

Language Resources and Evaluation manuscript No.
(will be inserted by the editor)

Is It Possible to Create a Very Large WordNet in 100 days? – an Evaluation

Krister Lindén · Jyrki Niemi

Received: date / Accepted: date

Abstract Wordnets are large-scale lexical databases of related words and concepts, useful for language-aware software applications. They have recently been built for many languages by using various approaches.

The Finnish wordnet, FinnWordNet (FiWN), was created by translating the more than 200,000 word senses in the English Princeton WordNet (PWN) 3.0 in 100 days. To ensure quality, they were translated by professional translators. The direct translation approach was based on the assumption that most synsets in PWN represent language-independent real-world concepts. Thus also the semantic relations between synsets were assumed mostly language-independent, so the structure of PWN could be reused as well. This approach allowed the creation of an extensive Finnish wordnet directly aligned with PWN and also provided us with a translation relation and thus a bilingual wordnet usable as a dictionary.

In this paper, we address several concerns raised with regard to our approach in one single paper, many of them for the first time. We evaluate the craftsmanship of the translators by checking the spelling and translation quality, the viability of the approach by assessing the synonym quality both on the lexeme and concept level, as well as the usefulness of the resulting lexical resource both for humans and in a language-technological task. We discovered no new problems compared with those already known in PWN. As a whole, the paper contributes to the scientific discourse on what it takes to create a very large wordnet.

As a side-effect of the evaluation, we extended FiWN to contain 208,645 word senses in 120,449 synsets, effectively making version 2.0 of FiWN the currently largest wordnet in the world by these statistics.

Keywords WordNet · quality assessment · knowledge representation · word-sense discovery

K. Lindén · J. Niemi

Department of Modern Languages, University of Helsinki, PO Box 24, FI-00014 University of Helsinki
E-mail: krister.linden@helsinki.fi, jyrki.niemi@helsinki.fi

1 Introduction

FinnWordNet (FiWN) is a wordnet for Finnish that conforms to the framework presented in Fellbaum (1998) and Vossen (1998). The synset structure of FiWN corresponds closely to that of the English Princeton WordNet (PWN), version 3.0. FiWN is open source.¹

A classic wordnet consists of synsets, or sets of partial synonyms whose shared meaning is described and exemplified by a gloss, a common part of speech and a hyperonym. Synsets in a wordnet are arranged in hierarchical partial orderings according to semantic relations such as hyponymy/hyperonymy. Together, the gloss, part of speech and hyperonym fix the meaning of a word and constrain the possible translations of a word in a given synset.

The Finnish group opted for translating PWN 3.0 synsets (word senses) wholesale into Finnish by professional translators, because the translation process could be controlled with regard to quality, coverage, cost and speed of translation. All the more than 200,000 word senses were translated, including multi-word phrases but excluding some proper names. The translation process, outlined and discussed by Lindén and Carlson (2010), was diligent, and the quality was on a par with PWN.

The direct translation approach was based on the assumption that most synsets in PWN represent language-independent real-world concepts. Thus also the semantic relations between synsets are assumed to be mostly independent of the language, so the structure of PWN can be reused as well. This approach made it possible to create an extensive Finnish wordnet directly aligned with PWN. The direct translation of PWN word senses from English into Finnish also provided us with a translation relation and thus a bilingual wordnet that can also be used as a bilingual dictionary.

Version 2.0 of FiWN, released in October 2012, is based on the evaluation results of the original version 1.0 and a few intermediate versions. FiWN 2.0 contains 120,449 synsets, 208,645 word senses, 142,078 unique word literals and 244,742 translation relations. As of version 2.0, FiWN has also been extended beyond being a translation of PWN by adding new synsets as hyponyms of existing synsets (without glosses and English translations). The new synsets correspond to senses of common Finnish compound words whose final part is in the hypernym synset, as suggested by Pääkkö and Lindén (2012).

The number of modified items after the evaluation process is shown in Table 1. A total of 18,898 (16.1%) of the 117,659 original synsets were affected. Some of the additions and corrections were based on suggestions received from users of FiWN.

The total number of synsets, word senses and other items in FiWN 2.0 compared with the original version 1.0 and with PWN 3.0 are shown in Table 2. After having been the second largest wordnet in the world after PWN, with version 2.0, FiWN is now effectively the largest one by the number of synsets and word senses. PWN 3.0 still wins in the number of unique words and lexical relations.

An effort to build wordnets for the Nordic and Baltic languages is also in progress, see Pedersen et al. (2012). The builders of these wordnets have applied different compilation strategies: The Danish, Icelandic and Swedish wordnets are being developed

¹ FiWN can be downloaded from <http://www.ling.helsinki.fi/en/lt/research/finnwordnet/>

Table 1 The number of items changed between FiWN version 1.0 and 2.0. In unique words, homographs with different parts of speech are counted as different

	added	removed	corrected
synsets	2,790	0	–
word senses	19,754	2,106	3,710
unique words	14,070	5,247	–
translation relations	22,635	6,490	–

Table 2 The number of items in FiWN 2.0 compared with version 1.0 and PWN 3.0

	FiWN 2.0	FiWN 1.0	increase	%	PWN 3.0
synsets	120,449	117,659	2,790	2.4	117,659
word senses	208,645	190,997	17,648	9.2	206,978
unique words	142,078	133,255	8,823	6.6	156,584
semantic relations	159,330	156,540	2,790	1.8	156,540
lexical relations	68,047	59,751	8,296	13.9	91,800
translation relations	244,742	228,597	16,145	7.1	–

via monolingual dictionaries and corpora and subsequently linked to PWN. In contrast, the Finnish and Norwegian wordnets are applying the expand method by translating from PWN and the Danish wordnet, DanNet, respectively. The Estonian wordnet was built as part of the EuroWordNet project (Vossen 1998) and by translating the base concepts from English as a first basis for monolingual extension. The wordnets are displayed in the WordTies² browser, which currently includes the Danish, Finnish, Swedish and Estonian wordnets, which have been linked to Princeton Core WordNet, thereby providing a linked coverage of almost 5,000 core synsets. The aim is to test the perspective of a multilingual linking of the Nordic and Baltic wordnets and via the linking to perform a tentative comparison and validation of the wordnets along the measures of taxonomical structure, coverage, granularity and completeness.

From a lexicographic point of view, our translation approach was considered impossible and unable to result in anything useful, whereas our main concern was that it would be a waste of several decades of lexicographic work to rebuild the concept relations from scratch monolingually. After the translation process was finished, we therefore set out to test the quality of FiWN in various ways. The present evaluation of FiWN covers various aspects of the craftsmanship, the viability of the translation approach and the usefulness of the result.

2 Evaluation of Craftsmanship

To evaluate the craftsmanship of the translators, we looked at possible spelling errors and part-of-speech mismatches as a first indicator of how carefully the work had been carried out. The translation quality also needed to be evaluated.

² <http://wordties.cst.dk/>

2.1 Evaluation of Spelling and Preserving Part of Speech

To evaluate the spelling and the part of speech. We ran the words in FiWN through a Finnish morphological analyzer. Unrecognized words were further inspected and the spelling was corrected or the words were added to the morphological analyzer. In the next round, we checked that each FiWN word had the expected base form and part of speech. Words with a mismatch with the expectation were manually corrected. We corrected 1,487 words on the first stage, and 3,217 on the second one; in total 4,707 (3.6%) of the original 130,412 words.

Not all words in the synsets were translatable by equivalent terms. This was foreseen by allowing the translators to use standardized comments to indicate how the translation relates to the original. A total of 3,230 translations (1.4% of all the 228,597 translation relations) were marked broader than (hyperonym of) the original, 165 (0.1%) narrower (hyponym), 2,471 (1.0%) approximate (near-synonym), 4,099 (1.8%) unconfirmed, 1065 (0.5%) as having an idiomatic part-of-speech difference, and 527 (0.2%) had been given a free-form comment.

The translators all worked on separate sections, so no interannotator agreement was measured. For other translation quality controls, see Lindén and Carlson (2010).

2.2 Synonym Translation Quality

In order to evaluate the translation quality, we extracted translation pairs from Wikipedia and Wiktionary (Lindén et al. 2012; Niemi et al. 2012). The potentially relevant ones had an English word found in PWN with a Finnish translation in Wikipedia or Wiktionary. Even though Wiktionary is a less extensive resource than Wikipedia, it provided almost twice the number of translation pair candidates, 58,110 and 25,062, respectively. To evaluate the quality of the manual translation work, we only considered the words in FiWN that may need to be replaced. The number of translations that should replace ones existing in FiWN was 104 (0.4% of 25,062) from Wikipedia and 171 (0.3% of 58,110) from Wiktionary. From the small number of words to be replaced, we can conclude that the translation quality of FiWN from PWN is high, i.e. there were few direct errors.

3 Evaluation of Viability of Approach

With the help of well-intended critique from Martola (2011), we evaluated the viability of the translation approach by investigating the synonym quality both on the concept and the lexeme level as well as some structural differences and potentially missing culture-specific words.

3.1 Concept Quality

The quality of concept linking was evaluated by rating 1,000 synsets linked to Finnish from Swedish and Danish via Core WordNet. The task was not only to evaluate the

Finnish linking to Core WordNet, but more importantly to evaluate if the translations using Core WordNet as a pivot had been understood in the same way in the three languages and therefore conveyed the same meaning in the other languages as well. This also evaluates the general approach of using English synsets as a basis for translation and synsets in other languages. The general finding was that from a Finnish perspective the links to Danish and Swedish via Core WordNet are almost all, i.e. 99.3% (993/1,000) of good quality (rated as 3 on a scale of 1 to 3) for both the Finnish–Danish and Finnish–Swedish translations. Some synsets, whose concept linking as a whole was considered good, still got suggestions for further improvement by adding or removing individual words. The remaining 7 words were rated as 2 out of 3.

3.2 Synonym Set Quality

For evaluating the quality of the synonym sets, we collected user feedback through a crowdsourcing procedure where the users were asked to rate (on a scale of 1 to 5) the quality of synonyms presented to them in the FiWN Web search interface with the context of an English gloss, part of speech and hyperonym. The ratings did not directly affect the data; instead, synonyms receiving poor grades were manually examined and corrected where appropriate.

Users submitted 1,237 synonym quality ratings between December 2010 and October 2012, when the evaluation form was available.³ The majority of the evaluations were positive: the users gave 758 (61.3%) good grades (4 or 5), more than twice that of poor grades (1 or 2), 317 (25.6%). The average of the grades was 3.6.

3.3 Structural Differences

Since FiWN is translated from PWN, they do not differ structurally from each other, except for some derivational links, removed between lexemes with no derivational relationship in Finnish (Niemi and Lindén 2012, 2443). It is a future task to add new such links between Finnish words with an obvious derivational relationship.

Already in their initial work, Lindén and Carlson (2010) pointed out some of the translation mismatches due to differing semantic fields in the languages, but many of them were dealt with through to the fine-grained structure of PWN.

Sometimes the translation has resulted in the same word being included in a synset and its hyperonym synset, effectively stating that a word is its own hyperonym. Sometimes this is called for when a word has a specific meaning as well as a collective meaning, but in other cases it remains a future task to go through the lexicon and to decide on the right level of abstraction for each word.

3.4 Culture-Specific Words

To evaluate the number of missing Finnish culture-specific words in FiWN, we took all the Finnish words from a text corpus of approx. 170 million words that were

³ The figure excludes the suggestions for additional synonyms and “not sure” answers.

equally frequent⁴ as the words already in FiWN. Among the 6,388 words missing from FiWN and not found in Wiktionary, we looked more closely at a sample and found that only a small portion of it was culture-specific.

The culture-specific words in our sample included words for inhabitants of Finnish towns and provinces (e.g. *helsinkiäinen*, ‘inhabitant of Helsinki’), names of or other words related to Finnish institutions (e.g. *lääninhallitus*, a former governing body of a Finnish province) and names of traditional Finnish dishes (e.g. *mämmi*, an Easter-time dessert). Based on the sample, we estimated that 309 ± 187 relatively frequent culture-specific words are missing from FiWN. Adding them would increase the number of words by $0.2 \pm 0.1\%$ from 142,078, see Lindén et al. (2012).

Rarer words may or may not be more culture-specific, e.g., old agricultural vocabulary. The importance of them depends on the intended use of the resource.

4 Usefulness of FinnWordNet

4.1 Coverage Evaluation

We tested the coverage of FiWN using a large text corpus of Finnish newspaper text. The words in FiWN 2.0 cover 57.3% of the all words in the running text, up from 54.5% in version 1.0. If we only count nouns, verbs, adjectives and adverbs and exclude proper names, the coverage is 82.4% of running text, up from 78.4% in version 1.0.

We also evaluated the coverage of FiWN 2.0 senses by manually disambiguating a sample corpus of 846 words of text annotated with FiWN senses. We found 579 word occurrences in this sample that were also in FiWN. Of these, 97 word occurrences (16.8%) had no appropriate meaning in context. This indicates that FiWN 2.0 covers approximately 57% (482 / 846) of all the word meanings occurring in running text.

4.2 Applicability in an NLP Task

To see if FiWN would be useful for distinguishing between word senses, we took the UKB software, which uses a wordnet (or similar) to perform word-sense disambiguation and word-sense tagging of running text (Agirre and Soroa 2009).

In addition, we took the small coverage test corpus of 846 words with the manually FiWN-annotated 579 words. The manual annotation was done independently by two linguists. Initially, they disagreed on 10% of the cases but after comparing notes, they reached on a common conclusion. This was used as the gold standard. The words in the gold standard corpus had 12.1 different senses on the average whereas the FiWN words in a newspaper text have 11.7.

Preliminary results show that using the personalized PageRank method with a sentence context in UKB using FiWN gives a disambiguation result of 57.0% for the first suggestion. The top two suggestions give 70.3% and the top three 81.4%. This

⁴ “Equally frequent” meant that the words in FiWN had an average frequency in the text corpus and that we used all the words in the text corpus with a least a quarter of this average frequency.

should be compared with the baseline of 23% assuming an equal distribution for all the word senses. The results are also on a par with those of Agirre and Soroa (2009) for English using UKB with PWN.

Since we have a bilingual wordnet, it is possible to use data in which English words can serve as disambiguation context, e.g. in parallel corpora if we consider the aligned translation. The 846-word gold standard text was chosen so that an aligned English translation was available. If the aligned English sentence is also considered as part of the context, the disambiguation result rises to 63% correct disambiguations in the first suggestion.

5 Discussion

Our main motivation for creating FiWN by translation was the considerable saving in construction time. Due to the objections by lexicographers regarding the viability of the translation approach, see Martola (2011), a sizable effort has now also gone into evaluating FiWN from as many angles as possible to see whether the lexicographic concerns are real or merely objections that one could raise against any new lexicon.

We supervised the craftsmanship of the translators already during the creation of FiWN, see Lindén and Carlson (2010). To further verify the basic quality, we looked at potential typos in Finnish as well as part-of-speech mismatches between translated synsets. We found no major problems and corrected the few remaining ones.

A greater effort went into evaluating the viability of our approach by studying the synonym and translation quality both on the lexeme and the concept level as well as the lexical coverage and some structural differences. Again we discovered no major problems but the effort provided us with valuable data for further extending FiWN. The missing culture-specific words were few and of low frequency. Rare words are by definition unlikely to be a major problem.

Depending on lexicographic perspective and tradition, lower and more course-grained lexical structure for the synonym hierarchy could perhaps be preferable. In addition, the adjectives could benefit from a treatment more similar to other words. This updating effort is under way in PWN, and a similar effort will then be used in FiWN. Also, the verbs could benefit from a more grammatically motivated restructuring of the hierarchy with regard to transitivity both in Finnish and English.

Initially, we also confirmed the usefulness of FiWN for humans as a large-scale bilingual dictionary, see Muhonen and Lindén (2011). To confirm the usefulness of our work for machines, we also needed to evaluate FiWN in some language-technology application. We used the all-words unsupervised word-sense disambiguation task for this purpose and found that the performance was on a par with similar experiments using PWN.

Further work to take advantage of the bilingual aspect of FiWN in word-sense disambiguation will be carried out. This will be pursued in order to provide FinnTree-Bank, the Finnish parse bank, with word-sense annotation, as it has parallel data in all the languages of the European Union. Another interesting application experiment left for further work would be to use FiWN for bilingual corpus alignment.

6 Conclusion

The overall conclusion is that it is possible to create a wordnet for a new language in 100 days, as we discovered no hidden problems compared with those already known in PWN. As a side-effect of the evaluation, we were able to extend FiWN to contain a total of 208,645 word senses in 120,449 synsets, i.e. after having been the world's second largest wordnet, it is now effectively the largest according to these statistics. In addition, we found a promising avenue of further research using FiWN for bilingual word-sense annotation.

Acknowledgements We are grateful to Mirka Hyvärinen, Kristiina Muhonen and Paula Pääkkö for checking the long lists of words with potential spelling errors and part-of-speech mismatches. We also thank Mirka Hyvärinen and Pinja Pennala for their valuable contribution to the creation of the word-sense disambiguated test corpus and for the many hours spent on evaluating sets of words extracted from Wikipedia and Wiktionary. Mirka Hyvärinen also conducted the crowdsourcing experiment.

This work was funded by the FIN-CLARIN and META-NORD projects. The META-NORD project has received funding from the European Union's ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme under grant agreement no. 270899.

References

- Agirre, E., & Soroa, A. (2009). Personalizing PageRank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2009)*, (pp. 33–41). Athens, Greece: Association for Computational Linguistics. <http://www.aclweb.org/anthology/E09-1005>.
- Fellbaum, C. (Ed.) (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Global WordNet Association (2012). *Proceedings of the 6th International Global Wordnet Conference (GWC 2012)*. Matsue, Japan: Global WordNet Association.
- Lindén, K., & Carlson, L. (2010). FinnWordNet – WordNet på finska via översättning. *LexicoNordica – Nordic Journal of Lexicography*, 17, 119–140.
- Lindén, K., Niemi, J., & Hyvärinen, M. (2012). Extending and updating the Finnish Wordnet. In D. Santos, K. Lindén, & W. Ng'ang'a (Eds.), *Shall We Play the Festschrift Game? Essays on the Occasion of Lauri Carlson's 60th Birthday*, (pp. 67–98). Berlin / Heidelberg: Springer. doi: 10.1007/978-3-642-30773-7_7.
- Martola, N. (2011). FinnWordNet och kulturbundna ord. *LexicoNordica – Nordic Journal of Lexicography*, 18, 111–133.
- Muhonen, K., & Lindén, K. (2011). Do wordnets also improve human performance on NLP tasks? In B. S. Pedersen, G. Nešpore, & I. Skadiņa (Eds.), *Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011, NEALT Proceedings Series*, volume 11, (pp. 146–152). Northern European Association for Language Technology (NEALT). <http://hdl.handle.net/10062/16955>.
- Niemi, J., & Lindén, K. (2012). Representing the translation relation in a bilingual wordnet. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, (pp. 2439–2446). Istanbul, Turkey. <http://www.lrec-conf.org/proceedings/lrec2012/summaries/194.html>.
- Niemi, J., Lindén, K., & Hyvärinen, M. (2012). Using a bilingual resource to add synonyms to a wordnet: FinnWordNet and Wikipedia as an example. In Global WordNet Association (2012), (pp. 227–231).
- Pääkkö, P., & Lindén, K. (2012). Finding a location for a new word in WordNet. In Global WordNet Association (2012), (pp. 286–293).
- Pedersen, B. S., Borin, L., Forsberg, M., Lindén, K., Orav, H., & Rognvaldsson, E. (2012). Linking and validating Nordic and Baltic wordnets: A multilingual action in META-NORD. In Global WordNet Association (2012), (pp. 254–260).
- Vossen, P. (Ed.) (1998). *EuroWordNet: A multilingual database with lexical semantic networks*. Dordrecht: Kluwer Academic Publishers.