
Språkbanken

Årsrapport 2012

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 6th International Global Wordnet Conference (GWC), Matsue, Japan/januari (1)
- The 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH) at EACL, Avignon, Frankrike/april (2)
- The World Wide Web Conference (WWW), Lyon, Frankrike/april (1)
- The 5th Geneva Conference on Person-centered Medicine, Geneve, Schweiz/april (1)
- The 8th International Conference on Language Resources and Evaluation (LREC), Istanbul, Turkiet/maj (7)
- The 3rd Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM) at LREC, Istanbul, Turkiet/maj (1)
- The 11th International Symposium on Functional and Logic Programming (FLOPS), Kobe, Japan/maj (1)
- 10:e nationella konferensen i cancervård, Stockholm/maj (1)
- The 7th International Natural Language Generation Conference (INLG), Starved Rock State Park, Utica IL, USA/maj (1)
- Joint Conference on Lexical and Computational Semantics (*SEM), Montreal, Kanada/juni (1)
- Workshop on Inducing Linguistic Structure Submission (WILS) at NAACL-HLT, Montreal, Kanada/juni (1)
- Lexical Functional Grammar Conference (LFG), Bali, Indonesien/juni (1)
- The 50th Annual Meeting of the Association for Computational Linguistics (ACL), Jeju, Korea/juli (1)
- The 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), Seoul, Korea/juli (1)
- The 6th Linguistic Annotation Workshop, Jeju, Korea/juli (1)
- Digital Humanities, Hamburg, Tyskland/juli (1)
- The 15th EURALEX International Congress, Oslo, Norge/augusti (4)
- Computational approaches to the study of dialectal and typological variation, ESSLLI workshop, Opole, Polen/augusti (1)
- The 24th Conference of the European Federation for Medical Informatics (MIE), Pisa, Italien/augusti (1)
- Joint Workshop on Intelligent Methods for Software System Engineering (JIMSE), Montpellier, Frankrike/augusti (1)
- EuroCALL, Göteborg/augusti (2)
- Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis (CLEFeHealth), Rom, Italien/september

- (1)
- The 15th International Conference on Text, Speech and Dialogue (TSD), Brno, Tjeckien/september (2)
 - Advances in Visual Methods for Linguistics (AVML), York, Storbritannien/september (1)
 - Workshop on Language Technology for Historical Text(s) at Konvens, Wien, Österrike/september (1)
 - The 11th Conference on Natural Language Processing (Konvens), Wien, Österrike/september (1)
 - The 5th International Symposium on Semantic Mining in Biomedicine (SMBM), Zürich, Schweiz/september (1)
 - The 5th international workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR), Maui, Hawaii/oktober (1)
 - Scandinavian Conference on Health Informatics (SHI), Linköping/oktober (1)
 - The 4th Swedish Language Technology Conference (SLTC), Lund/oktober (7)
 - Workshop on exploratory query-log analysis at SLTC, Lund/oktober (1)
 - Workshop on NLP for computer assisted language learning at SLTC, Lund/oktober (3)
 - The 20th International Conference on Computers in Education (ICCE), Singapore, Malaysia/november (1)
 - The 24th International Conference on Computational Linguistics (COLING), Mumbai, Indien/december (2)

Andra presentationer, plats/månad

- Centre for Language Technology (CLT) seminarium, Göteborg/under året (ca 8 seminarier)
- Centre for Language Technology (CLT) årlig workshop, Gullmarstrand/oktober (ca 15 bidrag)
- Linguistically Driven Methods for the Extraction of Fine-grained Opinion Structure, Uppsala/februari
- META-NORD: Main activities carried out since October 2011 & Deliverable 3.2 - Second batch of resources, including part of resources selected in D2.4, vid META-NORD consortium meeting, Vilnius/mars
- META-NORD: Task 3.1 - Upgrading resources to agreed standards, vid META-NORD consortium meeting, Vilnius/mars
- META-NORD: Task 4.2 - Harmonization of resource descriptions and related directory services, vid META-NORD consortium meeting, Vilnius/mars
- Medication Extraction from Clinical Data Using Frame Semantics, vid

4 *Språkbanken*

- Clinical NLP workshop, Stockholm/juni
- META-NORD: Information from UGOT, vid META-NORD consortium meeting, Reykjavík/augusti
- Svensk prosafiktion 1800-1900: en databas för tvärvetenskaplig metodutveckling, vid Digital humaniora - metodutveckling och möjligheter, Göteborg/oktober
- Om utvecklingen av Lärka, vid Seminarium om IT-undervisning i Nordens språk, Lund/oktober
- Something old, something new: Informativity and argument fronting in Spoken Dutch, Nijmegen, Nederländerna/november
- META-NORD, META-NET, META-SHARE: Vad har alla dessa META grupper att bjuda på? vid META-NORD workshop, Stockholm/november
- Svenska i den digitala tidsåldern - Hur mår svenskan i en flerspråkig värld? vid META-NORD workshop, Stockholm/november
- Grammatikseminarium: Two for the Price of One: An LFG treatment of sentence initial object es in German, Göteborg/september

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

- IR-seminarieserie, Göteborg/mar-maj (ca 10)
- Språkbankens årliga höstworkshop, Göteborg/oktober (ca 60)
- Workshop om NLP för CALL vid SLTC, Lund/oktober (ca 20)
- Workshop om SweFN++ vid SLTC, Lund/oktober (ca 10)
- META-NORD nationell workshop, Stockholm/november (ca 50)
- Workshop för Work Package 2 i Promise, Göteborg/december (8)

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 2); finansiär ISA
- Automatic identification of language change; egen finansiering
- A System Architecture for Intelligent Computer-Assisted Language Learning (år 2 av 3); finansiär NordPlus Sprog
- CLT:s textteknologilaboratorium (år 4 av 4); finansiär CLT (Göteborgs universitet)
- Diabase – Towards a diachronic BLARK; egen finansiering
- Digital areallingvistik (år 3 av 3); finansiär VR
- Förbättring av testkollektionen MedEval; finansiär CLT (Göteborgs universitet)

- GRUS – grundskoleelevers skrivande i ”en dator till en elev”-satsning (år 3); finansierad av LUN/ISA
- Improving semi-automatic selection of best corpus examples for Swedish; egen finansiering
- Litteraturbanken (permanent); finansierad av Svenska Akademien, Kungl. Vitterhetsakademien
- META-NORD (år 2 av 2); finansierad av EU
- MOLTO (år 3 av 3); finansierad av EU
- PROMISE (år 2 av 3); finansierad av EU
- SweFN++ (år 3); finansierad av CLT (Göteborgs universitet), VR
- Validating the Swedish Kelly-list following the CEFR agenda for vocabulary competence; egen finansiering

Flera projektansökningar har inlämnats för pågående och kommande projekt.

- Automatic identification of language change. Ansökt till Riksbankens Jubileumsfond (avslagen)
- Automatic identification of language change. Ansökt till Vetenskapsrådet (avslagen)
- Construction Linguistics in Natural Language Processing. Ansökt till ERC Starting Grant (avvaktar besked)
- Ett svenskt konstruktikon. Ansökt till Riksbankens Jubileumsfond (beviljad)
- Informationsåtkomst inom medicin med hänsyn till användares expertis och språkbruk. Ansökt till Riksbankens Jubileumsfond (avslagen)
- MAPiR: Methods for the automatic analysis of text in digital historical resources. Ansökt till Markus och Amalia Wallenberg (avvaktar besked)
- Mot kunskapsbaserad storskalig kunskapsutvinning ur svensk text. Ansökt till Vetenskapsrådet (beviljad)
- Nya metoder för utvärdering av syntaktisk uppmärkning i naturligt språk. Ansökt till Riksbankens Jubileumsfond (avslagen)
- Nya metoder för utvärdering av syntaktisk uppmärkning i naturligt språk. Ansökt till Vetenskapsrådet projektbidrag unga forskare (avslagen)
- PATSECURE - Improving Patient Safety by Mining for Adverse Events in Electronic Patient Records, COST (European Cooperation in Science and Technology) (avslagen)
- Skillnader i expert- och lekmanaspråk i medicinska dokument och i söfrågor och hur man kan använda sig av dem i informationsökning. 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.

SAMVERKAN

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

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

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

Nationella samarbeten

- Dramawebben
- Gothia Forum för klinisk forskning
- GPCC – Centrum för personcentrerad vård, Göteborgs universitet
- Institutet för svenska som andraspråk, Göteborgs universitet
- Kungliga biblioteket
- Lexikaliska institutet, Göteborgs universitet
- Riksarkivet mediakonverteringscentrum
- Sahlgrenska universitetssjukhuset
- Socialstyrelsen
- Språkrådet, Institutet för folk- och språkminnen
- Uppsala universitet

Internationella samarbeten

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

BESÖK

Ett flertal gästforskare och seminariegäster har besökt oss under året.

- Stefan Gries, Kaliforniens universitet i Santa Barbara
- Lars Hellan, Trondheims universitet
- Peter Ingwersen, Informationsvetenskapliga akademien i Köpenhamn
- Anni Järvelin, Tammerfors universitet
- Kalervo Järvelin, Tammerfors universitet
- Steven Krauwer, Utrechts universitet
- Hrafn Loftsson, Reykjaviks universitet
- Paul Meurer, Uni research
- Alessandro Moschitti, Trentos universitet
- Kristiina Muhonen, Helsingfors universitet
- John Nerbonne, Groningens universitet
- Bolette Pedersen, Köpenhamns universitet

8 *Språkbanken*

- Tanja Purtonen, Helsingfors universitet
- Victoria Rosén, Oslo universitet/Uni research
- Frédérique Segond, Viseo
- Anders Sjøgaard, Köpenhamns universitet
- Reut Tsarfaty, Uppsala universitet

NY PERSONAL

Under året har Språkbanken anställt flera nya medarbetare.

- Malin Ahlberg (systemutvecklare; 6+6 månader)
- Peter Andersson (forskare; 6 månader)
- Dan Rosén (systemutvecklare; 3 månader)
- Carl-Johan Schenström (systemadministratör; tills vidare)
- Kaarlo Voionmaa (forskare; 6+6 månader)
- Kenneth Wilhelmsson (forskare; 6 månader)

UTBILDNING

Grundläggande och avancerad nivå

- Statistical methods for NLP (masternivå) ht 2012
- Machine learning (masternivå) ht 2012
- Language technology resources (masternivå) ht 2012
- Information retrieval (masternivå) ht 2012

Forskarutbildning

- Doktorander:
 - Dana Dannélls, Multilingual text generation from structured formal representations, började 2007, handledare Lars Borin och Arne Ranta
 - Ann-Marie Eklund, Analysis of queries to medical web sites and databases as a basis for knowledge sharing, började 2010, handledare Dimitrios Kokkinakis och Jussi Karlgren
 - Katarina Heimann Mühlenbock, Enhancing readability by text simplification, började 2008, handledare Sofie Johansson Kokkinakis och Lars Borin
 - 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

PUBLIKATIONER

1. Yvonne Adesam, Malin Ahlberg och Gerlof Bouma. bokstaffua, bokstaffwa, bokstafwa, bokstaua, bokstawa... Towards lexical link-up for a corpus of Old Swedish. I: *Proceedings of the LTHist workshop at Konvens*, 2012.
2. Yvonne Adesam, Malin Ahlberg och Gerlof Bouma. Processing spelling variation in historical text. I: *Proceedings of the Fourth Swedish Language Technology Conference (SLTC)*, Lund, 2012.
3. Malin Ahlberg och Gerlof Bouma. A best-first anagram hashing filter for approximate string matching with generalized edit distance. I: *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, ss 13–22, Mumbai, Indien, 2012.
4. Malin Ahlberg och Ramona Enache. A type-theoretical wide-coverage computational grammar for Swedish. I: *Proceedings of the 15th International Conference on Text, Speech and Dialogue (TSD)*, ss 183–190, 2012.
5. Christiane Andersen och Markus Forsberg. Sibirientyska, 2012.
6. Amel Bennaceur, Falk Howar, Valérie Issarny, Richard Johansson, Alessandro Moschitti, Romina Spalazzese, Bernhard Steffen och Daniel Sykes. Machine learning for emergent middleware. I: *Proceedings of the Joint Workshop on Intelligent Methods for Software System Engineering (JIMSE)*, 2012.
7. Jenny Bergenmar och Leif-Jöran Olsson. Connecting European women writers. The Selma Lagerlöf archive and women writers database. I: *Digital Humanities 2012. Book of Abstracts*, Hamburg, Tyskland, 2012.
8. Dovile Bielevičiūtė, Aurelija Tamulionienė, Aivars Bērziņš, Lars Borin och Martha D. Brandt et al. META-NORD national workshops. Teknisk rapport, juli 2012.
9. Lars Borin. Core vocabulary: A useful but mystical concept in some kinds of linguistics. I: *Shall we play the festschrift game? Essays on the Occasion of Lauri Carlson's 60th Birthday*, ss 53–65. Springer, Berlin, Tyskland, 2012.
10. Lars Borin, Martha D. Brandt, Jens Edlund, Jonas Lindh och Mikael Parkvall. *Svenska språket i den digitala tidsåldern – The Swedish Language in the Digital Age*. META-NET White Paper Series. Georg Rehm och Hans Uszkoreit (red. serie). Springer, Berlin, Tyskland, 2012. Tillgänglig via <http://www.meta-net.eu/whitepapers/volumes/swedish>.
11. Lars Borin, Markus Forsberg, Karin Friberg Heppin, Richard Johansson

10 Språkbanken

- och Annika Kjellandsson. Search result diversification methods to assist lexicographers. I: *Proceedings of the 6th Linguistic Annotation Workshop*, ss 113–117, 2012.
12. Lars Borin, Markus Forsberg, Richard Johansson, Kristiina Muhonen, Tanja Purtonen och Kaarlo Voionmaa. Transferring frames: Utilization of linked lexical resources. I: *Proceedings of the Workshop on Inducing Linguistic Structure Submission (WILS)*, ss 8–15, 2012.
 13. Lars Borin, Markus Forsberg, Benjamin Lyngfelt, Julia Prentice, Rudolf Rydstedt, Emma Sköldberg och Sofia Tingsell. Growing a Swedish constructicon in lexical soil. I: *Proceedings of the Fourth Swedish Language Technology Conference (SLTC)*, ss 10–11, Lund, 2012.
 14. Lars Borin, Markus Forsberg, Leif-Jöran Olsson och Jonatan Uppström. The open lexical infrastructure of Språkbanken. I: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, ss 3598–3602, Istanbul, Turkiet, 2012.
 15. Lars Borin, Markus Forsberg och Johan Roxendal. Korp – the corpus infrastructure of Språkbanken. I: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkiet, 2012.
 16. Gerlof Bouma. Real-time persistent queues and dequeues with logic variables (declarative pearl). I: *Proceedings of the 11th International Symposium on Functional and Logic Programming (FLOPS 2012)*, ss 62–73, 2012.
 17. Martha D. Brandt, Lars Borin och Markus Forsberg et al. META-NORD: Awareness, mobilisation and dissemination actions. Teknisk rapport, januari 2012.
 18. Carina Carlund, Håkan Jansson, Sofie Johansson Kokkinakis, Julia Prentice och Judy Ribeck. An academic word list for Swedish - a support for language learners in higher education. I: *Proceedings of the SLTC 2012 workshop on NLP for CALL*, 2012.
 19. Frieda Charalabopoulou, Maria Gavrilidou, Sofie Johansson Kokkinakis och Elena Volodina. Building corpus-informed word lists for 12 vocabulary learning in nine languages. I: Linda Bradley och Sylvie Thouësny, redaktörer, *CALL: Using, Learning, Knowing. EuroCALL Conference*, Dublin, Irland, 2012. Research-publishing.net.
 20. Dana Dannélls. On generating coherent multilingual descriptions of museum objects from semantic web ontologies. I: *The 7th International Natural Language Generation Conference (INLG 2012)*, 2012.

21. Dana Dannélls och Lars Borin. Toward language independent methodology for generating artwork descriptions – exploring FrameNet information. I: *EACL 2012 workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, s 18–23, 2012.
22. Dana Dannélls, Ramona Enache, Damova Mariana och Chechev Milen. Multilingual online generation from semantic web ontologies. I: *The World Wide Web Conference (WWW2012), 16th-20th April 2012*, ss 239–242, 2012.
23. Ann-Marie Eklund. Are prepositions and conjunctions necessary in health web searches? I: *Proceedings of the Fourth Swedish Language Technology Conference (SLTC)*, ss 23–24, Lund, 2012.
24. Ann-Marie Eklund. Tracking changes in search behaviour at a health web site. I: *Quality of life through quality of information, XXIV Conference of the European Federation for Medical Informatics*, ss 858–862, Pisa, Italien, 2012.
25. Ann-Marie Eklund. Why query annotations may help in providing accurate public health information. I: *ESAIR'12: Proceedings of the fifth workshop on Exploiting Semantic Annotations in Information Retrieval*, Maui, Hawaii, 2012.
26. Ann-Marie Eklund och Dimitrios Kokkinakis. Drug interests revealed by a public health portal. I: *Proceedings of the SLTC-Workshop Exploratory Query-log Analysis*, s 2, Lund, 2012.
27. Martha D. Brandt et al. META-NORD: Awareness, mobilisation and dissemination actions - revised edition. Teknisk rapport, juni 2012.
28. Nicola Ferro, Richard Berentsen, Allan Hanbury, Mihai Lupu, Vivien Petras, Maarten de Rijke, Gianmaria Silvello, Maristella Agosti, Toine Bogers, Martin Braschler, Paul Buitelaar, Khalid Choukri, Giorgio Maria Di Nunzio, Pamela Forner, Karin Friberg Heppin, Preben Hansen, Anni Järvelin, Birger Larsen, Ivano Masiero, Henning Müller, Simone Peruzzo, Florina Piori, Giuseppe Santucci och Elaine Thoms. Promise retreat report prospects and opportunities for information access evaluation. Teknisk rapport, New York, NY, 2012.
29. Markus Forsberg och Torbjörn Lager. Cloud logic programming for integrating language technology resources. I: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkiet, 2012.
30. Karin Friberg Heppin och Håkan Friberg. Using FrameNet in communicative language teaching. I: *Proceedings of the XV EURALEX International Congress*, Oslo, Norge, 2012.

12 *Språkbanken*

31. Karin Friberg Heppin och Anni Järvelin. Towards improving search results for medical experts and laypersons. I: *Proceedings of CLEFeHealth 2012*, 2012.
32. Karin Friberg Heppin och Maria Toporowska Gronostaj. The rocky road towards a Swedish FrameNet – creating SweFN. I: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, ss 256–261, Istanbul, Turkiet, 2012.
33. Karin Friberg Heppin och Kaarlo Voionmaa. Practical aspects of transferring the English Berkeley FrameNet to other languages. I: *Proceedings of the Fourth Swedish Language Technology Conference (SLTC)*, ss 28–29, Lund, 2012.
34. Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi och Sara Tonelli. Improving the recall of a discourse parser by constraint-based post-processing. I: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, ss 2791–2794, Istanbul, Turkiet, 2012.
35. Sucheta Ghosh, Giuseppe Riccardi och Richard Johansson. Global features for shallow discourse parsing. I: *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIG-DIAL)*, ss 150–159, 2012.
36. Dorte Haltrup Hansen, Lene Offersgaard, Bolette Pedersen och Martha D. Brandt et al. META-NORD: Second batch of resources complying with the project's technical, linguistic, legal, etc. Teknisk rapport, juli 2012.
37. Aron Henriksson, Hans Moen, Maria Skeppstedt, Ann-Marie Eklund, Vidas Daudaravicius och Martin Hassel. Synonym extraction of medical terms from clinical text using combinations of word space models. I: *5th International Symposium on Semantic Mining in Biomedicine (SMBM)*, ss 10–17, Zürich, Schweiz, 2012.
38. Lars Höglund, Johan Eklund och Kenneth Wilhelmsson. Maskininlärningsbaserad indexerings av digitaliserade museiartefakter - projektrapport. Teknisk rapport, Institutionen för svenska språket, Göteborgs universitet, 2012.
39. Håkan Jansson, Sofie Johansson Kokkinakis, Judy Ribbeck och Emma Sköldberg. A Swedish academic word list: Methods and data. I: *Proceedings of the 15th EURALEX International Congress*, ss 955–960, Oslo, Norge, 2012.
40. Richard Johansson. Bridging the gap between two different Swedish treebanks. I: *Proceedings of the Fourth Swedish Language Technology*

- Conference (SLTC)*, Lund, 2012.
41. Richard Johansson. Non-atomic classification to improve a semantic role labeler for a low-resource language. I: *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM)*, ss 95–99, Montreal, Kanada, 2012.
 42. Richard Johansson, Karin Friberg Heppin och Dimitrios Kokkinakis. Semantic role labeling with the Swedish FrameNet. I: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, ss 3697–3700, Istanbul, Turkiet, 2012.
 43. Richard Johansson och Alessandro Moschitti. Relational features in fine-grained opinion analysis. *Computational Linguistics*, 2012.
 44. Sofie Johansson Kokkinakis, Emma Sköldberg, Birgit Henriksen, Kari Kinn och Janne Bondi Johannessen. Developing academic word lists for Swedish, Norwegian and Danish – a joint research project. I: R. V. Fjeld och J. M. Torjusen, redaktörer, *Proceedings of the 15th EURALEX International Congress*, ss 563–569, Department of Linguistics and Scandinavian Studies, University of Oslo, Norge, 2012.
 45. David Junger, Torbjörn Lager och Johan Roxendal. SCXML for building conversational agents in the dialog web lab. I: *Proceedings of the Fourth Swedish Language Technology Conference (SLTC)*, Lund, 2012.
 46. Dimitrios Kokkinakis. Initial experiments of medication event extraction using frame semantics. I: *Scandinavian Conference on Health Informatics (SHI)*, ss 41–47, 2012.
 47. Dimitrios Kokkinakis. The Journal of the Swedish medical association - a corpus resource for biomedical text mining in Swedish. I: *The Third Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM) at LREC*, Istanbul, Turkiet, 2012.
 48. Dimitrios Kokkinakis, Markus Forsberg, Sofie Johansson Kokkinakis, Frida Smith och Joakim Öhlén. Literacy demands and information to cancer patients. I: *Proceedings of the 15th International Conference on Text, Speech and Dialogue (TSD)*, ss 64–71, 2012.
 49. Dimitrios Kokkinakis, Eva Lidén, Staffan Svensson, Elisabeth Björk Brämberg och Sylvia Määttä. Contextualisation of functional symptoms in primary health care. I: *The 5th Geneva Conference on Person-Centered Medicine*, 2012.
 50. Dimitrios Kokkinakis och Daniela Oelke. Men, women and gods: Distant reading in literary collections - combining visual analytics with language technology. I: *Proceedings of the Advances in Visual Methods for Linguistics*, 2012.

14 *Språkbanken*

- istics (AVML)*, 2012.
51. Staffan Larsson och Lars Borin, redaktörer. *From Quantification to Conversation*. College Publications, London, Storbritannien, 2012.
 52. Benjamin Lyngfelt, Lars Borin, Markus Forsberg, Julia Prentice, Rudolf Rydstedt, Emma Sköldberg och Sofia Tingsell. Adding a constructicon to the Swedish resource network of Språkbanken. I: *11th Conference on Natural Language Processing (KONVENS) Proceedings*, ss 452–461, 2012.
 53. Benjamin Lyngfelt och Markus Forsberg. Ett svenskt konstruktikon. utgångspunkter och preliminära ramar. Teknisk rapport, Institutionen för svenska språket, Göteborgs universitet, 2012.
 54. Alessandro Moschitti, Qi Ju och Richard Johansson. Modeling topic dependencies in hierarchical text categorization. I: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*, ss 759–767, Jeju, Korea, 2012.
 55. Daniela Oelke, Ann-Marie Eklund, Svetoslav Marinov och Dimitrios Kokkinakis. Visual analytics and the language of web query logs - a terminology perspective. I: *The 15th EURALEX International Congress (European Association of Lexicography)*, s 8, Oslo, Norge, 2012.
 56. Daniela Oelke, Dimitrios Kokkinakis och Mats Malm. Advanced visual analytics methods for literature analysis. I: *Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH) at EACL*, s 10, Avignon, Frankrike, 2012.
 57. Bolette Sandford Pedersen, Lars Borin, Markus Forsberg, Krister Lindén, Heili Orav och Eiríkur Rögnvaldsson. Linking and validating Nordic and Baltic wordnets. I: *Proceedings of the 6th International Global Wordnet Conference*, ss 254–260, 2012.
 58. Jussi Piitulainen, Imre Bartis och Martha D. Brandt et al. META-NORD: Second upload of language resources. Teknisk rapport, juli 2012.
 59. Taraka Rama. N-gram approaches to the historical dynamics of basic vocabulary. I: *Preproceedings of Computational approaches to the study of dialectal and typological variation, ESSLLI workshop*, 2012.
 60. Taraka Rama och Lars Borin. Properties of phoneme N-grams across the world's language families. I: *Proceedings of the Fourth Swedish Language Technology Conference (SLTC)*, Lund, 2012.
 61. Taraka Rama och Prasanth Kolachina. How good are typological distances for determining genealogical relationships among languages? I: *Proceedings of the 24th International Conference on Computational Linguistics*

- istics (COLING)*, 2012.
62. Aarne Ranta och Markus Forsberg. *Implementing Programming Languages*. College Publications, London, Storbritannien, 2012.
 63. Eiríkur Rögnvaldsson, Aivars Bērziņš och Martha D. Brandt et al. META-NORD: Long-time viability and sustainability strategy - revised edition. Teknisk rapport, juni 2012.
 64. Emma Sköldbberg och Sofie Johansson Kokkinakis. A och O om akademiska ord. Om framtagning av en svensk akademisk ordlista. I: Birgit Eaker, Lennart Larsson och Anki Mattisson, redaktörer, *Nordiska studier i lexikografi 11. Rapport från Konferensen om lexikografi i Norden 24-27 maj 2011*, ss 575–585, Lund, 2012.
 65. Frida Smith, Joakim Öhlén, Eva Carlsson, Markus Forsberg, Dimitrios Kokkinakis och Febe Friberg. Hur kan vi förbättra skriftligt informations- och utbildningsmaterial för patienter som opereras elektivt för kolorektal cancer? I: *10:e nationella konferensen i cancervård*, Stockholm, 2012.
 66. Frida Smith, Joakim Öhlén, Eva Carlsson, Febe Friberg, Markus Forsberg och Dimitrios Kokkinakis. Ny studie visar hur information till patienter med kolorektal cancer kan förbättras. *Cancervården*, (5):18–21, 2012.
 67. Nadine Theiler och Gerlof Bouma. Two for the price of one: an LFG treatment of sentence initial object es in German. I: Miriam Butt och Tracy Holloway King, redaktörer, *Proceedings of the Lexical Functional Grammar Conference (LFG'12)*, ss 603–623, 2012.
 68. Andrejs Vasiljevs, Markus Forsberg, Tatiana Gornostay, Dorte H. Hansen, Kristín M. Jóhannsdóttir, Krister Lindén, Gunn I. Lyse, Lene Offersgaard, Ville Oksanen, Sussi Olsen, Bolette S. Pedersen, Eiríkur Rögnvaldsson, Roberts Rozis, Inguna Skadina och Koenraad De Smedt. Creation of an open shared language resource repository in the Nordic and Baltic countries. I: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkiet, 2012.
 69. Elena Volodina och Lars Borin. Developing an open-source web-based exercise generator for Swedish. I: Linda Bradley och Sylvie Thouésny, redaktörer, *CALL: Using, Learning, Knowing. EuroCALL Conference*, Dublin, Irland, 2012.
 70. Elena Volodina, Lars Borin, Hrafn Loftsson, Birna Arnbjörnsdóttir och Guðmundur Örn Leifsson. Waste not, want not: Towards a system architecture for ICALL based on NLP component re-use. I: *Proceedings of the SLTC 2012 workshop on NLP for CALL*, ss 47–58, Lund, 2012.

71. Elena Volodina, Richard Johansson och Sofie Johansson Kokkinakis. Semi-automatic selection of best corpus examples for Swedish: Initial algorithm evaluation. I: *Proceedings of the SLTC 2012 workshop on NLP for CALL*, ss 59–70, Lund, 2012.
72. Elena Volodina och Sofie Johansson Kokkinakis. Introducing Swedish Kelly-list, a new free e-resource for Swedish. I: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkiet, 2012.
73. Elena Volodina och Sofie Johansson Kokkinakis. Swedish Kelly. Teknisk rapport, Institutionen för svenska språket, Göteborgs universitet, 2012.
74. Elena Volodina, Hrafn Loftsson, Birna Arnbjörnsdóttir, Lars Borin och Guðmundur Örn Leifsson. Towards a system architecture for ICALL. I: G. Biswas et al., redaktör, *Proceedings of the 20th International Conference on Computers in Education*, band 2012, Singapore, Malaysia, 2012.
75. Kenneth Wilhelmsson. Adverbialkaraktistik för praktisk informations-extraktion i svensk text. Teknisk rapport, Institutionen för svenska språket, Göteborgs universitet, 2012.
76. Kenneth Wilhelmsson. Automatic question generation for Swedish: The current state. I: *Proceedings of the SLTC 2012 workshop on NLP for CALL*, Lund, 2012.