The Global Wordnet Formats: Updates for 2020

John P. McCrae[♣], **Michael Wayne Goodman,**[♢] **Francis Bond,**[♢]
Alexandre Rademaker,[♠] **Ewa Rudnicka**[♡] and **Luís Morgado da Costa**[♢]
[♣] Data Science Institute, NUI Galway, Ireland
[♢] Nanyang Technological University, Singapore
[♠]IBM Research and FGV/EMAp, Brazil
[♡]Wrocław University of Science and Technology, Poland
john@mccr.ae, goodmami@uw.edu, bond@ieee.org
alexrad@br.ibm.com, ewa.rudnicka@pwr.edu.pl, lmorgado.dacosta@gmail.com

Abstract

The Global Wordnet Formats have been introduced to enable wordnets to have a common representation that can be integrated through the Global WordNet Grid. As a result of their adoption, a number of shortcomings of the format were identified, and in this paper we describe the extensions to the formats that address these issues. These include: ordering of senses, dependencies between wordnets, pronunciation, syntactic modelling, relations, sense keys, metadata and RDF support. Furthermore, we provide some perspectives on how these changes help in the integration of wordnets.

1 Introduction

The introduction of the Global WordNet Grid (Vossen et al., 2016) and the Collaborative Interlingual Index (Bond et al., 2016) presented a need for greater compatibility between individual wordnet projects through a common format for the representation of wordnets. As such the Global WordNet Association introduced a format with several serialization methods[1] that have been used by several projects, including the new open English WordNet (EWN; McCrae et al., 2020, 2019), the Open Multilingual Wordnet (OMW; Bond and Foster, 2013) and the Wn Python library (Goodman and Bond, 2021). Along with the increased adoption came the perception of shortcomings in the format as it was initially defined, such as the inability to capture all of the information present in Princeton WordNet (PWN; Miller, 1995; Fellbaum, 2012) or to capture some key information that other projects wished to use in their modelling. It was therefore deemed necessary to extend the model and, for this reason, we have introduced a new extended version (v1.1) of the format that covers some of these use cases.

In this paper, we describe the model as a reference for users and then describe the extensions that have been made to the format. In particular, there are several main areas that have been improved. Firstly, the order of words in a synset was not being captured explicitly within the model, which was information that was present in PWN, but did not have a clear semantics. Secondly, as many projects build on other projects, either by adding new information (McCrae et al., 2017) or by translating an existing wordnet, it was felt that it was important to capture the dependencies between projects. In addition, pronunciation information was something that some wordnets have or are in the process of adding, so modelling for this was added. Furthermore, we have added some new semantic relations (mainly inspired by plWordNet (Piasecki et al., 2009).

Finally, there were some technical issues to do with the modelling of syntactic behaviours, and while the current formats could capture the information, they did so in a way that was quite verbose and lead to bloated files. In addition, we fixed a few minor issues related to the representation of lexicographer files, sense keys and metadata.

2 Background

The Global WordNet Association's formats are a common data model with three(-plus) serializations in XML, JSON and various RDF formats.[2] The XML format is based on the Lexical Markup Framework (Francopoulo et al., 2006) and in particular on the version developed in the Kyoto project (Soria and Monachini, 2008). This represents the wordnet as a `LexicalResource` with a number of `Lexicons`, one for each language, along with multiple metadata elements about the lexicon, including identifiers, version, language, license, contact email,

[1] https://globalwordnet.github.io/schemas

[2] Any RDF serialization is valid, but for this paper we consider the Turtle form of RDF.

citation, etc. The format splits the data into two distinct elements, the `LexicalEntry`, which contains the syntactic information about the usage of individual words, and the `Synset`, which provides the semantic information about the synset and the relations to other synsets.

```
<LexicalEntry id="ex-rabbit-n">
  <Lemma writtenForm="rabbit"
         partOfSpeech="n"/>
  <Sense id="ex-rabbit-n-1"
         synset="ex-s1"/>
</LexicalEntry>

<Synset id="ex-s1">
  <Definition>Example
    definition</Definition>
  <SynsetRelation
    relType="hypernym"
    target="ex-s2"/>
</Synset>
```

The JSON schema is very close to this and is defined by means of both a JSON Schema description and also a JSON-LD context, that means that it can be easily interpreted as an RDF file as well. The same example in JSON is rendered as follows:

```
{
  "entry": [{
    "@id": "ex-rabbit-n",
    "lemma": {
      "writtenForm": "rabbit" },
    "partOfSpeech": "noun",
    "sense": [{
      "@id": "ex-rabbit-n-1",
      "synset": "ex-s1"
    }]
  }],
  "synset": [{
    "@id": "ex-s1",
    "definition": [{
      "gloss":
"Example definition"
    }],
    "relations": [{
      "relType": "hypernym",
      "target": "ex-s2"
    }]
  }]
}
```

The RDF version of this as serialized in Turtle is very similar and uses the OntoLex-Lemon vocabulary (Cimiano et al., 2016) to express most of the elements along with a small wordnet specific ontology that is published at `https://globalwordnet.github.io/schemas/wn`.

```
<#ex-rabbit-n>
  a ontolex:LexicalEntry ;
  ontolex:canonicalForm [
    ontolex:writtenRep "rabbit"@en
  ] ;
  wn:partOfSpeech wn:noun ;
  ontolex:sense <#ex-rabbit-n-1> .

<#ex-rabbit-n-1>
  a ontolex:LexicalSense ;
  ontolex:reference <#ex-s1> .

<#ex-s1>
  a ontolex:LexicalConcept ;
  wn:definition [
    rdf:value "Example definition"@en
  ] .

[] vartrans:source <#ex-s1> ;
 vartrans:category wn:hypernym ;
 vartrans:target <#ex-s2> .
```

3 Updates to the WordNet Schemas

3.1 Ordering within/of Synsets

The ordering of words in a synset and, correspondingly, the order of synsets (senses) of a word can be used to model the relative importance of synsets and words. While many wordnets do not systematically order their senses or synsets, the ordering is something that we would like the format to be able to capture. The latter issue, the order of senses of a word, has been captured in the XML by means of the order of the `<Sense>` tags. However, the converse information was being lost in the format. Resources such as the open English WordNet were preserving this by means of encoding it within the sense identifiers with a new attribute `members`. In the following example, we see that the order of the senses of 'rabbit' is `ex-synset-1` followed by `ex-synset-2`, while the order of the lemmas in `ex-synset-1` is 'rabbit', 'bunny'.

```
<LexicalEntry id="ex-rabbit-n">
  <Lemma writtenForm="rabbit"
         partOfSpeech="n"/>
  <Sense id="ex-rabbit-n-1"
    synset="ex-synset-1"/>
```

```
  <Sense id="ex-rabbit-n-2"
    synset="ex-synset-2"/>
</LexicalEntry>
<Synset id="ex-synset-1"
  members="ex-rabbit-n
           ex-bunny-n"/>
```

3.2 WordNet Dependencies and Extensions

Most wordnets are not built in isolation, but expect and depend upon the entities and relationships of other wordnets. We acknowledge two categories of such dependencies: concept-relation dependencies and lexicon extensions. The first is for wordnets built by the *expand* methodology (Vossen, 1998) whereby lexical entries and senses in the new language are defined around the concept structure of a larger wordnet which is almost always PWN. The second is for supplementary resources that build on top of an existing wordnet, for instance to add new lexical entries, senses, synsets, or relations. As this is a new feature, we are keeping it simple and only allowing monotonic effects. Destructive extensions that, for instance, remove entries or senses from a lexicon or selective dependencies that exclude certain relations are left to future work.

3.2.1 Concept-Relation Dependencies

Wordnets included in the Open Multilingual Wordnet (OMW; Bond and Foster, 2013) are linked together using CILI IDs. This linking allows for cross-lingual searches and the sharing of wordnet structure through synset relations, but it also means that most wordnets, particularly smaller ones, are dependent on the others for their structure. This approach works well when the OMW is taken as a holistic, multilingual resource but, as it is left implicit which structure-providing wordnets are required, it is not straightforward to use a wordnet in isolation of the full OMW, e.g., for experimental purposes. This issue is even more pronounced for wordnets that are not included in the OMW. What we need, then, is a way for a wordnet to specify what other resources are required, much as how software projects specify their dependencies. We therefore introduce a new `Requires` element which selects the `id` and `version` attributes of an external lexicon that should be loaded along with the current lexicon for it to behave as expected. For example, the following specifies that the Japanese Wordnet (Isahara et al., 2008; Bond et al., 2008) depends on the PWN for its synset relations:

```
<Lexicon id="wnja" id="2.0">
  <Requires id="pwn" version="3.0"/>
</Lexicon>
```

The purpose is to declare what, exactly, is required so that an application that hosts the wordnets can signal to the user if dependencies are unmet, or to limit the wordnets that may be used when traversing external synset relations. It is left implicit which elements or kinds of elements from the external wordnet become available to the dependent wordnet but, following the OMW's behaviour, an application may choose to only allow synset relations and not, say, synsets or lexical entries. The `Requires` declarations are not only for expand-wordnets, but whenever a lexicon wants to reuse synset relations from another, as discussed in the next section.

3.2.2 Lexicon Extensions

A lexicon extension is an augmentation of an existing resource. For instance, someone may want to publish an extension providing domain-specific jargon, a list of common misspellings, or neologisms that may soon fall out of use (McCrae et al., 2017). An extension could even just provide additional relations between synsets. These entries and relations may not be a good fit for inclusion in the primary project, or perhaps the release cadence of the project is too slow for the user to wait for the entries to be added to the wordnet.

These situations would be well-served by the use of a partial wordnet that could be loaded alongside the primary wordnet and queried together. Unlike the concept-relation dependencies described in Section 3.2.1 where linking was implicit through the CILI, extensions require mechanisms for linking into the actual structures of a resource. Therefore we introduce a new lexicon element, `LexiconExtension`, which is similar to the `Lexicon` element of LMF, but requires an `Extends` element which specifies the `id` and `version` of the lexicon it extends. Under a `LexiconExtension`, lexical entries and synsets can be defined as normal, but in order to link them with primary wordnet through sense or synset relations, we need to introduce the identifiers of the external entities.[3] For these, we allow `ExternalLexicalEntry`, `ExternalSense`, and `ExternalSynset` elements. In addition to estab-

[3]This requirement is partially just to satisfy XML validators, but can also serve as a check on the dependent lexicon's assumptions about the structure of the primary wordnet.

lishing IDs for linking, these elements allow for augmenting the elements themselves, such as for adding senses to an existing lexical entry or relations to a synset. However, these elements do not allow one to change information in the provider wordnet, so lemmas on lexical entries, ILIs on synsets, and other required information may not be specified on the corresponding external elements.

For example, the Geonames Wordnet (Bond and Bond, 2019) provides additional synset relations on top of the PWN as well as an extended lexical hierarchy of location names in the PWN and many other wordnets. The extension would specify that it extends the PWN as follows:

```
<LexiconExtension id="geonames-pwn"
                  version="1.0">
  <Extends id="pwn" version="3.0"/>
</LexiconExtension>
```

In some cases it might make sense to use both the `Extends` and `Requires` elements. For instance, if we want to extend the Japanese Wordnet with its entries from the Geonames Wordnet and reuse the relations from the English Geonames extension, we could specify the relationships as follows:

```
<LexiconExtension id="geonames-wnja"
                  version="1.0">
  <Extends id="wnja" version="2.0"/>
  <Requires id="geonames-pwn"
            version="1.0"/>
</LexiconExtension>
```

3.3 Pronunciation

One of the extensions that has been requested by other projects (Declerck et al., 2020) is the ability to represent phonetic information giving the pronunciation of lemmas in a schema such as the International Phonetic Alphabet. As well as giving the IPA text, it was also desired that we should be able to provide information about the specific variety, as well as further notes about the form of the pronunciation. In addition, we want to indicate whether the transcription is phonemic or phonetic, that is whether it includes expected features of the language such as aspiration. For 'variety', we decided to support the use of IETF language tags to indicate dialect, for example encoding British English in IPA as `en-GB-fonipa`, and an additional `notes` field that can encode further information such as indicating a particular British English dialect. We also added a field allowing a URL to give an audio file of the word being pronounced. An example of encoding is given below:

```
<LexicalEntry id="ex-rabbit-n">
  <Lemma writtenForm="rabbit"
         partOfSpeech="n"/>
    <Pronunciation
     variety="en-GB-fonxsamp
              en-US-fonxsamp">
     'r\{bIt</Pronunciation>
    <Pronunciation
     variety="en-AU-fonxsamp"
     notes="weak vowel merger">
     'r\{b@t</Pronunciation>
  </Lemma>
</LexicalEntry>
```

3.4 Syntactic Behaviours

One weakness of the current format was that the representation of syntactic behaviours was quite verbose and required that all of the information about the syntactic behaviour was repeated for each entry. This meant that, even for simple generic frames like transitive verbs, you would have a different frame for each entry. We changed this with the current version by allowing each frame to appear only once at the lexicon-level and have an identifier which can be referenced by individual senses. For example:

```
<Lexicon id="ex">
  <LexicalEntry id="ex-play-n">
    <Lemma writtenForm="play"
      partOfSpeech="n"/>
    <Sense id="ex-play-n-1"
      subcat="transitive"/>
    <Sense id="ex-play-n-1"
      subcat="transitive
        intransitive-with"/>
  </LexicalEntry>
  <SyntacticBehaviour
    id="transitive"
    subcategorizationFrame=
  "Somebody ----s something"/>
  <SyntacticBehaviour
    id="intransitive-with"
    subcategorizationFrame=
  "Somebody ----s with something"/>
</Lexicon>
```

3.5 New Relations

The original inventory of semantic relations (between synsets) and sense relations (between senses) were mainly drawn from the Princeton WordNet (PWN) and Euro WordNet (EWN) (Fellbaum, 1998; Vossen, 1998). Up-to-date documentation of these resources is available at `https://globalwordnet.github.io/gwadoc/`. This is important as there have been changes in the interpretation of the meanings of particular relations over the lifetime of the various projects, and of course between projects. By maintaining documentation through one of the Global Wordnet Association working groups, we hope to keep it up-to-date. To keep it accessible we use a version control system to store the documentation, and release it it under an open license, rather than in journals, books and technical documentation.

However, there are some relations used in several wordnets not currently in our inventory. In order to make the resource more useful across languages, we propose to add some of them. They are listed in Table 1. All of these are used in the innovative plWordNet project (Piasecki et al., 2009) and many of them in other projects as well.

The first two relations are to do with aspect. Most Slavic languages have two forms for most verbs: perfective and imperfective, and these are linked with the `simple_aspect` relation. This is the same as the "pure aspect" in plWordNet where the two members of a "pure" aspectual pair are located in distinct synsets with no change in meaning (Piasecki et al., 2009). The Bulgarian Wordnet (BulNet) also marks these pairs as different synsets, but links them to common hypernyms (Koeva, 2008). Apart from pure aspectual pairs, many Slavic languages have other productive verb alternations rendered by the addition of various prefixes. plWordNet groups them under a common label "secondary aspect". To represent these we would like to include `secondary_aspect` relation. In order to show the direction, the actual relations will be in pairs: `simple_aspect_ip` "simple aspect, imperfective to perfective" and `simple_aspect_pi` "simple aspect, perfective to imperfective", and similarly for `secondary_aspect`.

The next five are for specific relations, normally derivational in Slavic languages. PWN marks these relations (where they exist) as `hyponym`. plWordNet and BulNet specialize `feminine_form`, `young_form`, `diminutive` and `augmentative`. The Czech wordnet also suggested two relations here `X_HAS_MALE` and `X_HAS_FEMALE` (Pala and Smř, 2004). Although these relations are relatively rare in English (we estimate around a hundred), in plWordNet there are almost 10,000 of these (mainly feminine form and diminutives)![4] For the masculine, feminine and young relations, we wish to capture both derivative relations like ***prince/princess*** but also purely semantic ones (like ***king/queen*** or ***kangaroo/joey***). For this reason we allow them both at the sense level (when there is a dervational relation) and the synset level.

Because some wordnets use these as sense relations and some as synset relations, we propose to allow them for both. Here we will also have two forms of each: e.g., `female` and `has_female`.

Next we propose to introduce three specializations of `antonym`. These are used in the plWordNet, but we follow the naming convention of Saeed (2009). The first, and most common, is gradable antonyms. Then there are simple antonyms (also known as complementary or binary antonyms) where the negative of one entails the positive of the other. Finally we add `converse`: these are those which describe a relationship between two entities from different points of view. Piasecki et al. (2009) argues that the converse relation is different enough from the other antonyms that it should be kept separate. However, linguists such as Saeed (2009) consider converse to be antonymy and in other wordnets, such as PWN, converses are treated as antonyms, so we decided to group them together.

The last relation we introduce is inter-register synonymy (`ir_synonym`), introduced by Maziarz et al. (2015). This is for synsets where the denotation is the same, but the connotation is different, for example for informal terms or honorific variants. This is a very common relation: there are over 12,000 examples of these in the plWordNet. Antonyms and synonyms are reflexive: they are their own reverse relation.

3.6 Other improvements

3.6.1 Lexicographer files and sense keys

One concern was with the modelling of lexicographer files and sense keys. These two aspects are part of the development of Princeton WordNet and it is not clear how many other wordnet projects use them. For the lexicographer files, it was previously recommended that they be modelled using Dublin Core (Weibel and Koch, 2000) metadata properties, in particular with the 'subject' property. It was de-

[4] `http://plwordnet.pwr.wroc.pl/wordnet/stats`

Relation	Example	Lang
`simple_aspect`	***czytać*** "read/be reading (habitual/progressive)" → ***przeczytać*** "have read"	pl
`secondary_aspect`	***kopać*** "dig/be digging" → ***nakopać*** "have dug out a lot of sth"	pl
`female`	***pig*** → ***sow***	en
`male`	***pig*** → ***boar***	en
`young`	***pig*** → ***piglet***	en
`diminutive`	***pig*** → ***piggy***	en
`augmentative`	**дом** "house" → **домище** "great house"	ru
`anto_gradable`	***hot*** ↔ ***cold***, ***warm*** ↔ ***cool***	en
`anto_simple`	***complete*** ↔ ***incomplete***	en
`anto_converse`	***wife*** ↔ ***husband***, ***employer*** ↔ ***employee***	en
`ir_synonym`	***money*** ↔ ***dough***, ***loot*** 《informal》,	en
	食べる *taberu* "eat"↔ 召し上がる *meshiagaru* "honored person eats" 《honorific》	ja

Table 1: Proposed new relations
Examples are in English (en), Japanese (ja), Polish (pl) and Russian (ru).

cided that for the 1.1 version of the schema,[5] we should allow a special property for these values that can be used by Princeton and other projects that make use of lexicographer files. The second issue was that the sense keys used in Princeton WordNet are sometimes used to map between other wordnets. This is problematic, as the principal methodshould be through the InterLingual Index and the sense IDs are limited to particular senses of PWN. This issue principally came from English WordNet, which mapped back to Princeton WordNet using sense keys represented in another Dublin Core property (in this case 'identifier'). The English WordNet project is now removing its own sense key schema and using sense identifiers that correspond in a one-to-one manner with Princeton identifiers. In a few cases, the sense keys have had to been changed, due to either changes of spelling in a lemma, changing part-of-speech from satellite to head adjective or changes in the structure of the wordnet. For these cases, we recommend the use of a stand-off annotation to provide mapping if it is necessary.[6]

3.6.2 Metadata improvements

Metadata about elements is an important part of the schema and as such we allowed any Dublin Core property to be represented. It was noted that the XML format we published did not follow the Dublin Core recommendations, in that it specified that the properties should be attributes, rather than independent elements. In order to maintain backwards compatibility, we updated the namespace for Dublin Core to one on our repository so that there is no issue with clashing XML schemas, while not leading to any need for users of the schema to update the data except for the XML header. In addition, a further metadata property was added for projects to give a logo that can be displayed on the Open Multilingual Wordnet.

3.6.3 Further RDF schemas

Following the increasingly popular way of addressing the issue of interoperability, the use of Linked Data and Semantic Web standards such as RDF and OWL (McGuinness et al., 2004) have led to the emergence of a number of Linked Data projects for lexical resources (De Melo, 2015; Cimiano et al., 2020). The adoption of such standards not only allows both the data model and the actual data to be published in the same format, they also provide for instant compatibility with a vast range of existing data processing tools and storage systems, triple stores, providing query interfaces based on the SPARQL standard (W3C SPARQL Working Group, 2013).

To encode any data in RDF, one needs to decide which classes and properties (vocabulary) will be used. The adoption of already defined vocabularies helps with data interoperability since these makes data easily integrate with other resources.

The first RDF vocabulary for wordnets encoding proposed by Van Assem et al. (2006) was based on Princeton WordNet 2.0. Their work includes

[5] Princeton WordNet's schema cannot be used directly as a sense ID, due to the '%' character

[6] English WordNet's file is at `https://github.com/globalwordnet/english-wordnet/blob/master/src/sensekey-maps.csv`

(1) a mapping of WordNet 2.0 concepts and data model to RDF/OWL; (2) conversion scripts from the WordNet 2.0 Prolog distribution to RDF/OWL files; and (3) the actual WordNet 2.0 data. The suggested representation stayed as close to the original source as possible, that is, it reflects the original WordNet data model without interpretation. The WordNet schema proposed by Van Assem et al. (2006) has three main classes: *Synset*, *WordSense* and *Word*. The first two classes have subclasses for each lexical group present in WordNet. Each instance of Synset, WordSense and Word has its own URI, sharing the same prefix, a project-specific namespace. Another RDF vocabulary for wordnet encoding is the already cited OntoLex-Lemon vocabulary (Cimiano et al., 2016).

Since Van Assem et al. (2006) was based on Princeton WordNet 2.0, its use required adaptations. The first decision was regarding the URIs. Each wordnet project should have their own base URIs (namespace) for instances of synsets, senses and words. Second, additional relations were added in the RDF vocabulary available at `https://github.com/globalwordnet/schemas`. In RDF, the support the interoperability between wordnets (see Section 3.2) is very natural. For instance, a synsets of a particular wordnet can be connected to any other wordnet synset instances through *owl:sameAs* relations, estabilishing the mapping. That is the approach adopted in the OpenWordnet-PT (de Paiva et al., 2012). The code for converting Princeton WordNet 3.0 database files to RDF following this vocabulary is provided at `https://github.com/own-pt/wordnet2rdf`.

4 Discussion and Future Work

In order for wordnets to continue to grow, we have to allow for changes in their structure. In the past, each project has gone ahead on its own, which has led to divergence, with similar changes being implemented in slightly different ways. Through the release of a community-driven schema, we can help to harmonise the various projects. This should also lead to the development of interoperable tools, allowing for more rapid development.

Ideally, we do not just want to make the new format available, but to help projects take advantage of it. For example, the open English WordNet may wish to specify its antonym links using the three types (simple, gradable, converse) from plWordNet. We can use the CILI to suggest these changes automatically.

We would also like to help grow a collection of wordnets available in the new format, both through the Open Multilingual Wordnet or as individual wordnets and extensions.

The formats we are proposing fit well with the standardisation initiatives that are on-going around the representation of lexicographic data. As described in this paper we take advantage of both the Lexical Markup Framework (Francopoulo et al., 2006), being developed by ISO as well as the OntoLex model (Cimiano et al., 2016) from the W3C. In addition, we are looking at other standardisation efforts such as the LEXIDMA model[7] from the OASIS standardisation body. We are also aware of and taking account of other formats and tools in use in the community including DebVisDic (Horák et al., 2006), WordNetLoom (Piasecki et al., 2013) and Mill.[8]

5 Conclusion

The formats proposed by the Global WordNet Association have already been adopted by some projects and this has provided valuable feedback on the quality. We have found that the open methodology we have adopted has allowed us to quickly address these changes (with some spirited debate). The changes that we have made should ensure that the format continues to be useful and relevant and helps in the integration of wordnets through the collaborative interlingual index.

Acknowledgements

This work is supported by the EU H2020 programme under grant agreements 731015 (ELEXIS - European Lexicographic Infrastructure). John McCrae is also supported by a research grant from Science Foundation Ireland, co-funded by the European Regional Development Fund, for the Insight Centre under Grant Number SFI/12/RC/2289. Ewa Rudnicka is supported by the CLARIN-PL project, which is part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education. Documentation for the semantic relations was supported by the Google Season of Docs (2020).[9]

[7] `https://www.oasis-open.org/committees/lexidma`

[8] `https://github.com/own-pt/mill/`

[9] `https://developers.google.com/season-of-docs/`

References

Francis Bond and Arthur Bond. 2019. GeoNames wordnet (GeoWN): extracting wordnets from GeoNames. In *Proceedings of the 11th Global Wordnet Conference (GWC 2019)*.

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual Wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362, Sofia, Bulgaria. Association for Computational Linguistics.

Francis Bond, Hitoshi Isahara, Kyoko Kanzaki, and Kiyotaka Uchimoto. 2008. Boot-strapping a WordNet using multiple existing WordNets. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Francis Bond, Piek Vossen, John P. McCrae, and Christiane Fellbaum. 2016. CILI: the Collaborative Interlingual Index. In *Proceedings of the Global WordNet Conference 2016*.

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

Philipp Cimiano, John P. McCrae, and Paul Buitelaar. 2016. Lexicon Model for Ontologies: Community Report.

Gerard De Melo. 2015. LexVo.org: Language-related information for the linguistic linked data cloud. *Semantic Web*, 6(4):393–400.

Thierry Declerck, Lenka Bajcetic, and Melanie Siegel. 2020. Adding pronunciation information to wordnets. In *Proceedings of the LREC 2020 Workshop on Multimodal Wordnets (MMW2020)*, pages 39–44.

Christiane Fellbaum. 2012. Wordnet. *The encyclopedia of applied linguistics*.

Christine Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Gil Francopoulo, Monte George, Nicoletta Calzolari, Monica Monachini, Nuria Bel, Mandy Pet, and Claudia Soria. 2006. Lexical markup framework (lmf). In *International Conference on Language Resources and Evaluation - LREC 2006*.

Michael Wayne Goodman and Francis Bond. 2021. Intrinsically interlingual: The Wn Python library for wordnets. In *11th International Global Wordnet Conference (GWC2021)*. (this volume).

Aleš Horák, Karel Pala, Adam Rambousek, and Martin Povolnỳ. 2006. Debvisdic–first version of new client-server wordnet browsing and editing tool. In *Proceedings of the Third International WordNet Conference-GWC*, pages 325–328.

Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. 2008. Development of the Japanese WordNet. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Svetla Koeva. 2008. Derivational and Morphosemantic Relations in Bulgarian Wordnet. *Intelligent Information Systems*, XVI:359–389.

Marek Maziarz, Maciej Piasecki, Stan Szpakowicz, and Joanna Rabiega-Wisniewska. 2015. Semantic relations among nouns in Polish WordNet grounded in lexicographic and semantic tradition. *Cognitive Studies*, 11:161–182.

John P. McCrae, Alexandre Rademaker, Francis Bond, Ewa Rudnicka, and Christiane Fellbaum. 2019. English WordNet 2019 – An Open-Source WordNet for English. In *Proceedings of the 10th Global WordNet Conference – GWC 2019*.

John P. McCrae, Ian Wood, and Amanda Hicks. 2017. The Colloquial WordNet: Extending Princeton WordNet with Neologisms. In *Proceedings of the First Conference on Language, Data and Knowledge (LDK2017)*, pages 194–202.

John Philip McCrae, Alexandre Rademaker, Ewa Rudnicka, and Francis Bond. 2020. English WordNet 2020: Improving and Extending a WordNet for English using an Open-Source Methodology. In *Proceedings of the Multimodal Wordnets Workshop at LREC 2020*, pages 14–19.

Deborah L McGuinness, Frank Van Harmelen, et al. 2004. OWL web ontology language overview. W3c recommendation, World Wide Web Consortium.

George A Miller. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. 2012. Openwordnet-pt: An open Brazilian Wordnet for reasoning. In *Proceedings of COLING 2012: Demonstration Papers*, pages 353–360, Mumbai, India. The COLING 2012 Organizing Committee. Published also as Techreport http://hdl.handle.net/10438/10274.

Karel Pala and Pavel Smř. 2004. Building Czech WordNet. *Romanian Journal of Information Science and Technology*, 7:79–88.

Maciej Piasecki, Michał Marcińczuk, Radosław Ramocki, and Marek Maziarz. 2013. Wordnetloom: a wordnet development system integrating form-based and graph-based perspectives. *International Journal of Data Mining, Modelling and Management*, 5(3):210–232.

Maciej Piasecki, Stan Szpakowicz, and Bartosz Broda. 2009. *A Wordnet from the Ground Up*. Wroclaw University of Technology Press. (ISBN 978-83-7493-476-3).

John I. Saeed. 2009. *Semantics*, 3 edition. Wiley-Blackwell.

Claudia Soria and Monica Monachini. 2008. Kyoto-LMF. Wordnet representation format. *KYOTO Working Paper WP02_TR002_V4_Kyoto_LMF*.

Mark Van Assem, Aldo Gangemi, and Guus Schreiber. 2006. Conversion of WordNet to a standard RDF/OWL representation. In *Language Resource and Evaluation Conference (LREC)*, pages 237–242.

Piek Vossen, editor. 1998. *Euro WordNet*. Kluwer.

Piek Vossen, Francis Bond, and John P. McCrae. 2016. Toward a truly multilingual Global Wordnet Grid. In *Proceedings of the Global WordNet Conference 2016*.

W3C SPARQL Working Group. 2013. SPARQL 1.1 overview. Technical report, World Wide Web Consortium.

Stuart L Weibel and Traugott Koch. 2000. The Dublin core metadata initiative. *D-Lib Magazine*, 6(12):1082–9873.