Toward a truly multilingual Global Wordnet Grid

Piek Vossen
VU University Amsterdam
Amsterdam, The Netherlands
piek.vossen@vu.nl

Francis Bond
Nanyang Technological University,
Singapore
bond@ieee.org

John P. McCrae
Insight Centre for Data Analytics
NUI Galway
Galway, Ireland
john@mccr.ae

Abstract

In this paper, we describe a new and improved Global Wordnet Grid that takes advantage of the Collaborative InterLingual Index (CILI). Currently, the Open Multilingal Wordnet has made many wordnets accessible as a single linked wordnet, but as it used the Princeton Wordnet of English (PWN) as a pivot, it loses concepts that are not part of PWN. The technical solution to this, a central registry of concepts, as proposed in the EuroWordnet project through the InterLingual Index, has been known for many years. However, the practical issues of how to host this index and who decides what goes in remained unsolved. Inspired by current practice in the Semantic Web and the Linked Open Data community, we propose a way to solve this issue. In this paper we define the principles and protocols for contributing to the Grid. We tested them on two use cases, adding version 3.1 of the Princeton WordNet to a CILI based on 3.0 and adding the Open Dutch Wordnet, to validate the current set up. This paper aims to be a call for action that we hope will be further discussed and ultimately taken up by the whole wordnet community.

1 Introduction

Princeton WordNet (PWN: Fellbaum, 1998) has existed for 25 years. It is a manually created resource that has proven its worth in many different aspects of linguistics, computational linguistics, industrial applications and last but not least lexicology and knowledge engineering, cited over 11,000 times in Google Scholar[1]. It models language based on a division between words and concepts (represented as synsets) and semantic relations between these synsets. WordNet provided a different perspective on lexical resources from the traditional view in which the lemmas are the basis for defining concepts. Since EuroWordNet Vossen (1998), the Princeton WordNet model has spread to many other languages all over the world and has been extended with inter-lingual relations through the InterLingual Index (ILI). By linking concepts across languages it became possible to compare wordnets across languages, raising fundamental issues with respect to the definition of a word and a concept.

Because synsets are based on sets of synonyms, they mainly represent concepts lexicalized in a particular language (although you can have a synset with a phrase rather than a single word). This implies that different language wordnets may define different concepts related to the network and in fact define semantic spaces that partially match and partially do not. Within EuroWordNet, two approaches were defined to build wordnets: **expand** and **merge**. The expand method takes the concepts from the Princeton WordNet (PWN) as a starting point and translates the synonyms in the synsets to equivalences in the target language. If the same word is a translation of synonyms in different synsets, this creates different senses for the translation. By default, the fund of concepts for this wordnet and the semantic space is identical to the PWN structure. Concepts that are not lexicalized in English cannot be represented, or will be added to the nearest possible synset, even if the denotation is slightly different. The merge method takes the words of a language as a starting point and independently creates the synsets and relations between them. This leads to an independently created semantic space, which can then be aligned with the PWN structure by providing equivalence relations. In the case of the merge approach, the spaces are usually partially aligned and

[1] 11,266 citations on 2015-09-12.

there may be concepts that are in the new wordnet but not in PWN.

Currently, there is no central registry for these new concepts. Wordnet builders for different languages have no control over the concepts included in PWN and cannot easily share their concepts to other wordnet builders. Some projects have created their own internal InterLingual indexes (for example MCR (Gonzalez-Agirre et al., 2012) and the Multilingual Wordnet (Pianta et al., 2002) but these have not been widely adopted, are also based on PWN and most importantly cannot be modified by the community.

The idea of a GlobalWordNet Grid (GWG): a platform for making all wordnets and their linkage available was proposed at the bi-annual business meeting of the Global Wordnet Conference in Jeju, Korea 2006. Such a platform would enable the discussion about what defines a word and a concept across the different wordnets and also enable concept-sharing in a more fundamental way.

The Open Multilingual Wordnet (OMW: Bond and Paik, 2012; da Costa and Bond, 2015) went a long way to making linked wordnets available. The key insight was that wordnets could only be legally linked if the data was freely available and allowed manipulation and redistribution. They showed that wordnets were cited more if released under an open license, and managed to persuade many projects to release under open licenses. As a result there are now open wordnets available for 33 languages, all linked to each other, as well as automatically constructed data for 150 languages. They postponed the question of how to link concepts across all languages by using PWN 3.0 as a *de facto* ILI, and dropping concepts that could not be linked.

Concluding: the essential problem of how to coordinate adding new concepts for multiple languages has not been realized until today. We want to start a new era for wordnets by establishing a framework so that the building and comparison of the different language wordnets may achieve another level: both theoretically and from an engineering point of view. This paper describes the details of this platform and opens up the discussion with the community how to proceed. The paper is further structured as follows. In Section 2, we give the background and motivation for the Grid, while section 3 describes the main principles for the GWG and Section 4 for the Collaborative ILI (CILI). Section 5 describes the procedures and the current status. In Section 6, we explain how the ILI is used to map to WordNet3.0 and WordNet3.1. We also discuss methods for gloss-comparison across synsets in wordnets to find matches and candidates for new concepts. Section 7 reports on an experiment to map the Open Dutch Wordnet to the Grid and the attempt to find new ILI concepts. Finally in Section 8, we discuss the future options to proceed and come to our conclusions.

2 Background and motivation

The Global Wordnet Association website currently lists 76 wordnet groups and projects for 47 languages and other initiatives such as IndoWordnet and Asian Wordnet with many more languages. Not all of these projects are at the stage where they have produced a working wordnet. These wordnets almost all have some relation with PWN, either through the expand method or through equivalence relations (merge). All wordnets implement the notion of a synset as the core structure with at least lexical semantic relations between these synsets. Although PWN has a well-defined structure, the development of wordnets for other languages shows a large variety of decisions and choices. Some of these choices relate to the content of the databases, whereas others apply to the way the resources are distributed. This variation seriously hampers the use and principled study of the wordnets, especially since it is not possible to obtain all wordnets and access them through a unified format and API, which is our main motivation for establishing the GWG platform. Further, different wordnet projects have extended the wordnet structure in different ways, adding different relations and using conventions. Because of this, it is hard to compare wordnets across languages.

In addition to these more fundamental problems, there are also various practical problems for usage of the collection of wordnets. Different wordnets are linked to different versions of the Princeton WordNet, released in different formats (e.g. Princeton offsets and sense-keys, EuroWordNet XML, Multiwordnet, WordnetLMF, RDF) and according to different licenses (from completely open source to commercially restricted). Further, many wordnets have added new concepts, but as there is no central ILI how many of these, if any,

are duplicates?

When Princeton releases a new version of WordNet, it immediately leads to a further decrease in compatibility of wordnets and all related tools and systems, in particular as synset identifiers cannot be preserved across versions, although sense keys are intended to be preserved. The fact that all wordnets and systems adhere to some version of the Princeton WordNet also means that the fund of concepts is biased towards an Anglo-Saxon worldview and is not open to concepts from other languages and cultures.

It could be argued that these problems would go away if a single multilingual database was developed instead. This would, in theory, solve problems of incompatible formats and coordination. In practice, however there is no single group that has expertise in all the world's languages. Further, much experimentation is done in the different projects; adding new relations (Vossen, 1998), adding richer domains (Bentivogli et al., 2004), adding new parts-of-speech (Seah and Bond, 2014) and so forth. This would be harder to do in one monolithic project.

As time passes, and PWN now celebrates its 25th anniversary, the need for implementing the GWG becomes more urgent. The GWG should be a platform for achieving linguistic and conceptual interoperability across wordnets and all related machinery. It should allow researchers to study the universals and idiosyncracies in lexicalisation across languages, to address fundamental questions about what is a word and what is a concept (Fellbaum and Vossen, 2010; Vossen and Fellbaum, 2011). Tools built on wordnets should enable the development of software that can process text in any language according to a common semantic backbone as was demonstrated by the KYOTO[2] (Vossen et al., 2013a) and NewsReader[3] (Vossen et al., 2014) projects.

3 The new Global Wordnet Grid

The global wordnet grid consists of:

- The individual wordnet **project**s
- The collaborative interlingual index (**CILI)**
- The platform that ties them together and allows for adaptation and collaboration

[2] `www.kyoto-project.eu`

[3] `www.newsreader-project.eu`

The projects contribute data for wordnets that they produce in an agreed upon format: WordnetLMF or a *lemon*-based WordnetRDF. These should be validated and checked by the projects, who will have the responsibility of clearly marking which synsets are ready to be included in the ILI (that is, hand checked to a good quality). Although most projects specialize in a single language, there are some that produce multiple languages: both LMF and *lemon* can handle this.

As the GWG will manipulate and redistribute the projects' data, it must be released under a suitable open license. The CILI is released under a the Creative Commons Attribution 4.0 (CC BY) license. However, some projects use the ShareALike license (CC BY SA). In order to keep compatibility across the grid, any projects in the Global Wordnet Grid must have a license compatible with CC BY SA (such as the original wordnet license, CC BY, MIT and many others), and the entire grid will be released under this license.

The individual projects, starting with PWN, are the foundations upon which the GWG is built, the CILI links them and the platform ties them together, allows for versioning and adaptation through the community.

4 The Collaborative ILI (CILI)

The Collaborative ILI is an extension of the ILI defined in EuroWordNet (see Bond et al., 2016, for more details). As a base for the CILI we take the synsets currently in Princeton Wordnet 3.0, the *de facto* ILI for the Open Multilingual Wordnet. This shows its central position in the current wordnet community. Each synset in PWN 3.0 gives rise to a concept in the CILI.

The CILI is just a collection of concepts to which all wordnets are linked. It does not duplicate the relations between these concepts as represented in any wordnet and it does not have any lexicalizations. Concepts and concept identifiers in the CILI are permanent. They will never be removed or changed. However, new concepts can be added to the CILI but only if:

- there is a synset in a wordnet in the GWG that represents this concept (that is, linked by a `owl:sameAs` relation)
- this synset is related to another concept in this wordnet that is already represented in the CILI with one of a set of known relations (`hypernymy`, `meronomy`, `antonymy`)

- It must have a unique English definition that complies with the definition guidelines

The CILI is expanded when a project commits a wordnet, or a new version of a wordnet, to the repository. A committed wordnet is analysed by the moderators of the site (the authors of this submission). If syntactically correct, we will update the ILI records for all synsets that have such a record as a value of the ILI-attribute so that the records get `owl:sameAs` mappings to the contributed wordnet.

All synsets without an CILI-attribute that fulfill the conditions given above (linked, uniquely defined) will generate a proposed new concept which is distributed to the wordnet community for feedback and voting.

Gloss similarity can be used to find CILI concepts that are similar, where we can limit the search space on the basis of the semantic relations (of any linked wordnet). This prevents orphan concepts to be added that cannot be positioned in the semantic space of any available wordnet. We will demonstrate this in the next sections for Princeton WordNet 3.1 and the Open Dutch Wordnet.

5 The community platform

The GWG platform consists of:

- the website providing the most important information and the status of the Grid: `http://globalwordnet.org/global-wordnet-grid/`
- the ILI hosted as an LOD repository with persistent identifiers for concepts: `globalwordnet.org/ili`
- the collection of wordnets in WordnetLMF, *lemon*-based WordnetRDF format in a version control platform (such as `https://github.com/globalwordnet`)

The versioning control system is used to keep track of changes and contributions.

Adapting the ILI within GWG is important to get a better mapping across wordnets, especially when following a merge approach. It enables us to bypass conceptual gaps in PWN and the English language and share related resources across languages such as parallel corpora, ontologies, terminologies and sense-tagged corpora. It should also tighten definitions of synonyms and relations through translation relations across texts in different languages or word embeddings derived for any language (such as Mikolov et al., 2013). Ultimately, it allows us to define what is a word and what is a concept across languages.

There should be no limit to the number of concepts. Phrasenets are equally legitimate as synsets to define a concept. We can allow for example for frequent adjective-noun, noun-prep-noun, verb-object combinations as well as for proverbs, idioms and compounds in languages. Whether and how these concepts are lexicalized is up to the wordnet builders in each language. Ultimately, we will be able to infer which and how many languages provide some type of lexicalization for these concepts. Concepts that are linked to many independently-built wordnets do matter, concepts linked to a single wordnet play a minor role within the Grid. The more `owl:sameAs` relations a concept gets, the more it is valued by the wordnet community. It is also possible to axiomatize concepts through any ontology, exploiting `owl:sameAs` relations between URIs. An ontology defines a semantic space just as any other wordnet, albeit more formally.

Given the fact that concepts with sufficiently different glosses can be added and can be adopted by others, we can imagine that the GWG forms different layers of concepts, starting from a core of concepts shared by many wordnets, possibly ontologized and applied to many different texts in different languages, up to concepts recently added and mapped to only a single wordnet. The linkage of the data can be seen as an *onion model*[4] of concepts based on:

- a kernel of fund consists of concepts that are:
  - shared by all associated wordnets
  - sufficiently voted for by different wordnets, built independently and with sufficient spread in language-families and cultures
  - axiomized through ontologies
  - passed various consistency checks
- an outer layer that contains:
  - most recently proposed new concepts with an `owl:sameAs` relation to a synset in a single wordnet that meets the minimal criteria described above

[4] as presented at the LREC-2014 workshop on Linked Data in Linguistics

- In between layers:
  - linked to more wordnets across languages and language families
  - while these wordnets express semantic relations for these concepts that are not in conflict
  - may have been moderated by the community for example through voting
- an external layer that contains:
  - synsets defined in project wordnets that do not fit the criteria for inclusion into the ILI (e.g. no English definition or unlinked). These concepts need more work to either link them or to be added as new concepts.

In addition to the CILI itself, we will host all public wordnets that are linked to the ILI and may have provided new concepts. We extended WordnetLMF (Vossen et al., 2013b) with some additional attributes to support the mappings of wordnets to the CILI. First of all, each synset element has an optional attribute *ili* for the CILI-record to which the synset is connected. Furthermore, the definition element has an obligatory language attribute and an optional provenance attribute to enable matching concepts. Below we show a WordnetLMF example for an Open Dutch Wordnet synset with a mapping to the CILI and different definitions:

```
<Synset id="eng-30-13956488-n" ili="i110277">
  <Definitions>
    <Definition gloss="overeenstemming met de werkelijkheid"
    language="nl" provenance="odwn"/>
    <Definition gloss="conformity to reality or actuality"
    language="en" provenance="pwn"/>
    <Definition gloss="agreement with reality" language="en"
    provenance="google-translate"/>
   </Definitions>
   <SynsetRelations>
     <SynsetRelation provenance="pwn" relType="has_hyperonym"
     target="eng-30-13954818-n"/>
   </SynsetRelations>
</Synset>
```

6 Mapping updates in PWN 3.1

One of our first checks was to ensure that the graph of the 3.1 version of PWN can be mapped to the CILI (which is based on version 3.0 of PWN). This should have been a trivial case as while the synset identifiers, which are based on the offset in a the release files, are not stable between versions, the sense keys used to identify the senses in Princeton WordNet should be. Using this as the basis of the mapping we found that 1,796 (1.5%) of all synsets were modified between version 3.0 and 3.1 and we manually mapped these synsets. The results were as follows:

- No equivalent in 3.1 (986 synsets):
  - Proper names, drug names, brand names and other proper nouns were systematically removed from 3.0
  - Many sexist, racist, homophobic etc. terms were removed (e.g., 'shirtlifter' and many much worse)
  - Some terms such as 'that much' or senses of terms were considered not be lexicalized concepts and thus erroneously introduced into PWN.

  In these cases, the concept in the CILI is marked as **deprecated**
- Multiple 3.0 synsets mapped to one in 3.1 (51 synsets)
  - Mostly duplicates, e.g., 'finish coat' (03342657-n and 03342863-n)

  In these cases, one of them (typically the one with a different definition from the one that was kept), should be marked as being **superceded** by the other and **deprecated**.
- Single 3.0 synset mapped to multiple in 3.1 (22 synsets)
  - In some cases a word is removed from a synset and put into a new synset, which may be either a hypernym, hyponym or co-hyponym of the previous synset. The definition of the original synset is preserved, e.g., the adjective 'documentary' was removed from the synset of 'objective' or 'documentary' ("emphasizing or expressing things as perceived without distortion of personal feelings, insertion of fictional matter, or interpretation") and given a new synset specifically stating that it must be a film or TV show.
  - In some cases an existing synset is split and both new meanings appear to be more specific, e.g., the heraldic terms 'annulet' and 'roundel' were given new synsets and the previous definition was removed.

  In the first case, a new concept is created and no other change is necessary. In the second

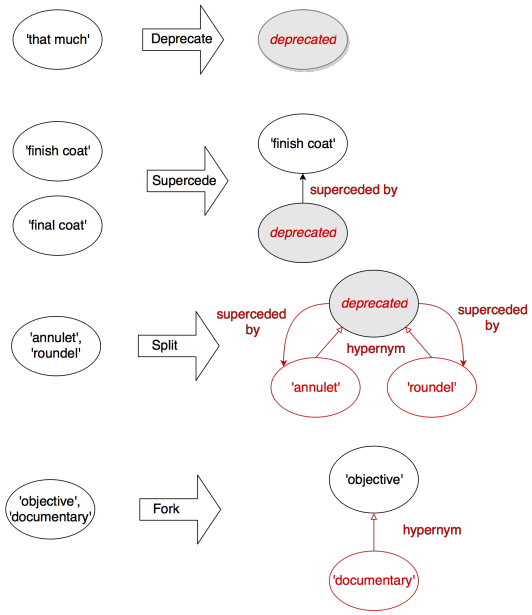

Figure 1: Examples of changes that can be made to the CILI

case, two new concepts need to be created and linked to the original one, which should be deprecated.

- The remaining 737 mappings were changes of part-of-speech between satellite and non-satellite adjectives. These require no changes to the ILI as it does not mark part-of-speech.

In summary, these changes should be the most common changes of the CILI as it develops.

Deprecate A synset may be flagged as deprecated, meaning that we no longer consider it a true lexical concept. This is primarily the case when a compound term has been introduced by mistake. The synset identifier is not removed from the ILI.

Supercede If a duplicate is detected we would choose one of the synsets to remain, and the second synset identifier is marked as deprecated and a link is introduced to the superceding synset, but this second synset is not removed from the CILI.

Split If a synset is considered to generalize two distinct concepts we split it into two new synsets and add these as hyponyms, which are marked as supercedents of the original synset. The original synset is marked as deprecated but not removed from the ILI.

Fork Alternatively if the original synset is still considered valid it is kept undeprecated and a new more specific and closely related synset is added.

Note, that a new wordnet version on its own does not give enough information to decide when a concept should be deprecated or superceded. The platform must therefore allow projects to suggest this as a separate operation.

7 The Open Dutch Wordnet

The Open Dutch Wordnet (ODWN, Postma et al. (2016)) was created from PWN through a mixture of expand and merge methods. PWN synset identifiers and relations have been re-used as much as possible. However, new concepts that originate from the Referentie Bestand Nederlands (RBN: Van der Vliet, 2007)) and have no equivalence relation to PWN synsets have been added. Table 1 shows the distribution of synsets with mappings to PWN (Dutch PWN synsets) and synsets without (Dutch ODWN synsets). To maintain the PWN hierarchy, the wordnet includes hypernym synsets from PWN even if they do not have any Dutch synonyms (English PWN synsets).

Table 1: Overview of the Open Dutch Wordnet

Open Dutch Wordnet	Total	Nouns	Verbs
Word forms	57,602	50,255	7,347
Lexical Units	94,140	78,612	15,528
Dutch ODWN synsets	21,636	15,992	5,644
Dutch PWN synsets	19,980	15,706	4,274
English PWN synsets	75,376	66,409	8,967
Total	116,992	98,107	18,885

In all cases that we could use a PWN synset, we could also map the concept to a CILI record. All synsets with an ODWN identifier were not mapped to the CILI. Consider the word *bierbuik* which is ambiguous between two senses: one for *a big belly because of drinking too much beer* and the second referring to *a person with such a belly*. Neither sense is currently in PWN3.0, and thus new CILI concepts would need to be created. The first sense is lexicalized in English (*beer belly, beer gut*, but has not yet been added to PWN. Both the concepts are linked (as hyponyms) to existing synsets in PWN, therefore to add the concepts to the CILI, the ODWN project would just need to write English glosses.

In total there are 21,636 synsets without a mapping to a PWN synset and therefore without a

Dutch	English
perzikhuid	peach skin
kalfskotelet	veal chop
natuurramp	natural disaster
verwachtingspatroon	expectations
sluikreclame	product placement

Figure 2: ODWN entries not in PWN

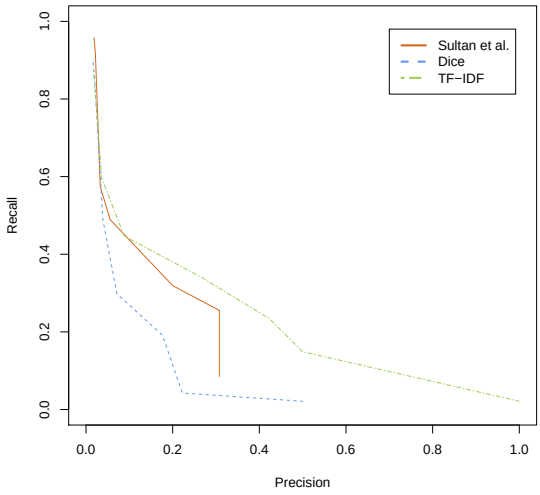

Figure 3: Precision-Recall curve of gloss matching between ODWN and PWN

mapping to the CILI. From these, there are about 5,067 synsets in which case the lemma as translated by Google Translate is not an entry in PWN and another 4,479 synsets for which the translations have a low similarity according to the PWN hierarchy, using the method described by Leacock and Chodorow (1998). We show some more examples of translations not in PWN in Figure 2.

We consider these 9,546 synsets potential new CILI concepts. To validate these as new, we need to ensure that they do not match an existing English gloss. This task is complicated by two main issues: firstly, semantic textual similarity is still a difficult task and secondly, we are using machine translations of the definitions, which introduces further error into the process. To investigate whether automatic methods would solve this task we translated the definitions of Dutch synsets which were already aligned to synsets in PWN and attempted to see if we can distinguish this gloss from similar glosses, in particular glosses of synsets that were up to 3 hyperonym/hyponym links from the target synset. We tried three similarity metrics, namely, the Dice co-efficient, the cosine of TF-IDF vectors of the glosses and an alignment method (Sultan et al., 2014), which had the strongest performance for the Semantic Textual Similarity Task at SemEval-2014. For each of these methods, we varied the acceptance threshold and calculated precision and recall in the usual manner and the results are presented in figure 3. A random baseline has an expected precision of 3.0% and the highest F-Measure was 35.5%: a strong improvement.

However, the performance of the semantic matching is still low, and while high recall can be achieved, which would allow us to select a list of potential duplicates this is only at very low precision, meaning that annotators may have to work through a very long list of candidates. We believe this in part due to the relatively short glosses in ODWN, for example, for ‘afweersystem’ (‘immune system’), the gloss is only ‘afweer tegen ziektes’ (‘defense against diseases’) where as the PWN gloss is 27 words long: “a system (including the thymus and bone marrow and lymphoid tissues) that protects the body from foreign substances and pathogenic organisms by producing the immune response”. As such, automatic systems can aid in the detection of duplicates in the CILI but must be considered along with guidelines that glosses must be submitted in English and not automatically translated and that glosses must conform to quality guidelines.

Our impression overall so far is that many of the ODWN synsets are already in PWN, although it is very difficult in some cases to find them. The best candidates for new concepts are actually translations of synonyms that could not be found as entries in PWN (5,067 in total). However, even these need critical review and their glosses should have zero scores compared with a wide range of candidate synsets. Concluding, we can say that extending the CILI with new concepts should be done conservatively and with great care.

8 Future work and conclusions

We presented the implementation of the Global Wordnet Grid, which has been pending for many years. We described the data structures and data points as well as the main principles, the protocols and the motivation. The success of the platform will depend on the community. We described two use cases. They made it clear that the process is not trivial and we still will need to discuss many

details. We welcome any further suggestions and contributions of wordnet builders and users.

Finally, the position of the Princeton WordNet in the Grid is essential. Reference to concepts, words, word senses and versions of resources is essential. We hope that future version of PWN will support the GWG and make reference to the ILI just as other wordnets should do.

References

Luisa Bentivogli, Pamela Forner, Bernardo Magnini, and Emanuele Pianta. 2004. Revising wordnet domains hierarchy: Semantics, coverage, and balancing. In *Coling 2004 Workshop on Multilingual Linguistic Resources*, pages 101–108. Geneva.

Francis Bond and Kyonghee Paik. 2012. A survey of wordnets and their licenses. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*. Matsue. 64–71.

Francis Bond, Piek Vossen, John McCrae, and Christiane Fellbaum. 2016. CILI: The collaborative interlingual index. In *Proceedings of the 8th Global Wordnet Conference (GWC 2016)*. (this volume).

Lus Morgado da Costa and Francis Bond. 2015. OMWEdit - the integrated open multilingual wordnet editing system. In *ACL-2015 System Demonstrations*.

Christiane Fellbaum and Piek Vossen. 2010. Connecting the universal to the specific: Towards the global grid. *Lecture Note in Computer Science (LNCS) Vol.4568*.

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

Aitor Gonzalez-Agirre, Egoitz Laparra, and German Rigau. 2012. Multilingual central repository version 3.0: upgrading a very large lexical knowledge base. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*. Matsue.

Claudia Leacock and Martin Chodorow. 1998. Combining local context and wordnet similarity for word sense identification. In Fellbaum (1998), chapter 11, pages 265–283.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *Proceedings of Workshop at International Conference on Learning Representations*.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. Multiwordnet: Developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, pages 293–302. Mysore, India.

Marten Postma, Emiel van Miltenburg, Roxane Segers, Anneleen Schoen, and Piek Vossen. 2016. Open Dutch wordnet. In *Proceedings of the 8th Global Wordnet Conference (GWC 2016)*. (this volume).

Yu Jie Seah and Francis Bond. 2014. Annotation of pronouns in a multilingual corpus of Mandarin Chinese, English and Japanese. In *10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*. Reykjavik.

Md Arafat Sultan, Steven Bethard, and Tamara Sumner. 2014. Back to basics for monolingual alignment: Exploiting word similarity and contextual evidence. *Transactions of the Association for Computational Linguistics*, 2:219–230.

Hennie Van der Vliet. 2007. The referentiebestand Nederlands as a multi-purpose lexical database. *International Journal of Lexicography*, 3(20):239–257.

P. Vossen, E. Agirre, G. Rigau, and A. Soroa. 2013a. *New Trends of Research in Ontologies and Lexical Resources*, chapter KYOTO: a knowledge-rich approach to the interoperable mining of events from text, pages 65–90. Springer Verlag, Heidelberg.

P. Vossen, G. Rigau, L. Serafini, P. Stouten, F. Irving, and W. Van Hage. 2014. Newsreader: recording history from daily news streams. *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*.

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

Piek Vossen and Christiane Fellbaum. 2011. *Multilingual FrameNets in Computational Lexicography, Methods and Applications*, chapter Universals and Idiosyncracies in Multilingual WordNets, pages 319–346. Mouton de Gruyter, Berlin.

Piek Vossen, Claudia Soria, and Monica Monachini. 2013b. LMF - lexical markup framework. In Gil Francopoulo, editor, *LMF - Lexical Markup Framework*, chapter 4. ISTE Ltd + John Wiley & sons, Inc.