

English WordNet 2020: Improving and Extending a WordNet for English using an Open-Source Methodology

John P. McCrae¹, Alexandre Rademaker², Ewa Rudnicka³, Francis Bond⁴

¹ Insight Centre for Data Analytics, NUI Galway, john@mccr.ae

² IBM Research and FGV/EMAP, alexrad@br.ibm.com

³ Wroclaw University of Science and Technology, ewa.rudnicka@pwr.edu.pl

⁴ Nanyang Technological University, bond@ieee.org

Abstract

The Princeton WordNet, while one of the most widely used resources for NLP, has not been updated for a long time, and as such a new project English WordNet has arisen to continue the development of the model under an open-source paradigm. In this paper, we detail the second release of this resource entitled “English WordNet 2020”. The work has focused firstly, on the introduction of new synsets and senses and developing guidelines for this and secondly, on the integration of contributions from other projects. We present the changes in this edition, which total over 15,000 changes over the previous release.

Keywords: WordNet, lexicons, open source, lexicography, NLP

1. Introduction

English WordNet (McCrae et al., 2019) is a fork of Princeton WordNet (Fellbaum, 2010; Miller, 1995), which aims to further the development of a wordnet for English. Wordnets are one of the most widely used resources in natural language processing¹ and as the English language is not static, it is necessary to continually update the resource so that it remains relevant for these tasks. Wordnets group words into sets of synonyms, called *synsets*, and each *sense* of a word corresponds to its membership in one synset. These synsets are then organized in a graph containing relationships such as *hypernym/hyponym* (broader or narrower), *antonym* and many more relations. The English WordNet has taken an open-source policy for this and the resource is available on GitHub,² where anyone can contribute to its development. A first release of this resource was made in 2019, although this release had some limitations in terms of the changes it made, in particular no new synsets were created in that release. In this paper, we describe the 2020 release, which provides a more thorough revision of the resource, including new synsets from other resources including Colloquial WordNet (McCrae et al., 2017), enWordNet (Rudnicka et al., 2015) and Open Multilingual WordNet (Bond and Paik, 2012). We also discuss some of the major challenges that we have encountered during the development of this resource. In particular, as the resource has introduced a large number of new synsets, we have had to develop guidelines for the significance of synsets to be included in the resource. We have also looked into making clearer guidelines for making sense distinctions between two meanings of the same word, as this seems to be a significant challenge for those who build systems on top of WordNet. Finally, we look at some of the challenges that we wish to address in the next year of the development, of which the most pressing is the adjective hierarchy, which is less dense and contains many unclear sense distinctions, as well as issues related to

improving the procedure for development of the resource, in particular with the format and the issue of ensuring backwards compatibility with Princeton WordNet.

2. Development Methodology

2.1. Open Source Development

Add Relation	5	Change Relation	18
Definition	44	Example	8
Delete Synset	8	New Synset	30
Synset Duplicate	32	Synset Member	19
Synset Split	1	Enhancements	8
Contribution	3	Bug	9

Table 1: The number of issues by type addressed in this release

English WordNet is based on an open source methodology and as such anyone can contribute to the development of this resource. We have developed a methodology as described previously (McCrae et al., 2019), that relies on **issues** and **pull requests** in order to manage requests for changes. While, there have been relatively few pull requests made directly to the project (in fact only 3 in the last year), issues have proven to be an effective method by which requests can be logged. In total 161 issues were created asking for changes in the WordNet, that have been closed as part of this release. The number of each type of issue is given in Table 1 where they are categorized according to the following scheme:

Add Relation A relation should be added between two synsets;

Change Relation A relation is of the wrong type or has the wrong target;

Definition The definition of a synset should be updated;

Example The examples of a synset should be updated;

¹The original paper has over 13,000 citations

²<https://github.com/globalwordnet/english-wordnet>

Delete Synset A synset represents a concept that should be removed from WordNet. There are few reasons for this: the concept cannot be found in any other reference material and there is no corpus evidence for its members; the synset refers to a compositional meaning; the synset exists twice in the wordnet;

New Synset A synset covering a new concept is being proposed;

Synset Duplicate Two synsets are not possible to distinguish or refer to the same concept. This is fixed by either creating a new concept for all synsets or by deleting all but one of the duplicates;

Synset Split A synset refers to two distinct concepts and should be split into two new synsets;

Synset Member A word in a synset should be added or removed;

Enhancement A request for an improvement in the tooling around English WordNet or for a new kind of data;

Contribution Issues related to large external contributions (see Section 3.);

Bug A technical flaw that needs to be addressed in the data files.

Once these issues have been logged then a solution is proposed by one of the team members and a pull request is made, and then accepted. The process is designed to give high visibility to the changes proposed in the wordnet (which has helped to detect minor errors) and to provide tracking so that the discussion and implementation can be easily connected through Git. This means that the changes are all well-documented and as such could easily be taken up by other projects or included back into Princeton WordNet.

2.2. Guidelines for new synsets

As this version of English WordNet has introduced new synsets it has been necessary to formalize the guidelines for the introduction of new synsets. These guidelines attempt to formalise best practices from Princeton WordNet and other projects and they are based on the principle that new synsets should be added with some caution. In fact, they are much stricter than the current set of synsets that are derived from Princeton WordNet, thus if applied retroactively would lead to the removal of many existing synsets, and is not planned for the foreseeable future.

We have defined five basic criteria that a new synset is required to pass before being introduced into the wordnet: (1) Significance; (2) Non-compositionality; (3) Distinction; (4) Well-defined; and (5) Linked.

2.2.1. Significance

A concept in English WordNet should be significant, this means that it should be possible to easily find *at least 100 examples* of the usage of the word with this meaning. This

can be done by using a search interface such as Sketch Engine³ or other corpus search interface. For future releases, we aim to integrate corpora tools into the GitHub instance. In the case that a new sense of an existing word is being proposed, then it should be possible to propose collocates that occur with this sense of the word and these can be used to find and distinguish examples.

English WordNet is a dictionary not an encyclopedia. For this reason, it should not contain long lists of people, places, organizations, etc. Proper nouns are generally not expected to be included in the resource and many kinds of common nouns for narrow domains or geographical usage should not be included, examples of this would include elements of different cuisines around the world. As a rule of thumb, if there is a Wikipedia page for this concept it should not be in English WordNet.⁴ For future releases a more complete alignment of the resource and Wikipedia is planned based on previous works (De Melo and Weikum, 2009; McCrae, 2018) to address the introduction of synsets already well-described in Wikipedia.

2.2.2. Non-compositionality

One of the goals of English WordNet is to support annotation. If a word (or multiword expression) is already covered by English WordNet it should not be added.

For multiword expressions (MWE), this means that the meaning of the term should not be derivable from its components, e.g., “French Army” could be tagged with the synsets for “French” and “Army”; in contrast “operational system” refers not to a system that is operational, but it is a computer science term for the system that runs on every computer. Another case of MWE is the conventionalized ones. Conventionalization refers to the situation where a sequence of words that refer to a particular concept is commonly accepted in such a way that its constituents cannot easily be substituted for near-synonyms, because of some cultural or historical conventions (Farahmand et al., 2015). Consider the expression “geologic fault”. It is compositional but no one would consider substituting it with “geologic defect”. There are many types of MWE and a extensive literature about them (Sag et al., 2002), here we just want to emphasize that expressions that could have their parts annotated with senses already in the resource don’t need to be explicitly added.

For single words, the word should not be derived in a systematic manner, these include:

- Converting a verb to a noun or adjective by adding ‘-ing’ or ‘-ed’
- Converting an adjective to an adverb by adding ‘-ly’
- Productive prefixes such as ‘non-’, ‘un-’
- Systematic polysemy: e.g., using a part to refer to a whole, for example: “congress” meaning the “members of congress”

³<http://sketchengine.eu>

⁴There is no plan to apply this retroactively to existing synsets at the moment

2.2.3. Distinction

The concept should be distinct from other concepts in the WordNet and care should be taken to check relevant synonyms. For each word in the synset, the sense should thus be distinct as described above. This is best considered in terms of a substitution check, e.g., “happy” and “felicitous” are synonyms, `ewn-01052105-s` and the examples can be substituted, e.g., “a happy life”/“a felicitous outcome”. This does not mean that they can be substituted in every sense, e.g., “happy to help” but not *“felicitous to help”.

2.2.4. Well-defined

It should be possible to easily write a definition for this concept that is distinct from other concepts in English WordNet. A good definition consists of a *genus* and a *differentia*.

Genus The type of the thing, often the hypernym,

Differentia Something that makes this word unique

An example of a good definition is:

a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs

Where a poor definition would be:

a piece of furniture

used for eating

In addition an example should be provided with a link to a website where the example is used as follows:

```
<Synset id="ewn-...">
...
<Example dc:source=
"https://en.wikipedia.org/wiki/Example.com">
The example domains have one subdomain
name defined in the Domain Name System
</Example>
...
</Synset>
```

2.2.5. Linked

The synset should be possible to link into the graph, more specifically:

Nouns A hypernym must be identified

Verbs A hypernym, entailment, cause or antonym must be identifier. Verbs should also have at least one subcategorization frame.

Adjectives They should be marked as similar to a non-satellite adjective (in which case they are satellites) **or** antonyms of a non-satellite adjective **or** hypernyms of an adjective

Adverbs No clear guidelines but at least one link should be proposed. Ideally a link for the corresponding adjective via `derivation` relation.

The more links that can be provided the better a synset is.

2.2.6. Sense keys and lexicographer files

Two key design aspects that are derived from Princeton WordNet are the use of lexicographer files and sense keys, however the changes in the development procedure for English WordNet (as opposed to Princeton WordNet) have made into necessary to update how these elements are used. English WordNet is divided into a number of source files that correspond to the original lexicographer files in WordNet, but are now in XML. New synsets proposed from issues should be assigned to one of these lexicographer files as they are created. For contributed resources (see below), we merged them into the original resource according to the hypernym.

Sense keys were a mechanism that provided stability between releases of WordNet, and sense keys were (mostly) stable identifiers between different versions of Princeton WordNet. Instead, English WordNet has adopted the CILI interlingual index (Vossen et al., 2016; Bond et al., 2016) as the principal method of providing cross-version stability. Moreover, for new senses the calculation of stable sense identifiers is complicated as the Princeton WordNet formula relied on information in lexicographer files that is no longer present. Initial proposals were just to jettison sense keys, however community feedback has encouraged the creation of new methodology for assigning sense keys.⁵ In addition, we now also track the changes of sense keys, caused for example changes in the spelling of a lemma or if a sense has been moved across lexicographer files.

2.2.7. Sense distinctions

One particular issue that has been common in the reported set of issues is the issue of sense distinction. WordNet has been criticized (Palmer et al., 2004; Snow et al., 2007) for a long time for issues related to its sense granularity. As such, there have been many issues claiming that synsets are duplicated as the meanings are quite hard to distinguish. In order to simplify these decisions, we have developed a few key principles that help us in distinguishing senses.⁶

Ontological Typing Two synsets that have difficult to distinguish senses may typically occur in different parts of the WordNet graph. This can often make the distinction clearer than the definitions as the two synsets refer to ontological distinct aspects. For example, for ‘rock’,⁷ the definitions were not clear however the structure clearly gave away that the two senses referred to ‘rock’ as a material and ‘rock’ as a physical object, that is the first sense was uncountable and the second countable.

Collocations Following methods in word sense induction (Klapaftis and Manandhar, 2008; Denkowski, 2009), one clear rule for distinguishing two senses is the existence of collocations that cannot be applied to both senses. We aim in the future to extend this basic principal with some quantitative scoring function that can help us in distinguishing senses based on corpus information. For example, ‘rock’ collocated with

⁵Issue #157.

⁶Track with Issue #243

⁷Issue #135

‘concert’ suggests a very different sense to a collocation with ‘metamorphic’.

Other dictionaries The final method we use for deciding whether to make a sense distinction is to look at other dictionaries. In cases, where a very subtle distinction is being discussed often comparison with other dictionaries can help to decide these issues.

3. Integration of Existing Resources

3.1. Colloquial WordNet

POS	Lemmas	Synsets
Noun	196	195
Verb	75	79
Adjective	34	36
Adverb	5	5
Total	310	315

Table 2: New synsets and lemmas introduced by Colloquial WordNet by part of speech

Colloquial WordNet (McCrae et al., 2017) was a resource developed to extend wordnet with recent slang terms. The resource included a number of changes that would not be in line with the existing wordnet, although may be later included as these features are added to the mainstream of wordnet. These include the marking of non-referential expressions (such as “ah!” or “haha”), the sense linking from a multiword expression to the senses of its individual words and the mark of words as loanwords from other languages. Once these had been removed the resource was integrated, which is relatively simple as the Colloquial WordNet uses the same format as English WordNet. However the new synsets introduced by this wordnet were given 8-figure numeric codes much like in the existing wordnet. As these cannot be based on the offset in a file, instead they were assigned based on the original identifiers with a code starting 90 or 91. For example, ‘adulting’⁸ is code `ewn-900004011-n`.

3.2. Open Multilingual WordNet

The Open Multilingual WordNet (Bond and Paik, 2012; Bond and Foster, 2013) project has also introduced new synsets and made changes related to the English WordNet. We are in the process of integrating these changes, one of the most major changes is the rewriting of definitions so as to ensure uniqueness. This change affects 1,673 synsets and most of these changes directly improve the definitions as given, for example, ‘Thai’ was previously defined as ‘a branch of the Tai languages’ and is now defined as ‘a branch of the Tai languages, spoken in central Thailand, centered in Bangkok’.

In addition, there are a large number of changes that introduce new synsets, mostly to cover concepts that are not already in the wordnet. We are currently in the process of identifying these changes and integrating those that meet the guidelines for new synsets.

⁸“acting like an adult”

3.3. enWordNet

The plWordNet team at Wrocław University of Science and Technology has also developed a number of extensions of the English WordNet (Zaśko-Zielińska and Piasecki, 2018; Dziob and Piasecki, 2018; Janz et al., 2017) to cover concepts not currently covered in English WordNet. We are integrating these changes into our format. In total, the enWordNet (as of version 4.0) has proposed 7,656 new synsets, however our analysis quickly deduced that many of them consist of concepts that are easily found in Wikipedia and are defined by sections of text copied from Wikipedia. We automatically reduced the set of proposed changes to 2,084 synsets by applying the guidelines in Section 2.2., in particular by looking for lemmas that match existing Wikipedia page titles. We then conducted a manual review of this, we found that 1,843 out of 2,084 (88.4%) synsets were of acceptable quality to be introduced in English WordNet. This represents a large part of the changes that have been made in the 2020 release.

4. Open Challenges

4.1. Satellite Adjectives

As previously discussed, sense distinctions have been an important difficulty in the development of the resource. For adjectives, most of these issues have not yet been solved as the structure of adjectives in WordNet is currently quite suboptimal. In particular, English WordNet distinguishes between two kinds of adjectives: ‘head’ adjectives and satellite adjectives. Head adjectives should have an antonym relation to another head adjectives, which satellite adjectives should be marked as similar to a head adjective; this is called the ‘dumbbell’ model. The distinction is made at the part-of-speech level in the resource, although no other part-of-speech catalogue or dictionary to our knowledge makes the distinction this way.⁹ This means that there is often fewer links to other synsets and also shorter definitions; in fact adjectives typically have 1.44 synset links against a general average of 2.43. The plan for a future version, is to revamp the adjective so that they follow a more conventional classification such as that proposed by (McCrae et al., 2014), where the formal categories are:

Intersective These refer to properties that the adjective indicates the presence of. The most significant group of these are pertainyms, which mean that a concept is of or pertaining to a noun, e.g., “French” pertaining to “France”. The existing pertainym relation marks many of these but can be expanded.

Gradable These adjectives refer to the value of a property on some scale, for example ‘hot’ is on a scale of ‘temperature’, a new property relating adjectives to their scales will be introduced and this will replace the ‘dumbbell’ model.

Operator This group will capture that final set of adjectives that have a meaning that modifies the meaning

⁹This is even though more widely-accepted distinction such as postpositive adjectives are distinguished at the sub-part-of-speech level

of the noun, such as ‘former’. We will look into new properties that could be introduced to help with connecting these concepts in the WordNet graph.

4.2. Format

The English WordNet is currently published under the GWA XML format, however there have been a number of issues related to this, most principally that the format is quite verbose as is typical for XML. Moreover, we have found that some aspects of the data contained in the original Princeton WordNet are not possible to represent in this format.

There are two proposals for moving from the XML format:¹⁰ the first is to stick with the GWA model but use a less verbose serialization (namely YAML) and reduce the amount of information represented in the dictionary files. The second option is to adopt the model presented in Muniz et al. (2018) and being investigated by the OpenWordnet-PT project (Paiva et al., 2012),¹¹ as there is a large amount of work already carried out with this model, however it is a non-standard serialization and due to its brevity it can be difficult to understand for those not used to it.

In addition, there are a number of problems related to the representation of existing data from Princeton WordNet, these include the morphosemantic relations that are provided in a second stand-off file¹² and these can be easily included in the main resource simply by extending the set of relations that are available in the WordNet. The next issue is related to adjective position,¹³ which was not captured in the previous release and cannot be encoded in the XML format as the part-of-speech categories are a closed group. We have added this as a new attribute, `adjposition`, on the `LexicalEntry` tags in the resource. Finally, there were some verb example sentences,¹⁴ that were not being captured. This was after it was discovered that the previous release was using the wrong file to generate the syntactic behaviour of the entries. As such, this will be added as new examples on the corresponding synset with an extra attribute to say that they were generated by sentence templates.

4.3. Backwards Compatibility

One of the key goals of the model is to ensure that there is backwards compatibility with previous Princeton WordNet releases, and as such, although the project has moved to the XML format entirely, we still make releases in the previous WNDB format. This leads to a number of issues, most notably that synset identifiers are based on file offsets in this format. In particular, as we do not wish to recalculate the identifiers used in the XML files at every release the identifiers in the WNDB release will not correspond to those in the XML. This is further exacerbated by the introduction of new synsets, whose identifier is set to be high enough

that it cannot correspond to a byte offset in the file. In addition, there have been a number of issues related to sense identifiers that have been improved in this release to provide more continuity for users of English WordNet in the WNDB format.

Finally, as the license of WordNet is unique to Princeton WordNet, we are moving to use a Creative Commons Attribution license to protect the changes made on top of Princeton WordNet. As the underlying resource (Princeton WordNet) has its own bespoke license, it is necessary to reproduce both licenses when deriving resources from English WordNet.¹⁵

4.4. Distributed model

It is not clear how domain-specific or goal-specific wordnets (such as the Colloquial WordNet) should be incorporated or linked to the English Wordnet. Regarding the data format, a linked-open data format such as RDF could help us in the definition of global identifiers (URI) that could help on the link of entities in different resources. But this is part of the problem, the maintainance of the links and the track of changes on these resources can be far from trivial. On the other way, incorporating domain-specific or goal-specific wordnets into English Wordnet would make the resource maintainance even harder with increasing difficult on the definition of guidelines such as the ones explained above.

5. Changes in 2020 Release

The total number of changes are detailed in Table 3, and as can be seen the largest number of changes are firstly to do with the definitions. This is due to the contribution of many new definitions from Colloquial WordNet and enWordNet and secondly, to do with the many changes proposed by the Open Multilingual WordNet project. Secondly, we see a large number of new lemmas and synsets proposed by both Colloquial WordNet and enWordNet, representing the largest number of changes. As many of these are single nouns whose lemma does not already occur in WordNet, the majority of the changes result in one new synset, one new lemma, one new sense and two more synset relations (typically a hypernym and a hyponym). While much effort has gone into the directly reported issues, most of these result in only small changes to the structure of the wordnet. We also see a lot of changes in the senses, this is primarily due to the change in the representation of adjective categories (e.g., postpositivity) as discussed above.

6. Conclusion

English WordNet is continuing to grow and meet the annual release schedule, to ensure that an up-to-date and accurate WordNet is available for the many users of WordNet in natural language processing. The open-source methodology that has been adopted has been generally successful so far and has provided impetus for the development of clear guidelines that are easy-to-follow. In this paper, we have discussed guidelines for new synsets and senses and detailed some of the open challenges that we are looking into, including the structuring of adjectives.

¹⁰Discussion is to be found at [Issue #31](#)

¹¹This format is called [Mill](#)

¹²[Issue #132](#)

¹³[Issue #180](#)

¹⁴[Issue #245](#) definition in [WNDB](#)

¹⁵[Issue #144](#)

	Princeton WordNet 3.1	English WordNet 2019	English WordNet 2020	Changed
Synsets	117,791	117,791	120,054	
Lemma	159,015	159,789	163,079	
Senses	207,272	208,353	211,864	
Synset Relations	285,668	285,666	291,299	
Sense Relations	92,535	92,535	92,526	
Definitions	117,791	117,791	120,059	1,587
Examples	47,539	48,419	49,675	151

Table 3: Comparative size of Princeton WordNet 3.1 and English WordNet 2019 and 2020

Acknowledgements

This work is supported by the EU H2020 programme under grant agreements 731015 (ELEXIS - European Lexical 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.

7. Bibliographical References

- Bond, F. and Foster, R. (2013). Linking and extending an open multilingual wordnet. Sofia.
- Bond, F. and Paik, K. (2012). A survey of wordnets and their licenses. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, Matsue. 64–71.
- Bond, F., Vossen, P., McCrae, J. P., and Fellbaum, C. (2016). CILI: the Collaborative Interlingual Index. In *Proceedings of the Global WordNet Conference 2016*.
- De Melo, G. and Weikum, G. (2009). Towards a universal wordnet by learning from combined evidence. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 513–522.
- Denkowski, M. (2009). A survey of techniques for unsupervised word sense induction. *Language & Statistics II Literature Review*, pages 1–18.
- Dziob, A. and Piasecki, M. (2018). Implementation of the verb model in plWordNet 4.0. In *Proceedings of the 9th Global Wordnet Conference*, pages 114–123.
- Farahmand, M., Smith, A., and Nivre, J. (2015). A multiword expression data set: Annotating non-compositionality and conventionalization for English noun compounds. In *Proceedings of the 11th Workshop on Multiword Expressions*, pages 29–33.
- Fellbaum, C. (2010). WordNet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.
- Janz, A., Kocoń, J., Piasecki, M., and Zaśko-Zielińska, M. (2017). plWordNet as a basis for large emotive lexicons of Polish. *Proceedings of Human Language Technologies as a Challenge for Computer Science and Linguistics Poznań: Fundacja Uniwersytetu im. Adama Mickiewicza w Poznaniu*, pages 189–193.
- Klapaftis, I. P. and Manandhar, S. (2008). Word sense induction using graphs of collocations. In *ECAI*, pages 298–302.
- McCrae, J. P., Unger, C., Quattri, F., and Cimiano, P. (2014). Modelling the Semantics of Adjectives in the Ontology-Lexicon Interface. In *Proceedings of 4th Workshop on Cognitive Aspects of the Lexicon*.
- McCrae, J. P., Wood, I., and Hicks, A. (2017). The Colloquial WordNet: Extending Princeton WordNet with Neologisms. In *Proceedings of the First Conference on Language, Data and Knowledge (LDK2017)*.
- McCrae, J. P., Rademaker, A., Bond, F., Rudnicka, E., and Fellbaum, C. (2019). English WordNet 2019 – An Open-Source WordNet for English. In *Proceedings of the 10th Global WordNet Conference GWC 2019*.
- McCrae, J. P. (2018). Mapping WordNet Instances to Wikipedia. In *Proceedings of the 9th Global WordNet Conference*.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.
- Muniz, H., Chalub, F., Rademaker, A., and de Paiva, V. (2018). Extending wordnet to geological times. In *Global Wordnet Conference 2018*, Singapore, January.
- Paiva, V. d., Rademaker, A., and Melo, G. d. (2012). OpenWordNet-PT: An open Brazilian WordNet for reasoning. Technical report, COLING 2012.
- Palmer, M., Babko-Malaya, O., and Dang, H. T. (2004). Different sense granularities for different applications. In *Proceedings of the 2nd International Workshop on Scalable Natural Language Understanding (ScaNaLU 2004) at HLT-NAACL 2004*, pages 49–56.
- Rudnicka, E., Witkowski, W., and Kaliński, M. (2015). Towards the Methodology for Extending Princeton WordNet. *Cognitive Studies*, 15(15):335–351.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). Multiword expressions: A pain in the neck for nlp. In *International conference on intelligent text processing and computational linguistics*, pages 1–15. Springer.
- Snow, R., Prakash, S., Jurafsky, D., and Ng, A. Y. (2007). Learning to merge word senses. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (emnlp-conll)*, pages 1005–1014.
- Vossen, P., Bond, F., and McCrae, J. P. (2016). Toward a truly multilingual Global Wordnet Grid. In *Proceedings of the Global WordNet Conference 2016*.
- Zaśko-Zielińska, M. and Piasecki, M. (2018). Towards emotive annotation in plWordNet 4.0. In *Proceedings of the 9th Global Wordnet Conference, Singapore*, pages 154–163.