

Tuning Hierarchies in Princeton WordNet

Ahti Lohk¹, Christiane D. Fellbaum² and Leo Vöhandu¹

¹Department of Informatics, Tallinn University of Technology, Tallinn, Estonia

²Department of Computer Science, Princeton University, New Jersey, USA

<ahti.lohk@ttu.ee, fellbaum@princeton.edu, leo.vohandu@ttu.ee>

Abstract

Many new wordnets in the world are constantly created and most take the original Princeton WordNet (PWN) as their starting point. This arguably central position imposes a responsibility on PWN to ensure that its structure is clean and consistent. To validate PWN hierarchical structures we propose the application of a system of test patterns. In this paper, we report on how to validate the PWN hierarchies using the system of test patterns. In sum, test patterns provide lexicographers with a very powerful tool, which we hope will be adopted by the global wordnet community.

1 Introduction and background

Many new wordnets in the world are constantly created and most take the original Princeton WordNet (PWN) as their starting point. This arguably central position imposes a responsibility on PWN to ensure that its structure is clean and consistent. This is particularly true for hierarchical relations, which are the most frequently encoded relations and which form the backbone of the network. To validate PWN hierarchical structures we propose the application of a system of test patterns developed in (Lohk, 2015). Importantly, all instances returned by the test pattern system were manually validated by two members of the Estonian Wordnet (EstWN) team (Kadri Vare and Heili Orav). The results were encouraging, and we applied the algorithms to PWN. We propose that after a few iterations on PWN other wordnets apply the algorithm on their resources and, after a couple of iterations, compare their structures with that of PWN, which can serve as some kind of Gold Standard for wordnets. Alternatively, the analysis is commercially available from the first author.

In this paper we report on how to validate the PWN hierarchies using the system of test patterns.

A test pattern is a description of a specific substructure in the wordnet hierarchy. The system of test patterns and the descriptions of all patterns are

found in (Lohk, 2015). This system consists of ten test patterns that all involve multiple inheritance, an important property that can point to different semantic inaccuracies going back to lexicographic errors. Because they are semantic, every test pattern applies cross-lingually and sheds new light on wordnets by examining their hierarchies and helping to detect and correct possible errors.

These patterns were used to validate the semantic hierarchies of Estonian Wordnet over four years (2011–2014) and on ten versions. During this time, the structure of Estonian Wordnet changed significantly, as described in Section 3.

The aim of this paper is to show that the same specific substructures that have been found in Estonian Wordnet also exist in Princeton WordNet. Moreover, some experiments on Princeton WordNet confirm the promising benefits of test pattern application (Section 4). Therefore, we propose test patterns as a method for validation and tuning hierarchies in PWN and all other wordnets.

This paper is structured as follows: Section 2 provides an overview of the validation methods applied to the wordnet hierarchies. Section 3 presents the results of using test patterns iteratively on EstWN. Section 4 demonstrates that the same pattern instances can be found in PWN as well as in other wordnets. Some experiments are described. We close with a conclusion and proposals for future work.

2 State of the art in validating the semantic hierarchies of wordnet

To give a better understanding of the test patterns approach we provide a short overview of the validation methods applied on the semantic hierarchies of wordnet. (Lohk 2015) argues that the methods can be divided into three groups based on two features, as shown in Table 1. These features can be formulated as questions as follows: *do they rely on corpus data and lexical resources? Do they make use the contents of a synset?*

Group of methods	use of corpus data, lexical resources	use the contents of a synset
Group I	+	+
Group II	–	+
Group III	–	–

Table 1: Features that classify a group of validating methods

Group I comprises all methods based on *lexical resources and corpora*; group II includes rules or rule-based methods, while group III consists of graph-based methods.

2.1 Corpus-based methods

The most frequently used validation methods for wordnet hierarchies rely on corpora and lexical resources. Different techniques for extracting the relevant information have been applied. Some of the well-known approaches include:

- Lexico-syntactic patterns (Hearst, 1992), (Nadig et al., 2008)
- Similarity measurements (Sagot and Fišer, 2012)
- Mapping and comparing to wordnet (Pedersen et al., others, 2013)
- Applying wordnet in NLP tasks (Saito et al., 2002)

Resources used in this group of methods are:

- Monolingual text corpora (Sagot and Fišer, 2012)
- Bilingual aligned corpora (Krstev et al., 2003)
- Monolingual explanatory dictionaries (Nadig et al., 2008)
- Wordnets (Peters et al., 1998; Pedersen et al., 2012)
- Ontologies (Gangemi et al., 2002)

2.2 Rule-based methods

These methods for validating hierarchies rely on lexical relations (word-word), semantic relations (concept-concept) and the rules among them. This includes the rules applied to the construction of WordNet (Fellbaum, 1998), and additional rules, such as the following:

- Metaproperties (*rigidity, identity, unity and dependence*) described in ontology construction (Guarino and Welty, 2002)

- Top Ontology concepts or “unique beginners” (*Object, Substance, Plant, Comestible, ...*) (Atserias et al., 2005; Miller, 1998)
- Specific rules for particular error detections (Gupta, 2002; Nadig et al., 2008). For instance, a rule proposed by (Nadig et al., 2008): “*If one term of a synset X is a proper suffix of a term in a synset Y, X is a hypernym of Y*”

2.3 Graph based methods

These methods are purely formal and do not take into account the semantics among word forms. Specific substructures of a wordnet’s hierarchies are checked and validated. Target substructures include:

- Cycles (Šmrz, 2004), (Kubis, 2012)
- Shortcuts (Fischer, 1997)
- Rings (Liu et al., 2004; Richens, 2008)
- Dangling uplinks (Koeva et al., 2004; Šmrz, 2004)
- Orphan nodes (null graphs) (Čapek, 2012).
- Small hierarchy (Lohk et al., 2014c)
- Unique beginners (Lohk et al., 2014c)

In addition, (Lohk, 2015) proposes different, yet undiscovered substructures and shows that the application of these substructures to validate the semantic hierarchies of wordnet may improve wordnet structure significantly. These substructures with a specific nature which are used in wordnet assessment are called **test patterns**. Next, we explain the concept of test patterns and demonstrate their efficient use with Estonian Wordnet.

3 A case study: applying test patterns to Estonian Wordnet

Since 2011, different types of test patterns have been developed and applied progressively to EstWN. Currently, ten test patterns exist. For every test pattern we implemented a program to find the relevant instances. Four programs are implemented for semi-automatic application (*closed subsets, closed subset with a root, the largest closed subset and connected roots*) and six for automatic use (the test patterns shown in italics in Table 2). Instances found with test patterns using programs for semi-automatic application were discussed elsewhere (Lohk, 2015). Test pattern instances found with programs for automatic use are employed in the constant validation process.

Version	Noun roots	Verb roots	Multiple inheritance cases	Shortcut	Ring	Synset with many roots	Heart-shaped substructure	Dense component	„Compound“ pattern
60	142	24	1,296	235	3,445	1,123	1,825	104	301
61	183	22	1,592	259	3,560	1,309	1,861	121	380
62	102	16	1,700	299	3,777	1,084	1,941	128	415
63	114	16	1,815	321	3,831	1,137	2,103	141	447
64	149	15	1,893	337	3,882	1,173	2,232	149	471
65	248	14	1,717	194	2,171	791	451	132	459
66	144	4	1,677	119	1,796	613	259	121	671
67	129	4	1,164	79	928	477	167	24	407
68	131	4	691	60	537	232	38	18	54
69	121	4	102	18	291	35	1	8	23
70	118	4	51	7	21	30	0	3	7

Table 2: A numerical overview of EstWN spanning eleven version

Table 2 shows the number of instances that each test pattern returned after its automatic application. The first two patterns (*shortcut* and *ring*) are inspired by (Fischer, 1997; Liu et al., 2004; Richens, 2008). There are also some cases of *synset with many roots*, called *dangling uplinks* in (Koeva et al., 2004) and (Šmrz, 2004). Bold font in the table shows when the test pattern was given to a lexicographer for verification. An example of this is the “shortcut” cases where lexicographers verified each instance manually in the 63rd version submitted to the EstWN. The effect, as reflected in the next version, can be clearly seen in the table. It is evident that the application of heart-shaped substructure and dense component patterns had a considerable effect on the lexicography.

As all instances of test patterns include multiple inheritance cases, the fourth column (Multiple inheritance cases) demonstrates the influence of using test patterns most clearly. For example, a comparison between versions 66 and 70 shows that the number of cases has gone down about 32 times (97%). Note that 118 hierarchies contain about 75% of shallow hierarchies where roots are connected to only one level of subordinates.

According to (Lohk, 2015), over ten versions of EstWN the most popular correction operation has been removing the hypernymy and hyponymy relations – 21,911 times. Secondly, the lexical units in synsets were changed 5,344 times (including deleted and added lexical units). Thirdly, 4,122

times hypernymy and hyponymy relations were replaced by another semantic relation, mainly by near *synonymy* and *fuzzynymy*.

4 Validating Princeton WordNet

Substructures connected with multiple inheritances have been used to validate PWN. (Fischer, 1997; Liu et al., 2004 and Richens, 2008) examined *shortcuts*; *rings* were suggested by (Koeva et al., 2004), and (Šmrz, 2004) examined *dangling uplinks*. There are also some examples of *closed subsets* in (Lohk et al., 2012) and one example of *heart-shaped substructure* in (Lohk and Vöhandu, 2014). Lohk gave an example of a *connected roots* case in his poster presentation at Estonian Applied Linguistics Conference in Tallinn in April 2013.

Next, we provide some examples of test pattern instances to grasp their structure and how they may help to discover specific inconsistencies in PWN semantic hierarchies. The complete overview of test patterns has been given in the dissertation of the first author (Lohk, 2015).

4.1 Shortcut

Shortcut is a pattern wherein a synset (based on **Figure 1**, {event}) is simultaneously connected to another synset ({group action}) both directly and indirectly. In that case, {group action} is not an ambiguous concept. Instead, it merely contains a redundant link (dotted line).

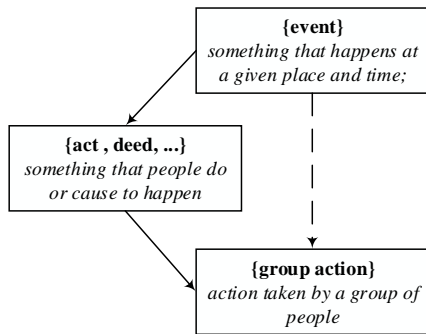


Figure 1. An instance of shortcut, PWN (version 3.1)

4.2 Heart-shaped substructure

In a heart-shaped substructure, two nodes (based on Figure 2, {hard drug} and {cannabis, ...}) have a direct connection through an identical parent ({controlled substance}) and an indirect connection through a semantic relation {soft drug} – {narcotic}) that links their second parent.

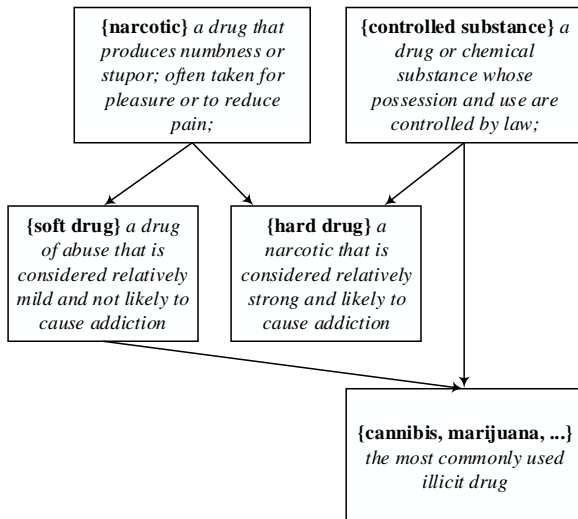


Figure 2. An instance of heart-shaped substructure, PWN (version 3.1)

In the case of PWN, we have seen that the instances of heart-shaped substructure tend to show the cases where instead of *role* or *type* relation hypernymy is used. An example of this is presented in Figure 2, where {hard drug} is actually a certain type of {narcotic} and as well as a {controlled substance}.

It is remarkable that when heart-shaped substructure was first used in EstWN, its number of instances was 451 (see Table 2) yet five versions later it had decreased to 0. Moreover, during the correction operations no *hypernymy/hyponymy* relation was changed to *role* or *type* (Lohk, 2015).

4.3 “Compound” pattern

“Compound” pattern is an exception among other test patterns as it considers the content of synsets.

More precisely, that kind of substructure satisfies the following two conditions:

Firstly, this substructure contains a case where a lexical unit of a superordinate (based on Figure 3 {ball}) is connected to two subordinates (1- {baseball}, 2- {basketball}... 24- {volleyball}) which contain that lexical unit (*ball*).

Secondly, at least one subordinate has an extra superordinate ({baseball equipment}, {basketball equipment}, ..., {golf equipment}).

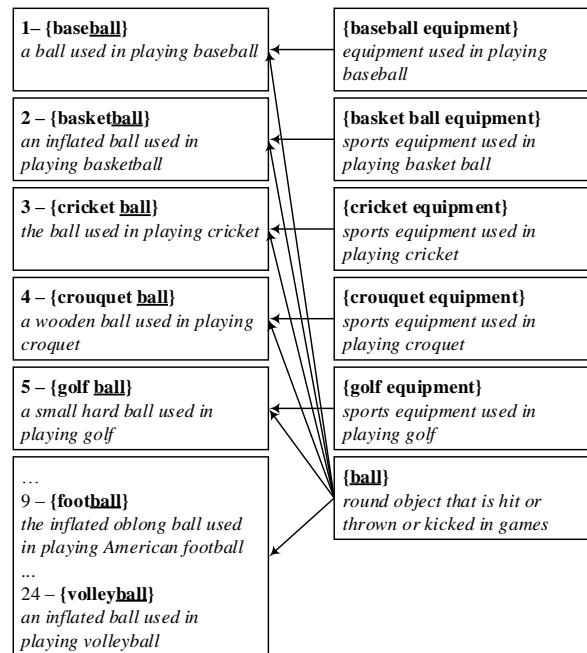


Figure 3. An instance of “compound” pattern, PWN (version 3.1)

To validate the kind of pattern instance shown in Figure 3, the lexicographer has to ask if subordinates 1 to 5 have an extra superordinate, and why it is not true about subordinates from 6 to 24. Studying this figure more carefully, we see that {basketball} is a {basketball equipment}. However, {football} and {volleyball} being quite similar in their definitions do not follow the same logic. That is to say, {football} and {volleyball} are not equipment.

4.4 Dense component

The *dense component* pattern provides an opportunity to uncover substructures where, due to *multiple inheritance*, the density of the interrelated concepts in the semantic hierarchy is higher (Lohk et al., 2014a), (Lohk et al., 2014b). This substructure (subgraph) consists of two *synsets* (nodes) (based on Figure {manicure} and {pedicure}) with at least two identical parents (it corresponds to *complete bipartite graph*) ({beauty treatment} and {aid, attention, care, ...}). The overall size of an instance of a *dense component*

depends on how many *synsets* (nodes) with at least two parents are interconnected through *multiple inheritance* and/or same parents (Lohk, 2015).

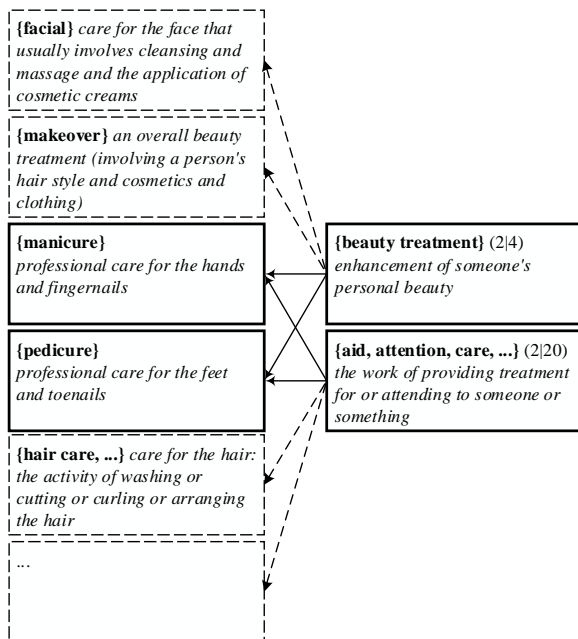


Figure 4. An instance of dense component, PWN (version 3.1)

In **Figure 4**, the dense component pattern is emphasized with bold lines. While this substructure contains at least two multiple inheritance cases, we see it as a case of the regularity of multiple inheritance. Herewith, the aim of the dense component is to help detect if this regularity is justified or alternatively, if this regularity has to be expanded.

In the case of **Figure 4**, the regularity of multiple inheritance has to be expanded. Two reasons for that are concepts {facial} and {hair care, ...}. In addition to {beauty treatment}, {facial} fits in with {aid, attention, care, ...}. Moreover, {hair care} is a {beauty treatment} besides being {aid, attention, care, ...}.

4.5 Connected roots

The connected roots test pattern involves different hierarchies through multiple inheritance cases.

This pattern helps to see how big and deep the connections between POS hierarchies are. Every

node acts as a unique beginner and is equipped with the number of hierarchy levels and the number of subordinates in the same hierarchy (Figure 1). The first number of the edge label indicates the number of common subordinates for two hierarchies. The next two numbers separated by “|” denote the hierarchy levels where the first common concept is located in both hierarchies.

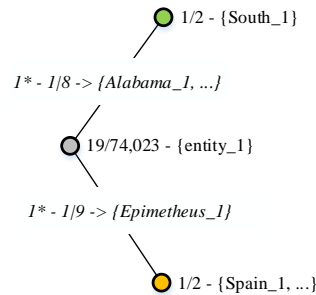


Figure 5. An instance of connected roots, PWN (version 3.1)

In Figure 1, there is only one large hierarchy with the unique beginner {entity}. It heads a 19-level hierarchy and 74,023 subordinates. By contrast, the two hierarchies ({South_1} and {Spain_1 ...}) are very small. They both dominate only one additional level. The edge labels reveal that the common concepts of both hierarchies are on the first lower levels in both of the smaller hierarchy cases. Both unique beginners ({South_1} and {Spain_1}) seem to be too specific to be the highest concepts.

Table 2 presents a comparison between PWN’s structure and that of other wordnets.

4.6 Short numerical overview of the test pattern instances

In **Table 3**, it is easy to see that the wordnets are very different. Finnish Wordnet was manually translated from PWN (Lindén and Niemi, 2014), hence it is not surprising that the first two rows are essentially identical.

The table shows a clear need for a deep structural analysis of all wordnets. Of course, it must be remembered that the hierarchies of different languages will never show a one-to-one correspondence, as the lexicons necessarily differ.

Version	Noun roots	Verb roots	Multiple inheritance cases	Short cut	Ring	Synset with many roots	Heart-shaped sub-structure	Dense component	„Compound“ Pattern
Princeton WordNet, v3.0	12	334	1,453	40	2,991	18	155	115	358
Finnish Wordnet, v2.0	12	334	1,453	40	2,991	18	155	115	394
Cornetto, v2.0	2	2	2,438	351	5,309	62	1,226	217	549
Polish Wordnet, v2.0	637	42	10,942	553	57,887	205,254	5,037	778	541
Estonian Wordnet, v70	118	4	51	7	21	30	0	3	7

Table 3: Five wordnets in comparison

5 Conclusions

Test patterns are a unique form of validating hierarchies. They are not language-specific and can be applied cross-lingually. Their value lies in aiding lexicographers in detecting and correcting errors and thus provide more accurate resources.

Every test pattern has the property of multiple inheritance. In most cases, except for the pattern of *shortcut* (Sec. 4.1), there is a lexical polysemy behind multiple inheritance.

Reference

- Atserias Batalla, J., Climent Roca, S., Moré López, J., Rigau Claramunt, G., 2005. A Proposal for a Shallow Ontologization of WordNet. *Proces. Leng. Nat.* N° 35 Sept 2005 Pp 161-167.
- Čapek, T., 2012. SENEQA-System for Quality Testing of Wordnet Data, in: *Proceedings of the 6th International Global Wordnet Conference*. Toyohashi University of Technology, Matsue, Japan, pp. 400–404.
- Fellbaum, C., 1998. *WordNet: An Electronic Lexical Database*, MIT Press. ed. Wiley Online Library, Cambridge, USA.
- Fischer, D.H., 1997. Formal Redundancy and Consistency Checking Rules for the Lexical Database WordNet 1.5, in: *Workshop Proceedings of on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Association for Computational Linguistics (ACL), Madrid, Spain, pp. 22–31.
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L., 2002. Sweetening Ontologies with DOLCE, in: *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*. Springer, pp. 166–181.
- Guarino, N., Welty, C., 2002. Evaluating Ontological Decisions with OntoClean. *Communications of the*

Multiple inheritance is not always wrong. However, PWN still contains many cases where instead of *role* or *type* relation the *hypernymy* relation has been used. This is one reason for why sometimes multiple inheritance cases are presented in PWN (see Figure 2).

In sum, the analysis of wordnet structures using test patterns provides lexicographers with a very powerful tool, which we hope will be adopted by the global wordnet community.

- ACM - Ontology: Different Ways of Representing the Same Concept 45, 61–65.
- Gupta, P., 2002. Approaches to Checking Subsumption in GermaNet, in: *Proceedings of the 3rd International Conference on Language Resources and Evaluation*. European Language Resources Association (ELRA), Las Palmas, Canary Islands, Spain, pp. 8–13.
- Hearst, M.A., 1992. Automatic Acquisition of Hyponyms from Large Text Corpora, in: *Proceedings of the 14th Conference on Computational Linguistics - Volume 2, COLING '92*. Association for Computational Linguistics (ACL), Stroudsburg, PA, USA, pp. 539–545.
- Koeva, S., Mihov, S., Tinchev, T., 2004. Bulgarian Wordnet–Structure and Validation. *Romanian J. Inf. Sci. Technol.* 7, 61–78.
- Krstev, C., Pavlović-Lažetić, G., Obradović, I., Vitas, D., 2003. Corpora Issues in Validation of Serbian Wordnet, in: Matoušek, V., Mautner, P. (Eds.), *Text, Speech and Dialogue, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 132–137.
- Kubis, M., 2012. A Query Language for WordNet-Like Lexical Databases, in: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (Eds.), *Intelligent Information and Database Systems, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 436–445.

- Liu, Y., Yu, J., Wen, Z., Yu, S., 2004. Two Kinds of Hypernymy Faults in WordNet: the Cases of Ring and Isolator, in: Proceedings of the 2nd Global Wordnet Conference. Brno, Czech Republic, pp. 347–351.
- Lohk, A., 2015. A System of Test Patterns to Check and Validate the Semantic Hierarchies of Wordnet-type Dictionaries. Tallinn University of Technology, Tallinn, Estonia.
- Lohk, A., Allik, K., Orav, H., Vöhandu, L., 2014a. Dense Component in the Structure of Wordnet, in: Proceedings of the 9th International Conference on Language Resources and Evaluation. European Language Resources Association (ELRA), Reykjavik, Iceland, pp. 1134–1139.
- Lohk, A., Norta, A., Orav, H., Vöhandu, L., 2014b. New Test Patterns to Check the Hierarchical Structure of Wordnets, in: Information and Software Technologies. Springer, pp. 110–120.
- Lohk, A., Orav, H., Vöhandu, L., 2014c. Some Structural Tests for WordNet with Results. Proceedings of the 7th Global Wordnet Conference 313–317.
- Lohk, A., Vare, K., Vöhandu, L., 2012. First Steps in Checking and Comparing Princeton Wordnet and Estonian Wordnet, in: Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH. Association for Computational Linguistics (ACL), pp. 25–29.
- Lohk, A., Vöhandu, L., 2014. Independent Interactive Testing of Interactive Relational Systems, in: Gruca, D.A., Czachórski, T., Kozielski, S. (Eds.), Man-Machine Interactions 3, Advances in Intelligent Systems and Computing. Springer International Publishing, pp. 63–70.
- Miller, G.A., 1998. Nouns in WordNet, in: WordNet: An Electronic Lexical Database. MIT Press, Cambridge, Massachusetts, USA, pp. 24–45.
- Nadig, R., Ramanand, J., Bhattacharyya, P., 2008. Automatic Evaluation of WordNet Synonyms and Hypernyms, in: Proceedings of ICON-2008: 6th International Conference on Natural Language Processing. CDAC Pune, India.
- Pedersen, B.S., Borin, L., Forsberg, M., Kahusk, N., Lindén, K., Niemi, J., Nisbeth, N., Nygaard, L., Orav, H., Rögnvaldsson, E., others, 2013. Nordic and Baltic Wordnets Aligned and Compared Through “WordTies,” in: The 19th Nordic Conference of Computational Linguistics. Linköping University Electronic Press, Oslo University, Norway, pp. 147–162.
- Pedersen, B.S., Forsberg, M., Borin, L., Lindén, K., Orav, H., Rögnvaldsson, E., 2012. Linking and Validating Nordic and Baltic wordnets, in: Proceedings of the 6th International Global Wordnet Conference. Matsue, Japan, pp. 254–260.
- Peters, W., Peters, I., Vossen, P., 1998. Automatic Sense Clustering in EuroWordNet, in: Proceedings of the 1st International Conference on Language Resources and Evaluation. European Language Resources Association (ELRA), Granada, Spain, pp. 409–416.
- Richens, T., 2008. Anomalies in the Wordnet Verb Hierarchy, in: Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics (ACL), pp. 729–736.
- Sagot, B., Fišer, D., 2012. Cleaning Noisy Wordnets, in: Proceedings of the 8th International Conference on Language Resources and Evaluation. European Language Resources Association (ELRA), Istanbul, Turkey, pp. 23–25.
- Saito, J.-T., Wagner, J., Katz, G., Reuter, P., Burke, M., Reinhard, S., 2002. Evaluation of GermaNet: Problem Using GermaNet for Automatic Word Sense Disambiguation, in: Proceedings of the LREC Workshop on WordNet Structure and Standardization and How These Affect WordNet Applications and Evaluation. European Language Resources Association (ELRA), Las Palmas, Canary Islands, Spain, pp. 14–29.
- Šmrz, P., 2004. Quality Control and Checking for Wordnet Development: A Case Study of BalkaNet. Science and Technology 7, 173–181.