Tradita innovare, innovata tradere[⋆]

The Gothenburg approach to computational lexicography

Lars Borin[1,*,†], Louise Holmer[1,*,†]

[1]*Department of Swedish, Multilingualism, Language Technology, University of Gothenburg, Box 200, SE-405 30, Gothenburg, Sweden*

Abstract

Swedish computational lexicography has a long history at the University of Gothenburg, both in its primary role as a central aspect of the scientific study of vocabulary and also as an infrastructural component for conducting research based on language data. Starting in the 1960s, the Språkdata research group pioneered corpus-supported lexicography for Swedish, forming the basis for successive editions of the two main descriptive dictionaries of contemporary Swedish, SAOL and SO. Language technological lexical resources for Swedish have been developed by the research unit/research infrastructure Språkbanken Text since the turn of the millennium, most recently in the framework of the *Swedish FrameNet++* initiative. After two decades of separation, these two largely mutually independently developed strands of computational lexicography have now joined forces under the umbrella of *Språkbanken's lexical research infrastructure* to advance the field technically, methodologically, and scientifically.

Keywords

Saldo, SAOL, SO, Språkbanken Text, lexicon, lexical resource, lexical infrastructure

1. Introduction

The combination of computers and lexicography has a long and distinguished history at the University of Gothenburg. Almost 60 years ago, in 1965, Sture Allén initiated the collection of digital texts for what was to become the first Swedish text corpus – the one-million word *Press 65*[1] – in order to be able to address research questions and aims such as "In a broad sense, what are the lexical units of Swedish as represented by a large corpus? How common are they, and how are they distributed over different text types? The results were primarily to be published in a frequency dictionary." [2, p. 61, our translation].

Språkdata, the research unit founded by Allén, pursued corpus-supported lexicography for many years, concurrently with activities aimed at promoting and developing computational lin-

Huminfra Conference 2024, Gothenburg, 10–11 January 2024.

[⋆]For the main title of the present paper we have borrowed the motto of our university *Tradita innovare, innovata tradere* 'Renew [our] heritage, [and] pass [it] on renewed', a description that we feel fits our approach to construction of a research infrastructure for computational lexicography like a glove.

[*]Corresponding author.

[†]These authors contributed equally.

lars.borin@svenska.gu.se (L. Borin); louise.holmer@svenska.gu.se (L. Holmer)

0000-0001-5434-9329 (L. Borin); 0009-0009-6763-6672 (L. Holmer)

[1]Thus, Press-65 appeared only a few years after what is usually considered the first modern text corpus, the one-million word *Brown Corpus* of American English [1].

guistics as a discipline in Gothenburg and Sweden. For various reasons, these two strands started to diverge in the early years of this millennium, but since 2021, they are again organizationally unified, with an expressed synergistic aim. The two strands have developed separately over the years, being somewhat non-communicating vessels considering researchers as well as databases and research output (graphically illustrated in Figure 1, where the two parallel strands cover roughly the twenty-year period 2002–2021).

An important historical milestone was the establishment in 1975 of *Språkbanken* ('the Swedish Language Bank') as a dedicated research infrastructure operated by Språkdata in support of Swedish linguistic research in general and the local lexicographical activities in particular. Språkbanken has grown considerably over the years from its humble beginnings almost 50 years ago. The main focus of its present incarnation – *Språkbanken Text* – is on language technological research rather than corpus linguistics as in the beginning of its existence. The lexicographical element has been very much present throughout its history, as described in more detail below.

The aim of this paper is to describe the background and current state of our computational lexicography infrastructure, and to point to some possible future directions for its development.

2. The *tradita*

The lexicographical projects in Gothenburg, aiming at publishing dictionaries, resulted in two print dictionaries in the 1980s, namely the first edition of the Swedish monolingual *Svensk ordbok* ('Swedish Dictionary' 1986 [3]; in short SOB) [4] and the 11th ed. of the *Swedish Academy Glossary* (1986 [5]; henceforth SAOL). These two datasets (SAOL and SOB, as well as their revised, later editions, respectively: [6, 7, 8]; [9], referred to as SO) have so far been treated as separate entities during the years of development. They have been revised and refined, one dictionary at a time, often by more or less the same lexicographers. SAOL and SO are financed by the Swedish Academy, and their editorial staff is employed by the University of Gothenburg [10].

2.1. SAOL: main features and functions

In this article, we mainly focus on SAOL. The first edition was published in 1874, and SAOL is therefore celebrating its 150th anniversary in 2024. SAOL is a contemporary, monolingual dictionary, aiming at providing information on orthography, inflectional patterns, and, to some extent, word formation and pronunciation [11, 12, 13].

SAOL has a unique position in Sweden – it is considered an unofficial norm, mostly regarding orthography and inflection.[2] Even so, many users tend to turn to SAOL for semantic information as well, although semantics is not one of the Glossary's main features [12]. Further, a common opinion among users is that SAOL contains (only) accepted Swedish words [11]. This is of course a misconception, since a dictionary by necessity at best contains a rich selection of words rather than "the definite vocabulary".

A ground principle for the editors of SAOL is to include new words in every new edition, preferably conventionalized ones with sufficient frequency in text, mainly contemporary newspaper text.

[2]There is no official normative dictionary of Swedish at the present time.

The word formation rules of Swedish allow an almost infinite number of solid compounds and derivatives, and one major editorial task in the revisional process is to decide which new words to include and which to exclude [14]. Although SAOL comprises about 126,000 headwords, it is inevitable that it will have lacunas with regards to neologisms and regularly formed derivatives [15]. It will also most certainly contain obsolete words, that ought to be excluded for various reasons [10].

The preliminary manuscript of SAOL 12 was transferred to the University of Gothenburg in 1984 [14, 4], previously being revised in Lund by the editorial staff of the historical *Swedish Academy Dictionary* (SAOB [16]). Since 2017, SAOL and SO, together with SAOB, can be accessed in the same dictionary web portal, Svenska.se. Users tend to consult SAOL mainly via the app version or the web version on Svenska.se nowadays. In the web version, search results show three (often) different lexicographical analyses for the same word: one with focus on orthography and inflection, and a more normative perspective (SAOL), one with focus on semantics, constructions and etymology and a more descriptive perspective (SO), and one providing exhaustive semantic information on a word and its other linguistic properties and their development over time (SAOB). This side-by-side presentation has made it evident to the lexicographers that coordination and harmonization between SAOL and SO are desirable. Classification in terms of parts of speech, lemma variants, lemma order between homographs etc. varies between the two dictionaries, and users tend to (with good reason) question the differences [17, 13].

Despite the advantages of Svenska.se, SAOL is one of the few contemporary dictionaries that is planned to be published in print, with the next (15th) edition slated to appear in late 2025. This should be regarded as quite exceptional since the major publishing houses in Sweden have by and large discontinued their lexicographic activities for commercial reasons over the last two decades [15].

2.2. SMDB: *a Swedish morphological database*

Starting with the 12th edition of SAOL (1998 [6]), a morphological database was created, SMDB, containing the 120,000 headwords and all their inflectional forms. In the print dictionaries, inflectional suffixes are presented in abbreviated form, but the SMDB allowed the inflectional paradigms to be presented in full with morphosyntactic descriptors [18]. The full inflectional paradigms of SMDB have, among other things, formed the basis of inflectional information in the e-versions of SAOL, from CD-ROM (2007), over smart phone apps (2011 and onwards) to the web version (2017).[3] SMDB has also been used as a lexicographical tool for corpus investigations; by comparing SAOL's headwords and their inflectional forms to modern texts, it is possible to tease out which forms (and lemmas) are in use in texts, and which have become obsolete and are no longer in use, hence being potential candidates for exclusion from the dictionary [14]. There was also an outspoken aim for the SMDB to be continously updated and connected to the corpora of Språkbanken [18].

[3]There are earlier e-versions as well, such as SAOL 11 on floppy disks etc.

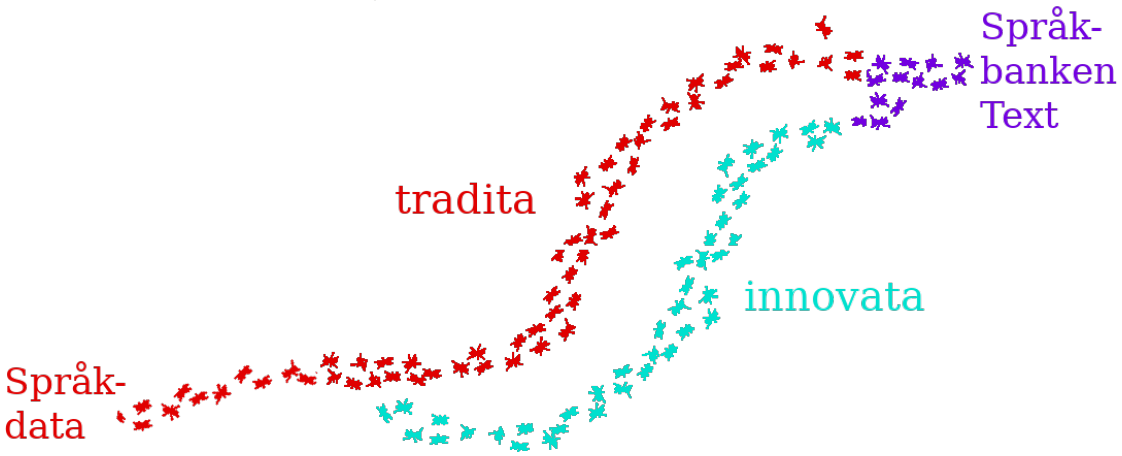

Figure 1: Computational lexicography at Gothenburg from Språkdata to Språkbanken Text. The *tradita* represents the continued development of corpus-supported lexicography as initiated by Språkdata, and the *innovata* the language technological approach to lexical resource building

3. The *innovata*

In the early years of this millennium, Språkbanken's focus started to shift from traditional corpus linguistics in the direction of mainstream language technology. One central aim in this connection was to develop computational tools for automatic linguistic annotation of the considerable amounts of text collected in Språkbanken's corpora,[4] in order to make them available for research in language technology as training and testing data. Following general practice in the field, the software and language resources used should be open and freely available, ideally for all purposes, not least in order to ensure reproducibility of research. At the time, there was no freely available full-sized Swedish digital lexical resource that could be used as the basis of morphological analysis and lemmatization of the corpora, and the content of the Swedish Academy dictionary databases developed in-house could not be made openly available because of commercial commitments.

3.1. Enter Saldo

Instead, Språkbanken in practice came to initiate a parallel computational lexicographic project, which took its point of departure in SAL (*Svenskt associationslexikon* 'Swedish Associative Thesaurus'), an onomasiological (semantic) Swedish dictionary developed by Lennart Lönngren at Uppsala University between 1987 and 1992 [19, 20]. There was still a clear connection to our local lexicographical projects, since Lennart Lönngren had purchased the lemma list of the recently published first edition of SO (SOB 1986 [3]) from the Språkdata group and used this as the backbone of SAL.

[4]With one exception (the part-of-speech tagged Parole corpus; https://spraakbanken.gu.se/en/resources/parole), the quite extensive text corpora of Språkbanken had no annotations at the time, thus allowing only for various forms of text-word and string searches.

The original SAL took the semantic-lexicon approach to an extreme, providing no information at all about linguistic form, except for using lemmas (with numerical indices in cases of colexification) as convenient labels of word senses. Thus, there was no information about part of speech, let alone inflection. In order to use it as the basis for linguistic analysis of arbitrary Swedish text, this formal linguistic information had to be added. For the inflectional information, we used a morphological processing application developed as a PhD project at nearby Chalmers University of Technology, Markus Forsberg's *functional morphology* (FM; [21, 22]). In practice, the FM processor is used to generate a full-form lexicon from Saldo after each update of its contents. This lexicon is made available as a separate lexical resource, *Saldom* (short for 'Saldo's morphology'). At corpus import time, text words are matched against the full-form list and their morphosyntactic description(s) retrieved from Saldom, rather than morphologically decomposed on the fly.[5] The only rule-based processing taking place at corpus import time is compound analysis, since the set of Swedish compounds is completely open-ended as mentioned above in Section 2.1.

The final product comprised a full-sized – containing slightly over 72,000 entries in its first release – semantic dictionary of Swedish with complete morphological specifications (inflectional paradigms plus compounding forms) provided for all entries [23], released under an open (CC-BY) license allowing all kinds of use, including for commercial purposes [24, 25]. The present development version of Saldo contains 147,650 entries, and the latest official release (Saldo 2.3, from 2015) holds slightly over 131,000 entries, i.e., Saldo is approximately comparable in size to SAOL (126,000 headwords).

Saldo differed from the print-dictionary project datasets in at least two important ways, viz. by its organization as an onomasiological lexicon and by its data model, which was explicitly designed as a formal language inspired by knowledge representation languages such as those used in the semantic web, to cater to the needs of both humans (mainly by having meaningful identifiers instead of e.g. numbers) and machines (by having an explicit syntax and compositional semantics). In the case of Saldo, the "database" is a deliberately designed intrinsic part of the lexical resource, whereas the databases used in the print-dictionary projects have always in practice been extrinsic to the lexical data: purely technical storage solutions, as it were.

There are also more subtle differences having to do with the differing aims of the two lexicographical undertakings, primarily concerning the role of easily inferrable (to a human native speaker of Swedish) information. Even though the current working version of Saldo contains more entries than the most recent edition of SAOL, the two sets of entries are not commensurable, since Saldo explicitly lists many items which are implicit in SAOL, e.g. participles – formally (deverbal) adjectives in Saldo, but inflected forms of verbs in SAOL – and verbal nouns in *-nde*. In both cases these are provided by SAOL only exceptionally. In order to serve the practical purpose of high-accuracy automatic text analysis, Saldo also includes a number of non-normative spellings and inflectional forms which frequently occur in real-world texts, but which normally are not listed in conventional reference dictionaries such as SAOL and SO.

[5] Saldom is thus broadly comparable to SMDB, mentioned in Section 2.2.

3.2. Towards Swedish FrameNet++

Even if there was no free Swedish computational lexicon available before Saldo which was both large and general enough for the intended purposes, the long history of lexicographical activity in Språkdata and Språkbanken had left behind a number of smaller and more specialized computational lexicons, resulting from various projects carried out through the years, to which could be added initiatives started elsewhere, such as the (partial) *Swedish WordNet* compiled at Lund University [26], or the crowdsourced *People's Synonym Lexicon* created and maintained at the Royal Institute of Technology in Stockholm [27]. The *Swedish FrameNet++* project was initiated around 2010, with two complementary and interlocking aims. One aim was to combine the rich, painstakingly compiled linguistic information hidden in these both formally and content-wise quite heterogeneous resources into one unified lexical macroresource, SweFN++. The other aim was to create a computational infrastructure facilitating development of the resources themselves as well as research based on their content [25, 28].

4. The present: *innovata tradere*

So, has the *tradita* been renewed? We like to think so, in many respects. Starting in 2021, the lexicographical projects formerly organized under the Centre for Lexicology and Lexicography at our department were formally made a part of Språkbanken Text, and merged with the SweFN++ activities under a new umbrella designation: *Språkbanken's lexical research infrastructure*. In a way, this move signalled a return to the pre-2000 organization, but at a considerably higher level of technical and methodological maturity, the former originating primarily in Språkbanken Text and the latter contributed in equal and complementary parts by the two strands of lexicographical R&D that have now joined forces.

Furthermore, the underlying databases of SAOL and SO have been migrated into the Karp lexical platform, which has been under active development for over a decade as a tool for working with formally structured language data [29], notably the computational lexical databases making up SweFN++ (in particular the Swedish FrameNet [30] and the Swedish constructicon [31]). The migration has also resulted in a long sought-after union, and to some extent harmonization, of the two sibling print-dictionary database structures (SAOL and SO).

5. The future: *tradita innovare*?

We see a bright future for computational lexicography in Gothenburg. With the recent developments described in the previous section, the strengths of the two strands that were pursued separately for two decades are synergistically combined. The result is a vibrant and multifaceted research environment intertwined with and supported by a closely integrated cutting-edge computational infrastructure for working with lexical data. This will advance Swedish computational lexicography technically, methodologically, and scientifically, and serve a broad range of R&D purposes, in particular in the humanities and social sciences.

We will now be able to draw both on highly information-rich Swedish lexical databases compiled and enriched over several decades by highly trained lexicographers and on the most

recent language technologies built on deep learning and AI. Some promising directions for the short and medium term future are development of new or improved sophisticated computational tools for mining very large text corpora in order to

- find evidence for new words and word usages, as well as obsolescing word usages [32];
- investigate phraseology and multi-word expressions [33, 34];
- track the historical development of the Swedish lexicon [35, 36, 37, 38];
- contribute to the state of the art of lexical typology [39, 40]; and, of course
- make better dictionaries (for human consumption) and lexical resources (for computer processing of Swedish text).

Acknowledgments

The work reported on here has received funding from a number of sources. Språkbanken Text is the coordinating partner of *Nationella språkbanken*, a national research infrastructure funded jointly by the Swedish Research Council and the 10 partner institutions (grant no. 2017-00626, 2018–2024). The print dictionary projects have for many years been financially supported by the Swedish Academy. In its various guises over the years, Språkbanken Text has received additional funding centrally from the University of Gothenburg, from its Faculty of Humanities, and from the Department of Swedish, Multilingualism, Language Technology, which hosts the infrastructure. For the extensive financial support that the SweFN++ initiative has received over the years, see [41, p. 8]

References

[1] W. N. Francis, H. Kučera, Computational Analysis of Present-Day American English, Brown University Press, Providence, 1967.

[2] S. Allén, Språkvetenskaplig databehandling, in: Personliga tillbakablickar över ämnesområden vid Göteborgs universitet: Seniorakademien Dokumentationsserie Del 2, Seniorakademien vid Göteborgs universitet, Gothenburg, 2014, pp. 61–66. http://hdl.handle.net/2077/51552.

[3] SOB, Svensk ordbok utgiven av Svenska Akademien, 1 ed., Esselte studium, Solna, 1986.

[4] S.-G. Malmgren, E. Sköldberg, The lexicography of Swedish and other Scandinavian languages, International Journal of Lexicography 26 (2013) 117–134. doi:10.1093/ijl/ect008.

[5] SAOL 11, Svenska Akademiens ordlista, 11 ed., Norstedts, Stockholm, 1986.

[6] SAOL 12, Svenska Akademiens ordlista, 12 ed., Norstedts, Stockholm, 1998.

[7] SAOL 13, Svenska Akademiens ordlista, 13 ed., Norstedts, Stockholm, 2006.

[8] SAOL 14, Svenska Akademiens ordlista, 14 ed., Norstedts, Stockholm, 2015.

[9] SO 1, Svensk ordbok utgiven av Svenska Akademien, 2 ed., Svenska Akademien, Stockholm, 2021.

[10] E. Sköldberg, L. Holmer, E. Volodina, I. Pilán, State-of-the-art of monolingual lexicography for Sweden, Slovenščina 2.0 7 (2019) 13–24. doi:10.4312/slo2.0.2019.1.13-24.

[11] M. Gellerstam, Vad är Svenska Akademiens ordlista?, in: M. Gellerstam (Ed.), SAOL och tidens flykt: Några nedslag i ordlistans historia, Norstedts, Stockholm, 2009, pp. 11–30.
[12] S.-G. Malmgren, Svenska Akademiens ordlista genom 140 år: Mot fjortonde upplagan, LexicoNordica 21 (2014) 81–98.
[13] K. Blensenius, L. Holmer, E. Sköldberg, SAOL 14 som rättesnöre – diskussion om den senaste upplagan, LexicoNordica 28 (2021) 39–58.
[14] S. Berg, L. Holmer, E. Sköldberg, Time to say goodbye? On the exclusion of solid compounds from the Swedish Academy Glossary (SAOL), in: A. Dykstra, T. Schoonheim (Eds.), Proceedings of the 14th EURALEX International Congress, Fryske Akademy, Leeuwarden/Ljouwert, 2010, pp. 567–576.
[15] L. Holmer, Neutrala substantiv på -ande i text och ordbok, 47, Meijerbergs arkiv för svensk ordforskning, Gothenburg, 2022.
[16] SAOB, Svenska Akademiens ordbok, Gleerups, Lund, 1898–. URL: https://www.saob.se/.
[17] E. Bäckerud, P. Nilsson, E. Sköldberg, Så används Svenska Akademiens ordböcker på nätet. Implicit och explicit feedback från användarna, in: C. Sandström, U.-M. Forsberg, C. af Hällström-Reijonen, M. Lehtonen, K. Ruppel (Eds.), Nordiska studier i lexikografi, volume 15, Nordisk Forening for Leksikografi, Helsinki, 2020, pp. 91–101.
[18] S. Berg, Y. Cederholm, Att hålla på formerna: Om framväxten av Svensk morfologisk databas, in: S. Allén, S. Berg, S.-G. Malmgren, K. Norén, B. Ralph (Eds.), Gäller stam, suffix och ord: Festskrift till Martin Gellerstam den 15 oktober 2001, 29, Meijerbergs arkiv för svensk ordforskning, Gothenburg, 2001, pp. 58–69.
[19] L. Lönngren, Lexika, baserade på semantiska relationer, in: Nordiske Datalingvistikdage og Symposium for datamatstøttet leksikografi og terminologi 1987. Proceedings, Handelshøjskolen i København, Institut for Datalingvistik, Copenhagen, 1988, pp. 229–236.
[20] L. Lönngren, A Swedish associative thesaurus, in: Euralex '98 proceedings, Vol. 2, Euralex, Liège, 1998, pp. 467–474.
[21] M. Forsberg, A. Ranta, Functional morphology, in: ICFP'04. Proceedings of the ninth ACM SIGPLAN international conference of functional programming, ACM, Snowbird, 2004, pp. 213–223. doi:10.1145/1016848.1016879.
[22] M. Forsberg, Three Tools for Language Processing: BNF Converter, Functional Morphology, and Extract, Ph.D. thesis, Göteborg University and Chalmers University of Technology, 2007.
[23] L. Borin, M. Forsberg, L. Lönngren, The hunting of the BLARK – SALDO, a freely available lexical database for Swedish language technology, in: J. Nivre, M. Dahllöf, B. Megyesi (Eds.), Resourceful language technology. Festschrift in honor of Anna Sågvall Hein, number 7 in Acta Universitatis Upsaliensis: Studia Linguistica Upsaliensia, Uppsala University, Department of Linguistics and Philology, Uppsala, 2008, pp. 21–32.
[24] L. Borin, M. Forsberg, L. Lönngren, SALDO: A touch of yin to WordNet's yang, Language Resources and Evaluation 47 (2013) 1191–1211. doi:10.1007/s10579-013-9233-4.
[25] L. Borin, M. Forsberg, L. Lönngren, N. Zechner, Swedish FrameNet++: Lexical samsara, in: D. Dannélls, L. Borin, K. Friberg Heppin (Eds.), The Swedish FrameNet++: Harmonization, Integration, Method Development and Practical Language Technology Applications, John Benjamins, Amsterdam, 2021, pp. 69–95. doi:10.1075/nlp.14.
[26] Å. Viberg, K. Lindmark, A. Lindvall, I. Mellenius, The Swedish WordNet project, in:

Proceedings of EURALEX 2002, University of Copenhagen, Copenhagen, 2002, pp. 407–412. URL: https://euralex.org/publications/the-swedish-wordnet-project/.

[27] V. Kann, M. Rosell, Free construction of a free Swedish dictionary of synonyms, in: Proceedings of Nodalida 2005, University of Joensuu, Joensuu, 2006, pp. 105–110.

[28] D. Dannélls, L. Borin, K. Friberg Heppin (Eds.), The Swedish FrameNet++: Harmonization, Integration, Method Development and Practical Language Technology Applications, John Benjamins, Amsterdam, 2021. doi:10.1075/nlp.14.

[29] L. Borin, M. Forsberg, L.-J. Olsson, J. Uppström, The open lexical infrastructure of Språkbanken, in: Proceedings of LREC 2012, ELRA, Istanbul, 2012, pp. 3598–3602.

[30] D. Dannélls, L. Borin, M. Forsberg, K. Friberg Heppin, M. T. Gronostaj, Swedish FrameNet, in: D. Dannélls, L. Borin, K. Friberg Heppin (Eds.), The Swedish FrameNet++: Harmonization, Integration, Method Development and Practical Language Technology Applications, John Benjamins, Amsterdam, 2021, pp. 37–65. doi:10.1075/nlp.14.

[31] B. Lyngfelt, L. Bäckström, L. Borin, A. Ehrlemark, R. Rydstedt, Constructicography at work: Theory meets practice in the Swedish constructicon, in: B. Lyngfelt, L. Borin, K. Ohara, T. T. Torrent (Eds.), Constructicography: Constructicon Development Across Languages, John Benjamins, Amsterdam, 2018, pp. 41–106. doi:10.1075/cal.22.03lyn.

[32] M. Forsberg, J. Sikora, E. Sköldberg, Words unboxed: Discovering new words with Kubord, 2023. KBLab blog post: https://kb-labb.github.io/posts/2023-08-29-kubord/.

[33] L. Borin, Multiword expressions: A tough typological nut for Swedish FrameNet++, in: D. Dannélls, L. Borin, K. Friberg Heppin (Eds.), The Swedish FrameNet++: Harmonization, Integration, Method Development and Practical Language Technology Applications, John Benjamins, Amsterdam, 2021, pp. 221–259. doi:10.1075/nlp.14.

[34] E. Sköldberg, Phraseological theory, evidence in corpora and lexicographical practice: On collocations in a monolingual dictionary of Swedish, in: K. Blensenius (Ed.), Valency and constructions: Perspectives on combining words, 46, Meijerbergs arkiv för svensk ordforskning, Gothenburg, 2022, pp. 155–182.

[35] J. Viklund, L. Borin, How can big data help us study rhetorical history?, in: Selected Papers from the CLARIN Annual Conference 2015, LiUEP, Linköping, 2016, pp. 79–93.

[36] E. Sköldberg, L. Holmer, Ordböcker som språkhistoriska källor, Svenskläraren: Tidskrift för svenskundervisning 61 (2017) 20–21.

[37] S. Petersson, E. Sköldberg, Semantic change in Swedish – from a lexicographic perspective, in: N. Tahmasebi, L. Borin, A. Jatowt, Y. Xu, S. Hengchen (Eds.), Computational approaches to semantic change, Language Science Press, Berlin, 2013, pp. 149–167. doi:10.5281/zenodo.5040308.

[38] Y. Adesam, P. Andersson, L. Borin, G. Bouma, A lexical resource for computational historical linguistics, in: D. Dannélls, L. Borin, K. Friberg Heppin (Eds.), The Swedish FrameNet++: Harmonization, Integration, Method Development and Practical Language Technology Applications, John Benjamins, Amsterdam, 2021, p. 98–121. doi:10.1075/nlp.14.

[39] L. Borin, Core vocabulary: A useful but mystical concept in some kinds of linguistics, 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, Springer, Berlin, 2012, p. 53–65.

[40] L. Borin, B. Comrie, A. Saxena, The Intercontinental Dictionary Series – a rich and principled database for language comparison, in: L. Borin, A. Saxena (Eds.), Approaches to Measuring Linguistic Differences, De Gruyter Mouton, Berlin, 2013, pp. 285–302.
[41] L. Borin, D. Dannélls, K. Friberg Heppin, Introduction, in: D. Dannélls, L. Borin, K. Friberg Heppin (Eds.), The Swedish FrameNet++: Harmonization, Integration, Method Development and Practical Language Technology Applications, John Benjamins, Amsterdam, 2021, pp. 3–35. doi:10.1075/nlp.14.

A. Online Resources

- Svenska.se
- Språkbanken Text
- Språkbanken Text: Press 65
- Språkbanken Text: Saldo
- Språkbanken Text: Saldom
- Swedish FrameNet++
- Språkbanken Text: Karp (lexical platform)
- Språkbanken Text: Korp (corpus platform)
- Språkbanken Text: lexical datasets