Proceedings of the Workshop on Multimodal Wordnets (MMWN-2020), pages 39–44
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Adding Pronunciation Information to Wordnets

Thierry Declerck[1,2], Lenka Bajčetić[2], Melanie Siegel[3]
[1]German Research Center for Artificial Intelligence, Multilinguality and Language Technology Lab
Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany
[2]Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences
Sonnenfelsgasse 19, 1010 Vienna, Austria
[3]Hochschule Darmstadt, University of Applied Sciences
Max-Planck-Str. 2, 64807 Dieburg, Germany
declerck@dfki.de, Lenka.Bajcetic@oeaw.ac.at, melanie.siegel@hda.de

Abstract

We describe on-going work consisting in adding pronunciation information to wordnets, as such information can indicate specific senses of a word. Many wordnets associate with their senses only a lemma form and a part-of-speech tag. At the same time, we are aware that additional linguistic information can be useful for identifying a specific sense of a wordnet lemma when encountered in a corpus. While work already deals with the addition of grammatical number or grammatical gender information to wordnet lemmas, we are investigating the linking of wordnet lemmas to pronunciation information, adding thus a speech-related modality to wordnets.

Keywords: Wordnet, Pronunciation, OntoLex-Lemon

1. Introduction

Wordnets are well-established lexical resources with a wide range of applications. For more than twenty years they have been elaborately set up and maintained by hand, especially the original Princeton WordNet of English (PWN) (Miller, 1995; Fellbaum, 1998). In recent years, there have been increasing activities in which open wordnets for different languages have been automatically extracted from various resources and enriched with lexical semantics information, building the so-called Open Multilingual Wordnet (OMW) (Bond and Paik, 2012). OMW brings together wordnets in different languages, harmonizing them in a uniform tabular format that lists synsets IDs and the associated lemmas, and linking them to PWN (Bond and Foster, 2013; Bond et al., 2016). Additionally, XML versions of LMF and *lemon* representations[1] of the data are provided.

A starting motivation for our work was to investigate if and how specific Wordnet senses can be restricted to what appears to be morphological variations of a lexical entry. The question touched also the issue on how to encode this information. (Gromann and Declerck, 2019) describe a first experiment done for English, looking at specific Princeton WordNet senses associated with word forms that look like regular plural forms of a lexical entry, but which rather need to be considered as separate lexical entries, due to the specific sense(s) they carry. And PWN is indeed introducing plural forms as "lemmas" in its inventory, when those are related to specific synsets. An example of this is given by the WordNet entry "silks" with the sense of "the brightly colored garments of a jockey; emblematic of the stable", which is distinct from the synsets associated to the two singular form entries included in PWN.[2]

The work described in the present article is an extension of recent experiments done in linking wordnets with additional lexical and morphological information, including grammatical number in the case of PWN (Gromann and Declerck, 2019), grammatical number and grammatical gender in the case of a German lexical semantics resource (Declerck et al., 2019) and of wordnets for Romance languages that are included in OWN (Racioppa and Declerck, 2019). In this context, we note that the Dutch WordNet was from its beginning including full lexical information for a large number of its entries (Vossen et al., 2008; Postma et al., 2016).

In the present work, we investigate the linking of pronunciation information to wordnets, dealing first with the German language. The pronunciation information is extracted from the corresponding German edition of Wiktionary.[3]

2. Pronunciation as Indicator of Senses

We are aware that different senses of a word, also within a shared part-of-speech category, can be marked by a distinctive pronunciation, like for example for the German substantive "Boot" (in IPA[4] notation [buːt]: *boot*) versus "Boot" ([boːt]: *boat*).[5] This phenomenon, also called heteronymy, can be relevant for a variety of speech-based ap-

[1]LMF stands for "Lexical Markup Framework", an ISO standard. See (Francopoulo et al., 2006) and `http://www.lexicalmarkupframework.org/` for more details. *lemon* stands for "LExicon MOdel for oNtologies". See (McCrae et al., 2012) and `https://lemon-model.net/` for more details.

[2]This information is retrieved from the PWN Web interface, accessible at `http://wordnetweb.princeton.edu/perl/webwn`.

[3]See `https://www.wiktionary.org/` and for the German edition `https://de.wiktionary.org/wiki/Wiktionary:Hauptseite`.

[4]IPA stands for "International Phonetic Alphabet". See `https://www.internationalphoneticassociation.org/content/ipa-chart` for more details.

[5]The pronunciation information is taken from `https://de.wiktionary.org/wiki/Boot`.

plications. Therefore, this type of information should be added to wordnets, so that they can help to disambiguate words in spoken utterances.

We need to make this linking of Wordnet entries to pronunciation information explicit, and for this we are adapting the approach described in (Racioppa and Declerck, 2019), and which is dealing with the linking of Wordnet lemmas to morphological information. We thus again chose the OntoLex-Lemon model (Cimiano et al., 2016)[6] as the representation formalism, since this model has proven to be able to accommodate both "classical" lexicographic descriptions (McCrae et al., 2017) as well as lexical semantics networks like wordnets (McCrae et al., 2014).

In the next sections, we give first some background description on the extraction of pronunciation information from Wiktionary sources. We continue with a section on OntoLex-Lemon, followed by a section that describes how OntoLex-Lemon supports the linking of lemmas in wordnets resources to pronunciation information.

3. Extracting Pronunciation Data from Wiktionary

It has been shown that the access and use of Wiktionary can be helpful in a series of Natural Language Processing (NLP) applications. (Kirov et al., 2016), for example, describe work to extract and standardize data contained in Wiktionary and to make it available for a range of NLP tasks, while the authors focus on extracting and normalizing a huge number of inflectional paradigms across a large selection of languages. This effort contributed to the creation of the UniMorph data (`http://unimorph.org/`). The UniMorph project was focusing on (scraping) the HTML representation of Wiktionary (mostly the English version, but also looking at other language editions). (Metheniti and Neumann, 2018) and (Metheniti and Neumann, 2020) describe a related approach, but making use of a combination of the HTML pages and the underlying XML dump of the English edition of Wiktionary, which is covering also 4,050 other languages, some of them with a very low number of entries.[7] The English edition of Wiktionary has of today a number of 6,262,000 pages, whereas 734,130 pages are dealing with English words.

BabelNet[8] is also integrating Witkionary data,[9] with a focus on sense information, in order to support, among others, word sense disambiguation and tasks dealing with word similarity and sense clustering (Camacho-Collados et al., 2016).

Many language specific editions of Wiktionary contain also pronunciation information, mostly encoded with the help of the IPA notation. (Jouvet et al., 2011) show that pronunciation information encoded in (the French edition of) Wiktionary can be "used efficiently for building a pronunciation lexicon for a speech transcription system". (Schlippe et al., 2010) assess the quality of pronunciation information in Witkionary for four languages (English, French, German, and Spanish) and come to satisfying results, especially in the case of French, when it comes to the evaluation of the coverage and also to the impact on automatic speech recognition (ASR) systems, especially in the case of Spanish. Those already older studies comforted us in the opinion that extracting pronunciation information from Wiktionary can deliver a relevant source of data for our experiment consisting in equipping wordnets with pronunciation information.

4. Extracting Pronunciation Information from the German Edition of Wiktionary

We display in Figure 1 below as an example the pronunciation information for the German substantive "Januar" (*january*) as represented in the XML dump of the German edition of Wiktionary.[10] As the reader can see, the

```
{{Aussprache}}

:{{IPA}} {{Lautschrift|ˈjanuaːɐ̯}}

:{{Hörbeispiele}} {{Audio|De-Januar.ogg}}
```

Figure 1: The Wiktionary markup encoding of the pronunciation of the German word "Januar" (*january*).

information on the pronunciation is encoded in the wiki markup language, and the element names are in German ("Aussprache" standing for *pronunciation*, "Lautschrift" for *phonetic script* and "Hörbeispiele" for *audio samples*). This means that for every language edition of Wiktionary a specific script has to be written for extracting the desired information. Also the use of the wiki markup is not consistent across language editions, so that the scripts have also to be adapted for dealing with the various templates in use in the different language editions.

A first version of our extraction program allowed us to detect a (provisional, as the extraction script can still be improved) list of 150 German substantives that have two or more pronunciations.[11] We are extending this list to other categories, also looking for words belonging to more than one category, as for example "modern" (adjective, [moˈdɛʁn], *modern*) versus "modern" (verb, [ˈmoːdn], *moulder*). But this cross-categories extension is less relevant, as wordnets would anyway introduce different lemmas for a word belonging to distinct categories.

An example of a German substantive having two different pronunciations is "Vollzug", with the stress put either at the

[6] See also `https://www.w3.org/2016/05/ontolex/` for more details.

[7] A possibly tentative list of entries in the different languages contained in the English Wiktionary is given here: `https://en.wiktionary.org/wiki/Special:Statistics?action=raw`.

[8] See (Navigli and Ponzetto, 2010) and `https://babelnet.org/`.

[9] As far as we are aware of, BabelNet integrates only the English edition of Wiktionary, but includes all the languages covered by this edition.

[10] XML dumps of the various editions of Wiktionary are available at `https://dumps.wikimedia.org/backup-index.html`.

[11] In parallel, we are extracting a list of German substantives that have different genders (502 entries detected) or different plural forms (440 entries detected), each with specific senses.

beginning or at the end of the word, as shown in Figure 2 and Figure 3, which are displaying screen shots from the Wiktionary page, and where the reader can see the meanings (encoded as the values of the key word "Bedeutungen") associated with the distinct pronunciations.[12]

Aussprache:
IPA: ['fɔl,ʦuːk]
Hörbeispiele: Vollzug (Info)

Bedeutungen:
[1] (U- oder S-)Bahn-Garnitur, welche die übliche Länge aufweist
[2] Güterzug, der mit Fracht beladen ist

Figure 2: The German word " Vollzug" in Wiktionary, with the meanings of *train set* and *charged freight train*.

Aussprache:
IPA: [fɔl'ʦuːk]
Hörbeispiele: Vollzug (Info)
Reime: -uːk

Bedeutungen:
[1] Umsetzung in die Tat, das Ausführen
[2] *kurz für:* Strafvollzug
[3] Einrichtung, in der Verurteilte ihre Freiheitsstrafe absitzen

Figure 3: The German word " Vollzug" in Wiktionary, with the meanings of *execution* [1] and *enforcement*, *penal system*, *prison* [2],[3].

Our internal representation for the pronunciation information, together with the associated meanings, extracted from the XML dump of Wiktionary is displayed in Figure 4. This is the type of data to be linked to synsets for German, making use for this of the OntoLex-Lemon representation model.

```
Vollzug
  ['fɔl't͡su:k']
    ["Umsetzung in die Tat, das Ausführen
       \n:2 ''kurz für:'' Strafvollzug
       \n:3  Einrichtung, in der Verurteilter|
             Verurteilte ihre Freiheitsstrafe
             absitzen\n\n{"]
  ['ˈfɔlˌt͡su:k']
    ['(U- oder S-)Bahn-Garnitur, welche die
             übliche Länge aufweist
       \n:2 Güterzug, der mit Fracht beladen
             ist\n\n{']
```

Figure 4: Our internal representation of the extracted pronunciation information, with the associated meanings, from Wiktionary for the word "Vollzug".

5. OntoLex-Lemon

OntoLex-Lemon is a further development of the "Lexicon Model for Ontologies" (*lemon*) (McCrae et al., 2012). Both *lemon* and the OntoLex-Lemon model, which is resulting from a W3C Community Group,[13] were originally developed with the aim to provide a rich linguistic grounding for ontologies, meaning that the natural language expressions used in the labels, definitions or comments of ontology elements are equipped with an extensive linguistic description.[14] This rich linguistic grounding includes the representation of morphological and syntactic properties of lexical entries as well as the syntax-semantics interface, i.e. the meaning of these lexical entries with respect to an ontology or to specialized vocabularies.

The main organizing unit for those linguistic descriptions is the *LexicalEntry* class, which enables the representation of morphological patterns for each entry (a multi word expression, a word or an affix). The connection of a lexical entry to an ontological entity is marked mainly by the denotes property or is mediated by the *LexicalSense* or the *LexicalConcept* classes, as this is represented in Figure 6, which displays the core module of the model.

A major difference between *lemon* and OntoLex-Lemon is that the latter includes an explicit way to encode conceptual hierarchies, using the SKOS[15] standard. As can be seen in Figure 6, lexical entries can be linked via the *ontolex:evokes* property to such SKOS concepts, which can represent Wordnet synsets. This structure is paralleling the relation between lexical entries and ontological resources, which is implemented either directly by the *ontolex:reference* property or mediated by the instances of the *ontolex:LexicalSense* class.

As can be seen in Figure 6, there is a property called `ontolex:phoneticRep` which is introduced for the class `ontolex:Form`. This property is used in the model for representing the pronunciation information, which is thus encoded at the level of morphological forms and not at the level of lexical entries, as this is shown in Figure 5 for the example entry "privacy":

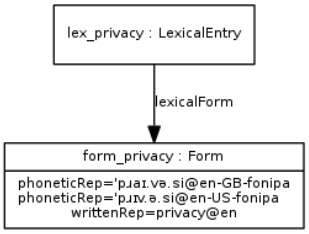

Figure 5: The graphical representation of the place of the "ontolex:phoneticRep" property in the OntoLex-Lemon model. Taken from `https://www.w3.org/2016/05/ontolex/#forms`

[12] `https://de.wiktionary.org/wiki/Vollzug`.

[13] See `https://www.w3.org/2016/05/ontolex/`

[14] See (McCrae et al., 2012) and (Cimiano et al., 2016).

[15] SKOS stands for "Simple Knowledge Organization System". SKOS provides "a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary" (`https://www.w3.org/TR/skos-primer/`).

More recently, OntoLex-Lemon has been used also as a de-facto standard in the field of digital lexicography and is being applied for example in the European infrastructure project ELEXIS (European Lexicographic Infrastructure).[16]

Our present goal is to integrate synsets, lemmas, morphological and pronunciation descriptions in the extended ontological framework specified by OntoLex-Lemon. Updating also past work on mapping some wordnets onto the former *lemon* model (McCrae et al., 2014). This work was done following the guidelines[17] for mapping Global WordNet formats onto *lemon*-based RDF.[18]

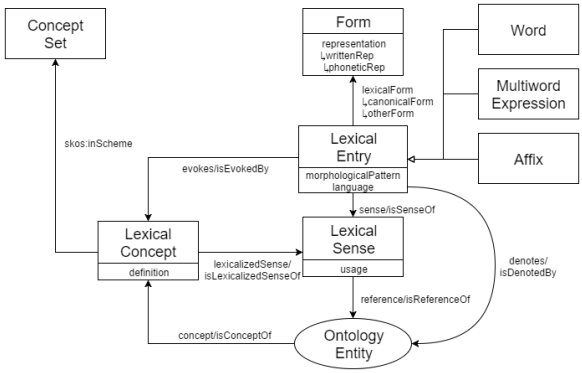

Figure 6: The core module of OntoLex-Lemon. Graphic taken from `https://www.w3.org/2016/05/ontolex/`.

6. The integrated Encoding in OntoLex-Lemon

We display in code listing 1 the (still tentatve) way we can express the phonetic restriction for a sense of an OdeNet[19] concept that points to the word "Vollzug".

Listing 1: The OntoLex-Lemon representation of the OdeNet synset for the concept associated with *Vollzug* pointing to all listed entries senses and a corresponding form

```
:synset_odenet-2345-n
  rdf:type ontolex:LexicalConcept ;
  wn:ili ili:i41311 ;
  skos:inScheme :OdeNet ;
  ontolex:isEvokedBy :entry_w10755 ;
  ontolex:isEvokedBy :entry_w11251 ;
  ontolex:isEvokedBy :entry_w11252 ;
  ontolex:isEvokedBy :entry_w11253 ;
  ontolex:isEvokedBy :entry_w11254 ;
  ontolex:isEvokedBy :entry_w11255 ;
  ontolex:isEvokedBy :entry_w11256 ;
  ontolex:isEvokedBy :entry_w11257 ;
  ontolex:isEvokedBy :entry_w11258 ;
  ontolex:isEvokedBy :entry_w11259 ;
  ontolex:isEvokedBy :entry_w11260 ;
  ontolex:isEvokedBy :entry_w7091 ;
  ontolex:lexicalizedSense
              :sense_w10755_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11251_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11252_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11253_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11254_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11255_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11256_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11257_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11258_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11259_2345-n ;
  ontolex:lexicalizedSense
              :sense_w11260_2345-n ;
  ontolex:lexicalizedSense
              :sense_w7091_2345-n ;
.

:entry_w11258
  rdf:type ontolex:Word ;
  wn:partOfSpeech wn:noun ;
  ontolex:canonicalForm :form_w11258 ;
  ontolex:evokes :synset_odenet-2345-n ;
  ontolex:evokes :synset_odenet-3815-n ;
  ontolex:sense :sense_w11258_2345-n ;
  ontolex:sense :sense_w11258_3815-n ;
.

:sense_w11258_2345-n
  rdf:type ontolex:LexicalSense ;
  ontolex:isLexicalizedSenseOf
              :synset_odenet-2345-n ;
  ontolex:isSenseOf :entry_w11258 ;
  lexicog:restrictedTo
        :form_w11258_Restriction_2 .
.

:form_w11258
  rdf:type ontolex:Form ;
  ontolex:writtenRep "Vollzug"@de ;
.

:form_w11258_Restriction_2
  rdf:type ontolex:Form ;
  ontolex:phoneticRep "vollZUG"@de ;
.
```

The most important part of this encoding is the property `lexicog:restrictedTo` added to the one Lex-

[16] See `http://www.elex.is/` for more detail.

[17] See `https://globalwordnet.github.io/schemas/#rdf`.

[18] RDF stands for "Resource Description Framework". See `https://www.w3.org/RDF/` for more details.

[19] "OdeNet" stands for "Open-de-WordNet". See (Declerck et al., 2019) for more info on OdeNet, a lexical semantics resource for German. The original resource (still under development) can be downloaded here: `https://github.com/hdaSprachtechnologie/odenet`.

icalSense that is relevant in our case. This property has been defined in a recent extension to the core module of OntoLex-Lemon: the "lexicog" module, which has been developed for covering specific aspects of Lexicography.[20] We then introduce a specific object called "form_w11258_Restriction_2", which encodes for the `:form_w11258` the special case of the second pronunciation for "Vollzug", as displayed in Figure 3.[21] This way we can not only add pronunciation information to wordnets, but also express the restriction that a specific meaning is dependant on a specific pronunciation.

7. Conclusion

We described work in progress consisting in adding pronunciation information to wordnets, as this information can be very relevant in making wordnets usable for sense disambiguation in speech applications. Using for this purpose the OntoLex-Lemon model allows us not only to encode this linking from original wordnets to pronunciation information extracted from Wiktionary dictionaries, but this supports also the possibility to express restrictions on senses, stating that a specific sense can be only selected in case a specific pronunciation is given.

8. Acknowledgements

This paper is supported by the Horizon 2020 research and innovation programme with the projects Prêt-à-LLOD (grant agreement no. 825182) and ELEXIS (grant agreeement no. 731015). It is also partially based upon work from COST Action CA18209 - NexusLinguarum "European network for Web-centred linguistic data science". We also thank the anonymous reviewers for their helpful comments.

9. Bibliographical References

Bond, F. and Foster, R. (2013). Linking and extending an open multilingual wordnet. In *ACL (1)*, pages 1352–1362.

Bond, F. and Paik, K. (2012). A survey of wordnets and their licenses. *Small*, 8(4):5.

Bond, F., Vossen, P., McCrae, J. P., and Fellbaum, C. (2016). Cili: the collaborative interlingual index. In *Proceedings of the Global WordNet Conference*, volume 2016.

Bosque-Gil, J., Lonke, D., Gracia, J., and Kernerman, I. (2019). Validating the OntoLex-lemon lexicography module with K Dictionaries' multilingual data. In *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference.*, pages 726–746, Brno, Czech Republic, October. Lexical Computing CZ s.r.o.,.

Camacho-Collados, J., Pilehvar, M. T., and Navigli, R. (2016). Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence*, 240:36–64.

[20] See `https://www.w3.org/2019/09/lexicog/` and (Bosque-Gil et al., 2019) for more details.

[21] We mark with capital letters the fact that the stress is on the second part of the word.

Cimiano, P., McCrae, J. P., and Buitelaar, P. (2016). Lexicon Model for Ontologies: Community Report.

Declerck, T., Siegel, M., and Gromann, D. (2019). Ontolex-lemon as a possible bridge between wordnets and full lexical descriptions. In Christiane Fellbaum, et al., editors, *Proceedings of the Tenth Global Wordnet Conference*, pages 264–271, wyb. Stanisława Wyspiańskiego 27 50-370 Wrocław Poland, 7. Oficyna Wydawnicza Politechniki Wrocławskiej, Oficyna Wydawnicza Politechniki Wrocławskiej.

Christiane Fellbaum, editor. (1998). *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.

Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M., and Soria, C. (2006). Lexical markup framework (lmf). In *International Conference on Language Resources and Evaluation-LREC 2006*, page 5.

Gromann, D. and Declerck, T. (2019). Towards the detection and formal representation of semantic shifts in inflectional morphology. In Maria Eskevich, et al., editors, *2nd Conference on Language, Data and Knowledge (LDK)*, volume 70 of *OpenAccess Series in Informatics (OASIcs)*, pages 21:1–21:15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 5.

Jouvet, D., Fohr, D., and Illina, I. (2011). Building a Pronunciation Lexicon for a Speech Transcription System from Wiktionary Pronunciations only. In *XIV International Conference "Speech and Computer" (SPECOM'2011)*, Kazan, Russia, September.

Kirov, C., Sylak-Glassman, J., Que, R., and Yarowsky, D. (2016). Very-large scale parsing and normalization of wiktionary morphological paradigms. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

McCrae, J., de Cea, G. A., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., and Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*, 46(6):701–709.

McCrae, J. P., Fellbaum, C., and Cimiano, P. (2014). Publishing and linking wordnet using lemon and rdf. In *Proceedings of the 3 rd Workshop on Linked Data in Linguistics*.

McCrae, J. P., Buitelaar, P., and Cimiano, P. (2017). The OntoLex-Lemon Model: Development and Applications. In Iztok Kosem, et al., editors, *Proceedings of eLex 2017*, pages 587–597. INT, Trojína and Lexical Computing, Lexical Computing CZ s.r.o., 9.

Metheniti, E. and Neumann, G. (2018). Wikinflection: Massive semi-supervised generation of multilingual inflectional corpus from wiktionary. In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018)*, Linköping Electronic Conference Proceedings. Linköping University Electronic Press, Linköpings universitet, 12.

Metheniti, E. and Neumann, G. (2020). Wikinflection corpus: A (better) multilingual, morpheme-annotated inflectional corpus. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*. LREC.

Miller, G. A. (1995). Wordnet: A lexical database for english. *COMMUNICATIONS OF THE ACM*, 38:39–41.

Navigli, R. and Ponzetto, S. P. (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, July. Association for Computational Linguistics.

Postma, M., van Miltenburg, E., Segers, R., Schoen, A., and Vossen, P. (2016). Open dutch wordnet. In *Proceedings of the Eight Global Wordnet Conference*, Bucharest, Romania.

Racioppa, S. and Declerck, T. (2019). Enriching open multilingual wordnets with morphological features. In Raffaella Bernardi, et al., editors, *Proceedings of the Sixth Italian Conference on Computational Linguistics*. CEUR, 10.

Schlippe, T., Ochs, S., and Schultz, T. (2010). Wiktionary as a source for automatic pronunciation extraction. In *11th Annual Conference of the International Speech Communication Association, Makuhari, Japan*. Interspeech 2010.

Vossen, P., Maks, E., Segers, R., and van der Vliet, H. (2008). Integrating lexical units, synsets and ontology in the cornetto database. In European Language Resources Association (ELRA), editor, *Proceedings of LREC 2008, Marrakech*.