percentages of assignments (84% and 60%, respectively); adjectives, on the other hand, appear to have been more challenging, as the percentages of assignments on which the raters moderately and strongly agreed are roughly the same (44% and 48%, respectively). Finally, against our expectations, despite the comparatively lower number of synsets, 48% of adverbial assignments were met with moderate agreement. Low agreement values might be caused by incorrect assignments (as was the case of the NEW assignment ‘with honesty; he was rightly considered the greatest singer of his time’

to *proprie*, approved by only one rater) or, more problematically, differences of opinion on subtle semantic differences. A close examination of the data, and, specifically, of the adverbs with agreement values below 60% (6 out of 25), points to the latter. A clear example is given by the adverb *brevisiter*, whose lowest rated assignment ‘with rapid movements; he works quickly’ (COM) was approved by two raters only. The adverb’s primary sense is ‘shortly, in a brief space of time’, and while ‘rapidly’ might, in some cases, reasonably be equated to ‘shortly’, three raters discarded the assignment as the senses conveyed by the terms ‘rapid’ and ‘quickly’ are better expressed by the Latin adverb *celeriter*. Similarly, in the case of *subtus* ‘below, underneath, in a lower position, beneath’, the NEW assignment ‘at a later place; see below’ was also rejected by three raters, despite it being a potentially valid sense. It is worth noting that in these and other arguable cases, synsets carrying temporal meanings tend to show lower agreement rates than those associated with space (i.e., ‘rapidly’ and ‘later’ are temporal equivalents of ‘short’ and ‘below’). The higher agreement rate on the spatial dimension resonates with cognitive linguistic theories on spatial semantics, according to which “Space is at the heart of all conceptualization” (Pütz and Dirven, 1996, xi), as its concreteness over temporal or more abstract meanings induces us to map its structure onto other semantic domains (Lakoff and Johnson, 1980; Lakoff, 1987). The validity of these theories in the context of LWN evaluation remains to be verified.

type	\bar{x} syn/lemma	IRR agreement in %		
		moderate	strong	\approx perfect
VERB	51.32	32	60	8
NOUN	46.56	0	84	16
ADJ	42.04	44	48	8
ADV	8.84	48	28	20

Table 7: IRR agreement rates per PoS relative to the number of synsets.

6 Conclusion and future work

This paper describes a preliminary assessment of the implications of evaluating the LWN carried out in the context of the *LiLa: Linking Latin* project. The objective of LiLa is to connect linguistic resources and NLP tools for Latin with a view to supporting different lines of linguistic and corpus-based research and to connecting Latin to other

languages. Owing to its automatic process of creation, the LWN is lexically and semantically limited, as well as noisy, subjecting its inclusion in LiLa to qualitative revision. For a close understanding of the implications of evaluating the LWN, we formulated a first experiment combining a small, automated extension of the sense coverage on the basis of 100 selected LWN lemmas with a follow-up manual revision of their corresponding synset assignments. The purpose of this experiment was to measure the reliability and feasibility of these two approaches in order to identify the most effective evaluation compromise.

Our synset recommender system produced many false positives, with only 0.18% (7) machine suggestions approved by all five raters. Even if the precision of the synset-recommendation algorithm were to be improved, recall would likely still be high due to the unavoidable assignment of modern senses to a historical resource. If applied to the entire LWN, the evaluation method described here, coupled with the additional evaluation of the relations between synsets, would turn this process of revision into an unsustainable effort or, at the very least, one that is not achievable within the scope and duration of LiLa.¹²

Moving forward, our plan for the improvement of the LWN will develop into various tasks. The first, ongoing effort is the manual removal of the modern senses originally inherited by the LWN. Next, once cleaned, we will extend the sense coverage of the LWN by manually adding the missing senses recorded by the raters for the 100 evaluated lemmas, careful not to introduce too much granularity (i.e., too many senses with only subtle semantic differences); extract hypernyms, synonyms and bags of words from dictionary definitions (Nadig et al., 2008), as well as lemma groups from three Latin synonym dictionaries: the Latin-English *Hand-book of Latin Synonymes* (Döderlein et al., 1875), the Latin-English *The synonymes of the Latin language* (Hill, 1804) and the Latin-Czech *Latinská synonymika pro školu i dům* (Skřivan, 1890).¹³ These are all freely available online in XML dictionary format (XDXF) and, combined, can supply the LWN with some

1,050 additional lemmas.

Thirdly, connect a graph version of the LWN to textual resources in LiLa to acquire lexical knowledge, and explore the possibility of extracting hypernym/hyponym pairs using syntactic patterns (Snow et al., 2004). Finally, extend the LWN with Named Entities extracted from the morphological analyser LEMLAT (Budassi and Passarotti, 2016).

The data and code repository for this paper are available at: <https://github.com/CIRCSE/latinWordnet-evaluation>

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme Grant Agreement No. 769994. The authors also wish to thank William Michael Short of the University of Exeter for the productive discussions leading up to this article, and the anonymous reviewers for their helpful suggestions.

¹²In an unlikely scenario of uninterrupted evaluation, our method applied to the entire LWN would indicatively require 64.65 months to complete.

¹³Available from: <https://nikita-moor.github.io/dictionaries/dictionaries.html>

References

- Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso. 2013. On the evaluation and improvement of Arabic WordNet coverage and usability. *Language Resources and Evaluation*, 47(3):891–917. <https://doi.org/10.1007/s10579-013-9237-0>.
- Enrico Bianchi, Raffaello Bianchi, and Onorio Lelli. 1972. *Dizionario illustrato della lingua latina*. Le Monnier, Firenze.
- Yuri Bizzoni, Federico Boschetti, Riccardo Del Gratta, Harry Diakoff, Monica Monachini, and Gregory Crane. 2014. The making of Ancient Greek WordNet. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 9th edition of the Language Resources and Evaluation Conference, 26-31 May, Reykjavik, Iceland*, pages 318–325. http://www.lrec-conf.org/proceedings/lrec2014/pdf/1071_Paper.pdf.
- Olivier Bodenreider, Anita Burgun, and Joyce A. Mitchell. 2003. Evaluation of wordnet as a source of lay knowledge for molecular biology and genetic diseases: a feasibility study. *Studies in health technology and informatics*, 95:379–384. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1893008/>.
- Marco Budassi and Marco Passarotti. 2016. Nomen Omen. Enhancing the Latin Morphological Analyser Lemlat with an Onomasticon. In *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 90–94. ACL. <http://www.aclweb.org/anthology/W16-2110>.
- Giuseppe Campanini and Giuseppe Carboni. 1993. *Nuovo Campanini-Carboni vocabolario latino-italiano italiano-latino, con appendice suddivisa in 11 glossari*. Paravia, Torino.
- Luigi Castiglioni and Scevola Mariotti. 1966, 1979, 1996, 2007. *IL Vocabolario della Lingua Latina,, Latino - Italiano . Italiano - Latino*. Loescher, Torino.
- Christian Chiarcos, John McCrae, Philipp Cimiano, and Christiane Fellbaum, 2013. *Towards Open Data for Linguistics: Linguistic Linked Data*, pages 7–25. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31782-8_2.
- L. Von Döderlein, S. H. Taylor, and H. H. Arnold. 1875. *Döderlein's Hand-book of Latin Synonymes*. Warren F. Draper. <https://archive.org/details/dderleinshandbo00arnogooq>.
- Christiane Fellbaum. 1998. Towards a Representation of Idioms in WordNet. In *Proceedings of the workshop on the Use of WordNet in Natural Language Processing Systems (Coling-ACL)*, pages 52–57. <http://ai2-s2-pdfs.s3.amazonaws.com/280b/cd6c4f1e3b9f9abb32a0c510614e5128d9df.pdf>.
- Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76:378–382. <http://dx.doi.org/10.1037/h0031619>.
- Greta Franzini, Marco Passarotti, Maria Moritz, and Marco Böhler. 2018. Using and evaluating TRACER for an Index fontium computatus of the Summa contra Gentiles of Thomas Aquinas. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), 10-12 December, Torino, Italy*. <http://ceur-ws.org/Vol-2253/paper22.pdf>.
- P. G. W. Glare. 1982. *Oxford Latin Dictionary*. Oxford University Press, Oxford.
- Jan Hajič, Martin Holub, Marie Hučínová, Martin Paylík, Pavel Pecina, Pavel Straňák, and Pavel Šidák. 2004. Validating and Improving the Czech WordNet via Lexico-Semantic Annotation of the Prague Dependency Treebank. In *Proceedings of LREC 2004*, pages 25–30. https://www.researchgate.net/publication/237100339_Validating_and_Improving_the_Czech_WordNet_via_Lexico-Semantic_Annotation_of_the_Prague_Dependency_Treebank.
- F. R. S. E. John Hill. 1804. *The synonymes of the Latin language*. Edinburgh. <https://archive.org/details/synonymesoflatin00hilluoft>.
- George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. The University of Chicago Press, Chicago.
- George Lakoff. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. The University of Chicago Press, Chicago.
- Charlton T. Lewis and Charles Short. 1879. *Harpers' Latin Dictionary: A New Latin Dictionary Founded on the Translation of Freund's Latin-German Lexicon Edited by E. A. Andrews*. Harper and Brothers, New York.
- Mary L. McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica*, 22:276–282. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/>.
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4):235–244. <https://academic.oup.com/ijl/article/3/4/235/923280>.

- Stefano Minozzi. 2017. Latin WordNet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'Information Retrieval. In Paolo Mastandrea, editor, *Strumenti digitali e collaborativi per le Scienze dell'Antichità*, number 14 in *Antichistica*, pages 123–134. <http://doi.org/10.14277/6969-182-9/ANT-14-10>.
- Raghuvar Nadig, J. Ramanand, and Pushpak Bhattacharyya. 2008. Automatic Evaluation of Wordnet Synonyms and Hypernyms. In *Proceedings of the Sixth International Conference on Natural Language Processing (ICON-2008)*. <https://www.cse.iitb.ac.in/~pb/papers/icon08-wn-validation.pdf>.
- Marco Passarotti, Flavio M. Cecchini, Greta Franzini, Eleonora Litta, Francesco Mambrini, and Paolo Ruffolo. 2019. LiLa: Linking Latin A Knowledge Base of Linguistic Resources and NLP Tools. In Thierry Declerck, editor, *Proceedings of the 2nd Conference on Language, Data and Knowledge (LDK 2019), 20-23 May, Leipzig, Germany*. <https://doi.org/10.5281/zenodo.3358550>.
- Emilio Pianezzola, Giuliano Ranucci, and Gian Biagio Conte. 2001. *Il dizionario della lingua latina*. Edmund Le Monnier.
- Martin Pütz and Renè Dirven, editors. 1996. *The construal of space in language and thought*. Mouton de Gruyter, Berlin.
- A. Skřivan. 1890. *Latinská synonymika pro školu i dům*. Chrudim. <https://archive.org/details/SkivanLatinskSynonymika>.
- Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. 2004. Learning syntactic patterns for automatic hypernym discovery. *Advances in neural information processing systems*, 17:1297–1304. http://ai.stanford.edu/~rion/papers/hypernym_nips05.pdf.
- Piek Vossen. 2002. EuroWordNet General Document. <http://dare.ubvu.vu.nl/bitstream/handle/1871/11116/EWNG?sequence=1>.