CILI: the Collaborative Interlingual Index

Francis Bond,[♠] **Piek Vossen,**[♢] **John P. McCrae**[♣] and **Christiane Fellbaum**[♡]
[♠]Nanyang Technological University, Singapore
[♢]VU University Amsterdam, The Netherlands
[♣]Insight Centre for Data Analytics, NUI Galway, Galway, Ireland
[♡]Princeton University, U.S.A.
<bond@ieee.org,piek.vossen@vu.nl,john@mccr.ae,fellbaum@princeton.edu>

Abstract

This paper introduces the motivation for and design of the Collaborative InterLingual Index (CILI). It is designed to make possible coordination between multiple loosely coupled wordnet projects. The structure of the CILI is based on the Interlingual index first proposed in the EuroWordNet project with several pragmatic extensions: an explicit open license, definitions in English and links to wordnets in the Global Wordnet Grid.

1 Introduction

Within 10 years of the release of Wordnet (Miller, 1990) researchers had started to extend it to other languages (Vossen, 1998). Currently, the Open Multilingual Wordnet (OMW: Bond and Paik, 2012; da Costa and Bond, 2015) has brought together wordnets for 33 languages that have released open data,[1] and automatically produced data for 150. There are even more wordnets than this: some large projects have released non-open data, notably German (Kunze and Lemnitzer, 2002) and Korean (Yoon et al., 2009) and many projects have yet to release any. This activity shows that the structure of wordnets is applicable to many languages.

All the wordnets are based on the basic structure of the Princeton wordnet (PWN: Fellbaum, 1998): synonyms grouped together into **synsets** and linked to each other by semantic relations.

The majority of wordnets have been based on the **expand** approach, that is adding lemmas in new languages to existing PWN synsets (Vossen, 1998, p83), boot-strapping from the structure of English. 28 out of 33 of the wordnets in OMW take this approach. A few wordnets are based on the **merge** approach, where independent language specific structures are built first and then some synsets linked to the PWN. In OMW, only five projects take this approach: Chinese (Taiwan), Danish, Dutch, Polish and Swedish (Huang et al., 2010; Pedersen et al., 2009; Postma et al., 2016; Piasecki et al., 2009; Borin et al., 2013).

To investigate meaning across languages, we need to link synsets cross-lingually. It is easy to link expand-style wordnets: they all link to PWN and it can be used as a **pivot** to link them together. This is one of the attractions of using the expand approach, you immediately gain multilingual links. The disadvantage is that concepts not in PWN (either because they are not lexicalized in English or just because they have not been covered yet) cannot be expressed. Because of this, many expand-style wordnets also define some new, language-specific synsets, typically a few tens or hundreds (Arabic, Chinese, Italian, Japanese, Catalan, Spanish, Galician, Finnish, Malay/Indonesian, Bulgarian, Greek, Romanian, Serbian and Turkish all do so)(Pianta et al., 2002; Tufiş et al., 2004; Elkateb and Fellbaum, 2006; Gonzalez-Agirre et al., 2012; Wang and Bond, 2013; Bond et al., 2014; Seah and Bond, 2014; Postma et al., 2016).

It is harder to link merge-style wordnets. The projects need to somehow identify links to PWN, and as a result, only a small subset of the language specific synsets are linked to PWN. Examining the unlinked synsets, this seems to be principally due to the lack of resources to link them than semantic incompatibility. For example, Danish and Polish (Pedersen et al., 2009; Piasecki et al., 2009) have many synsets which can be linked but are not currently.

Currently, when projects create their own synsets, there is no coordination between these projects. This means that similar or even identical concepts may be introduced in multiple places.

[1] We use the definition from the Open Knowledge Foundation: `http://opendefinition.org/`: ``anyone is free to use, reuse, and redistribute it --- subject only, at most, to the requirement to attribute and/or share-alike".

For example, most South East Asian languages distinguish between cooked and uncooked rice: these concepts have been added independently to the Korean and Japanese wordnets. Typically, clusters of projects have tried to coordinate, such as EuroWordNet, the Multilingual Central Repository for Basque, Catalan, Galician and Spanish (Gonzalez-Agirre et al., 2012), the MultiWordNet for Italian and Hebrew, (Pianta et al., 2002), Balkanet (Tufiş et al., 2004), the Wordnet Bahasa for Malay and Indonesian (Bond et al., 2014), the IndoWordnet project (Bhattacharyya, 2010).

Clearly, there is a need for a single shared repository of concepts. In this paper, we propose to build one: the **Collaborative InterLingual Index** (CILI). We base the index on the technical foundations laid down in EuroWordNet: a single list that is the union of all the synsets in all the wordnets (Peters et al., 1998; Vossen et al., 1999). To this we add ideas from the best-practice of the Semantic Web: a shared easily accessible resource with a well defined license; from open-source software: build a community of users who will co-develop the resource; and from experiences in many multilingual lexical projects: accept the *de facto* use of English as a common language of communication.

In the following sections we discuss the motivation further (§ 2), then describe in detail the structure of the CILI (§ 3), list some open issues (§ 4) and finally conclude.

2 Motivation

Wordnets have been built with different methods and from different starting points: expand or merge, manually or semi-automatically and based on pre-existing monolingual resources or using available bilingual resources to translate English synsets to words in the target language. Furthermore, it is up to the wordnet builders to make decisions about which words are synonyms, what are the semantic relations between the synsets and how to interpret each semantic relation. We can observe very large synsets in one wordnet being linked through PWN to small synsets in another language. Different granularities of synsets brings into questions the notion of the same concept existing across these wordnets. PWN uses 44 semantic relations (if separated by part-of-speech) but in EuroWordNet 71 relations were defined that partially overlap. Even if two wordnets use the same relation name, there is no guarantee that it is interpreted in the same way. In fact, different wordnet editors and algorithms may interpret relations differently. Even the symbols used for parts of speech differ in different projects (**adverb** is 'r' in PWN but 'b' in some projects). Finally, one can observe large differences in coverage of the vocabulary and in the degree of polysemy. Vocabularies and concepts differ in size but also in terms of genre, pragmatics, the inclusion of multiword expressions as ``phrase sets" (Bentivogli and Pianta, 2003) and specific domains and areas. Choices for distinguishing senses lead to fine-grained and coarse-grained polysemy, where the latter may lead to multiple hypernyms that can be modeled as complex types (Pustejovsky, 1995). Finally, the glosses for synsets play an underestimated role in addition to the synsets and the relations, but no formal structuring is defined for these glosses. As a result, glosses are not sufficiently descriptive to precisely identify the meaning of a concept. Such differences across wordnets make it difficult to establish the proper relations to the ILI and thus to compare and exploit wordnets across languages. Further, if a synset is not realized in a language it is not clear if that is because the concept is not lexicalized in that language, or if it is merely not realized yet (the compilers may just not have got round to it).

To solve these problems, we need to not just define an interlingual index, but also shared guidelines for relations, how to write definitions, standard data formats and so forth.

3 The Collaborative Interlingual Index

In this section we describe the core properties of CILI. To coordinate an index among all the different wordnet projects, we propose that it should, ideally, have the following properties (building on 1--5 from Fellbaum and Vossen, 2008):

1. The Interlinear Index (ILI) should be a flat list of concepts.
2. The semantic and lexical relations should mean the same things for all languages.
3. Concepts should be constructed for salient and frequent lexicalized concepts in all languages.
4. Concepts linked to Multiword units (MWUs) in wordnets should be included.
5. A formal ontology could be linked to but separate from the wordnets.

6. The license must allow redistribution of the index

7. ILI IDs should be persistent: we never delete, only **deprecate** or **supercede**; we should not change the meaning of the concept

8. Each new ILI concept should have a definition in English, as this is the only way we can coordinate across languages. The definition should be unique, which is not currently true, and preferably also parse and sense tag information should be included. Definition changes will be moderated.

9. Each new ILI concept should link to a synset in an existing project that is part of the GWG with one of a set of known relations (`hypernymy`, `meronomy`, `antonymy`, ...)

10. This synset should link to another synset in an existing project that is part of the GWG and links to an ILI concept.

    $\Rightarrow$ each concept is linked to another concept through at least one wordnet in the grid

11. Any project adding new synsets should first check that they do not already exist in the CILI

    - New concepts are added through their existing in a wordnet
    - If something fulfills the criteria is proposed
    - If no objections after three months then it is added

Property 6, an open license, is a necessary condition for groups to be able to use the ILI within their own project. To be maximally compatible, the license should place as few restrictions as possible, ideally requiring only that the source of the resource be mentioned: it should be either the wordnet license itself, Creative Commons Attribution (CC-BY) or the MIT license. We choose to use CC-BY, as the license has been well written and documented and is widely used.

Property 7, persistent identifiers, is an important criterion for stability. If the ILI changed its IDs, projects without the resources to maintain compatibility would fall behind. If a project changes its hierarchy, then it will need to add new nodes and delink the old ones. To keep backwards compatibility, even if a concept is deemed problematic, it will remain in the CILI, and marked as **deprecated**, preferably with a link to the concept that **supercedes** it.

Property 8, that all synsets should have a definition in English, recognizes that, in practice, the only language shared by all groups is English. Here we are inspired by experience with the CICC project, a multilingual machine translation project linking Thai, Chinese, Japanese, Malay and Indonesian (but not English) (CICC, 1994). No members spoke all five languages, but someone in each group spoke English, so all dictionary entries also had an English translation or definition. Having a universally understood definition is a prerequisite in avoiding redundant creation of new senses. This creates a burden on non-English speakers, which we will try to lighten by giving clear guidelines for writing definitions (see section 3.3). Note, that while the definition must be in English, the concept is not necessarily lexicalized in English, in contrast to Princeton WordNet.

Properties 9 and 10 make sure that all new concepts link to something, there should be no orphaned concepts. Exactly which links are acceptable is still a matter of research.[2]

The final point (11) is about coordination. Practically, it will not be possible to have a single moderator who can check new synsets in every language. We therefore propose that the burden of checking for duplication with existing synsets should be placed on the project wanting to add new synsets. As new concepts should be linked to existing concepts through relational links in a wordnet, and definitions in English will exist for all entries, checking for a compatible entry in the ILI should not be too burdensome. Project members with wordnets in the shared multilingual index would gain write privileges to the ILI, of course anyone should be able to read it. We will build automated tools that warn if definitions are too similar (for details see Vossen et al., 2016).

For the ILI to be successful there will be an initial cost to combine all existing non-English synsets, adding English definitions for all and merging duplicates. It would also require buy-in from all participating projects, but fortunately most non-English wordnets contain few synsets that do not correspond to an English synset, so this first step should not be too burdensome. For wordnets

[2] Many wordnets, including PWN, currently contain some orphans (e.g. ***uphill***$_{r:1}$), these would not be added to the ILI unless they are linked to something.

built with the merge approach there will be many more new synsets, these should be checked carefully and validated against corpora before being included in the ILI. We will support this with workshops at relevant conferences (such as the 16th Global Wordnet Conference).

In the long run, we hope that external resources will link to the ILI's persistent IDs (things like SUMO, TempoWordnet (Dias et al., 2014), the many Sentiment wordnets (Baccianella et al., 2010; Cruz et al., 2014).

3.1 Format

The ILI will be represented as RDF. Our reference implementation will be in Turtle (Terse RDF Triple Language: W3C, 2012) a compact format for RDF graphs.

It includes its own metadata, based on the Dublin Core, shown in Figure 1. As far as possible, triples are defined using existing schema (referenced in the preamble). The individual entries are designed to be extremely simple. Unlike synsets in individual wordnets, ILI concepts do not have explicit parts-of-speech. No further semantics is imposed within the ILI.

Each concept in the ILI has the following simple structure:

- A unique ID: `i1, i2, i3, …`
- A type: `Concept` or `Instance`
- A gloss in English: `skos:definition`
- A link to the synset that first motivated the ILI concept: `dc:source`
- Links to all current wordnets in the GWG that use this concept: `owl:sameAs`
- Optionally a `deprecate/supercedes` link

We give an example in Figure 2, which also shows the relevant prefixes.

Information about provenance (who added the entry, when it was made and so forth) are left to the version control system, for which we have chosen to use (`git: http://git-scm.com/`). When commits are made, the project will be added as the **author** so a record is kept of who is responsible for which change without making it visible in the ILI.

Note that the concept is defined not just by the written definition but by the links to the wordnets and the lemmas in those wordnets: the definition is a crucial tool for coordinating across languages, but is not meant to be the sole determiner of the ILI concept's meaning. The ILI concepts will always be linked to the global wordnet grid (Fellbaum and Vossen, 2007; Vossen et al., 2016).

Labels for the concepts can be produced automatically, as it is probably that different languages would want different labels. The easiest approach would be to take the most frequent lemma in the language of choice, backing off to the most frequent lemma in the language that introduced it (which can be obtained from the `dc:source`).

3.2 The WordNet Schema

In order to ensure that WordNets may be submitted in a form that is compatible with the ILI, we have developed two specific schemas, namely an XML schema based on the Lexical Markup Framework (Vossen et al., 2013, LMF) and the second in JSON-LD (Sporny et al., 2014) using the Lexicon Model for Ontologies (McCrae et al., 2012, *lemon*). These models are structured as follows:

LexicalResource The root element of the resource is the *lexical resource*

Lexicon Each WordNet has a *lexicon* for each resource, which has a name, an ID and a language. The language is given as a BCP 47 tag .

Lexical Entry Each 'word' is termed a *lexical entry*, it has exactly one lemma, at least one sense and any number of syntactic behaviors.

Lemma The lemma has a *written form* and part-of-speech, which may be one of **n**oun, **v**erb, **a**djective, adve**r**b, **p**hrase, **s**entence or **u**nknown.

Sense The *sense* has any number of *sense relations* and a synset.

Synset The *synset* has an optional definition and any number of *sense relations*.

Definition The definition is given in the language of the WordNet it came from as well as the ILI definition (in English). A definition may also have a *statement* that gives an example

Synset/Sense Relation A relation from a given list of relations such as synonym, hypernym, antonym. This list defines the relations used

```
<> a voaf:Vocabulary ;
  vann:preferredNamespacePrefix "ili" ;
  vann:preferredNamespaceUri "http://globalwordnet.org/ili" ;
  dc:title "Global Wordnet ILI"@en ;
  dc:description "The shared Inter-Lingual Index for the global wordnets.
   It consists of a list of concepts of instances with definitions,
   and their links to open wordnets."@en ;
  dc:issued "2015-07-30"^^xsd:date ;
  dc:modified "2015-07-30"^^xsd:date ;
  owl:versionInfo "0.1.1"@en ;
  dc:rights "Copyright Global Wordnet Association" ;
  cc:license <http://creativecommons.org/licenses/by/4.0> ;
  cc:attributionName "Global Wordnet Association";
  cc:attributionURL <http://globalwordnet.org>;
  dc:contributor  <http://www3.ntu.edu.sg/home/fcbond/>,  <http://john.mccr.ae> ,
                 <http://vossen.info/> ;
  dc:publisher <http://globalwordnet.org> .
```

Figure 1: ILI metadata

```
@prefix pwn30: <http://wordnet-rdf.princeton.edu/wn30/> .
@prefix jwn12: <http://compling.hss.ntu.edu.sg/omw/wns/jpn/> .
@prefix ili: <http://globalwordnet.org/ili/> .
@base <http://globalwordnet.org/ili/ili#>.

<i71370> a <Concept> ;
        dc:source       pwn30:06639428-n ;
        skos:definition "any of the machine-readable lexical databases
                         modeled after the Princeton WordNet"@en ;
        owl:sameAs      jwn12:jpn-06639428-n ;
        owl:sameAs      pwn30:06639428-n .
```

Figure 2: Example ILI entry for the concept of a *wordnet*

by the Global Wordnet Grid, and all the relations are documented on the Global Wordnet Association website.

Syntactic Behavior A syntactic behavior (verb frame) gives the subcategorization frame in plain text, such as ``Sam and Sue %s the movie".

Meta Dublin Core properties may be added to lexicons, lexical entries, senses and synsets.

Either format can be used to describe a WordNet and it is simple to convert between either. An example of the LMF form is given in figure 3 and in WN-JSON in figure 4

3.3 Guidelines for Definitions

In any given wordnet, the definition is only one of the things that helps to tell the meaning of a word, it is accompanied by the semantic relations, part of speech information, examples and so forth. The ILI is situated in the global wordnet grid, so this information should also be available. However the definition is the only thing guaranteed to be in the ILI, and the accompanying information may only be from a wordnet whose language is not comprehensible to another user. Moreover, as these definitions are given in natural language it is important to ensure that they are as unambiguous as possible, and can clearly identify the concepts, without the additional mechanisms of semantic relations. For these reasons strong guidelines for definitions are of primary importance.

There are already good general guidelines for writing dictionary definitions (Landau, 1989, Chapter Four). Almost all of these apply to wordnets in general, and the CILI in particular, with the exception that **brevity** is less important in an electronic resource.

There are some extra constraints for the CILI. First, definitions should be unique and there should be enough information to minimally distinguish one concept from all others. This was not the case in the wordnets, PWN has over 1,629 synsets with non-unique definitions, and there are similar numbers in other wordnets (1,362 in Japanese, 418 in Indonesian, 211 in Greek, 104 in Albanian and so on). For example it would not be sufficient to describe ***paella***$_{n:1}$ as ``a Spanish dish" as

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM "http://globalwordnet.github.io/schemas/WN-LMF.dtd">
<LexicalResource>
    <Lexicon label="Princeton WordNet" language="en">
        <LexicalEntry id="w1">
            <Lemma writtenForm="wordnet" partOfSpeech="n"/>
            <Sense id="106652077-n-1" synset="106652077-n"/>
        </LexicalEntry>
        <Synset id="106652077-n" ili="s35545">
            <Definition
              gloss="any of the..."
              iliDef="any of the..."/>
            <SynsetRelation relType="hypernym" target="106651393-n"/>
        </Synset>
        <Meta publisher="Princeton University"
              rights="http://wordnet.princeton.edu/wordnet/license/"/>
    </Lexicon>
</LexicalResource>
```

Figure 3: Example of WordNet entry in WN-LMF

```
{
  "@context": [ "http://globalwordnet.github.io/schemas/wn-json-context.json",
                { "@language": "en" } ],
  "@id": "pwn30",
  "label": "Princeton WordNet",
  "language": "en",
  "publisher": "Princeton University",
  "rights": "wordnetlicense:",
  "entry": [{
    "@id" : "w1",
    "lemma": { "writtenForm": "wordnet" },
    "partOfSpeech": "wn:noun",
    "sense": [{
      "@id": "106652077-n-1",
      "synset": {
        "@id": "106652077-n",
        "ili": "s35545",
        "definition": {
          "gloss": "any of the..." ,
          "iliDef": "any of the..."
         },
         "hypernym": ["106651393-n"]
      }
    }]
  }]
}
```

Figure 4: Example of an entry in WN-JSON

this is not sufficiently distinctive. For the wordnets, the combination of definition and lemmas is normally enough to distinguish a word, but for the ILI, if necessary, one of the English lemmas must be included in the definition (for example, including the species name in the definition). This conflicts somewhat with the best practice for individual wordnets, where in general we want to avoid redundancy: if the synset is linked through domain-category to e.g. mathematics, we would normally not start the definition with ``(mathematics)". A case in point is the definitions for PWN30:13223710-n *ground fir, princess pine, tree clubmoss, Lycopodium obscurum* and and PWN:13223588-n *ground cedar, staghorn moss, Lycopodium complanatum* which are both defined as ``a variety of club moss". In this case, amending the definition to ``a variety of club moss (Lycopodium obscurum)" and ``a variety of club moss (Lycopodium complanatum)" makes the definitions unique (at the cost of some redundancy. We propose using some of the wide array of brackets available to show the redundant information in the ILI definition: ``⟪plant⟫ a variety of club moss ⟦Lycopodium complanatum⟧". Doing this reduces the number of non-unique definitions by over 50%. The ILI definitions are thus produced automatically from PWN 3.0, without always being identical to them.

We also place some limitations on the format.

The definition should consist of one or more short utterances, separated by semicolons. Semicolons should not be used within each utterance, use comma or colon instead. Definitions will be split on semicolons before being parsed, so it is important to be consistent here. We also do not allow the use of ASCII double quotes instead preferring Unicode left and right (double) quotes to aid parsing.

In general, we need to be very conservative in changing the definitions of concepts in the ILI. When first written, we should try not to make the definition too restricted, for example, prefer for *angel, backer* instead of ``invests in a theatrical production", prefer ``someone who invests in something, typically a theatrical production". This makes it easier to avoid having to make multiple very similar synsets.

Definitions should use standard patterns, especially for the first utterance in a definition. Ideally, the definition should consist of a **genus** (the hypernym, not necessarily the immediate hypernym) and **differentiae**, e.g.,

> ***wordnet*** *(lemma)* ``any of the machine-readable lexical databases *(genus)* modeled after the Princeton WordNet" *(differentiae)*

Adjectives and adverbs are exceptions, in that they are often defined using prepositional phrases.

Finally we make a simple requirement that definitions have a minimum length of 20 characters or 5 words.

In future work we will produce a tool to parse the definition and automatically identify the hypernym (Nichols et al., 2005), sense tag the definition (Moldovan and Novischi, 2004) and report on this to the definition writer, as well as compare the definition to definitions from similar concepts. This can help identify infelicitous definitions.

4 Open Issues

There are a few cases where it was hard to decide whether a concept should be represented in the InterLingual Index.

One example is named entities. Roughly 6.6% of the entries in PWN are linked by the `instance` relation (including the names of people, places, planets, gods and many more). Named entities are much more numerous than words and these concepts and their relations are better captured by other kinds of resources. However, some named entities can be considered part of the lexicon as well as names for objects, for example ***Glaswegian***$_{a:1}$ ``of or relating to or characteristic of Glasgow or its inhabitants", which is also used in the definition of other concepts. Thus, we retain a small number of named entities, especially geographic terms but further discussion is required to refine an exact policy.

It could also be argued that some of the derived forms (for example ***quickly***$_{r:1}$ from ***quick***$_{a:1}$) are unnecessary: as the meaning change is generative, there is no point in having two concepts. These kind of changes can be applied later by means of superceding other concepts, and for the moment we apply the distinctions made by Princeton WordNet.

5 Conclusions

We have introduced and motivated the collaborative interlingual index (CILI). Its simple design allows us to link wordnets with a minimum of extra work. Once concepts are added to the CILI, they will get a persistent ID and thereafter should not be deleted or change in meaning. We propose that the task of checking the validity of new concepts is taken up by the individual wordnet projects, with only a light layer of moderation.

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7thh International Conference on Language Resources and Evaluation (LREC 2010)*. Valletta, Malta.

Luisa Bentivogli and Emanuele Pianta. 2003. Beyond lexical units: Enriching wordnets with phrasets. In *Proceedings of the Research Note Sessions of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL 03)*, pages 67--70. Budapest.

Pushpak Bhattacharyya. 2010. Indowordnet. In *Proceedings of 7th Language Resources and Evaluation Conference (LREC 2010)*. La Valletta. URL `http://www.cfilt.iitb.ac.in/indowordnet/`.

Francis Bond, Lian Tze Lim, Enya Kong Tan, and Hammam Riza. 2014. The combined wordnet Bahasa. *Nusa: Linguistic studies of languages in and around Indonesia*, 57:83--100.

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.

Lars Borin, Markus Forsberg, and Lennart Lönngren. 2013. Saldo: a touch of yin to WordNet's yang. *Language Resources and Evaluation*, 47(4):1191--1211. URL `dx.doi.org/10.1007/s10579-013-9233-4`.

CICC. 1994. Research on Malaysian dictionary. Technical Report 6---CICC---MT54, Center of the International Cooperation for Computerization, Tokyo.

Fermín L Cruz, José A Troyano, Beatriz Pontes, and F Javier Ortega. 2014. Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*.

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

Gaël Harry Dias, Mohammed Hasanuzzaman, Stéphane Ferrari, and Yann Mathet. 2014. Tempowordnet for sentence time tagging. In *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, WWW Companion '14, pages 833--838. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland. URL `http://dx.doi.org/10.1145/2567948.2579042`.

William Black Horacio Rodríguez Musa Alkhalifa Piek Vossen Adam Pease Elkateb, Sabri and Christiane Fellbaum. 2006. Building a wordnet for Arabic. In *In Proceedings of The fifth international conference on Language Resources and Evaluation (LREC 2006)*.

Christiane Fellbaum and Piek Vossen. 2007. Connecting the universal to the specific: Towards the global grid. In *First International Workshop on Intercultural Collaboration (IWIC-2007)*, pages 2--16. Kyoto.

Christiane Fellbaum and Piek Vossen. 2008. Challenges for a global wordnet. In Webster J., Nancy Ide, and A.Chengyu Fang., editors, *Online Proceedings of the First International Workshop on Global Interoperability for Language Resources (ICGL 2008)*, pages 75--82. City University of Hongkong.

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.

Chu-Ren Huang, Shu-Kai Hsieh, Jia-Fei Hong, Yun-Zhu Chen, I-Li Su, Yong-Xiang Chen, and Sheng-Wei Huang. 2010. Chinese wordnet: Design and implementation of a cross-lingual knowledge processing infrastructure. *Journal of Chinese Information Processing*, 24(2):14--23. (in Chinese).

C. Kunze and L. Lemnitzer. 2002. Germanet --- representation, visualization, application. In *LREC*, pages 1485--1491.

Sidney I. Landau. 1989. *Dictionaries: The Art and Craft of Lexicography*. Cambridge University Press, Cambridge, UK.

John McCrae, Philipp Cimiano, and Elena Montiel-Ponsoda. 2012. Integrating wordnet and wiktionary with lemon. In Christian Chiarcos, Sebastian Nordhoff, and Sebastian Hellman, editors, *Linked Data in Linguistics*. Springer.

George A. Miller. 1990. WordNet: An online lexical database. *International Journal of Lexicography*, 3(4). (Special Issue).

Dan Moldovan and Adrian Novischi. 2004. Word sense disambiguaition of WordNet glosses. *Computer Speech and Language*, 18:301--317.

Eric Nichols, Francis Bond, and Daniel Flickinger. 2005. Robust ontology acquisition from machine-readable dictionaries. In *Proceedings of the International Joint Conference on Artificial Intelligence IJCAI-2005*, pages 1111--1116. Edinburgh.

BoletteSandford Pedersen, Sanni Nimb, Jørg Asmussen, NicolaiHartvig Sørensen, Lars Trap-Jensen, and Henrik Lorentzen. 2009. DanNet --- the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation*, 43(3):269--299.

Wim Peters, Piek Vossen, Pedro Díez-Orzas, and Geert Adriens. 1998. Cross-linguistic alignment of wordnets with an inter-lingual-index. In Vossen (1998), pages 149--251.

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

Maciej Piasecki, Stan Szpakowicz, and Bartosz Broda. 2009. *A Wordnet from the Ground Up*. Wroclaw University of Technology Press. URL `http://www.plwordnet.pwr.wroc.pl/main/content/files/publications/A_Wordnet_from_the_Ground_Up.pdf`, (ISBN 978-83-7493-476-3).

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).

James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press, Cambridge, MA.

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.

Manu Sporny, Dave Longley, Gregg Kellogg, Markus Lanthaler, and Niklas Lindström. 2014. Json-ld 1.0: A json-based serialization for linked data. W3C recommendation, The World Wide Web Consortium.

Dan Tufiş, Dan Cristea, and Sofia Stamou. 2004. BalkaNet: Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science and Technology*, 7(1--2):9--34.

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

Piek Vossen, Francis Bond, and John McCrae. 2016. Toward a truly multilingual global wordnet grid. In *Proceedings of the 8th Global Wordnet Conference (GWC 2016)*. (this volume).

Piek Vossen, Wim Peters, and Julio Gonzalo. 1999. Towards a universal index of meaning. In *Proceedings of ACL-99 Workshop, Siglex-99, Standardizing Lexical Resources*, pages 81--90. Maryland.

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

World Wide Web Consortium W3C. 2012. Turtle --- terse RDF triple language. `http://www.w3.org/TR/2012/WD-turtle-20120710/`.

Shan Wang and Francis Bond. 2013. Building the Chinese Open Wordnet (COW): Starting from core synsets. In *Proceedings of the 11th Workshop on Asian Language Resources, a Workshop at IJCNLP-2013*, pages 10--18. Nagoya.

Aesun Yoon, Soonhee Hwang, Eunroung Lee, and Hyuk-Chul Kwon. 2009. Construction of Korean wordnet KorLex 1.5. *Journal of KIISE: Software and Applications*, 36(1):92--108.