The Linguistic Linked Open Data through the Linguists' Lens

Pasquale Esposito[1,*]

[1]Dipartimento di Studi Umanistici, Università degli Studi di Salerno, via Giovanni Paolo II, 132, 84084 Fisciano (SA), Italy

Abstract

The Linguistic Linked Open Data (LLOD) movement represents an extraordinary point of reference for linguistics in the Semantic Web, as data are openly published and structured according to the linked data principles for interoperability. The LLOD Cloud has continuously grown, embracing cross-domain linguistic studies such as general linguistics, lexicology, psycholexicology, and computational linguistics. The effort invested in proposing ontologies that are halfway between linguistics and information technology communities is outstanding.

This article proposes a qualitative observation of the *register labeling* for lexicons and ontologies of the LLOD cloud to assess its potential through a linguistic lens. More specifically, it compares linguistic expectations with aspects already covered by ontologies commonly used to model LLOD to determine in which direction improvements might be helpful. As a result, formality detection and weighted lexicons have been explored for analyzing texts in Natural Language Processing. However, attention to *formality relevance* and *register labeling* ought to be increased in the design of LLOD-related ontologies. Corpus-based approaches are instrumental in detecting variance in formality and register. The W3C community is demonstrating a significant interest in this approach with the implementation of the FrAC module. In conclusion, developing a metrics-based formality labeled lexicon could be a game-changer for LLOD resources. It could serve as a valuable linguistic research reference and enhance the accuracy of formality calibration in speech for machine learning operations.

Keywords

Semantic Web, Linguistics, Register Variation, Qualitative analysis, Linguistic Linked Open Data

1. Introduction

Linguistic Linked Open Data (LLOD)[1] [1, 2] is a movement led by the Open Linguistics Working Group and aims to publish data for linguistics and Natural Language Processing using the linked data (LD) principles[2]. The LLOD movement considers the publishing of LD as a possibility to allow resources to be globally and uniquely identified such that they can be retrieved through standard Web protocols and be easily linked to one another in a uniform fashion and become structurally interoperable [3]. LLOD practically represent the cut-out of linguistic expertise from the multi-domain Linked Open Data Cloud[3].

DQMLKG workshop at ESWC 2024

*Corresponding author.

pasesposito@unisa.it (P. Esposito)

0009-0006-3464-5861 (P. Esposito)

[1]LLOD: https://linguistic-lod.org

[2]Linked Data principles from https://www.w3.org/wiki/LinkedData

[3]Linked Open Data Cloud: https://lod-cloud.net

Data are meant to be machine readable but also available for user interaction as a heterogeneous set of resources for linguistics purposes and language description which comprehends corpora, lexicons, dictionaries, terminologies, thesauri, knowledge bases, linguistic metadata, data categories, typological databases.

Efforts have been made to offer systems and directives which are convenient for linguistic research [4, 5] in metadata enrichment [6] and dictionary generation from encyclopedic knowledge [7]. Data have also been kept machine-readable and exploitable for computer science's common task such as Large Language Models (LLMs) and word embeddings improvement [8], word-sense disambiguation [9, 10, 11, 12], meaning representation [13], personal knowledge-graph representation [14], cross language linking [15, 16, 17], Natural Language Processing operations and question answering [18]. The utility of LLOD in LLMs and, in general, Artificial Intelligence (AI) applications, requires a deep focus on lexical resources and the way they are structured to be fully and properly exploited in natural language related tasks.

Our intent is to verify the remarkable potential of LLOD through a qualitative assessment of the presence or absence of traditional lexicographic features among different-structured resources of the LLOD Cloud. This paper sets out to investigate the extent to which LLOD modeling approaches meet the expectation of linguistics. In more detail, it qualitatively compares features expected by linguists and if and to what extent they are covered by the ontologies used by the LLOD as a reference. This study is primarily viewed through the lens of linguistics, proposing LLOD ontology extensions that could potentially enhance AI-driven applications. By framing it as a Research Question (RQ), our intention can be modeled as follows:

Do LLOD represent (all) expected features of a traditional lexicographic resource?

As a result, the ontological design, and the different typologies of LLOD cloud resources seem broad enough to cover the whole range of linguistic domains. Regardless of resource data structure, consulting linguistic resources as lexicographic references can result in a series of missing expectations on the linguistic side.

The rest of the article is structured as follows. Section 2 overviews linguistic features expected by linguistics when approaching lexicographic resources. Section 3 retraces the sensitivity of ontologies used by the LLOD as a reference to cover linguistic expectations. Section 4 discusses what is missing in the LLOD ontologies through the linguistics' lens. Finally, the article concludes with final remarks and suggesting future directions.

2. Expected Linguistic Features

Studies in the field of lexicology define the parameters to lexical resources compilation. Both users and linguists place a range of expectation when consulting a lexicographic resource.
Over time, research has seen different ways to categorize lexicons and linguistics elements. Lexicographic resources can be mainly divided into traditional-like lexicons which are thought to give a unitary representation of the *lemma* and synset-based resources, which aggregate words following synonymy relation. By consulting a traditional lexicographic, parameters expected to be found for a single entry are the following: entry, form, definition, senses, phonetic transcription, morphological pattern, domain label, different usages, register label,

style label, relevance, animacy, aspect, case, clitic, definiteness, degree, finiteness, gender, number, modification type, part of speech, person, tense. In linguistic terms, speakers implicitly deal with the choice of a certain register and style are in relation to different communicative situations. Three main and general directions in register localization can be identified: neutral, formal and informal. Style-related studies are mainly performed on corpora. As Biber states, corpus-based studies generally have one of two primary research goals [19]: to describe the variants and use of a word or linguistic structure and to describe differences among texts and text varieties, such as registers or dialects.

3. OntoLex - Register and formality sensitivity

The OntoLex[4] model found its final specification in May 2016, published by the Ontology-Lexicon community Group under the W3C Community Final Specification. The main aim of the ontology has been to establish a standard for the publication of lexicons and representation of dictionaries [20]. OntoLex reports also a painstaking attention to recreate a mold though which one can filter lexicographic data [5]. The basic model is composed of a core module, four additional modules and two emerging as follow: **OntoLex-Lemon:** Core module, **OntoLex-SynSem:** OntoLex module for Syntax and Semantics, **OntoLex-Decomp:** OntoLex module for Decomposition, **OntoLex-VarTrans:** OntoLex module for Variation and Translation, **OntoLex-LiMe:** OntoLex module for Linguistic Metadata, **OntoLex-Lexicog:** OntoLex module for Lexicography, **OntoLex-Morph:** *emerging* OntoLex module for Morphology, **OntoLex-FrAC:** *emerging* OntoLex module for Frequency, Attestation and Corpus-Based Information.

3.1. Language Variation and Paradigmatic Description in OntoLex

In linguistic terms, pragmatics is mainly concerned with how users make use of the linguistic competence to interact with other users as Austin's book title *How to do Things with words* [21]. Lexicon represents one of pragmatics' interfaces. Speaker's lexical choice can assign a different weight or degree to the utterances, based on the words selected from the inventory. Lexical selection is based on an inventory which can be identified as lexicon. Languages are subjected to changes in each of their facets and are essentially shaped following speakers' arbitrariness. The field which deals with language change is the *historical linguistics* which identifies sound changes, lexical changes, semantic changes, and syntactic changes as the main interested to shifts. This perspective identifies language more like a dynamic system, rather than a static image circumscribed and spaced out in a series of recurrent snapshots.

The **Variation & Translation (vartrans)** module [6] is designated to provide the structure around which diatopic, diaphasic, diachronic, diastratic and dimensional variants can be formalized. Given that, the model shows a clear sensitivity towards the linguistic change and offers the structure to shape data around. However, this module lacks modeling of the temporal marks. Linguistically, lexical change and semantic shift are recognizable through time with the addition or disuse of new words. This implies the implication of the expansion of vocabularies. From a

[4]OntoLex: https://www.w3.org/community/ontolex
[5]https://www.w3.org/community/ontolex/wiki/Main_Page
[6]https://www.w3.org/ns/lemon/vartrans

diatopic perspective, terms used to denote concepts can certainly vary. To cope with the lack of temporal marks in the ontology used as a reference, **LemonDIA extension** [22] has been proposed to track shifts in meaning of a word in a diachronic perspective. It represents a clear intention to have a module able to keep track and report semantic change in data description from a diachronic (time) perspective LemonDIA achieves it by adding a temporal dimension to the word and its related senses [22].

3.2. Formality in Speech

Besides tracking denotative and connotative change in meaning for terms trough time, lexicographic resources need to be a reference for formality relevance in lexical speakers' choices. To this end, Heylighen and Dewaele [23] attempted to identify a computational measure to quantify formality in speech through the formula:

> F [*formality measure*] = (noun frequency + adjective frequency + preposition frequency + article frequency − pronoun frequency − verb frequency − adverb frequency − interjection frequency + 100)/2

The formula is mainly based on frequencies expressed as percentages of the number of words belonging to a particular category with respect to the total number of words in an excerpt. *F* will then vary between 0 and 100% based on the formality of the language. The higher the value of F, the higher is the formality of the excerpt [23]. The *F* formula can be applied to different communicative situations, genre, speakers. This applicability confirms, for the authors, that formality is the most fundamental and most universal dimension of stylistic variation [23].
Another approach to model formality has been proposed by Brooke and others [24] who move the focus on the quantification of formality by assuming the single lexical item as a unit, rather than approaching the weighting of the F on the whole text. It lets them distinguish the formality of near-synonym pairs. Following this approach, Eder et al. [25] propose a weighted lexicon for German language which computes F for each lexical item. Computation of the measure is obtained through a corpus-based approach, as a result of a manually phase, vectors and text-score. It is worth noting that the corpus-based approach reflects the state of the language production of speakers in the exact moment it is collected and can be both a quantitative and qualitative reference for linguistic observations.

3.3. Register Labels in Traditional OR Synset-based resources

While traditional-like lexicons use lemmas as unitary representation, the LLOD Cloud identifies also synset-based resources tend represent data in a blended representation of paired lemmas and descriptive dictionary. We take into account different resources with different technical infrastructure and ontological design. We consider the earlier version and the correspondent last linked data version, to investigate how they model the formal relevance report and register labeling and if there is an evolution and a rising attention towards this component. Among the most used linguistic resources, we selected the following leading resources:

- WordNet[7]/Open English WordNet[8] (*synset-based*)
- Wiktionary[9]/DBnary[10] (*non-synset*)
- ConceptNet[11] (*non synset and machine learning/embeddings dependent*)

Wordnet. The representative of synset-based resources is Princeton WordNet [26] as it represented the innovation and the bulwark upon which the largest part of current data modeling has been built. According to its authors, WordNet was compiled following psycholinguistic principles and by posing the attention towards synchronic lexical-semantics to that defined as *psycholexicology*. It represents a hybrid framework to organize units of meaning and model a lexical semantic network in a practical way. As defined by Millert et al., WordNet is a proposal for a more effective combination of traditional lexicographic information and modern high-speed computation [26]. The WordNet's synset was thought to be monolingual and aggregate synonyms in a single "node". Miller defines that this organization makes words denotationally equivalent and can be substituted for one another in many, but not all, contexts [27].

Formality relevance measures are also absent for lexical units in the synset-based resources such as WordNet, as well as in WordNet fork, such as, Open English WordNet (OEW).

Wiktionary. It is identified as a collaborative-built resource, intended to be a free and web alternative to traditional dictionaries and so-defined expert-built lexicons [28]. From a lexicographic perspective, Wiktionary has a traditional descriptive approach which is different from WordNet's psycholinguistic one and leaves room for a more accurate description of each lexeme and detaches different formal representations of the same meaning from being paired. Consequently, as also reported in the presentation article [28], Wiktionary embraces and explicitly reports the linguistic labels for *domain* and *style and register*, providing the related and expected information for the lexeme. Similarly, DBnary is the ontology-based representation of Wiktionary modeled according to a modified version of the OntoLex model[12]. Conversely from Wiktionary, DBnary does not report registers or domain labels.

ConceptNet. The one-lexeme one-node structure is also applied in ConceptNet [29], a non-synset-based resource defined as a semantic network designed to help computers understand the meanings of words that people use[13]. ConceptNet is part of the ecosystem of the LLOD and has its strength in the embedding-based structure, which integrates the hybrid framework between distributional semantics and relational knowledge through ConceptNet Numberbatch [29] to retrieve the best of both worlds. Regrettably, ConceptNet lexical description does not foresee any formality relevance measure.

4. Discussion

It is worth recalling that the RQ at the base of this contribution is to verify if *LLOD represent expected features of a traditional lexicographic resource.* This section first summarizes working

[7]WordNet: https://wordnet.princeton.edu
[8]Open English WordNet: https://en-word.net
[9]Wiktionary: https://www.wiktionary.org
[10]DBnary: https://kaiko.getalp.org/about-dbnary
[11]ConceptNet: https://conceptnet.io
[12]The DBnary Ontolex Extension Data Model: http://kaiko.getalp.org/static/datamodel/2.1.2/index-en.html
[13]ConceptNet: https://conceptnet.io

mechanisms adopted by the community to blend linguistic needs and semantic web technologies. Then, it emphasizes what is still missing in the LLOD to satisfy linguistic requirements fully. Finally, it overviews the impact of correctly modeling formality for AI-driven applications.

Language variations require to be better modeled. Previous sections outline how linguistics and computer science have put together efforts to find common ground in the description and representation of data. Since linguistics and languages are a vast field to circumscribe, the results thus far seem more than valuable. Data produced and modeled are, in most cases, satisfactory for linguists' expectations. However, some features still need to be included if we qualitatively analyze ontologies and LLOD from the linguistic point of view. Linguistics cannot sacrifice formality characterization. Several works in the literature paved the way towards a relevant framework of studies and resource building in terms of formality evaluation and computation [23, 24, 25]. What is still missing is an effort to obtain a computational and *weighted* formality relevance attached to LLOD, similar to the one experienced traditional dictionaries.

Collaborative effort to give voice to speakers. WordNet has been widely used in several projects and inspired several elaborations. To name a few, it inspired BabelNet[14] [30] and OEW. In particular, OEW solved many errors, such as spelling mistakes, increased the quantity of considered synsets by engaging a community of collaborators, and improved the quality of the original resource [31]. Guidelines for the addition of new synsets in OEW is based on a crowd-sourced method which needs of 100 examples in corpus-based research.

Similarly, Wiktionary and ConceptNet rely on collaboration and crowd-sourcing as a common and fruitful practice. In this direction, the description and the modeling of formality can take advantage of crowd-sourcing approaches to attach formality weights to words.

Utility of the formality measures for AI-driven applications. Previous research demonstrates how impactful the *USE* of KGs for Deep Learning Models can be, as in the case of Graph Neural Networks (GNNs) [32], translational model [33] and comprehensible AI [34]. Improvements and integration of computational measures would benefit LLOD to become more effective resources for linguistic research and for the *USE* of KGs as training for machine learning operations. Generally speaking, implementing metrics-based labeled resources that are sensitive to features like register variation can certainly raise the level of sensitivity of tools that use KGs as training references. Having a formality labeled resource based on speakers' perceptions can certainly represent a valuable point of reference for improving the AI reproduction of human speech in accordance with individual linguistic expectations.
Reasonably, the common ground between linguistics, Natural Language Processing, Semantic Web, and AI would be the usage of *corpora* as a qualitative and quantitative reference to analyze and then computing the authentic reproduction of speech. Consequently, computational measures can be applied to extract weighted lexicons to be integrated into KG's linguistic resources, aiming to depict a clear 'state of perception' of the lexical items. The consequent benefit would be the possibility of having an approach that does not only work on frequency and words' co-occurrence to determine a qualitative phenomenon such as the one object of this analysis.

[14]BabelNet: https://babelnet.org

Hence, adding a qualitative 'linguistic supervision' passed through computational application can improve the Language Model's sensitivity in this direction. Finally, when KGs are exploited for machine learning, the weighted correlation between the elements will likely increase the precision of managing the linguistic register.

5. Conclusion and Future Directions

Linguistics and computer science have created an extended collaboration environment that has certainly quantitatively and qualitatively produced fine-designed resources for research purposes of both worlds. The LLMs take advantage of pure natural language processing operations and implement the linguistics side with data management and AI principles to offer systems that can interact and automatically generate language in a human-like way.
Language and the representation of linguistics features in the Semantic Web field are this paper's main focus. In particular, we have focused on the LLOD Movement, which has widely and valuable posed relevant attention to linguistics and invested immense efforts in improving and finding common ground between computer science and linguistics in order to satisfy requests and tweak data for linguistic purposes. To further enhance the LLOD coverage of linguistic expectations, we suggest a hybrid corpus-based/crowd-sourced approach to detect formality *weights* for lexicons of different languages as a starting point towards the integration of computational formality measures to linguistics resources of the LLOD cloud. As we mentioned before, speakers arbitrarily and implicitly shape language. Linguistic material collected in corpora becomes data for machine learning and distributional semantics operations. Consequently, observations drawn from corpora and statistical methods to analyze the word behavior in contexts strongly influence AI-driven systems [35]. As a result, having a weighted and precise description of formality relevance for LLOD satisfies linguistic expectations and can produce a wide range of benefits in several (semi-) automatic AI-driven applications. This is because formality can be considered while generating content.

Future directions. Collaborative improvement of resources appears to be a beneficial way to improve KG's quality and quantity of data, as demonstrated by the success stories of OEW and ConceptNet. As a common practice in linguistics, the Semantic Web community might take advantage of manually annotated and linguistically validated corpora to include formality weights and model them in data. The **FrAC (frequency, Attestation and Corpus information) module** is emerging as a reference to integrate corpus-based information into lexical resources, including frequency information, attestations which correspond to corpus examples, collocation scores, embeddings and similarity metrics [36]. The specifications offered by the FrAC module in terms of (*frac:Similarity*) computation open the way to a weighted computation of linguistic aspects, valuable both for linguistics and computer-based operations. A similar approach might be applied to model the register labeling and formality metrics. We acknowledge that this suggestion is based on a qualitative assessment from a linguistic point of view. Further studies are required to quantify technical challenges to make this proposal concrete.

References

[1] C. Chiarcos, S. Hellmann, S. Nordhoff, Linking Linguistic Resources: Examples from the Open Linguistics Working Group, Springer Berlin Heidelberg, 2012, pp. 201–216. doi:10.1007/978-3-642-28249-2_19.

[2] C. Chiarcos, Ontologies of linguistic annotation: Survey and perspectives, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC), European Language Resources Association (ELRA), 2012, pp. 303–310. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/911_Paper.pdf.

[3] C. Chiarcos, P. Cimiano, T. Declerck, J. P. McCrae, Linguistic linked open data (LLOD). introduction and overview, in: Proceedings of the 2nd Workshop on Linked Data in Linguistics (LDL): Representing and linking lexicons, terminologies and other language data, Association for Computational Linguistics, 2013, pp. 1–. URL: https://aclanthology.org/W13-5501.

[4] A. Khan, C. Chiarcos, T. Declerck, D. Gifu, E. García, J. Gracia, M. Ionov, P. Labropoulou, F. Mambrini, J. Mccrae, É. Pagé-Perron, M. Passarotti, S. Muñoz, C.-O. Truică, When linguistics meets web technologies. recent advances in modelling linguistic linked data, Semantic Web 13 (2022) 987–1050. doi:10.3233/SW-222859.

[5] Profiling Linguistic Knowledge Graphs, Zenodo, 2022. URL: https://doi.org/10.5281/zenodo.6827645. doi:10.5281/zenodo.6827645.

[6] M. P. Di Buono, H. Gonçalo Oliveira, V. Mititelu, B. Spahiu, G. Nolano, Paving the way for enriched metadata of linguistic linked data, Semantic Web 13 (2022) 1–25. doi:10.3233/SW-222994.

[7] L. Bajcetic, T. Declerck, Using wiktionary to create specialized lexical resources and datasets, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, S. Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference, LREC 2022, Marseille, France, 20-25 June 2022, European Language Resources Association, 2022, pp. 3457–3460. URL: https://aclanthology.org/2022.lrec-1.370.

[8] A. Celikyilmaz, D. Hakkani-Tür, P. Pasupat, R. Sarikaya, Enriching word embeddings using knowledge graph for semantic tagging in conversational dialog systems, AAAI - Association for the Advancement of Artificial Intelligence, 2015.

[9] M. AlMousa, R. Benlamri, R. Khoury, A novel word sense disambiguation approach using wordnet knowledge graph, Computer Speech Language 74 (2022) 101337. URL: https://www.sciencedirect.com/science/article/pii/S0885230821001303. doi:https://doi.org/10.1016/j.csl.2021.101337.

[10] A. Gangemi, M. Alam, V. Presutti, Word frame disambiguation: Evaluating linguistic linked data on frame detection, in: LD4IE@ISWC, 2016. URL: https://api.semanticscholar.org/CorpusID:11945337.

[11] S. D. Roy, W. Zeng, Cognitive canonicalization of natural language queries using semantic strata, ACM Trans. Speech Lang. Process. 10 (2014). URL: https://doi.org/10.1145/2539053. doi:10.1145/2539053.

[12] K. Elbedweihy, Using babelnet in bridging the gap between natural language queries and linked data concepts, 2013.

[13] G. A. Burns, U. Hermjakob, J. L. Ambite, Abstract meaning representations as linked data, in: P. Groth, E. Simperl, A. Gray, M. Sabou, M. Krötzsch, F. Lecue, F. Flöck, Y. Gil (Eds.), The Semantic Web – ISWC 2016, Springer International Publishing, Cham, 2016, pp. 12–20.
[14] X. Li, G. Tur, D. Hakkani-Tür, Q. Li, Personal knowledge graph population from user utterances in conversational understanding, in: 2014 IEEE Spoken Language Technology Workshop (SLT), 2014, pp. 224–229. doi:10.1109/SLT.2014.7078578.
[15] M. Chen, W. Shi, B. Zhou, D. Roth, Cross-lingual entity alignment with incidental supervision, in: P. Merlo, J. Tiedemann, R. Tsarfaty (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online, 2021, pp. 645–658. URL: https://aclanthology.org/2021.eacl-main.53. doi:10.18653/v1/2021.eacl-main.53.
[16] F. Narducci, M. Palmonari, G. Semeraro, Cross-language semantic matching for discovering links to e-gov services in the lod cloud, CEUR Workshop Proceedings 992 (2013) 21–32.
[17] Q. Miao, H. Lu, S. Zhang, Y. Meng, Cross-lingual link discovery between chinese and english wiki knowledge bases, 27th Pacific Asia Conference on Language, Information, and Computation, PACLIC 27 (2013) 374–381.
[18] H. Li, F. Xu, Question answering with dbpedia based on the dependency parser and entity-centric index, in: 2016 International Conference on Computational Intelligence and Applications (ICCIA), 2016, pp. 41–45. doi:10.1109/ICCIA.2016.10.
[19] D. Biber, Register as a predictor of linguistic variation, Corpus Linguistics and Linguistic Theory 8 (2012) 9–37. URL: https://doi.org/10.1515/cllt-2012-0002. doi:doi:10.1515/cllt-2012-0002.
[20] C. D. F. John P. McCrae, P. Cimiano, Publishing and linking wordnet using lemon and rdf, 2014. URL: https://api.semanticscholar.org/CorpusID:60275443.
[21] J. L. Austin, How to do things with words, William James Lectures, Oxford University Press, 1962. URL: http://scholar.google.de/scholar.bib?q=info:xI2JvixH8_QJ:scholar.google.com/&output=citation&hl=de&as_sdt=0,5&ct=citation&cd=1.
[22] F. Khan, F. Boschetti, F. Frontini, Using lemon to model lexical semantic shift in diachronic lexical resources, 2014.
[23] F. Heylighen, J. Dewaele, L. Apostel, Formality of language: definition, measurement and behavioral determinants, 1999. URL: https://api.semanticscholar.org/CorpusID:16450928.
[24] J. Brooke, T. Wang, G. Hirst, Automatic acquisition of lexical formality, in: C.-R. Huang, D. Jurafsky (Eds.), Coling 2010: Posters, Coling 2010 Organizing Committee, Beijing, China, 2010, pp. 90–98. URL: https://aclanthology.org/C10-2011.
[25] E. Eder, U. Krieg-Holz, U. Hahn, Acquiring a formality-informed lexical resource for style analysis, in: P. Merlo, J. Tiedemann, R. Tsarfaty (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online, 2021, pp. 2028–2041. URL: https://aclanthology.org/2021.eacl-main.174. doi:10.18653/v1/2021.eacl-main.174.
[26] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, K. J. Miller, Introduction to WordNet: An On-line Lexical Database*, International Journal of Lexicography 3 (1990) 235–244. doi:10.1093/ijl/3.4.235.
[27] G. A. Miller, C. Fellbaum, Wordnet then and now, Language Resources and Evaluation 41 (2007) 209–214. URL: http://www.jstor.org/stable/30200582.

[28] C. M. Meyer, I. Gurevych, Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography, na, 2012.
[29] R. Speer, J. Chin, C. Havasi, Conceptnet 5.5: An open multilingual graph of general knowledge, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI Press, 2017, p. 4444–4451. doi:10.5555/3298023.3298212.
[30] R. Navigli, S. P. Ponzetto, Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, Artificial Intelligence 193 (2012) 217–250. doi:10.1016/j.artint.2012.07.001.
[31] J. P. McCrae, A. Rademaker, F. Bond, E. Rudnicka, C. Fellbaum, English WordNet 2019 – an open-source WordNet for English, in: Proceedings of the 10th Global Wordnet Conference, Global Wordnet Association, 2019, pp. 245–252. URL: https://aclanthology.org/2019.gwc-1.31.
[32] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Transactions on Neural Networks 20 (2009) 61–80. doi:10.1109/TNN.2008.2005605.
[33] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, Advances in neural information processing systems 26 (2013).
[34] S. Schramm, C. Wehner, U. Schmid, Comprehensible artificial intelligence on knowledge graphs: A survey, Journal of Web Semantics 79 (2023) 100806. doi:https://doi.org/10.1016/j.websem.2023.100806.
[35] A. Lenci, J. S. Littell, Distributional semantics in linguistic and cognitive research, The Italian Journal of Linguistics 20 (2008) 1–32.
[36] C. Chiarcos, E. S. Apostol, B. Kabashi, C.-O. Truică, Modelling frequency, attestation, and corpus-based information with ontolex-frac, in: International Conference on Computational Linguistics, 2022.