

Problems and Procedures to Make Wordnet Data (Retro)Fit for a Multilingual Dictionary

Martin Benjamin

École Polytechnique Fédérale de Lausanne
IC LSIR, BC 114, Station 14, CH 1015 Lausanne, Switzerland

`martin.benjamin@epfl.ch`

Abstract

The data compiled through many Wordnet projects can be a rich source of seed information for a multilingual dictionary. However, the original Princeton WordNet was not intended as a dictionary *per se*, and spawning other languages from it introduces inherent ambiguity that confounds precise inter-lingual linking. This paper discusses a new presentation of existing Wordnet data that displays joints (distance between predicted links) and substitution (degree of equivalence between confirmed pairs) as a two-tiered horizontal ontology. Improvements to make Wordnet data function as lexicography include term-specific English definitions where the topical synset glosses are inadequate, validation of mappings between each member of an English synset and each member of the synsets from other languages, removal of erroneous translation terms, creation of own-language definitions for the many languages where those are absent, and validation of predicted links between non-English pairs. The paper describes the current state and future directions of a system to crowdsource human review and expansion of Wordnet data, using gamification to build consensus validated, dictionary caliber data for languages now in the Global WordNet as well as new languages that do not have formal Wordnet projects of their own.

1. Introduction

When viewed from the perspective of creating a concept-based multilingual dictionary, the Global WordNet (GWN) is filled with both treasure and risk. The Kamusi Project has imported the freely available data from the Open Multilingual Wordnet (OMW) as seed for further dictionary development. In doing so, we have encountered issues with current Wordnet implementations¹ that we hope to contribute toward resolving.

¹ This paper uses “Wordnet” as a collective noun to signify the web of projects that adopt the synset and ontological approach, and that largely adhere to the same concept set,

Section 2 describes the work we have done to make existing OMW data available in a format that might add value for the public over previous distributions. Section 3 discusses problems encountered with using Wordnet data as the basis for detailed lexicography. Section 4 details the systems we are implementing to (1) offer improved data for current Wordnets and to (2) use as a basis for building parallel data for many more languages.

2. Converting synsets to concept-specific lemmas.

In structuring a multilingual dictionary, Kamusi has determined that each concept/spelling pair within a language should be a distinct node; “light” (not heavy) is different from “light” (not dark) is different from “light” (not serious). This arrangement is compatible overall with the Princeton WordNet (PWN), which separates each sense it has identified for a given English spelling. However, PWN clusters other terms with the same general meaning in the same “synset”, such as {*cloth, fabric, material, textile*}, so part of the conversion of PWN to the Kamusi structure is to make each member a separate node, each linked as a synonym to all others, while retaining for each the Wordnet working definition.

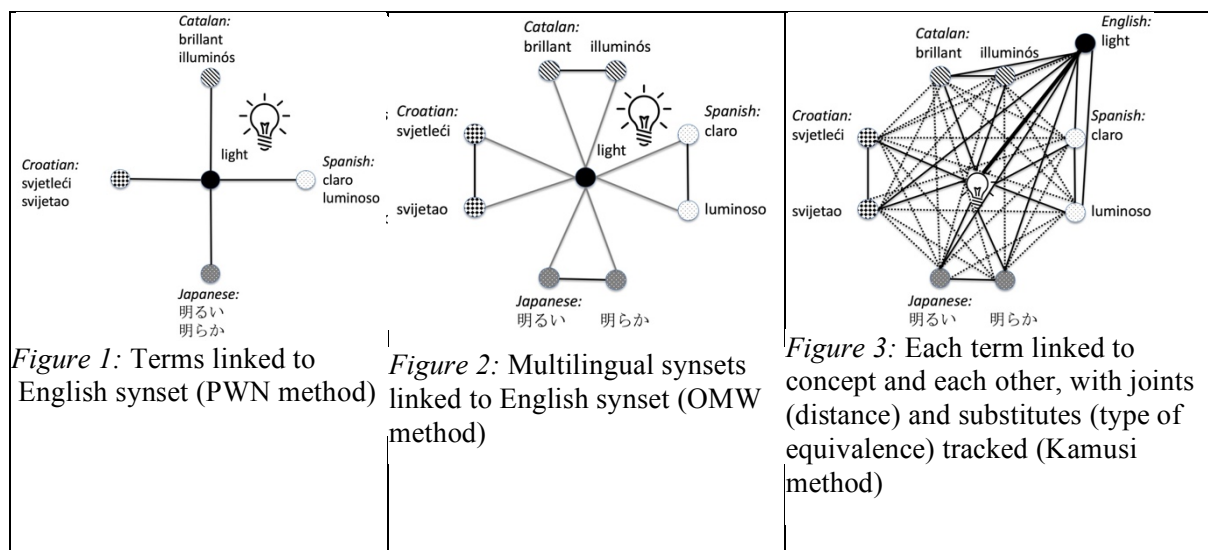
Wordnets for different languages are matched to PWN by synset (Bond and Foster 2013). PWN’s own search engine shows the terms in the OMW that correspond to a synset, marked by language, with no further navigation possible between languages (see figure 1). The OMW search interface better shows the different synsets that are linked to the English concept (see figure 2), and also allows users to seek synsets in a second language that match through English to a search term in a first. For Kamusi, by contrast, the

while also referring to individual Wordnets that exist for specific languages.

matrix of relationships between the individual terms within Wordnet synsets is the multilingual problematic. With English concepts and translation equivalents granted a debatable assumption of validity, Kamusi has now linked the individual terms in the synsets in each language independently, with the matches inferred through English shown as second degree. In the example of “light” (not dark) in figure 3, the concept as defined in English links to two different nodes in Catalan, “brillant” and “illuminós”, and two nodes in Spanish, “claro” and “luminoso”. These particular senses of “claro” and “luminoso” in turn link individually to “brillant” and “illuminós”, and all five of the preceding terms have independently negotiable relationships with Japanese “明るい” and “明らか”, Croatian “svjetleći” and “svijetao”, and

onward through the languages available in OMW. When new terms are matched to the concept in Kamusi for non-Wordnet languages, for example a Quechua equivalent matched to Spanish, links are formed, with degree of separation indicated, to all of the existing terms within the multilingual relation set.

The data from OMW includes 117,659 synsets from PWN, matched to varying amounts among 26 languages and two variants (for Chinese and Norwegian), resulting in approximately 1.2 million individual nodes. Some large relation sets include 150 or more terms as equivalents among languages, which can produce upwards of 11,000 individual links; while server resources have not been expended to tally the total links in the data, at least ten million term pairs have been mapped.



3. Problems with Wordnet as lexicography

Having thus worked at length with the data in OMW, we have encountered a number of limitations that bear mentioning and further work.

It is important to acknowledge that Wordnet was never intended to be a definitive dictionary, for English or any other language. The intent of the word list was to provide data for non linguistic research, initially in psychology (Miller et al 1990, Miller and Fellbaum 2007). It is thus not a criticism to state that it does not fulfill a role it was not designed for. However, in the absence of a better large and well organized set of freely available terms and definitions, it has taken on the de facto role of a universal lexicon, linked not only across languages but also across numerous projects related to computational linguistics. We

suggest that Wordnet can be retrofitted for incorporation within a more lexicographically oriented resource, without losing its strong bonds across languages and projects.

The first problem is that many of the English definitions in the PWN data are inadequate, some to the point of error. Many of the definitions were written by the founder of the project, who was not a lexicographer and was faced with the immense task of producing good-enough ways of understanding tens of thousands of terms. The data is thus peppered with definitions such as “elevator car: *where passengers ride up and down*”; the sense is clear to a knowledgeable speaker, but would not suffice for a credible dictionary. Sometimes the definition is a problem for one member of a synset, either because the terms do not have identical meanings (e.g., verb

“eat, feed: *take in food; used of animals only*” is valid for “feed” but not for “eat”) or because that term forms the nub of the explanation used to define the group (e.g., verb “visit, call, call in: *pay a brief visit*” functions for “call” and “call in”, but is a tautology for “visit”). Some definitions are simply wrong; a law practice, as a lexicalizable multiword expression, is not “*the practice of law*”, but a business through which lawyers conduct their profession.

The consequence of a wrong definition is that the errors propagate through reproductions, projects, and languages. Fixing mistakes is thus an opaque journey through long-completed Wordnet projects that are unlikely to be reopened, in languages that can only be corrected by their speaker communities if they are alerted to the issues and provided with the tools to make the necessary changes. All three languages that attempt an equivalent for “law practice” completely miss the true English sense (perhaps the other 25 groups were too stymied by the tautology to attempt a translation), so Finnish, Thai, and Spanish parties must somehow be alerted that the PWN definition has been modified, and given the platform to review and revise the term in their language. Further, the original PWN definition must be maintained with an indication that it had been deprecated, so projects like BabelNet² and VisuWords³ that link to or build upon it (Navigli and Ponzetto 2010) can see the adjustments flagged, and update themselves accordingly. Unfortunately, numerous websites have replicated the existing PWN data in apparently static form (e.g., vocabulary.com⁴), so the current data will live in many places forever.

The second problem is that many errors exist in the equivalents that other languages map to English. For example, the French word “*lumière*”, always a noun, translates to a few senses of English “light”, mostly in regard to things that shine and figuratively in respect to illuminating knowledge. As rendered in the WOLF French Wordnet, however, “*lumière*” is mapped to 45 senses of “light”, as a noun, verb, or adjective, with meanings such as “insubstantial”, “less than the full amount”, and “alight from (a horse)”. Of similar concern, “light” as visible radiation is mapped to 24 different terms in Polish, and the synset with “illuminate” is given 20 equivalents in

both Indonesian and Malaysian. While most languages have a lively list of expressions for some common concepts such as “goodbye”, large sets of synonyms for most concepts indicate an overly broad brush in the Wordnet compilation. In the Polish example, the purported synonyms include a range of things related to brightness, such as “*zaciemienie*”, which is an eclipse. As with poor English definitions, poor translations and clustering are unlikely to be fixed because their compilation projects have expired with no system in place for updating data.

These issues point to a third problem, a conceptual limitation that our concept-specific rearrangement of the data described above in section 2 seeks to address. A strength of Wordnet, and indeed its main organizing principle, is the highly detailed ontologies through which concepts are related (Vossen et al 1998, Vossen 1998)), such as hyponymy (this is a type of that) and meronymy (this is a part of that), e.g. a ship is a type of vessel and a deck is a part of a ship (Fellbaum 1998). These precise vertical ontologies are not matched, however, with a method for understanding horizontal distinctions within a synset (Derwojedowa et al 2008). Every term within a synset is defined as “this” same thing, e.g. E={approximate, estimate, gauge, guess, judge}, “judge tentatively or form an estimate of (quantities or time),” is all one notion.⁵ Moreover, every term in every synset linked from every other language in GWN is bequeathed with the same meaning, in this example including 6 terms in Croatian, 11 in Japanese including orthographic variations, 20 in Arabic, 22 in Indonesian, and 24 in Malaysian; any term in {ثَمَنٌ , قَضَاءٌ يَأْخُذُكُمْ , رَأْيَا كَانُ , ثَمَنٌ , قَوْلٌ , عَلَى حَكْمٍ , قَوْمٌ , حَزْرٌ , خَمْنٌ , خَمْنٌ , فَصْلٌ , تِبْ أَرْ , قَدْرٌ , قِيمٌ , إِحْكَامٌ , قَدْرٌ , ظَنٌّ , مَاشِي عَسْعَةَ عَيْنٍ , قَلْبٌ , اسْتَنْجٌ} is equivalent to any term in {見立てる, 見積る, 予算+する, 目算, 積もる, 目算+する, 見積もる, 予算, 積る, 推算, 推算+する}. Where the English synset elides the large difference between guessing and gauging, the multilingual composite compounds the weakness of the assumption of strict equivalence. The Arabic terms do not all share a meaning with each other, nor are all the Japanese terms internal synonyms, leaving no way to determine whether اسْتَنْجٌ is a viable translation for 積もる.⁶ Any term produced by a

² <http://babelnet.org/synset?word=bn:00050277n&details=1&orig=law%20practice&lang=EN>

³ <http://visuwords.com/law%20practice>

⁴ <http://www.vocabulary.com/dictionary/law%20practice>

⁵ <http://wordnet-rdf.princeton.edu/wn31/200674352-v>

⁶ To evaluate these two blindly-chosen terms, bilingual informants translated both synsets, yielding information similar to what the processes in section 4 are designed to

contributor in one language has a $1/E$ chance of being a direct translation of one of the English synset members, so any two cross-language terms in GWN have a $1/E^2$ chance of corresponding via the English intermediary with each other; in the example, $E=5$, any thoughtfully-produced term has a 20% of matching a specific term pertaining to assessing amounts, and any two non-English terms have a 4% chance of having been selected as best equivalents of the same English term. Linking the terms computationally is a prodigious shortcut to find likely pairs, but it is not lexicography.

If, however, we see the synset as a grouping of things that share a topical relationship rather than a strict meaning, we can resolve the problem by adding levels of detail similar to the vertical Wordnet ontologies. Kamusi splits the topical lumping of synonymy into what what can be seen as a two-tier horizontal ontology, joints and substitutes, that extends the conceptualization of a multilingual lexicon from a grid (Fellbaum and Vossen 2007) to a matrix.

1. “Joints” is the relationship that shows that terms have been linked transitively as synonyms (synset members) or translations. Joints are evaluated numerically by the degree of separation between links that have, in principle, some element of human confirmation.⁷ A first generation joint indicates that two terms have been manually paired, a second generation joint links though one pivot term, third generation has two intermediary terms, etc. With data from GWN, the presumption of manual linking is cloudy; all members of an English synset have been manually linked to each other, all members of internal synsets for most other languages have been manually linked unless the Wordnet was assembled computationally, and most other-language synsets have been manually linked to the English synset, but that does not mean that

استدنج or 積もる have been manually linked to “guess” or “gauge”. In the current import, joints within a language are all shown as first generation (to be re-filtered as “synonyms” in due course), and joints between each term in an English synset and each member of a linked synset are also shown as first generation, i.e., استدنج is said to be a first generation joint with both guess and gauge, as is 積もる, with the Arabic and Japanese terms therefore set as second generation. A future method to validate joints is described below in section 4.8.

2. “Substitutes” speaks to the degree of equivalence between terms. Whether in-language synonyms or cross-language translations, terms are either “parallel” or “similar”, with the additional possibility that a translation is an “explanatory phrase” invented in one language to fill a lexical gap for a concept that is indigenous to another (Benjamin 2014b). Pending programming will provide fields on Kamusi similar to those for definitions. These fields provide space for the differences between “similar” substitutes to be elaborated, such as the distinction between “arm” in English that is the body part from the shoulder to the wrist versus “mkono” in Swahili that extends from the shoulder to the fingertips. Substitution relationships can in principle be followed across joint relationships, so that the degree of equivalence can be tracked along with the degree of separation, a task for future coding. For the data imported from OMW, all substitution relations have been set initially to “parallel”, putting aside judgments about equivalence for a more distant future.

A fourth limitation with using Wordnet as a dictionary end-product is that it is incomplete in some essential ways. Wordnet cannot be faulted for not including every sense of every English term, much less every term from other languages,

elicit. The Arabic term is substantially more definitive (“concluded”) than the Japanese (“pile up like discussions during an absence”). {1. تُعَن , evaluated; 2. على حكم , judged; 3. قلوب , compared; 4. ثمن , price; 5. رأيا كان , had an idea about; 6. ياحكم , verdict; 7. قِيم , evaluated; 8. قدر , considered; 9. تدار , focused; 10. فصل , separated; 11. خمن , guessed; 12. حَظَن , quantified; 13. حزر , guessed; 14. قوم , measured; 15. استدنج , concluded; 16. قلن , measured; 17. ماشى عسعة عين , set capacity of; 18. ظن , doubted; 19. قتر , evaluated; 20. اچ كم , put to trial}; {1. 見立てる to judge or diagnose [kanji for see and stand up] (make a visual estimation such as a physical exam, or take measurements for clothing); 2. 見積る , 3. 見積もる to estimate [kanji for see and stack] (predict price and time for a job); 4. 予算+する , 5. 予算 to estimate or budget [kanji for calculate and

beforehand] (calculate anticipated expenses); 6. 目算 , 7. 目算+する to estimate [kanji for calculate and look] (an inexact number such as ml in a cup or remaining moves in Go); 8. 積もる , 9. 積る to estimate [kanji for stack] (uncountable things such as snow or emotions); 10. 推算 , 11. 推算+する estimation [kanji for calculate and guess] (less-knowable or unknowable things such as a coin flip, the size of a crowd, or evaluation of a crime scene)}.

⁷ This assumption does not necessarily hold, as some Wordnets are built using automatic generation techniques (Atserias et al 1997, de Melo and Weikum 2008, Oliver 2014). The tendency for error in computationally-derived datasets is amply displayed WOLF French Wordnet (Wordnet Libre du Français) (Sagot and Fišer 2012, <http://alpage.inria.fr/~sagot/wolf-en.html>)

as that was never its mission. However, terms or senses that are not in Wordnet, such as “light” as a traffic signal, or “lightsaber”, should be included – or at least includable – in a dictionary that aspires toward a thorough representation of a language. If a concept is missing in PWN, moreover, it stands little chance of appearing in other language Wordnets, and conversely there is no chance for a concept indigenous to another language to join the global Wordnet concept set. Within the scope of the Wordnet vision, relationships that have not been found by Wordnet editors cannot be forged by readers, such as proposing that “boat” and “ship” be joined in a synset. Further, the lack of own-language definitions in most languages leaves the impression that the meaning of each term can be encapsulated in the English definition of the corresponding synset, to the extent that the attributed definition for “zaćmienie” is, exactly and erroneously, “electromagnetic radiation that can produce a visual sensation”. Finally, and again because it is out of scope, Wordnet does not include a great deal of information that is relevant for dictionary or data purposes, such as word forms (Spanish “invitado” does not indicate an association with “invitada”, “invitados”, and “invitadas”).

A final limitation with Wordnet is that projects for many languages have licenses that restrict the use of the data, if the data can be located at all. For example, the Romanian Wordnet is distributed with a “no derivatives” license. This means that the data cannot be imported into the multilingual structure described above, because linking Romanian to Slovenian would be a derivative product. Nor could the data be expanded, with Romanian definitions or with information such as the female form “invitată” corresponding to the given masculine “invitat”. Furthermore, the Romanian data has a “no redistribution” restriction, so its use in a project that makes its data shareable or downloadable seems proscribed. GermaNet is even more restrictive, only allowing the data to be used for internal research within an institution. The openness or lack thereof of Wordnets is indicated at <http://globalwordnet.org/wordnets-in-the-world>. Bringing restricted Wordnets into a dictionary project does not offer new technical challenges, but is only possible if the creators choose to amend their licenses.

4. Tools and techniques for adding and improving Wordnet data

Wordnet’s popularity stems in part from its openness to the mash-ups others create from the core PWN data. In that spirit, Kamusi has developed tools that will transform the open Wordnet data into data that is appropriate for dictionaries and additional technological applications, using automated procedures as a starting point for human lexicographic review (Pianta, Bentivogli, and Girardi 2002). At the same time, these tools are designed to keep the data in synch with existing Wordnet instances, in such a way that transformations generated by Kamusi can be reincorporated in PWN or other language projects when and if their maintainers desire.

The primary new tools developed by Kamusi that can transform Wordnet data are a set of crowd-sourcing applications that include games embedded within Facebook and (still in alpha development) on mobile devices (Benjamin 2014a, Benjamin 2015). These games ask players to answer targeted questions about their language, for which they receive various rewards when their answers adhere to the consensus. The games build data progressively, such that a definition that has been approved for English can be shown to people producing equivalents or definitions for other languages.

These systems can transform Wordnet seed data into dictionary data, in several ways:

1. Each English definition will be reviewed as it pertains to the individual members of a synset, and improved when the participants find it appropriate. Players are shown the existing Wordnet “working definition”, and given the opportunity to either suggest their own definition, vote for the Wordnet definition, or vote for a contribution from another player. Once a definition passes the consensus threshold, it is published to Kamusi and used for subsequent game modes. If the Wordnet definition has been replaced, it is shown on Kamusi as deprecated.
2. Definitions in their own languages for terms from other Wordnets will be generated using the same procedure. This feature will be introduced after players have had the chance to validate existing translations against a critical mass of finalized English definitions, e.g. a new English definition for “law practice” will first be given to Spanish speakers to verify or replace the current matched Spanish term, and only afterwards will the approved Spanish term be advanced to the definition game.

3. Existing translations of PWN will be validated term by term. For example, Polish players will assuredly approve “światło” for the sense of visible light, but reject “zaćmienie”. This mode has not been developed at time of writing, the need only becoming evident through examination of the data imported in mid 2015, but is anticipated for quick completion. Terms that are evicted from a defined synset, like “zaćmienie”, will be moved through a sequence of games to produce definitions, translations, and sense matches.

4. Concepts from PWN that are not already matched in other Wordnet languages will be elicited. For example, the Arabic WordNet has only 10,000 synsets, so more than 100,000 concepts remain untouched. In the game, players are shown a defined English term and asked to provide an equivalent term in their language. Terms that pass the consensus threshold are added to Kamusi, while non-winning terms are passed to another mode to see whether they are synonyms for the concept.

5. Languages that do not have existing Wordnet projects will be opened to their speakers, using the improved English definition set and the game modes described above. Because the elicitation list used in the games is inherently linked to Wordnet, Wordnets for these other languages will be created as a default outcome. This opens GWN to languages that do not have formal organizations to take on the trouble of creating a Wordnet project, including building tools from scratch (e.g. Wijesiri et al 2014), but do have passionate speakers who will contribute through crowd methods.

6. Languages that have existing but restricted Wordnet projects, like German, will be opened for their speakers to start from scratch. This is a phenomenal waste of time and energy, if one can speak frankly in an academic paper, but, barring changes in license restrictions, may be the fastest way to acquire reliable data that can be used in an open resource.

7. One already-developed game calls on players to judge whether usages gleaned from Twitter or more formal corpora (currently configured for Wikipedia and the Helsinki Corpus of Swahili, but the technique can be applied more widely) are good examples to illustrate a particular sense. Most Wordnets lack usage examples, so this game can fill that gap for many languages. Future game modes will elicit additional lexical and ontological information, some of which falls

within the scope of what is sought within Wordnets.

8. A future game mode, which will be activated after languages have sufficient numbers of defined entries, will ask users to confirm joints established through English for their language pairs. For example, “światło” and “lumière” will be shown with their respective own-language definitions, and a registered Polish/ French speaker will vote whether the two concepts match. This game can only be played after sufficient data for the concerned languages has been gathered in the English-confirmation mode described above in paragraph 4.3. The result will be validated aligned Wordnets for numerous language pairs.

9. Work on other tracks within Kamusi will introduce many terms and senses that are not part of PWN or other Wordnets. These concepts will be made available to language teams, and some could form part of an extended multilingual Wordnet desiderata.

5. Conclusions

This paper has discussed two difficulties with using Global Wordnet as the source for a formal multilingual dictionary. First, Wordnet does not do things it was not intended to do, but that are needed for lexicography, such as differentiation of terms grouped topically in synsets and matching those concept distinctions across languages. Second, some of the things it does do bear improvement, either in quantity (completion of the full PWN set of synsets in other languages, production of own-language definitions), quality, or access. Fortunately, the open approach with which Wordnet was designed makes it possible to retrofit the data with English definitions that may be more sensible than those initially drafted, and with revised equivalents in other languages when necessary, without severing the bonds that have already been built across languages and projects. The broad inter-lingual predictions made possible by GWN have been refined by charting the joints between members of a topical group, and will further show the degree to which confirmed pairs can substitute for each other. The work will not be easy, involving recruiting many crowd members from many languages, as well as oversight from authoritative arbiters. However, many of the tools have already been developed, and are being rolled out gradually as Kamusi musters the resources to foster speaker communities and manage the incoming data flow. As time goes on, the data produced by various Wordnet projects will lie at

the core of a more comprehensive multilingual dictionary, and the data from the dictionary project will be available for the further refinement of existing and future Wordnets.

References

- Jordi Atserias, Salvador Climent, Xavier Farreres, German Rigau, and Horacio Rodriguez. 1997. Combining multiple methods for the automatic construction of multi-lingual WordNets. In *Recent Advances in Natural Language Processing II. Selected papers from RANLP*, volume 97, pages 327–338.
- Martin Benjamin. 2014a. Collaboration in the Production of a Massively Multilingual Lexicon. In *LREC 2014 Conference Proceedings*. Reykjavik (Iceland).
- Martin Benjamin. 2014b. Elephant Beer and Shinto Gates: Managing Similar Concepts in a Multilingual Database. In *Proceedings of the 7th International Global WordNet Conference*. Tartu (Estonia)
- Martin Benjamin. 2015. Crowdsourcing Microdata for Cost-Effective and Reliable Lexicography. In *AsiaLex 2015 Conference Proceedings*, Hong Kong.
- Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *51st Annual Meeting of the Association for Computational Linguistics: ACL-2013*, pages 1352–1362.
- Magdalena Derwojedowa, Maciej Piasecki, Stanisław Szpakowicz, Magdalena Zawislavska, and Bartosz Broda. 2008. Words, Concepts and Relations in the Construction of Polish WordNet. In *Proceedings of the Global WordNet Conference 2008*, Szeged (Hungary), pages 167–68.
- Christiane Fellbaum. 1998, ed.. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Christiane Fellbaum and Piek Vossen. 2007. Connecting the Universal to the Specific: Towards the Global Grid. In *Proceedings of the First International Workshop on Intercultural Communication*, Kyoto (Japan).
- Gerard de Melo and Gerhard Weikum. 2008. On the utility of automatically generated wordnets. In *Proceedings of 4th Global WordNet Conference, GWC 2008*, Szeged, Hungary. University of Szeged. Pages 147–161.
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to wordnet: an on-line lexical database. *International Journal of Lexicography*, 3(4):235–244.
- George A. Miller and Christiane Fellbaum. 2007. WordNet then and now. In *Language Resources and Evaluation*, 41:209–214.
- Roberto Navigli and Simone Paolo Ponzetto. 2010. *BabelNet: building a very large multilingual semantic network*. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL '10, pages 216–225, Stroudsburg, PA, USA. Association for Computational Linguistics. ACM ID: 1858704.
- Antoni Oliver. 2014. WN-Toolkit: Automatic generation of WordNets following the expand model. In *Proceedings of the 7th International Global WordNet Conference*, Tartu (Estonia), pages 7–15.
- Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: Developing an aligned multilingual database. In *Proceedings of the 1st Int'l Conference on Global WordNet*, Mysore (India), 293–302.
- Benoît Sagot and Darja Fišer. 2012. Automatic Extension of WOLF. In *Proceedings of the 6th International Global Wordnet Conference*, Matsue, (Japan)
- Piek Vossen, Laura Bloksma, Horacio Rodriguez, Salvador Climent, Nicoletta Calzolari, Adriana Roventini, Francesca Bertagna, Antonietta Alonge, and Wim Peters. 1998. *The EuroWordNet Base Concepts and Top Ontology*. Deliverable D017 D, 34:D036.
- Piek Vossen. 1998. Introduction to EuroWordNet. *Computers and the Humanities*, 32(2-3):73–89.
- Indeewari Wijesiri et al. 2014. Building a WordNet for Sinhala. In *Proceedings of the 7th International Global WordNet Conference*. Tartu (Estonia)