
Språkbanken

Årsrapport 2014

ÖVERSIKT

I DENNA ÅRSRAPPORT redovisas merparten av de språkteknologiverksamheter som bedrivs vid institutionen för svenska språket. I vårt identitetsarbete för vi samman dessa verksamheter under rubriken ”Språkbanken” (1) för att markera att de aktiviteter som beskrivs här hör så nära ihop att de i praktiken utgör en forskningsmiljö, samt (2) därför att Språkbanken är en väl inarbetad benämning (som går tillbaka till 1970-talet) på dessa aktiviteter.

När vi talar om Språkbankens verksamhet menar vi således språkteknologisk forskningsverksamhet vid institutionen för svenska språket som finansieras av fakulteten under benämningarna *Språkbanken* och *språkvetenskaplig databehandling* (forsknings- och forskarutbildningsämne som deltar i grundutbildningen inom masterprogrammet i språkteknologi), samt av Göteborgs universitet genom en riktad satsning på styrkeområdet språkteknologi under rubriken *Centre for Language Technology* (CLT), samt ett antal externfinansierade forskningsprojekt inom språkteknologi.

KONFERENSER/PRESENTATIONER

Under denna rubrik har vi samlat en viktig del av vår utåtriktade verksamhet. Denna del består i att vi presenterar vårt arbete i form av föredrag eller posterpresentationer vid konferenser och workshoppar. Dessa har en öppen inbjudan att inkomma med bidrag (inklusive sådana som vi själva arrangerar eller är med om att arrangera). Inom området språkteknologi är sådana konferenser huvudpubliceringskanalen. Bidrag skickas först in i fulltext (typiskt 4–8 sidor) och bedöms i normalfallet dubbelt anonymt (anonymiserade bidrag bedöms av anonyma granskare) av två till tre fackgranskare. Sådana presentationer återfinns i nästa avsnitt.

En annan viktig del av den utåtriktade verksamheten består i presentationer efter inbjudan eller egen anmälan vid seminarier, projektmöten o.dyl. Huvudskillnaden mot föregående kategori är avsaknaden av explicit fackgranskning. Naturligtvis avspeglar mängden och den geografiska spridningen av den här sortens presentationer ändå forskargruppens aktivitet och rykte. Dessa redovisas i avsnittet *Andra presentationer* nedan.

2 *Språkbanken*

Konferens/workshop, plats/månad (antal presentationer)

- The Seventh Global WordNet Conference (GWC 2014), Tartu, Estland/januari (1)
- The 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC), Göteborg/april (1)
- The Fifth International Workshop on Health Text Mining and Information Analysis (Louhi), Göteborg/april (1)
- The 10th Workshop on Multiword Expressions (MWE 2014), Göteborg/april (1)
- The 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Göteborg/april (1)
- The Digital Access to textual Cultural Heritage (DATECH), Madrid, Spanien/maj (1)
- Svenska språkets historia, Umeå/maj (2)
- The Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Island/maj (8)
- The Conference on Communication, Medicine and Ethics (COMET), Lugano, Schweiz/juni (2)
- The Ninth Workshop on Innovative Use of NLP for Building Educational Applications, Baltimore, USA/juni (1)
- Digital Humanities, Lausanne, Schweiz/juli (1)
- The Workshop on Lexical and Grammatical Resources for Language Processing, Dublin, Irland/augusti (1)
- The 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Irland/augusti (1)
- The 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italien/oktober (2)
- Workshop on Natural Language Interface for Web of Data, Trentino, Italien/oktober (1)
- Svenskans beskrivning, Lund/oktober (1)
- FSCONS, Göteborg/november (1)
- The Fifth Swedish Language Technology Conference (SLTC), Uppsala/november (4)
- The 3rd NLP4CALL workshop, Uppsala/november (1)

Andra presentationer, plats/månad

- Centre for Language Technology (CLT) seminarium, Göteborg/under året (1 seminarium)
- 'GRUS' Frölundaskolan, Göteborg/januari
- 'Training parsers on incompatible treebanks', IMS, Stuttgart, Tyskland/januari

- ‘When the FrameNet suit does not fit the Swedish language’, Texas universitet, Austin, Texas, USA/februari
- ‘Using framenets for language teaching and learning - Don’t say you will walk to Hawaii’, Texas universitet, Austin, Texas, USA/februari
- Grammatikfestivalen, Göteborg/mars (1 presentation)
- Presentationer av Språkbankens arbete, Kulturomikmöte, Lund/mars
- ‘Swedish multiword expressions and sublanguage parsing’, COST PARSEME-möte, Aten, Grekland/mars
- Workshop om sammansättningar, Göteborg/april (1 presentation)
- Vetenskapsfestivalen, Göteborg/maj (4 presentationer)
- ‘Improving Parsing Medical Discourse using Cascades of Multiword Expression Recognition’, COST PARSEME-möte, Frankfurt, Tysland/september
- Presentationer av Språkbankens arbete, besök från DSL, Göteborg/september
- Workshop on Computational Processing of Historical Texts, Helsingfors, Finland/september (4 presentation)
- ‘KORP an Open Corpus Infrastructure Environment for . . . Export’, Athens universitet, Aten, Grekland/oktober
- Centre for Language Technology (CLT) årlig workshop, Uddevalla/oktober (15 presentationer/postrar)
- Swedish FrameNet++ Workshop, Göteborg/november (8 presentationer)
- Presentationer av Språkbankens arbete, Kulturomikmöte, Göteborg/november
- Idéseminarium (doktorand), Göteborg/november
- Workshop om semantisk annotering, Köpenhamn, Danmark/november (2 presentationer)

Organiserad konferens/workshop/seminarium, plats, månad (totalt antal deltagare)

- eXist-db pre-conference day (XML-Prague), Prag, februari, (ca 40)
- Introduction to Text Analysis with Applications in R (Matthew Jockers), Göteborg, mars (ca 20)
- Introduction to Topic Modeling with Applications in R (Matthew Jockers), Göteborg, mars (ca 20)
- LTLOD@SB workshop, Göteborg, april (ca 10)
- 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Göteborg, april (ca 500)
- International workshop on linguistic microareas in South Asia, Uppsala, maj (ca 35)
- Korpworkshop, Göteborg, oktober (ca 30)
- Höstworkshop, Göteborg, oktober (ca 60)
- The first Swedish national SWE-CLARIN workshop, Uppsala, november (ca 40)

4 *Språkbanken*

- 3rd workshop on NLP for computer-assisted language learning (NLP4CALL), Uppsala november (ca 30)
- Swedish Framenet++ workshop, Göteborg, november (ca 20)
- Workshop on semantic technologies for research in the humanities and social sciences (STRiX), Göteborg, november (ca 40)

PROJEKT

Språkbankens arbete utförs inom en mängd olika projekt. Några av dessa projekt är interna, medan flertalet innebär samarbeten på nationell eller internationell nivå. Här redovisar vi de forskningsprojekt som vi deltog i under året.

- AO – en akademisk ordlista för svenska (år 4); finansiär ISA
- Automatic identification of language change; egen finansiering
- CLT:s textteknologilaboratorium (år 6 av 7); finansiär CLT (Göteborgs universitet)
- Distributional Methods to Represent the Meaning of Frames and Constructions (år 1 av 5); finansiär VR
- Diabase – Towards a diachronic BLARK; egen finansiering
- Digital areallingvistik (år 5); finansiär VR
- En fri molntjänst för OCR (år 1 av 1); finansiär Kungliga biblioteket, Göteborgs universitet
- GRUS – Gymnasie- och grundskoleelevers skrivande i ”en dator till en elev”-satsning (år 4); finansiär ISA
- Koala – Korps lingvistiska annotationer (år 1 av 3), finansiär RJ
- Kulturomik: Mot kunskapsbaserad storskalig kunskapsutvinning ur svensk text (år 3 av 5); finansiär VR
- Litteraturbanken (permanent); finansiär Svenska Akademien, Kungl. Vetternhetsakademien
- LTLOD@SB – Language Technology Linked Open Data at Språkbanken (år 1 av 1); finansiär VINNOVA
- MApIR – Metoder för automatisk Analys av Text i digitala Historiska Resurser (år 1 av 3), finansiär MAW
- SweCxn – ett svenskt konstruktikon (år 2 av 3); finansiär RJ
- Swe-Clarin (år 1 av 5); finansiär VR
- SweFN++ (år 4); finansiär CLT (Göteborgs universitet), VR

Flera projektansökningar har inlämnats för pågående och kommande projekt.

- Att återge tidsföljd i språket: konstruktioner med temporala konjunktioner i kontrastiv belysning. Ansökt till Riksbankens Jubileumsfond (avslagen)
- Automatic Linguistic Analysis for the Nordic languages In Social media.

- Ansökt till NOS-HS (avslagen)
- Concessive constructions in change. Grammaticalization meets construction grammar. Ansökt till Riksbankens Jubileumsfond (avslag)
 - Constructions in change. Tracing concessives through the history of Swedish. Ansökt till Vetenskapsrådet (avslagen)
 - Constructions in change. Tracing concessives through the history of Swedish. Ansökt till Marcus och Amalia Wallenberg (avslagen)
 - Diameme. Ansökt till Riksbankens Jubileumsfond (avslagen)
 - DIRECT – Digital Humanities Research Platform for Corpus Analysis and Text Mining. Ansökt till EU Horizon 2020 (avslagen)
 - Dramawebben – infrastruktur för teoretisk och kostnärilig forskning. Ansökt till Riksbankens Jubileumsfond (avslagen)
 - Emotions in Swedish microblogs – Creation, annotation and analysis of a Swedish Tweeter corpus with focus on emotion-related words, phrases and emoticons. Ansökt till Riksbankens Jubileumsfond (avslagen)
 - Incorporating Swedish in the InterSystems iKnow Technology. Ansökt till Intersystems Corporation (beviljad)
 - Känslor i svenska mikrobloggar. Ansökt till Torsten Söderbergs Stiftelse (avslagen)
 - Känslor i svenska mikrobloggar. Ansökt till Internetfonden (avslagen)
 - Linguistic and extra-linguistic parameters for early detection of cognitive impairment. Ansökt till Riksbankens Jubileumsfond (under bedömning)
 - Linguistic Encoding of Time Order: Constructions with Temporal Conjunctions in Contrast. Ansökt till Vetenskapsrådet (avslagen)
 - MULaN – MUlti-LinguAl FrameNet Coordination. Ansökt till Vetenskapsrådet International Career Grant (avslagen)
 - OCR correction of older Swedish texts – A case study. Ansökt till Åke Wibergs Stiftelse (avslagen)
 - Profiling Swedish as a Second Language: corpus-driven text studies based on Common European Framework of Reference. Ansökt till Vetenskapsrådet (avslagen)
 - SPARV Swedish pattern resource for verbs – a large database of integrated semantic and syntactic knowledge linked to corpora for Swedish language technology, digital humanities and language studies. Ansökt till Riksbankens Jubileumsfond (avslagen)
 - SPARV - A Swedish Pattern Resource for Verbs - Integrating semantic and syntactic knowledge linked to corpora for Swedish language technology. Ansökt till Vetenskapsrådet (avslagen)
 - SweLL - Swedish Learner Language Analysis and Annotation. Developing an infrastructure for text-based research on Swedish as a second

6 *Språkbanken*

- language. Ansökt till Riksbankens Jubileumsfond (avslagen)
- SweLL - Swedish Learner Language: Developing methods for semi-automatic analysis of Swedish as a second language. Ansökt till Marcus och Amalia Wallenberg (avslagen)
- Sydasien som språkförbund? Explorativa dataintensiva metoder för areallingvistik och genetisk språkjämförelse. Ansökt till Vetenskapsrådet (beviljad)
- Towards large scale, semantically and domain-driven event extraction for Swedish using frame semantic verb descriptions and a BioLexicon-like repository of biomedical terms. Ansökt till Vetenskapsrådet (avslagen)

INFRASTRUKTUR

Språkbanken arbetar aktivt för att utveckla en språkteknologisk infrastruktur. Detta arbete omfattar i dagsläget följande komponenter:

- lexikal infrastruktur,
- korpusinfrastruktur, samt
- metadata.

Uppbyggnaden av infrastruktur för lexikala resurser och korpora handlar för närvarande om att harmonisera och standardisera så många fria språkliga resurser som möjligt, samt att göra dem tillgängliga för forskningsvärlden. Vi skapar dessutom verktyg för att utforska dessa resurser, t ex Karp (tidigare SBLEX) för lexikala resurser och Korp för korpora. Målet är vidare att alla dessa resurser ska vara väl beskrivna i ett metadata-repositorium som följer de standarder som finns inom området.

Öppenhet är ett av våra ledord, en filosofisk ståndpunkt vi försöker att tillämpa i största möjliga mån. Vi anser att forskning ska utföras öppet, för att tillåta granskning och samarbete. I denna öppenhet ingår att använda öppna standarder och licenser samt att använda och skapa verktyg med öppen källkod. Även om alla äldre resurser inte kan göras fritt tillgängliga, p.g.a. restriktiva licenser, strävar vi efter att samla fler och större fria resurser, för att främja språklig forskning och utveckling av språkteknologiska tillämpningar, i Sverige och världen.

Resurserna används flitigt, och t ex Korp har fler än 1000 användare varje månad. Majoriteten av dessa finns av naturliga skäl i Sverige, men många sökningar görs även från andra delar av Europa och resten av världen.

SAMVERKAN

En central del i samverkan med det omgivande samhället består i att vi är representerade i relevanta externa organisationer. Här redovisas denna medverkan.

- CLT:s styrgrupp – Centre for Language Technology
- Dramawebbens referensgrupp
- Litteraturbankens styrelse
- MLT:s styrgrupp – Masterprogrammet i språkteknologi
- SIS TK115 – Swedish Standards Institute, Terminologi och språkliga resurser
- SND:s styrelse – Svensk Nationell Datatjänst
- Språkrådets rådgivargrupp för språkteknologi

Utöver medverkan i dessa organisationer samarbetar vi, genom Språkbanken som helhet eller genom projekt, med följande organisationer, institutioner, och företag.

Nationella samarbeten

- DART kommunikations- och dataresurscenter
- Digisam
- Dramawebben
- Gothia Forum för klinisk forskning
- GPCC – Centrum för personcentrerad vård, Göteborgs universitet
- Institutet för svenska som andraspråk, Göteborgs universitet
- Kungliga biblioteket
- Kungliga tekniska högskolan
- Lexikaliska institutet, Göteborgs universitet
- Linköpings universitet
- Lunds universitet
- Riksarkivet mediakonverteringscentrum
- Sahlgrenska universitetssjukhuset
- SND – Svensk Nationell Datatjänst
- Socialstyrelsen
- Språkrådet, Institutet för folk- och språkminnen
- Stockholms universitet
- Uppsala universitet

Internationella samarbeten

- Berkeley FrameNet
- Centrum för internationalisering och parallellspråklighet, Köpenhamns universitet

8 *Språkbanken*

- CLARIN – Common language resources and technology infrastructure
- Kelly – Keywords for language learning for young and adults alike
- Max Planck-institutet för evolutionär antropologi
- META-NORD – The multilingual Europe technology alliance, specifically the Baltic and Nordic parts
- MOLTO – Multilingual online translation
- Tekstlaboratoriet, UiO – Universitetet i Oslo

Priser

Våren 2014 fick Markus Forsberg och Lars Borin ta emot Erik Wellanders pris för framstående forskning inom språkvårdens område för sitt arbete med det digitala språkverktyget Korp.

BESÖK

Ett flertal gästforskare och seminariegäster har besökt oss under året.

- Thierry Declerck, DFKI
- Patrick Hanks, Västra Englands universitet/Wolverhamptons universitet
- Måns Huldén, Helsingfors universitet
- Matthew Jockers, Nebraskas universitet
- Peter Leonard, Yales universitetsbibliotek
- Krister Lindén, Helsingfors universitet
- Johann-Mattis List, Marburgs universitet
- Joakim Nivre, Uppsala universitet
- John McCrae, Bielefelds universitet
- Detmar Meurers, Tübingens universitet
- Miriam R L Petruck, International Computer Science Institute (ICSI)
- Erik Tjong Kim Sang, Meertens-institutet
- Nathan Schneider, Edinburghs universitet
- Caroline Sporleder, Triers universitet
- Joel Tetreault, Yahoo!
- Staffan Truvé, Recorded future
- Sowmya Vajjala, Tübingens universitet

BEDÖMNINGSUPPDRAG

Utöver deltagande med egna presentationer vid konferenser och workshoppar deltar vi även som fackgranskare av andra bidrag. Ofta tilldelas man flera bidrag att bedöma. Här redovisas sådana bedömningsuppdrag.

- Tidskriftsartiklar
 - Journal of Logic and Computation
 - Computational Linguistics
 - Journal of Artificial Intelligence Research
 - CALICO journal
 - BMC Medical Informatics and Decision Making
 - Special issue of the Artificial Intelligence in Medicine
 - The Semantic Web Journal (SWJ)
- Konferensbidrag
 - Conference on Building and Exploring Web Based Environments (WEB)
 - Conference of the Association for Computational Linguistics conference (ACL)
 - Conference of the European Chapter of the Association for Computational Linguistics (EACL)
 - Conference on Computational Linguistics (Coling)
 - Conference on Computational Natural Language Learning (CoNLL)
 - Digital Humanities (DH)
 - European Conference on Information Retrieval (ECIR)
 - Italian Conference on Computational Linguistics (CLIC)
 - Language Resources and Evaluation Conference (LREC)
 - Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)
 - Semantic technologies for research in the humanities and social sciences (STRiX)
 - Swedish Language Technology Conference (SLTC)
 - Swedish national SWE-CLARIN workshop
 - Workshop on Building and Evaluating Resources for Health and Biomedical Applications
 - Workshop on Health Text Mining and Information Analysis (Loughi)
 - Workshop on Histoinformatics
 - Workshop on Innovative Use of NLP for Building Educational Applications (BEA)
 - Workshop on Natural Language Processing for computer-assisted language learning (NLP4CALL)
 - Workshop on Natural Language Processing Techniques for Educational Applications (NLP-TEA)

10 *Språkbanken*

- Övrigt
 - Betygskommitté, licentiatseminarium, Göteborgs universitet
 - Extern examiner, disputation, Manchesters universitet, Storbritanien
 - Opponent, slutseminarium, Uppsala universitet

NY PERSONAL

Under året har Språkbanken anställt en ny doktorand i Språkvetenskaplig databehandling, Luis Nieto Piña, två systemutvecklare/forskningsingenjörer, Anne Schumacher och Roland Schäfer, samt en forskare/projektsamordnare, Nina Tahmasebi.

UTBILDNING

Personal från Språkbanken har deltagit i undervisning om olika aspekter av språkteknologi, ibland gällande enstaka föreläsningar, ibland med kursansvar, vilket innefattar planering och examination.

- Kurser:
 - Medicinsk terminologi ur ett språkteknologiskt perspektiv (grundnivå) vt2014
 - Skriva (grundnivå) vt2014
 - Svenskans struktur i ett andraspråksperspektiv (grundnivå) vt2014
 - Språkteknologi för språkvetare (grund-/masternivå) vt2014 (kursansvar)
 - Statistiska metoder inom språkteknologi (masternivå) vt2014 (kursansvar)
 - Maskininlärning inom språkteknologi (forskarnivå) vt2014 (kursansvar)
 - Problemlösning (grundnivå) ht2014
 - Skriva (grundnivå) ht2014
 - Textanalys (grundnivå) ht2014
 - Språkteknologi för språkvetare (grund-/masternivå) ht2014 (kursansvar)
 - Information retrieval (masternivå) ht2014
 - Introduction to formal linguistics (masternivå) ht2014
 - Introduktion till programmering (masternivå) ht2014 (kursansvar)
 - Maskininlärning inom språkteknologi (masternivå) ht2014 (kursansvar)
 - Natural language processing (masternivå) ht2014
 - Språkvetenskapliga teorier (masternivå) ht2014

- Litteraturteoretisk och metodisk fördjupning: Digital humaniora (forskarnivå) ht2014
- Doktorander:
 - Malin Ahlberg, Computational historical linguistics, började 2013, handledare Lars Borin, Markus Forsberg och Henrik Rosenkvist
 - Ann-Marie Eklund, Analysis of queries to medical web sites and databases as a basis for knowledge sharing, började 2010, handledare Dimitrios Kokkinakis, Lars Borin och Jussi Karlgren
 - Luis Nieto Piña, Distributional methods to represent the meaning of frames and constructions, började 2014, handledare Richard Johansson
 - Ildikó Pilán, Swedish second language readability from a semantic perspective, började 2013, handledare Lars Borin och Elena Volodina
 - Taraka Rama Kasichyanula, Computational historical linguistics, började 2010, handledare Lars Borin och Søren Wichmann
 - Judy Ribbeck, Syntactic analysis of textbooks used in Sweden when teaching mathematics and natural science, började 2010, handledare Emma Sköldberg och Lars Borin
- Disputationer:
 - Taraka Rama, Vocabulary lists in computational historical linguistics, licentiatseminarium januari 2014
 - Ann-Marie Eklund, The Game of Health Search, disputation november 2014

PUBLIKATIONER

1. Yvonne Adesam, Malin Ahlberg, Peter Andersson, Gerlof Bouma, Markus Forsberg och Mans Hulden. Computer-aided morphology expansion for old swedish. I: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ss 1102–1105, Reykjavik, Island, maj 2014.
2. Yvonne Adesam, Lars Borin, Gerlof Bouma, Markus Forsberg och Richard Johansson. Koala – korp's linguistic annotations developing an infrastructure for text-based research with high-quality annotations. I: *Proceedings of the Fifth Swedish Language Technology Conference (SLTC)*, Uppsala, november 2014.
3. Malin Ahlberg, Lars Borin, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj, Karin Friberg Heppin, Richard Johansson, Dimitri-

12 Språkbanken

- os Kokkinakis, Leif-Jöran Olsson och Jonatan Uppström. Swedish FrameNet++ – the beginning of the end and the end of the beginning. I: *Proceedings of the Fifth Swedish Language Technology Conference (SLTC)*, Uppsala, november 2014.
4. Malin Ahlberg, Markus Forsberg och Mans Hulden. Semi-supervised learning of morphological paradigms and lexicons. I: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, ss 569–578, Göteborg, april 2014.
 5. Lars Borin, Jens Allwood och Gerard de Melo. Bring vs. mtroget: Evaluating automatic thesaurus translation. I: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ss 2115–2121, Reykjavik, Island, maj 2014.
 6. Lars Borin, Dana Dannélls, Markus Forsberg och John P. McCrae. Representing swedish lexical resources in rdf with lemon. I: *Proceedings of the ISWC 2014 Posters & Demonstrations Track within the 13th International Semantic Web Conference (ISWC 2014)*, ss 329–332, Riva del Garda, Italien, oktober 2014.
 7. Lars Borin, Dana Dannélls och Leif-Jöran Olsson. Geographic visualization of place names in swedish literary texts. *Literary & Linguistic Computing*, 29(3):400–404, 2014.
 8. Lars Borin, Gerard de Melo, Karin Friberg Heppin och Tiago Timponi Torrent. Introduction: Constructions and frames meet language technology. *Constructions and Frames*, 6(1):1–8, 2014.
 9. Lars Borin och Markus Forsberg. Swesaurus; or, the frankenstein approach to wordnet construction. I: *Proceedings of the Seventh Global WordNet Conference (GWC 2014)*, Tartu, Estland, januari 2014.
 10. Lars Borin och Richard Johansson. Kulturomik: Att spana efter språkliga och kulturella förändringar i digitala textarkiv. *Historia i en digital värld*, 2014.
 11. Lars Borin, Anju Saxena, Taraka Rama och Bernard Comrie. Linguistic landscaping of south asia using digital language resources: Genetic vs. areal linguistics. I: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ss 3137–3144, Reykjavik, Island, maj 2014.
 12. Dana Dannélls, Ramona Enache och Mariana Damova. A multilingual sparql-based retrieval interface for cultural heritage objects. I: *Proceedings of the ISWC 2014 Posters & Demonstrations Track within the 13th International Semantic Web Conference (ISWC 2014)*, ss 205–208, Riva del Garda, Italien, oktober 2014.

13. Dana Dannélls, Karin Friberg Heppin och Anna Ehrlemark. Using language technology resources and tools to construct swedish framenet. I: *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing*, ss 8–17, Dublin, Irland, augusti 2014.
14. Dana Dannélls och Normunds Gruzitis. Controlled natural language generation from a multilingual framenet-based grammar. I: *Lecture Notes in Computer Science*, band 8625, ss 155–166, 2014.
15. Dana Dannélls och Normunds Gruzitis. Extracting a bilingual semantic grammar from framenet-annotated corpora. I: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ss 2466–2473, Reykjavik, Island, maj 2014.
16. David Dupplaw, Michael Matthews, Richard Johansson, Giulia Boato, Andrea Costanzo, Marco Fontani, Enrico Minack, Elena Demidova, Roi Blanco, Thomas Griffiths, Paul Lewis, Jonathon Hare och Alessandro Moschitti. Information extraction from multimedia web documents: an open-source platform and testbed. *International Journal of Multimedia Information Retrieval*, 2014.
17. Markus Forsberg, Richard Johansson, Linnéa Bäckström, Lars Borin, Benjamin Lyngfelt, Joel Olofsson och Julia Prentice. From construction candidates to construction entries: An experiment using semi-automatic methods for identifying constructions in corpora. *Constructions and Frames*, 6(1):114–135, 2014.
18. Karin Friberg Heppin och Miriam R.L. Petruck. Encoding of compounds in swedish framenet. I: *Proceedings of the 10th Workshop on Multiword Expressions (MWE 2014) at EACL*, ss 67–71, Göteborg, april 2014.
19. Karin Friberg Heppin och Maria Toporowska Gronostaj. Exploiting framenet for swedish: Mismatch? *Constructions and Frames*, 6(1):52–72, 2014.
20. Inga-Lill Grahn och Dimitrios Kokkinakis. Legitimizing the visit a recurrent challenge among patients with medically unexplained symptoms (mus). I: *Proceedings of the Conference on Communication, Medicine and Ethics (COMET)*, Lugano, Schweiz, juni 2014.
21. Tobias Günther, Jean Vancoppenolle och Richard Johansson. Rtrgo: Enhancing the gu-mlt-It system for sentiment analysis of short messages. I: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, ss 497–502, Dublin, Irland, augusti 2014.
22. Helge Holzmann, Nina Tahmasebi och Thomas Risse. Named entity evolution recognition on the blogosphere. *International Journal on Digital Libraries*, ss 1–27, 2014.

14 *Språkbanken*

23. Richard Johansson. Automatic expansion of the swedish framenet lexicon. *Constructions and Frames*, 6(1):92–113, 2014.
24. Dimitrios Kokkinakis och Inga-Lill Grahn. A corpus-based approach to the identification of non-literal language in a medical setting. I: *Proceedings of the Conference on Communication, Medicine and Ethics (COMET)*, Lugano, Schweiz, juni 2014.
25. Dimitrios Kokkinakis, Ann Ighe och Mats Malm. Vocation identification in swedish fiction. I: *Proceedings of the Fifth Swedish Language Technology Conference (SLTC)*, Uppsala, november 2014.
26. Dimitrios Kokkinakis, Mats Malm, Jenny Bergenmar och Ann Ighe. Semantics in storytelling in swedish fiction. I: *Proceedings of the Digital Access to textual Cultural Heritage (DATeCH)*, Madrid, Spanien, maj 2014.
27. Dimitrios Kokkinakis, Jyrki Niemi, sam hardwick, Krister Lindén och Lars Borin. Hfst-swener – a new ner resource for swedish. I: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, s 6, Reykjavik, Island, maj 2014.
28. Mikael Kågebäck, Olof Mogren, Nina Tahmasebi och Devdatt Dubhashi. Extractive summarization using continuous vector space models. I: *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC) at EACL*, ss 31–39, Göteborg, april 2014.
29. Benjamin Lyngfelt, Lars Borin, Linnéa Bäckström, Markus Forsberg, Leif-Jöran Olsson, Julia Prentice, Rudolf Rydstedt, Emma Sköldberg, Sofia Tingsell och Jonatan Uppström. Ett svenskt konstruktikon – grammatik möter lexikon. I: *Svenskans beskrivning : Förhandlingar vid Trettiofjärde sammankomsten för svenskans beskrivning. Helsingfors den 15–17 maj 2013*, band 33, ss 268–279, 2014.
30. Dimitrios Kokkinakis Mats Malm, Jenny Bergenmar och Peter Leonard. The inherited self: Reappraising literary cultural heritage through digital methods. I: *Digital Humanities*, Lausanne, Switzerland, juli 2014.
31. Farnaz Moradi, Ann-Marie Eklund, Dimitrios Kokkinakis, Tomas Olovsson och Philippas Tsigas. A graph-based analysis of medical queries of a swedish health care portal. I: *Proceedings of the Fifth International Workshop on Health Text Mining and Information Analysis (Louhi) at EACL*, ss 2–10, Göteborg, april 2014.
32. Ildikó Pilán och Elena Volodina. Reusing swedish framenet for training semantic roles. I: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ss 1359–1363, Reykjavik, Island, maj 2014.

33. Ildikó Pilán, Elena Volodina och Richard Johansson. Rule-based and machine learning approaches for second language sentence-level readability. I: *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications at EACL*, ss 174–184, Baltimore, USA, juni 2014.
34. Taraka Rama. *Vocabulary lists in computational historical linguistics*. Licentiatuppsats, Göteborgs universitet, Göteborg, 2014.
35. Taraka Rama och Lars Borin. N-gram approaches to the historical dynamics of basic vocabulary. *Journal of Quantitative Linguistics*, 21(1):50–64, 2014.
36. Georg Rehm, Hans Uszkoreit, Sophia Ananiadou, Núria Bel, Audrone Bieleveciene, Lars Borin, António Branco, Gerhard Budin, Nicoletta Calzolari, Walter Daelemans, Radovan Garabík, Marko Grobelnik, Carmen Garcia-Mateo, Josef Van Genabith, Jan Hajic, Inma Hernaez, John Judge, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Joseph Mariani, John Mcnaught, Maite Melero, Monica Monachini, Asuncion Moreno, Jan Odijk, Maciej Ogrodniczuk, Piotr Pezik, Stelios Piperidis, Adam Przepiórkowski, Eiríkur Rögnvaldsson, Michael Rosner, Bolette Sandford Pedersen, Inguna Skadina, Koenraad De Smedt, Marko Tadić, Paul Thompson, Dan Tufiş, Tamás Váradi, Andrejs Vasiljevs, Kadri Vider och Jolanta Zabarskaite. The strategic impact of meta-net on the regional, national and international level. I: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ss 1517–1524, Reykjavik, Island, maj 2014.
37. Judy Carola Ribeck och Lars Borin. Lexical bundles in swedish secondary school textbooks. I: *Human Language Technology Challenges for Computer Science and Linguistics*, nummer XVI i Lecture Notes in Computer Science, ss 238–249. Springer International Publishing, 2014.
38. Judy Carola Ribeck, Håkan Jansson och Emma Sköldberg. Från aspekt till övergripande - en ordlista över svensk akademisk vokabulär. I: R. Vatvedt Fjeld och M. Hovdenak, redaktörer, *Rapport fra Konferense om leksikografi i Norden Oslo 13. - 16. august 2013*, nummer 12 i Nordiske studier i leksikografi, ss 370–384, 2014.
39. Elena Volodina, Lars Borin och Ildikó Pilán, redaktörer. *Proceedings of the third workshop on NLP for computer-assisted language learning at SLTC 2014*. Linköping University Press, Linköping, 2014.
40. Elena Volodina och Therese Lindström Tiedemann. Evaluating students' metalinguistic knowledge with lärka, november 2014.
41. Elena Volodina, Ildikó Pilán, Lars Borin och Therese Lindström Tiede-

mann. A flexible language learning platform based on language resources and web services. I: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ss 3973–3978, Reykjavik, Island, maj 2014.

42. Elena Volodina, Ildikó Pilán, Stian Rødven Eide och Hannes Heidars-son. You get what you annotate: a pedagogically annotated corpus of coursebooks for swedish as a second language. I: *Proceedings of the third workshop on NLP for computer-assisted language learning*, band NEALT Proceedings av 22, s 128–144, 2014.