

Swesaurus;
or,
The Frankenstein Approach to Wordnet Construction

Lars Borin
Språkbanken
Department of Swedish
University of Gothenburg
Gothenburg, Sweden

lars.borin@svenska.gu.se markus.forsberg@svenska.gu.se

Markus Forsberg
Språkbanken
Department of Swedish
University of Gothenburg
Gothenburg, Sweden

Abstract

Swesaurus is a freely available (under a CC-BY license) Swedish wordnet under construction, built primarily by scavenging and recycling information from a number of existing lexical resources. Among its more unusual characteristics are graded lexical-semantic relations and inclusion of all parts of speech, not only open-class items.

The materials at present within my command hardly appeared adequate to so arduous an undertaking, but I doubted not that I should ultimately succeed. I prepared myself for a multitude of reverses; my operations might be incessantly baffled, and at last my work be imperfect, yet when I considered the improvement which every day takes place in science and mechanics, I was encouraged to hope my present attempts would at least lay the foundations of future success. Nor could I consider the magnitude and complexity of my plan as any argument of its impracticability. [...] After having formed this determination and having spent some months in successfully collecting and arranging my materials, I began.

(Shelley, 1818, Ch. 4)

1 Introduction: Swesaurus – towards a quiltwork wordnet

Swesaurus is a Swedish open-source “proto-wordnet” under active development. The main novel methodological aspect of this development is its “quiltwork” – or “Frankenstein” – character. Swesaurus is being constructed mainly by scavenging and recycling lexical-semantic information from a number of existing lexical resources. Other noteworthy features of Swesaurus which distinguish it from most other wordnets is the fact that it does not practice “part-of-speech discrimination”; it constitutes a lexical-semantic resource encom-

passing all parts of speech (POS),¹ and its *graded* lexical-semantic relations.

In the literature we find basically two main approaches to (manual) wordnet construction (Vossen, 1998), viz. (1) the merging approach – based on language-specific lexicographical knowledge and where the synsets and their interrelations consequently respect the structure of the language in question, and where any linking to the English Princeton WordNet (PWN) is made subsequently to the wordnet construction – or (2) the extension approach – where the structure is imported wholesale using bilingual resources (dictionaries or translation corpora) from another language, typically English (PWN), and where the linking to the source wordnet consequently is part of the design. Good examples of wordnets built using the merging approach are the Danish wordnet DanNet (Pedersen et al., 2009) and the Polish wordnet plWordNet (Piasecki et al., 2009), while the extension approach can be well illustrated with the Finnish wordnet FinnWordNet (Lindén and Niemi, 2013), where a translation agency was employed for translating PWN into Finnish, or the two Norwegian wordnets (for the two written standards of Norwegian, Bokmål and Nynorsk), which were translated using partly automatic corpus-based methods, although not from PWN, but from DanNet.²

True to its quiltwork character, Swesaurus incorporates elements of both approaches. For most of the Swesaurus components, the merging approach has been the only choice, utilizing as they did information from pre-existing monolingual

¹Recall that Princeton WordNet covers only the open parts of speech, the “content words”, specifically nouns, verbs, adjectives, and adverbs. However, because of what seems to be a specific Anglo-Saxon lexicographical practice (Apresjan, 2002), numerals are also included in WordNet, classified as nouns (cardinals) or adjectives (ordinals).

²This makes eminent sense, given that one of the two Norwegian written standards is historically based on Danish.

Swedish resources, but in the case of the Core WordNet component, the extension approach has been used (see section 3.3).

The motivation for building Swesaurus is manifold. The utility of having a PWN-like resource for a language is often stated in the literature. Further, the work is driven in part by opportunity: The general resource harmonization and standardization activities described below in section 2 open the possibility of mining and compiling similar kinds of information from originally quite heterogeneous lexical resources, both as to their content, and above all with respect to their format.

2 The prerequisites: A unified lexical infrastructure

The construction of Swesaurus would not be possible without the groundwork laid in the Swedish FrameNet++ project (Borin et al., 2010), the goals of which are threefold: (1) creation of an integrated lexical macro-resource; (2) construction of a Swedish framenet; and (3) creation of open resources.

Crucially, the first goal has been implemented on the content level through a decision on resource interlinking on the word sense level. As a consequence of this, a structured system of persistent identifiers has been designed for word senses, which are used as links among resources. This does not mean that the only possible kind of link is based on (word sense) identity. Especially with diachronic and cross-language links, other relations are needed, at least hyponymy/hyperonymy, expressed using `skos:narrower` and `skos:broader` (Miles and Bechhofer, 2009).

The macro-resource, of which Swesaurus is a part, is large and diverse, consisting of 23 lexical resources, ranging from the Swedish FrameNet to Old Swedish dictionaries, containing a total of 686,237 lexical entries at the time of writing. To be able to work productively with this macro-resource, we need good tools for interacting with the data, for abstracting, ordering, searching and visualizing the data itself, for inferring and presenting relations among data items, and for editing the data. To meet these demands, a generalized lexical infrastructure is under development (Borin et al., 2012a; Borin et al., 2013b), geared towards dealing with large networks of interconnected lex-

icons (Borin, 2010; Borin et al., 2010) that have been encoded in the LMF format (Lexical Markup Framework; see ISO 2008; Francopoulo 2013).

One essential component of the lexical infrastructure is a generic search interface that provides a plug-and-play search tool for resources already in LMF, where the LMF format is employed both internally within the infrastructure and, trivially, as an export format.

The lexical infrastructure also maintains a strong bidirectional connection to a general and flexible corpus infrastructure (Borin et al., 2012b). For example, the lexical information in the macro-resource is used in annotating the corpora, and the language examples for the lexical resources are retrieved from the corpus infrastructure.

A pervasive theme for both infrastructures is *openness*, which for the lexical infrastructure is demonstrated through its utilization of open standards and open-content licenses, as well as the daily publication of not only the resources but everything else that is available in-house, such as formal test protocols, change history and the tools themselves. The tools are available through a set of web services, which are open for others to use, and which provide a convenient way of accessing the lexical information programmatically.

3 The lexical resources

Below, we discuss the existing lexical resources underlying the component parts of Swesaurus and how they are processed for inclusion in Swesaurus.

3.1 SALDO

The lexical macro-resource described in section 2 is topologically a hub-and-spokes structure. There is one primary lexical resource, a pivot, to which all other resources are linked. This is SALDO (Borin et al., 2013a), a large (130K entries and 1.9M wordforms), freely available (under a Creative Commons Attribution license) morphological and lexical-semantic lexicon for modern Swedish. It has been selected as the pivot partly because of its size and quality, but also because its form and sense units are identified by carefully designed unique persistent identifiers (PIDs) to which the lexical information in other resources are linked.

The standard scenario for a new resource to be integrated into the macro-resource is to (par-

tially) link its entries to the sense PIDs of SALDO. This cannot be done automatically on the level of word senses in the general case. However, like many other linguistic phenomena, the distribution of senses over lemmas in lexical resources is roughly Zipfian (Moon, 2000; Borin, 2010). Thus, the vast majority of the lemmas are monosemous, reducing the sense mapping problem to the much simpler problem of pairing up forms between lexical resources. This is ultimately what makes an endeavor such as Swesaurus feasible.

Doing this automatic pairing of forms typically has the effect that the ambiguity of a resource becomes explicit: the bulk of the resources associate lexical information to part-of-speech-tagged base forms, information not always valid for all senses of that base form. This is natural since most of the resources have initially been created for human consumption, and a human can usually deal with this kind of underspecification without problem. Some of these ambiguities can be resolved automatically – especially if information from several resources is combined – but in the end, manual work is required for complete disambiguation.

SALDO is a kind of lexical-semantic network, superficially similar to PWN, but quite different from it in the principles by which it is structured. SALDO is about the same age as PWN, and it was developed completely independently of the latter, inspired more by a Russian tradition of lexical description, rather than an Anglo-Saxon one; cf., for example, Igor’ Mel’čuk’s *Meaning – Text Model* (Mel’čuk, 1974).

The basic linguistic idea underlying the structure of SALDO is that, semantically speaking, the whole vocabulary of a language can be described as having a center – or core – and (consequently) a periphery. The notion of *core vocabulary* is familiar from several linguistic subdisciplines (Borin, 2012). In SALDO this idea is consistently applied down to the level of individual word senses. Thus, every entry in SALDO – representing a word sense – has one or more semantic descriptors, which are themselves also entries in the dictionary. All entries in SALDO (with one sole exception; see below) are actually occurring words or conventionalized or lexicalized multi-word units of the language.

One of the descriptors, called *primary*, is obligatory. The primary descriptor is the entry which better than any other entry fulfills two re-

quirements: (1) it is a semantic neighbor of the entry to be described; and (2) it is more central than it. However, there is no requirement that the primary descriptor is of the same part of speech as the entry itself. Thus, the primary descriptor of *kniv* ‘knife (n)’ is *skära* ‘cut (v)’, and that of *lager* ‘layer (n)’ is *på* ‘on (p)’.

Through the primary descriptors SALDO is a single tree, rooted by assigning an artificial top sense (called PRIM) as primary descriptor to the 42 topmost word senses.

That two words are semantic neighbors means that there is a direct semantic relationship between them (such as synonymy, hyponymy, meronymy, argument-predicate relationship, etc.). As could be seen from the examples given above, SALDO includes not only open-class words, but also pronouns, prepositions, conjunctions etc. In such cases closeness must sometimes be determined with respect to function or syntagmatic connections, rather than (“word-semantic”) content.

Centrality is determined by means of several criteria: frequency, stylistic value, word formation, and traditional lexical-semantic relations all combine to determine which of two semantically neighboring words is to be considered more central.

Relevant to the Swesaurus endeavor, the primary descriptor will in practice quite often be either a hyperonym or synonym of the keyword. Thus, SALDO was mined for Swesaurus candidates by extracting all same-POS entry–primary descriptor pairs. In the process, some important special cases were recognized which require very little manual post-processing, such as noun compound entries where the form of the primary descriptor corresponds to the last member of the compound, e.g., *livförsäkring* : *försäkring* ‘life insurance : insurance’, and where the entry in the overwhelming majority of cases is a hyponym of the primary descriptor. In this way, a large number of synonyms, near-synonyms, hyperonyms, antonyms, and related senses could be extracted from SALDO, representing all parts of speech.

3.2 Synlex

Synlex (the People’s Synonym Lexicon; Kann and Rosell 2006) is a lexical resource that has been created by asking members of the public – users of an online Swedish-English dictionary – to judge the degree of synonymy of a random, automati-

cally generated synonym pair candidate, on a scale from 0 (not synonyms) to 5 (fully synonymous). A synonym pair list containing all pairs that average 3.0 or more on a large number of judgements is available for download under an open-source license. The latest version of the list at the time of writing is dated 2013-05-23, and contains 19,269 graded synonym pairs (38,538 if symmetry of synonymy is not taken into account).

The members of these pairs are words (i.e., text word forms) – not even part of speech is indicated – mainly dictionary base forms (lemmas), but sometimes inflected forms, and in some cases multi-word expressions. One problem then becomes, in the case of a word having as synonyms several other words – because of homonymy and polysemy – to determine how many senses we are dealing with. Also, for those familiar with PWN, we should add that synonymy relations in Synlex are sometimes between words with different POS, just as in EuroWordNet. Although in EuroWordNet this kind of synonymy is still formally distinct from within-POS synonymy, bearing the label XPOS_NEAR_SYNONYM (Alonge et al., 1998).

Since Synlex gives us access to graded synonymy, we may introduce the notion of fuzzy synsets into Swesaurus (Borin and Forsberg, 2010), i.e., synsets where a word's membership is a matter of degree (see section 4 for a discussion about synsets in Swesaurus).

3.2.1 Wiktionary

Wiktionary is an undertaking similar to Wikipedia, but for collaborative writing of dictionaries rather than encyclopedias. The Swedish Wiktionary,³ is a downloadable free resources that, among other things, contains some lexical-semantic relations. The work of extracting such relations from Wiktionary is hampered by the fact that the data set is only partially encoded with a formal structure. It is the responsibility of the writer to encode the different information categories in a lexical entry in the correct wiki format that was intended by the creator of Wiktionary, but no automatic check of the encoding is actually done. Since the result of a faulty encoding may actually look correct for the human eye, there are in practice a number of errors in Wiktionary that complicate the automatic information extraction.

³See <http://sv.wiktionary.org>

We have experimented with extracting synonymy relations between words, with a resulting set of 10,529 synonymy pairs, of which 3,857 of the word pairs have members with only one sense in SALDO. Hence, no manual disambiguation is needed, so they may be incorporated immediately into Swesaurus. Some of the pairs are wrong, since some lexical entries contain information from other languages and relation within them. This results in a few cases where, e.g., a Swedish word is linked to a Polish one. In practice, this is not a major problem, since the linking to SALDO filters out those words that are not in SALDO.

The synonymy relations in Wiktionary are in general of higher quality than those in Synlex, which is to be expected since the author of a lexical entry in Wiktionary makes a conscious choice when assigning synonyms to a word, but Synlex, on the other hand, builds upon automatically generated word pairs, with the consequence that words that is not normally judged synonymous are sometimes assigned a degree greater than zero. For example, consider the pair *förlovning* : *förpliktelse* 'engagement to be married : obligation', the members of which are normally not considered to be synonymous, but when presented together and you are asked to quantify their synonymy degree, you may be tempted to give them at least a small degree of synonymy.

3.3 Core WordNet

As part of the EC-funded META-NORD project (2011–2013), a linking of the Princeton Core WordNet (CWN) to Swedish was completed and included in Swesaurus. The linkage was bootstrapped by using the Lexin basic Swedish-English dictionary (~25,000 entries). Swedish lemmas in Lexin were automatically linked, in an overgenerating manner, to SALDO sense identifiers, giving us a set of senses for every lemma. The glosses of CWN were subsequently, via Lexin, linked to these sense sets. CWN has 5,000 entries, of which around 89% were covered by Lexin. Furthermore, 23% had a unique link to one SALDO sense, and the remaining an average ambiguity of 4.4 (a rather high ambiguity, but not unexpected for a core vocabulary).

3.4 The Gothenburg Semantic Database

The Gothenburg Semantic Database (SDB; Järborg 2001) is a lexical database for modern

Swedish covering 61,000 entries with an extensive description inflection, morphology and meaning. Originally building on a lexicographical database that has been used in producing two modern Swedish reference dictionaries, SDB has been enriched with a deeper semantic description where many of the verb senses have been provided with semantic valency information using a set of about 40 general semantic roles and linked to example sentences in a corpus.

SDB holds two kinds of relevant lexical-semantic information: (1) explicit lexical semantic relations cross-referencing among different lexical entries (lemmas); and (2) implicit in its hierarchical organization of lexical entries into main senses and subsenses, typically corresponding to a superordinate–hyponym relation.

The linking of SDB senses to SALDO sense identifiers is ongoing. An initial automatic linking is now being manually checked and corrected. For those senses that are already processed in this way, the explicit lexical semantic relations have been included in Swesaurus, and some of the derived relations calculated (see section 4), while the entry-internal hierarchical relations present in SDB have not yet been extracted. In the process, it has become clear that the explicit relations are not consistent, and will need a good deal of manual correction, which is ongoing.

3.4.1 Bring’s thesaurus

The author of what is probably the first Swedish thesaurus, Sven Casper Bring (1842–1931) worked as a lawyer, district judge and translator. Besides practicing law, he published several translations from French, Italian and English to Swedish. His final work was an adaptation to Swedish of Roget’s well-known *Thesaurus* (Bring, 1930). He writes in his preface to the book that he was inspired by similar adaptations that had taken place of Roget’s *Thesaurus* to German.

Bring’s thesaurus was digitized in the early 1990s and has since been made available under an open-content license. Work is ongoing to create a modernized version of Bring by using SALDO and other modern lexical resources in order to semi-automatically add modern vocabulary to it.

Like in Roget, the vocabulary included in Bring is divided into slightly over 1,000 “conceptual classes”. Each class consists of a list of words, where, when there are enough relevant words, nouns are listed first, followed by

verbs, and finally a mixed group containing adjectives, adverbs, interjections and phrases. Semicolons, together with paragraph structure, group words together, which are thought to be more closely semantically related. Semicolon groups often contain synonym clusters, with distance between words in a cluster roughly correlating to degree of synonymy, and we plan to explore how the semicolon groups can be used as a source for yet another Swesaurus component.

4 Some matters of method

Following a long tradition in lexicography and lexical semantics, we posit as primary semantic entities in all our lexical resources *word senses*, i.e., roughly the content side of the Saussurean linguistic sign, paired with a form side on the word level (a word, a conventionalized or lexicalized multiword expression, or, rarely, a sub-word-unit). Importantly in this connection, *synsets* are not primary entities in our resources.

As a corollary to the above, all lexical-semantic relations are between word senses only. Synonymy is simply one of these relations among many others. A PWN-style wordnet, on the other hand, does not have the synonymy relation at all. Synsets are defined through (one construal of) synonymy, but the relation itself is not present as such in the wordnet.

The decision not to allow synsets into Swesaurus as first-class citizens rests partly on tradition. Importantly, however, in doing this, we also avoid building in a strong assumption about the nature of synonymy into the foundations of our resource. Even though synonym dictionaries are among the oldest products of lexicography – even the Sumerians and Akkadians compiled them (Civil, 1990) – in practice synonymy has turned out to be a most slippery notion: While synonyms are self-evidently a central feature of language according to Lieber (1841, vii), they are “morally impossible” to Döderlein (1863, xii). Thus, in constructing Swesaurus we have opted for treating synsets as derived, from a possibly varying or changing definition of synonymy.

Ockham’s razor also enters into the picture: Since word senses seem to be needed in any case, and to be in some sense more basic than synsets – more than half (54%) of the synsets in PWN have only one member, arguably a word sense – we see

no pressing need to adopt the synset as basic notion.

This makes the basic information unit in Swesaurus the (word-sense) *relational triple*, whose three components are: (1) a source word sense; (2) a *graded* lexical-semantic relation; and (3) a target word sense. In addition, each triple has provenance information, i.e., from which resource it originates and whether it is primary or derived. The relations used so far in Swesaurus are the ones listed in Table 1.

At present, all derived relations except related-sense are generally taken to hold only within a part of speech – i.e., source and target words must have the same part of speech – although this may change in the future.⁴ The related-sense relation is a catchall label covering a mixed bag of semantic and formal relations among word senses, both more loose “evocation” (Boyd-Graber et al., 2006) or “associative” (Borin et al., 2013a) semantic relations, and formal derivational relations, e.g., verbs and the corresponding deverbal nouns, nouns and their denominal adjectives, etc.

The resources generally have fragmentary information, for various reasons. From the logical properties of the relations follow certain inference rules which allow us to partly ameliorate this situation.

For example, the transitivity of most of the relations allows us to add many derived relational triples to Swesaurus. If we have the information that A is-a-synonym-of B and B is-a-synonym-of C, we can infer that A is-a-synonym-of C even in the absence of this explicit information. More subtle inferences are also possible, for example: If we have the explicit information that A is-a-hyponym-of B and C is-a-hyponym-of D and further that A is-a-cohyponym-of C, we can then infer that B is-a-synonym-of D.

⁴Some of the original synonym pairs in Synlex already cross part-of-speech boundaries, and even SDB has a small number of such examples, e.g. some color adjectives are listed as hyponyms of the noun *grundfärg* ‘primary color’. Further, we note that especially in linguistic descriptions of languages with rich derivational morphological systems it is often taken for granted that, e.g., a verb and the corresponding deverbal action noun express the same concept – are synonymous – so that the difference between *to eat* (v.) and *the eating* (n.) is on a par with the tense difference between *eats* and *ate*. Both express the concept of eating, but in forms determined by the syntactic frame in which they are made to function (see, e.g., Fellbaum 2005). The differences are in both cases purely formal, not conceptual.

There are also some less obviously useful entailment relations, which however should be recognized both for completeness’ sake and for implementing correct behavior in search and browsing tools, such as: Synonymy entails cohyponymy; and all other relations entail related-sense.

The consequence of this is that Swesaurus contains two kinds of relational triples: (1) primary triples, explicitly present in the sources; and (2) derived triples, automatically computed using the inference rules for triples.

“Wordnetified” versions of Swesaurus in addition contain synsets constructed through the transitive closure of the synonymy relation.

Graded relations complicate this picture, and it is not completely clear how to best use the degree information in computing derived relations. Consequently, we must be careful when deriving new synonym pairs in Synlex, especially if we iterate over already derived ones. A few pairs like the already mentioned *förlovning* : *förpliktelse* ‘engagement to be married : obligation’ may give rise to a large number of questionable synonymy pairs. A more conservative approach than general transitivity is to use the existing synonymy cliques in the derivation process, and only derive new synonyms if we create a new clique by deriving that synonym. This has been the strategy chosen for deriving new synonym pairs from Synlex.

According to the website of the *Global WordNet Association*,⁵ “resources that follow the wordnet design” must include

- links to WordNet (Princeton or others that are linked to PWN)
- WN structure (minimally: synset, hyponymy)

Swesaurus marginally fulfills the first criterion – only the CWN component is linked to PWN – although we acknowledge the usefulness of such a linking, and are planning to extend it to the other components of Swesaurus. It also fails the second criterion, since there are no synsets at all in Swesaurus. However, as we have argued and shown above, A PWN-style wordnet – in fact, many different PWN-style wordnets – can be completely mechanically derived from Swesaurus, with synset sizes dependent on the synonymy degree threshold chosen for synset assembly. The synsets can then inherit selected lexical semantic relations from their member word senses.

⁵See http://globalwordnet.org/?page_id=38

Relation	Logical properties
synonymy	symmetric, transitive
antonymy	symmetric
related-sense	symmetric, transitive(?)
hyponymy/subordinate sense	transitive, inverse of hyperonymy
hyperonymy/superordinate	transitive, inverse of hyponymy
cohyponymy	symmetric, transitive
partonymy	transitive(?), inverse of holonymy
holonymy	transitive(?), inverse of partonymy

Table 1: Lexical-semantic relations used in Swesaurus

5 Conclusions and future work

All the activities listed in section 3 are ongoing to various degrees. In summary, approximate current numbers of primary and derived relational triples in the different Swesaurus components are as follows:

Component	Primary	Derived
Synlex	19,000	9,500
Wiktionary	4,000	–
CWN	4,500	–
SDB	10,000	13,500
SALDO	32,500	–

All numbers are for *normalized relational triples*, which means that symmetric relations are counted only once for a given word-sense pair, and that for relations with an inverse, only one of the two is present in the data. Thus, A is-an-antonym-of B will exclude the presence of B is-an-antonym-of A, and A is-a-hyperonym-of B will be transformed into B is-a-hyponym-of A.

We are already starting to see how genuine synergy could arise from the work described above. The flow of information is not one-way; instead, the derived lexical-semantic information made possible through the construction of Swesaurus may in its turn be used to enrich the original lexical resources. Synonyms may be a good source of new lexical units in a framenet, for instance, and the modernization of Bring’s Thesaurus will probably be easier to accomplish using the lexical-semantic information from Swesaurus. We have already mentioned that semicolon groups in Bring are often made up of synonym clusters, but like its predecessor and model Roget, Bring, too, organizes many of its conceptual classes according to antonymies, making the antonymy information

in Swesaurus a potential source of enrichment of Bring.

So far our work on Swesaurus has focused on the crosslinking and consequent synergistic enrichment of heterogeneous lexical resources. Another important line of research found in the literature on wordnet construction, but that we have not touched upon in this paper, concerns corpus-driven, machine-learning based methods for wordnet building. We have conducted some initial experiments based on a large Swedish corpus collection, and this is a direction which we plan to pursue further in the future. In this connection, a particularly intriguing question is to what extent near synonymy of the kind found in Synlex can be discovered automatically in corpora.

Acknowledgments

The research presented here has been made possible through financial support from the Swedish Research Council (*Swedish FrameNet++* project, contract no 2010-6013), from the University of Gothenburg through its support of the Centre for Language Technology and through its support of Språkbanken (the Swedish Language Bank), and from the European Commission through its support of the META-NORD project under the ICT PSP Programme, grant agreement no 270899.

References

- Antonietta Alonge, Nicoletta Calzolari, Piek Vossen, Laura Bloksma, Irene Castellon, Maria Antonia Marti, and Wim Peters. 1998. The linguistic design of the EuroWordNet database. In Piek Vossen, editor, *EuroWordNet: A Multilingual Database with Lexical Semantic Networks for European Languages*. Kluwer, Dordrecht. Pages 19–43.
- Yuri D. Apresjan. 2002. Principles of systematic lexicography. In Marie-Hélène Corréard, editor, *Lex-*

- icography and Natural Language Processing. *A Festschrift in Honour of B. T. S. Atkins*. Euralex, Grenoble. Pages 91–104.
- Lars Borin and Markus Forsberg. 2010. From the people's synonym dictionary to fuzzy synsets - first steps. In *Proceedings of the LREC 2010 workshop Semantic relations. Theory and Applications*, pages 18–25.
- Lars Borin, Dana Danélls, Markus Forsberg, Dimitrios Kokkinakis, and Maria Toporowska Gronostaj. 2010. The past meets the present in Swedish FrameNet++. In *14th EURALEX International Congress*. Leeuwarden. EURALEX. Pages 269–281.
- Lars Borin, Markus Forsberg, Leif-Jöran Olsson, and Jonatan Uppström. 2012a. The open lexical infrastructure of Språkbanken. In *Proceedings of LREC 2012*. Istanbul. ELRA. Pages 3598–3602.
- Lars Borin, Markus Forsberg, and Johan Roxendal. 2012b. Korp – the corpus infrastructure of Språkbanken. In *Proceedings of LREC 2012*. Istanbul. ELRA. Pages 474–478.
- Lars Borin, Markus Forsberg, and Lennart Lönngren. 2013a. SALDO: a touch of yin to WordNet's yang. *Language Resources and Evaluation*, May. Online first publication; DOI 10.1007/s10579-013-9233-4.
- Lars Borin, Markus Forsberg, Leif-Jöran Olsson, Olof Olsson, and Jonatan Uppström. 2013b. The lexical editing system of karp. In *Kosem, I., Kallas, J., Gantar, P., Krek, S., Langemets, M., Tuulik, M. (eds.) 2013. Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia.*, volume 2013, pages 503–516.
- Lars Borin. 2010. Med Zipf mot framtiden – en integrerad lexikonresurs för svensk språkteknologi. *LexicoNordica*, 17:35–54.
- Lars Borin. 2012. Core vocabulary: A useful but mystical concept in some kinds of linguistics. In Diana Santos, Krister Lindén, and Wanjiku Ng'ang'a, editors, *Shall we play the Festschrift game? Essays on the occasion of Lauri Carlson's 60th birthday*. Springer, Berlin. Pages 53–65.
- Jordan Boyd-Graber, Christiane Fellbaum, Daniel Osherson, and Robert Schapire. 2006. Adding dense, weighted connections to WordNet. In *GWC 2006 Proceedings*. Brno. Masaryk University. Pages 29–35.
- Sven Casper Bring. 1930. *Svenskt ordförråd ordnat i begreppsklasser*. Hugo Gebers förlag, Stockholm.
- Miguel Civil. 1990. Sumerian and Akkadian lexicography. In Oskar Reichmann Hausmann, Franz Josef and, Herbert Ernst Wiegand, and Ladislav Zgusta, editors, *Wörterbücher: Ein internationales Handbuch zur Lexikographie. Zweiter Teilband / Dictionaries: An international encyclopedia of lexicography. Second volume / Dictionnaires: Encyclopédie internationale de lexicographie. Tome second*. Walter de Gruyter, Berlin. Pages 1682–1686.
- Ludwig Döderlein. 1863. The author's preface. In *Döderlein's hand-book of Latin synonymes. Translated by Rev. H.A. Arnold, B.A., with an introduction by S.H. Taylor, LL.D.* Warren F. Draper, Andover. Pages ix–xvi.
- Christiane Fellbaum. 2005. Co-occurrence and antonymy. *International Journal of Lexicography*, 8(4):281–303.
- Gil Francopoulo, editor. 2013. *LMF: Lexical Markup Framework*. ISTE/Wiley, London/Hoboken, NJ.
- ISO. 2008. Language resource management – Lexical Markup Framework (LMF). International Standard ISO 24613:2008.
- Jerker Järborg. 2001. Roller i Semantisk databas. Research Reports from the Department of Swedish No. GU-ISS-01-3. University of Gothenburg, Department of Swedish.
- Viggo Kann and Magnus Rosell. 2006. Free construction of a free Swedish dictionary of synonyms. In *Proceedings of the 15th NODALIDA conference, Joensuu 2005*. Department of Linguistics, University of Joensuu. Pages 105–110.
- Francis Lieber. 1841. Preface of the translator. In *Dictionary of Latin synonymes, for the use of schools and private students, with a complete index. By Lewis [Ludwig] Ramshorn. From the German by Francis Lieber*, Charles C. Little and James Brown. Pages iii–viii.
- Krister Lindén and Jyrki Niemi. 2013. Is it possible to create a very large wordnet in 100 days? An evaluation. *Language Resources and Evaluation*, July. Online first publication; DOI 10.1007/s10579-013-9245-0.
- Igor' A. Mel'čuk. 1974. *Opyt teorii lingvističeskikh modelej «Smysl ↔ Tekst»*. Nauka, Moscow.
- Alistair Miles and Sean Bechhofer. 2009. SKOS Simple Knowledge Organization System Reference. W3C Recommendation. <http://www.w3.org/TR/skos-reference/>.
- Rosamund Moon. 2000. Lexicography and disambiguation: The size of the problem. *Computers and the Humanities*, 34(1–2):99–102.
- Bolette Sandford Pedersen, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig 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.

Maciej Piasecki, Stanisław Szpakowicz, and Bartosz Broda. 2009. *A Wordnet from the Ground Up*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.

Mary Wollstonecraft Shelley. 1818. *Frankenstein; or, The Modern Prometheus*. Lackington, Hughes, Harding, Mavor & Jones, London.

Piek Vossen. 1998. Introduction to EuroWordNet. In Piek Vossen, editor, *EuroWordNet: A Multilingual Database with Lexical Semantic Networks for European Languages*. Kluwer, Dordrecht. Pages 1–17.