

The lexical editing system of Karp

**Lars Borin, Markus Forsberg, Leif-Jöran Olsson,
Olof Olsson and Jonatan Uppström**

Språkbanken, Department of Swedish, University of Gothenburg, Sweden
{lars.borin, markus.forsberg, leif-joran.olsson, olof.olsson2, jonatan.uppstrom}@gu.se

Abstract

Karp is the open lexical infrastructure of Språkbanken (the Swedish Language Bank). The infrastructure has three main functions: (1) to support the work on creating, curating, and integrating our various lexical resources; (2) to publish the resources, making them searchable and downloadable; and (3) to offer advanced editing functionalities. An important feature of the lexical infrastructure is also that we maintain a strong bidirectional connection to our corpus infrastructure. At the heart of the infrastructure is the SweFN++ project with the goal to create free Swedish lexical resources geared towards language technology applications. The infrastructure currently hosts 23 Swedish lexical resources. The resources are integrated through links to a pivot lexical resource, SALDO, a large morphological and lexical-semantic resource for modern Swedish.

Keywords: lexicon, editing, infrastructure, Swedish language resources, language technology, LMF

1. Introduction

The research and development unit Språkbanken¹ (the Swedish Language Bank) at the University of Gothenburg has since its establishment in 1975 accumulated a large variety of language resources, including corpora of over two billion words of modern and historical Swedish text and a multitude of lexical resources. Some of the lexical resources are digitized dictionaries describing older forms of Swedish, but most of them are contemporary resources intended for NLP use. For most of these, the development of an adequate technical support infrastructure has been hampered by limited research funding, thus leading to the adoption of suboptimal technical solutions such as simple form-based frontends to relational databases or even tab-separated text files, saddling the lexicographers with the responsibility for making sure that any formal requirements are met and for manually weeding out any inconsistencies.

The SweFN++ project (Borin et al., 2010a) had the objective to create, curate, and integrate free Swedish lexical resources with the explicit goal of making them usable for language technology applications. META-NORD² is a broad EC-funded European collaboration with the aim of upgrading and harmonizing language resources and

¹ <<http://spraakbanken.gu.se>>

² <<http://www.meta-nord.eu>>

tools for the Nordic and Baltic languages and making them available across Europe. Thanks to these two, and other externally funded infrastructure initiatives,³ Språkbanken has had the opportunity to focus on safeguarding its existing language technology resources, as well as to develop a generalized lexical infrastructure, referred to as Karp (Borin et al., 2012a). The heart of the lexical infrastructure is a large network of interconnected lexicons (Borin, 2010; Borin et al., 2010a), all encoded in the LMF format (Lexical Markup Framework; see ISO, 2008; Francopoulo, 2013).

Even though our digital lexical resources are primarily intended for use in NLP applications, they are still very much lexicographical entities. Thus, from a linguistic point of view, the work on individual resources as well as on their integration is at heart a genuinely lexicographical activity, to boot one with considerable potential to make significant theoretical contributions to lexicology, lexical semantics and lexical typology because of the large-scale empirical nature of our endeavor and the diversity of the lexical resources involved. In general, working with large amounts of data as we do, requires good tools for interacting with the data, for abstracting, ordering, searching and visualizing the data, for inferring and presenting relations among data items, and for editing the data. The Karp component of our lexical infrastructure has been designed with these aims in mind.

An important feature of the lexical infrastructure is that we maintain a strong bidirectional connection to our corpus infrastructure Korp (Borin et al., 2012b). For example, the corpora are annotated with the lexical information in Karp, and the language examples for the lexical resources in Karp are retrieved from Korp.

A pervasive theme of the infrastructure is openness, which may be seen as a philosophical stance – we believe that research should be carried out in the open to enable inspection and increased collaboration. Openness pervades the infrastructure, in the use 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.

One essential part of this infrastructure is a generic search interface, <<http://spraakbanken.gu.se/karp>>, which provides a plug-and-play search tool for

³In this context, an important initiative is CLARIN <<http://www.clarin.eu>>. Although Sweden is not yet a member of CLARIN, Språkbanken is involved in some CLARIN activities and is also the coordinating partner of a recently submitted funding application for Swedish membership. In the development of Karp and other infrastructure components, we pay close attention to the standards and best practices defined by CLARIN, in order to be able to quickly set up a CLARIN service center when Sweden decides to join CLARIN.

resources already in LMF, where the LMF format is employed both internally within the infrastructure and, trivially, as an export format. As a next logical step, we have augmented the search interface of Karp with editing functionalities, where authorized users may edit and create new lexical entries.

2. The lexical resources

The lexical infrastructure has one primary lexical resource, a pivot, to which all other resources are linked. This is SALDO⁴ (Borin and Forsberg, 2009; Borin et al., 2013b), a large (130K entries and 1.9M wordforms), freely available 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 have been assigned persistent identifiers (PIDs) to which the lexical information in other resources are linked.

Some of the 23 resources (including the pivot resource SALDO) have been created from scratch using existing free resources, both external and in-house. For example, Swesaurus, a Swedish wordnet (Borin and Forsberg, 2010; Borin and Forsberg, 2011b), is being built using not only in-house but also external resources, such as Synlex (Kann and Rosell, 2006), the Swedish Wiktionary,⁵ and more indirectly, using semantic relations extracted from Princeton WordNet (Fellbaum, 1998b) through links between SALDO and Core Princeton WordNet (Boyd-Graber et al., 2006).

Other resources are the result of digitization and (manual and automatic) post-processing of existing paper dictionaries. This holds generally for the historical lexicons and their associated (partial) morphologies (Borin and Forsberg, 2008; Borin and Forsberg, 2011a; Borin et al., 2010b).


As an illustration of the diversity of the resources, here follows a selection of the results of the word form query *springa* in Karp. The selection consists of 13 out of 62 results in seven out of 23 resources.

springa in SALDO

The word sense *springa* ‘run’ in SALDO is described with two semantic relations, the primary relation *röra sig* ‘move’ and the secondary relation *fort* ‘fast’. Furthermore, we have relations where *springa* acts as the primary or secondary, i.e. the reverse relations collectively referred to as children. SALDO has two more word senses of *springa*, one noun and one verb, not shown here, and *springa* is also a component of 11 particle verbs, e.g. *springa bort* ‘run away’.

⁴ <<http://spraakbanken.gu.se/saldo>>

⁵ <<http://sv.wiktionary.org>>


Sense	POS	Primary	Secondary	Children (primary)	Children (secondary)
<i>springa</i> 	verb	<i>röra sig</i>	<i>fort</i>	<ul style="list-style-type: none"> ○ <i>lubba –</i> ○ <i>kesa kalv</i> ○ <i>ifråsprungen ifrån</i> Show all (30)	<i>springa ihop² kollidera</i>

springa in SALDO morphology

The morphological description of SALDO is a separate resource that lists lemmagrams associated with word senses, where lemmagrams are form units denoting inflection tables.⁶

The lemmagram *springa* (noun), which denotes the inflection table of the two verb senses of *springa*, illustrates the connection to the corpora infrastructure: next to the small raven we see the number of hits in Korp’s corpora collection (307,539 hits), and the table shows which of the word forms are attested: only *sprunges* (passive past subjunctive) is unattested.

springa (verb)

 (307,539)

springa ...

pres ind aktiv	<i>springer</i>
pres ind s-form	<i>springes</i>
pres ind s-form	<i>springis</i>
pres konj aktiv	<i>springe</i>
pres konj s-form	<i>springes</i>
pret ind aktiv	<i>sprang</i>
pret ind s-form	<i>sprangs</i>
pret konj aktiv	<i>sprunge</i>
pret konj s-form	<i>sprunges</i>
imper	<i>spring</i>
inf aktiv	<i>springa</i>
inf s-form	<i>springas</i>

springa in Swedish FrameNet

The word sense *springa* ‘run’ is a lexical unit in the frame Self_motion in the Swedish FrameNet. A click on the frame name takes us to the full description of the frame.

Sense	Frame
<i>springa</i>	<i>Self_motion</i>

A frame is a large information unit: only part of the entry is shown here. Self_motion is a frame from the Berkeley FrameNet (Baker et al., 1998; Ruppenhofer et al., 2005)

⁶ More specifically, a lemmagram is a unique combination of a citation form and certain other formal characteristics, in SALDO pronunciation, part of speech, inflectional paradigm and compounding behavior. This corresponds to one usage of the term lemma, but unfortunately this term also is used in other meanings, e.g. ‘citation form’, which is why we have opted for coining a new, unambiguous term.

that the Swedish FrameNet is built upon, and the core and non-core elements have been directly imported from that resource. The word sense *springa* ‘run’ occurs in past tense, *sprang*, in the first annotated example: *Två hästar på rymmen sprang ...* ‘Two horses on the loose ran (on the roadway in southern Södertälje in the afternoon.)’, but is also listed among the lexical units (not shown).

Self_motion

Domain: **Gen**

Semantic Type: **Move**

Core Elements: **Area Direction Goal Path Self_mover Source**

Peripheral Elements: **Concessive Coordinated_event Cotheme Depictive Distance Duration External_cause Internal_cause Manner Means Path_shape Place Purpose Reason Result Speed Time**

Examples:

Två hästar på rymmen *sprang ute på vägbanan i södra Södertälje på eftermiddagen*.

springa in Swesaurus

The word sense *springa* ‘run’ has seven graded synonymy relations in Swesaurus, all extracted from Synlex (Kann and Rosell, 2006). They are all manners of running, such as *rusa* ‘rush’, roughly corresponding to the troponyms of Princeton WordNet (Fellbaum, 1998a).

Sense	Degree	Sense	Type	Source
<i>springa</i>	↖80	↗ <i>kila</i> ²	syn	fsl
<i>springa</i>	↖92	↗ <i>kuta</i>	syn	fsl
<i>springa</i>	↖94	↗ <i>lubba</i>	syn	fsl
<i>springa</i>	↖88	↗ <i>löpa</i>	syn	fsl
<i>springa</i>	↖80	↗ <i>rusa</i>	syn	fsl
<i>springa</i>	↖90	↗ <i>ränna</i> ²	syn	fsl
<i>springa</i>	↖62	↗ <i>skubba</i>	syn	fsl

springa in IDS/LWT

The IDS/LWT list is a massively multilingual vocabulary of 1,460 word senses used for typological studies. The basic list, 1,310 entries, was first compiled in the Intercontinental Dictionary Series project (Borin et al., 2013a), and new languages are continually being added to the IDS archive at the Max Planck Institute for Evolutionary Anthropology in Leipzig.⁷ Another 150 entries were added when the IDS


⁷ <<http://lingweb.eva.mpg.de/ids/>>

list was used in the recent Loanword Typology project (Haspelmath and Tadmor, 2009).⁸ The new information provided in our version of the Swedish IDS/LWT list is the link between *springa* ‘run’ and the IDS/LWT id S10.460, thereby providing a link from our Swedish lexical resources to a basic vocabulary in over 200 languages.

Sense	FormRepresentation			
<i>springa</i>	LWT ID	English	Definition	Example
	S10.460	to run	(intransitive)	"They ran all the way to school."

springa in Schlyter

Schlyter (1887) is an Old Swedish dictionary describing the vocabulary of Old Swedish law texts, which becomes clear in the definition text: it describes the expression *springa af kaghen* ‘run of the scaffold’, which is a punishment involving a pillory on an elevated platform for public shame or whipping.

springa (verb)
 (40)

springa v. n.

1. *springa*. Chr. * s. af kaghen, o: stå vid skampålen på den upphöjda schavotten för att skämmas l. hudstrykas, och sedan *springa* ned därifrån (jfr. Kgh, och VSt. o. kac), Sk.*; detta straff beskrives tydligare i Dr. Margaretas D. stadsrätt c. 55, där det säges att brottslingen schall settis paa kaget (läs kagen) och self needsprunge (KR. GDL . V. 511; jfr. s. 596).

springa in Diapivot

The Diapivot resource (Borin and Forsberg, 2011a; Andersson and Ahlberg, 2013) provides diachronic links between the lemgrams of the four morphological resources: SALDO, the SALDO morphology (the pivot); Dalin, a 19th century morphology; Swedberg, a 17th century morphology; and finally, an Old Swedish morphology. The linking is done using SKOS (Simple Knowledge Organization System; see Miles and Pérez-Agüera, 2007) where the linking relations are either equivalence or a broader-narrower relationship.

In the lexical entry below we have three lemgrams *springa* (verb), which are not formally the same: they all live in different namespaces. The first one is from the pivot SALDO morphology, followed by a lemgram in Dalin, and three lemgrams in the Old Swedish morphology. All are linked with the equivalence relation.

<i>springa</i> (verb)  (307,539)	<i>springa</i> (verb)  (1) Dalin	<i>löpa</i> (verb)  (158) Old Swedish	<i>skumpa</i> (verb)  (0) Old Swedish	<i>springa</i> (verb)  (40) Old Swedish
--	---	--	--	--

⁸ <<http://wold.livingsources.org/>>

The links between the lemgrams are clearly providing under-specified information. The links are on the sense level where it is proper to talk about equivalence, not on the form level. However, since most words in any of the lexical resources are monosemous (Borin, 2010), most lemgram links are in fact also sense links. Because of this, we have at the moment settled for accepting some degree of under-specification in the Diapivot to allow the resource to grow quickly. Establishing proper word sense links is, of course, part of the ever-growing future work.

3. Search in Karp

Arguably the biggest motivation for building the editing system on top of the existing Karp database is to make use of the extensive and already existing search functionalities. There are four ways to search the Karp lexicons, as described in the following sections. The different ways of searching are available in Karp's search interface and through its web services.

3.1 Basic search

The basic search accepts a wordform, a sense identifier, or a lemgram. The lexical entries containing the requested information are returned.

In addition, the basic search supports full text search in the textual parts of the lexical resources, such as examples and definitions. The full text search, beyond extending the search capabilities, also makes the lexical information lacking wordforms, senses, and lemgrams discoverable.

3.2 Pivot search

The pivot search accepts a wordform that is looked up in all selected morphologies. If one or more lemgrams are found, the lexical entries containing the lemgrams or any of their associated senses are returned.

For example, a search for *katter* 'cats'

- ⇒ finds the lemgram *katt..nn.1* in the SALDO morphology and *katter..nn.1* in the Old Swedish morphology.
- ⇒ finds all senses of *katt..nn.1* and *katter..nn.1*
 - ⇒ searches for *katt..nn.1* and *katter..nn.1* and their senses in the current lexicon selection.

3.3 Diapivot search

As previously mentioned, there is a diachronic pivot resource that links the lemgram units of different morphologies – typically reflecting different historical stages of Swedish, hence the name 'Diachronic pivot' or 'Diapivot' – and thus acts as a middle-layer allowing the location of diachronic lexical information related to the

current search, e.g. a spelling variation or, moving backwards in time, a completely different form unit related to the current search.

Consider *springa* (verb) in the Diapivot resource that was exemplified in section 2. A diapivot search for *springa* (verb) would trigger a search for all lemgrams to which *springa* (verb) is linked in the Diapivot resource, i.e. Dalin *springa* (verb), Old Swedish *löpa* (verb), etc.

The diapivot search has been incorporated into the corpus search interface Korp (Borin et al., 2012b), so that, e.g. a search for *räv* (noun) ‘fox’ also finds words like *räf* (noun) ‘fox’ (a 19th century spelling variant).

3.4 Extended search

The extended search enables search in any of the data fields occurring in the resources of Korp. It uses CQL (Contextual Query Language)⁹ as the query language, which supports complex queries using logical operators, regular expressions, sorting, and more.

The extended search is represented graphically in the search interface. When a query is submitted the interface maps the graphical representation of the query onto a CQL expression that is sent to the Korp web service. For example, in figure 1, we search for the word forms *torsk* ‘cod’ or *långa* ‘ling’ with an exclusion of adjectives (in order to avoid the adjective form *långa* ‘long DEF/PL’).

Search for entries that match

The screenshot shows a search interface with the following elements:

- A search bar with a minus sign on the left and a plus sign on the right.
- Two search criteria:
 - Criterion 1: **wordform** (dropdown), **equals** (dropdown), **torsk** (text input).
 - Criterion 2: **wordform** (dropdown), **equals** (dropdown), **långa** (text input).
- A plus sign (+) between the two criteria, indicating an OR relationship.
- An **except** (dropdown) button.
- A search criterion: **part of speech** (dropdown), **equals** (dropdown), **adjective** (dropdown).
- A plus sign (+) below the 'except' criterion, indicating an exclusion.
- A **Sort by** dropdown set to **wordform** and a **ascending** dropdown.
- A **Search** button.

Figure 1: Extended search

⁹ <<http://www.loc.gov/standards/sru/specs/cql.html>>

4. Towards a generic lexicon editor

Even though all 23 lexicons currently in the Karp system are in LMF format, they contain, as shown in section 2, varying kinds of linguistic information. For example, some contain only morphological descriptions while others contain syntactic and semantic information of different kinds. In order to be useful, a general editing system has to provide synergies but still handle the particularities of each resource and not limit their expressiveness. The editing system should provide methodological support such as additional suggestions and consequence analysis, i.e. the effects one lexical judgement may have on related lexical information. For example, a new synonymy relation may trigger a suggestion in another resource or flag something as being in conflict with the new relation.

To elaborate further, access to the lexical information in all other resources while editing one resource provides rich background information for the lexicographer who is about to make a lexical decision. Moreover, formal inter/intra-resource dependencies can be verified on the fly, and new entries may be derived (semi-)automatically from other resources.

Statements of inter-resource dependencies also function as hypothesis testing: what intuitively seems true, and hence stated as a formal requirement, may instead illustrate important yet subtle differences in the resources. For example, it may seem intuitively evident that the frame hierarchy in the Swedish FrameNet should respect the hyponym relations in Swesaurus, such that if w_1 is a hyponym of w_2 , then w_1 should never occur higher in the frame hierarchy than w_2 . However, if this is a reasonable assumption or not is an empirical question.

Another important challenge is to allow lexical editing systems to take advantage of the massive amounts of linguistically-annotated text which are available in the corpora infrastructure Korp (Borin et al., 2012b) at Språkbanken, for example, when annotating examples or writing sense definitions. The information is, in principle, already available since the corpora have been annotated with lemmata and sense identifiers occurring in the resources of Karp, but it is still an open question how this information is best utilized and presented in the editor.

4.1 Editing the Swedish Constructicon

As an example of the current state of affairs of the lexicon editor, let us consider how editing of the Swedish Constructicon (Lyngfelt et al., 2012) is performed in Karp.

To start editing an existing lexicon entry in the Swedish Constructicon, the user has to log in and look up the particular entry using any of the provided search tools (see section 3). In the presentation view, the user clicks a button to open up a new editor tab with a slightly different presentation more suited for editing. Having different modes for normal presentation and for editing has the advantage that the editing mode can be generalized for all lexicons, while the presentation view may be tailored

for a specific audience or presentation style.

In the editor mode, the different fields of the lexicon are presented as a list. The exact presentation can be specified in a configuration file for each lexicon since the kinds of data can differ considerably. In figure 2, an entry of the Swedish Constructicon is being edited: the REFLEXIVE RESULTATIVE construction. It is a partially schematic construction expressed formally as VB REFL AP and semantically as ACTION ACTOR RESULT, with constructs such as *äta sig mätt* 'eat oneself full' and *skrika sig hes* 'shout oneself hoarse'. Moreover, the example section contains annotated authentic example sentences that illustrate the construction, e.g. *Drick dig smal i vår* 'Drink yourself thin this spring'. For more detailed information about this entry, please consult Lyngfelt et al. (2012).

All entries in this resource are quite similar in structure and have a small number of fields, making it sensible to also show the unfilled ones (toned down in gray). For other resources, however, the set of possible fields is much larger and there is more hierarchy in the entries. For such resources, fields can instead be added to the view upon request.

4.2 Technical details

The editing functionality of Karp is divided into three technical components: a backend with a REST-based web service API, a user authentication service, and a graphical frontend.

The backend recognizes a set of generic commands for adding, removing, updating lexical entries, and for manipulating the edit queue. A special **updates** command enables multiple actions at once, making it easier to log and backtrack changes by session.

Since the operations are generic, it is possible to use the Karp API for other applications as well. One such example is the language exercise platform Lärka 'lark' (Volodina et al., 2012) which uses the Karp editing API for logging user input.

The backend hosts two different database layers. The first one gets updated directly when a user edits a lexicon. If the user is fully authorized, the modified entries will eventually be copied to the main database layer, but only after they have been batch processed on a regular interval to ensure global consistency. If the user is not fully authorized, however, the changes will be put into the edit queue, waiting for a fully authorized user to accept or reject the changes.

Having multiple users working with the same lexicon may lead to the same problems as for any multi-user project. Changes may need to be undone while not altering other changes. Instead of reinventing the wheel, Karp makes use of an off-the-shelf version control system (VCS) inside the database. With each update the particular

lexicon change is checked into the VCS repository.

Although no two lexicons have exactly the same structure, they typically have certain traits in common that manifest themselves as similar frontend requirements, such as sharing the same settings or editing logic. For that purpose, Karp uses a class hierarchy for handling data structures in the frontend. The most basic class type is a string which is represented to the user as a simple text field. This can be extended to handle more complicated data structures, and modify the graphical user interface for editing the data. For example, the basic text widget can be subclassed to allow the user to select from a drop-down value list, that can be further subclassed to add consistency checks and other functionalities.

5. Conclusions and future work

We have briefly presented the ongoing work on adding editing functionality to the open lexical infrastructure at Språkbanken. It is still under active development, but is already a versatile tool for our work on the lexical resources.

The technical foundation is now in place, so our next step is to make all lexical resources of Karp editable. We will also explore the methodological details to ensure that the lexicographic work becomes as efficient as possible, and to secure the consistency and completeness of each resource by employing both internal and external lexical information.

6. Acknowledgements

The research presented here was supported by the Swedish Research Council (the projects Safeguarding the future of Språkbanken, VR dnr 2007-7430 and Swedish FrameNet++, VR dnr 2010-6013), by the University of Gothenburg through its support of the Centre for Language Technology and of Språkbanken (the Swedish Language Bank), and by the European Commission through its support of the META-NORD project under the ICT PSP Programme, grant agreement no 270899.

7. References

- Peter Andersson and Malin Ahlberg. 2013. Towards automatic tracking of lexical change: Linking historical lexical resources. *NEALT Proceedings Series*, 18.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 17th international conference on Computational linguistics*, pages 86–90, Morristown, NJ, USA.
- Lars Borin and Markus Forsberg. 2008. Something old, something new: A computational morphological description of Old Swedish. In *LREC 2008 Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, pages 9–16.
- Lars Borin and Markus Forsberg. 2009. All in the family: A comparison of SALDO and WordNet. In *Proceedings of the Nodalida 2009 Workshop on WordNets and*

- other Lexical Semantic Resources – between Lexical Semantics, Lexicography, Terminology and Formal Ontologies, Odense.
- 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, Valletta. ELRA.
- Lars Borin and Markus Forsberg. 2011a. A diachronic computational lexical resource for 800 years of Swedish. In Language technology for cultural heritage, pages 41–61. Springer, Berlin.
- Lars Borin and Markus Forsberg. 2011b. Swesaurus – ett svenskt ordnät med fria tyglar. *LexicoNordica*, 18:17–39.
- Lars Borin, Dana Danélls, Markus Forsberg, Dimitrios Kokkinakis, and Maria Toporowska Gronostaj. 2010a. The past meets the present in Swedish FrameNet++. In 14th EURALEX International Congress, pages 269–281, Leeuwarden. EURALEX.
- Lars Borin, Markus Forsberg, and Dimitrios Kokkinakis. 2010b. Diabase: Towards a diachronic BLARK in support of historical studies. In Proceedings of LREC 2010, pages 35–42, Valletta. ELRA.
- Lars Borin, Markus Forsberg, Leif-Jöran Olsson, and Jonatan Uppström. 2012a. The open lexical infrastructure of Språkbanken. In Proceedings of LREC 2012, pages 3598–3602, Istanbul. ELRA.
- Lars Borin, Markus Forsberg, and Johan Roxendal. 2012b. Korp – the corpus infrastructure of Språkbanken. In Proceedings of LREC 2012, Istanbul. ELRA.
- Lars Borin, Bernard Comrie, and Anju Saxena. 2013a. The Intercontinental Dictionary Series – a rich and principled database for language comparison. In Lars Borin and Anju Saxena (eds.), *Approaches to measuring linguistic differences*, pages 285–302. Mouton de Gruyter, Berlin.
- Lars Borin, Markus Forsberg, and Lennart Lönngren. 2013b. SALDO – a touch of yin to WordNet's yang. *Language Resources and Evaluation*.
- Lars Borin. 2010. Med Zipf mot framtiden – en integrerad lexikonresurs för svensk språkteknologi. *LexicoNordica*, 17:35–54.
- Jordan Boyd-Graber, Christiane Fellbaum, Daniel Osherson, and Robert Schapire. 2006. Adding dense, weighted connections to WordNet. In Proceedings of the Third International WordNet Conference.
- Christiane Fellbaum. 1998a. A semantic network of English verbs. In Christiane Fellbaum, editor, *WordNet: An electronic lexical database*, pages 69–104. MIT Press, Cambridge, Mass.
- Christiane Fellbaum, editor. 1998b. *WordNet: An electronic lexical database*. MIT Press, Cambridge, Mass.
- Gil Francopoulo, editor. 2013. *LMF: Lexical Markup Framework*. ISTE/Wiley, London/Hoboken, NJ.
- Martin Haspelmath and Uri Tadmor, editors. 2009. *Loanwords in the World's Languages: A Comparative Handbook*. Mouton de Gruyter, Berlin.
- ISO. 2008. *Language resource management – Lexical Markup Framework (LMF)*. International Standard ISO 24613:2008.

- Viggo Kann and Magnus Rosell. 2006. Free construction of a free Swedish dictionary of synonyms. In Proceedings of the 15th NODALIDA conference, Joensuu 2005, pages 105–110. Department of Linguistics, University of Joensuu.
- Benjamin Lyngfelt, Lars Borin, Markus Forsberg, Julia Prentice, Rudolf Rydstedt, Emma Sköldberg, and Sofia Tingsell. 2012. Adding a constructicon to the Swedish resource network of Språkbanken. In Proceedings of KONVENS 2012 (LexSem 2012 workshop), pages 452–461, Vienna.
- Alistair Miles and José R. Pérez-Agüera. 2007. SKOS: Simple knowledge organisation for the web. *Cataloging & Classification Quarterly*, 43(3):69–83.
- Josef Ruppenhofer, Michael Ellsworth, R. L. Miriam Petruck, R. Christopher Johnson, and Jan Scheffczyk. 2005. *FrameNet II: Extended Theory and Practice*. ICSI, Berkeley.
- C.J. Schlyter. 1887. *Ordbok till Samlingen af Sweriges Gamla Lagar*. (Saml. af Sweriges Gamla Lagar 13). Lund, Sweden.
- Elena Volodina, Lars Borin, Hrafn Loftsson, Birna Arnbjörnsdóttir, and Guðmundur Örn Leifsson. 2012. Waste not, want not: Towards a system architecture for ICALL based on NLP component re-use. In Proceedings of the SLTC 2012 Workshop on NLP for CALL, Lund, 25th October, 2012, number 80 in Linköping Electronic Conference Proceedings, pages 47–58.

✖  **ID** reflexiv_resultativ

Type ▾ ✖  cx, refl
+

✖  **Cat** VP

✖  **Evokes** Causation_scenario

✖  **Definition** [Någon]_{Actor} eller [något]_{Theme} utför eller undergår [en aktion]_{Activity} som leder (eller antas leda) till att [aktören]_{Actor} / [temat]_{Theme}, uttryckt med reflexiv, uppnår [ett tillstånd]_{Result}.

✖  **Structure** [V refl AP]

 Inheritance

Cee ▾ ✖  refl
+

Coll ▾ ✖  {äta¹ : mätt¹}

✖  {supa¹ : full²}

✖  {skrika¹ : hes¹}

✖  springa¹
+

Internal construction elements ▾ ✖  name=Activity, cat=vb

✖  cx=refl, name=Actor

✖  cx=refl, name=Theme

✖  name=Result, cat=AP
+

External construction elements ▾ ✖  name=Actor, cat=NP

✖  name=Theme, cat=NP
+

Example ▾ ✖  [Vi åskådare]_{Actor} [[springer]_{Activity} [oss]_{Actor} inte [varma]_{Result}]resultativ_reflexiv direkt.

✖  [Kornet och havren]_{Theme} får [[frysa]_{Activity} [sig]_{Theme} [mogen]_{Result}]resultativ_reflexiv ·

✖  [[Drick]_{Activity} [dig]_{Actor} [smal]_{Result}]resultativ_reflexiv i vår.
+

✖  **Comment** Det finns också en PP-variant med resultativ betydelse, t.ex. "träna sig i form", som ev. bör inkorporeras här - alt. betraktas som en metaforisk utvidgning av någon rörelse-cx.

 Internal comment (hidden)

✖  **Reference** Jansson, Håkan (2006): Har du ölat dig odödlig? En undersökning av resultativkonstruktioner i svenskan. (D-uppsats, även publicerad som MISS 57) <http://hdl.handle.net/2077/19000> Lyngfelt, Benjamin (2007): Mellan polerna. Reflexiv- och deponenskonstruktioner i svenskan. Språk och stil NF 17: 86–134. <http://hdl.handle.net/2077/21731>

 Status

Figure 2: Editing the SweCxn entry reflexiv_resultativ in Karp