

---

# **Språkbanken**

Årsrapport 2011

---



---

## 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 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 workshopar. 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)**

- Quantitative Investigations in Theoretical Linguistics (QITL 4), Berlin, Tyskland/mars (1)
- Nordic Conference of Computational Linguistics (Nodalida 2011), Riga, Lettland/maj (7)
- 11 konferensen om lexikografi i Norden, Lund/maj (3)
- Workshop on Health Document Text Mining and Information Analysis (LOUHI 2011), Bled, Slovenien/juli (1)
- 20th International Conference on historical Linguistics (ICHLXX), Osaka, Japan/juli (1)
- Conference of the European Federation for Medical Informatics (MIE 2011), Oslo, Norge/augusti (2)
- 16th World Congress on Applied Linguistics (AILA 2011), Beijing, Kina/augusti (1)
- Scandinavian Conference on Health Informatics (SHI 2011), Oslo, Norge/augusti (1)
- Conference on Software, Services & Semantic Technologies (S3T 2011), Bourgas, Bulgarien/september (1)
- Recent Advantages in Natural Language Processing (RANLP 2011), Hissar, Bulgarien/september (3)
- ECML/PKDD-PASCAL Workshop on Large-Scale Hierarchical Classification, Aten, Grekland/september (1)
- Fifth IEEE International Conference on Semantic Computing (ICSC 2011), Palo Alto, USA/september (1)
- Learner Corpus Research 2011, Louvain-la-Neuve, Belgien/september (1)
- Conference of the German Society for Computational Linguistics and Language Technology (GSCL), Hamburg, Tyskland/september (1)
- Electronic Lexicography in the 21st Century (eLex 2011), Bled, Slovenien/november (1)
- Language & Technology Conference (LTC 11), Poznan, Polen/november (1)
- Svenska Läkaresällskapets Riksstämman, Stockholm/november (2)
- Terminologisk workshop, Karlstad/november (1)
- Conference on Natural Language Processing, Chiang Mai, Thailand/november (1)
- Workshop on Language Resources, Technology and Services in the Sharing Paradigm, Chiang Mai, Thailand/november (1)

### Andra presentationer, plats/månad

- Centre for Language Technology (CLT) seminarium, Göteborg/under året (ca 5 seminarier)
- Presentation och demonstration vid Clarin CMDI workshop, Nijmegen, Nederländerna/januari
- 'Svenskt frasnät++, en bro mellan det förflutna och framtiden', Universitetet i Oslo/mars
- 'Svenskt frasnät', Universitetet i Oslo/mars
- 'WP8 Case study: Cultural Heritage', vid MOLTO meeting 2, Göteborg/mars
- Presentation av Språkbanken och Korp för SND:s vetenskapliga råd, Göteborg/mars
- Presentation och demonstration vid Clarin European demonstrator workshop, Nijmegen, Nederländerna/maj
- Presentation och demonstration vid International Association for Social Science Information Services & Technology (IASSIST) Conference, Vancouver, Kanada/juni
- 'WP8 Case study: Cultural Heritage', vid MOLTO meeting 3, Helsingfors, Finland/augusti
- 'Metadata discussion', vid META-NORD Metadata workshop i Helsingfors, Finland/september
- 'Hands-on session', vid META-NORD Metadata workshop i Helsingfors, Finland/september
- 'META-NORD and metadata', vid META-NORD Metadata workshop i Helsingfors, Finland/september
- 'Mot en webbaserad svensk ordlista', vid Med svenska språket som redskap för akademiska studier, Stockholm/september
- Centre for Language Technology (CLT) årlig workshop, Gullmarstrand/oktober (ca 15 bidrag)
- Presentation av Språkbanken för en delegation från Norges språkråd, Göteborg/oktober
- 'En infrastruktur för språkforskning och språkundervisning, Språkbankens nya korpus- och lexikoninfrastruktur', vid Sproglig og kulturel mangfoldighed i Norden. Milepæl i nordisk fremtidsforskning på sprog- og kulturområdet, Islands Universitet, Reykjavík/oktober
- 'Visual Exploration of Web Session Logs', presentation på FindWise AB, Göteborg/oktober
- 'Task 3.1 - Upgrading resources to agreed standards - Metadata, META-SHARE editor tool and UGOT upload environment', vid META-NORD Consortium meeting i Berlin, Tyskland/oktober

#### 4 *Språkbanken*

- ‘En webbaserad svensk akademisk ordlista’, vid Nätverkskonferens för universitetens och högskolornas språk- och studieverkstäder, Göteborg/november
- ‘Resursbehovet för svensk språkteknologi’, vid Seminarium om nationell språkdatabank, Stockholm/november
- Presentation och demonstration vid 3rd Annual European DDI User Group Meeting (EDDI11), Göteborg/december

#### **Organiserad konferens/workshop, plats/månad (antal deltagare)**

- Workshop i samband med lansering av korpussökgränssnittet Korp (Göteborg)/oktober (ca 70)
- Workshop on comparing approaches to measuring linguistic differences (Göteborg)/oktober (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 2); finansiär ISA
- CLARIN (år 4 av 4); EU-projekt med egen finansiering
- CLT:s textteknologilaboratorium (år 3 av 4); finansiär CLT (Göteborgs universitet)
- CONPLISIT (år 2 av 2); egen finansiering
- Digital areallingvistik (år 2 av 3); finansiär VR
- Effektiv informationsförädling i sjukvården (år 2 av 2); finansiär VGR regionala utvecklingsmedel
- GRUS – grundskoleelevers skrivande i ”en dator till en elev”-satsning (år 2); finansiär LUN/ISA
- IKT i lärarutbildningen, anknytning mellan forskning och undervisning, datorbaserad textanalys och ordförrådsbedömning (år 2); finansiär LUN
- KELLY (år 2 av 2); finansiär EU
- Litteraturbanken (permanent); finansiär Svenska Akademien, Kungl. Vitterhetsakademien
- META-NORD (år 1 av 2); finansiär EU
- MOLTO (år 2 av 3); finansiär EU
- Mätning av ordförrådet i andraspråket (år 3), finansiär RJ
- SPF (DB1800) (år 3 av 3); finansiär VR
- SweFN++ (år 2); finansiär CLT (Göteborgs universitet), VR

Flera projektansökningar har inlämnats för pågående och kommande projekt.

- EU: Mitras (avslagen)
- Göteborgs universitet: Akademisk ordlista för svenska (beviljad)
- Längmanska Kulturfonden: Medeval (avslagen)
- Nordforsk: DiaSynCorp (beviljad)
- Nordforsk: Nordisk workshop om Akademiska ordlistor (avslagen)
- Nordplus: A system architecture for intelligent computer-assisted language learning (beviljad)
- NOS-HS: Nordisk workshop om Akademiska ordlistor (avslagen)
- Stint, SweFN++ (avvaktar besked)
- Söderbergs Stiftelser: IPKC: Information för Patienter med Kolorektal Cancer - är texter skrivna på ett begripligt och lättläst språk? (avslagen)
- Vetenskapsrådet: SweCLARIN (avslagen)
- Vinnova: Effektiv informationsförädling i sjukvården (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
- Institutionen för historiska studier, Göteborgs universitet
- Kungliga biblioteket
- Lexikaliska institutet, Göteborgs universitet
- Riksarkivet mediakonverteringscentrum
- Sahlgrenska universitetssjukhuset
- Socialstyrelsen
- 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.

- Anni Järvelin, Tammerfors universitet/SICS
- Detmar Meurers, Tübingens universitet
- Maarten de Rijke, Amsterdams universitet
- Adam Kilgariff, Lexical Computing Ltd
- Pierre Nugues, Lunds universitet
- Peter Exner, Lunds universitet
- Daniela Oelke, Konstanz universitet
- Marina Santini, Artificial solutions
- Nancy Chang, Berkeley universitet
- Julia Pennlert, Umeå universitet
- Sverker Lundin, Göteborgs universitet
- Daniel Keim, Konstanz universitet
- Annette Hulth, Smittskyddsinstitutet
- Angus Roberts, Sheffields universitet

## NY PERSONAL

Under året har Språkbanken anställt flera nya medarbetare.

- Gerlof Bouma (postdok, 2 år)
- Johan Roxendal (systemutvecklare; tills vidare)
- Karin Friberg Heppin (postdok, 2 år)
- Martha Dís Brandt (systemutvecklare; tills vidare)
- Martin Hammarstedt (systemutvecklare; tills vidare)
- Richard Johansson (postdok, 2 år)
- Svetoslav Marinov (forskningsingenjör; tills vidare)
- Yvonne Adesam (projektsamordnare; 6+6 mån)

## UTBILDNING

- Kurser:
  - Korpuslingvistiska metoder och verktyg (master- och forskarnivå) vt 2011 (kursansvar och examination)
  - Introduction to programming (master- och forskarnivå) ht 2011 (kursansvar och examination)
  - Language Technology Resources (master- och forskarnivå) ht 2011 (kursansvar och examination)

## 8 Språkbanken

- Doktorander:

- Dana Danélls, Multilingual text generation from structured formal representations, började 2007, handledare Lars Borin och Aarne 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 Kasicheyanula, Computational historical linguistics, började 2010, handledare Lars Borin och Søren Wichmann
- Carola Ribeck, Syntactic analysis of textbooks used in Sweden when teaching mathematics and natural science, började 2010, handledare Sofie Johansson Kokkinakis och Lars Borin

## PUBLIKATIONER

1. Helen Allvin, Elin Carlsson, Hercules Dalianis, Riitta Danielsson-Ojala, Vidas Daudaravicius, Martin Hassel, Dimitrios Kokkinakis, Heljä Lundgren-Laine, Gunnar Nilsson, Øystein Nytrø, Sanna Salanterä, Maria Skeppstedt, Hanna Suominen och Sumithra Velupillai. Characteristics of Finnish and Swedish intensive care nursing narratives: a comparative analysis to support the development of clinical language technologies. *Journal of Biomedical Semantics*, 2(Suppl 3):S1, juli 2011.
2. Imre Bartis, Aivars Bērziņš, Martha D. Brandt och Leif-Jöran Olsson. First upload of language resources. Teknisk rapport, december 2011.
3. Lars Borin, Martha D. Brandt, Jens Edlund, Jonas Lindh och Mikael Parkvall. Languages in the European Information Society - Swedish. Teknisk rapport, Berlin, juni 2011.
4. Lars Borin och Markus Forsberg. A Diachronic Computational Lexical Resource for 800 Years of Swedish. I: Caroline Sporleder, Antal Bosch och Kalliopi Zervanou, redaktörer, *Language Technology for Cultural Heritage, Theory and Applications of Natural Language Processing*, ss 41–61. Springer Berlin Heidelberg, 2011.
5. Lars Borin och Markus Forsberg. Swesaurus – ett svenskt ordnät med fria tyglar. *LexicoNordica*, 18:17–39, 2011.
6. Lars Borin, Markus Forsberg och Christer Ahlberger. Semantic Search in Literature as an e-Humanities Research Tool: CONPLISIT – Consumption Patterns and Life-Style in 19th Century Swedish Literature. I: Bo-

- lette Sandford Pedersen, Gunta Nešpore och Inguna Skadina, redaktörer, *Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011*, band 11 av *NEALT Proceedings Series*, ss 58–65. Northern European Association for Language Technology (NEALT), 2011.
7. Lars Borin, Jonas Lindh, Martha D. Brandt och Leif-Jöran Olsson. Metadata descriptions and other interoperability standards. Teknisk rapport, maj 2011.
  8. Dana Dannélls. D8.1. Ontology and corpus study of the cultural heritage domain, first deliverable for MOLTO workpackage 8. <http://www.molto-project.eu/node/1372>, september 2011.
  9. Dana Dannélls och Mariana Damova. Reason-able View of Linked Data for Cultural Heritage. I: *Proceedings of the Third International Conference on Software, Services and Semantic Technologies (S3T'2011)*, Advances in Intelligent and Soft Computing, Bourgas, Bulgaria, september 2011. Springer Verlag.
  10. Dana Dannélls, Mariana Damova, Ramona Enache och Milen Chechev. A Framework for Improved Access to Museum Databases in the Semantic Web. I: *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage in conjunction with the Recent Advances in Natural Language Processing (RANLP)*, Hissar, Bulgaria, 2011.
  11. Ann-Marie Eklund. Medical Knowledge Evolution - Query Constraining Aspects. I: *User Centred Networked Health Care - Proceedings of the XXIII International Conference of the European Federation for Medical Informatics (MIE)*, band 169, ss 549–553, Oslo, Norway, 2011.
  12. Ann-Marie Eklund. Query Constraining Aspects of Knowledge - A Case Study. I: Bolette Sandford Pedersen, Gunta Nešpore och Inguna Skadina, redaktörer, *Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011*, band 11 av *NEALT Proceedings Series*, ss 279–282. Northern European Association for Language Technology (NEALT), 2011.
  13. Ann-Marie Eklund. Relational Annotation of Scientific Medical Corpora - A Case Study. I: *Proceedings of LOUHI 2011, Third International Workshop on Health Document Text Mining and Information Analysis*, ss 27–34, Bled, Slovenia, juli 2011.
  14. Markus Forsberg. Green resources in plain sight: opening up the SweFN++ project. I: Volume Editors Sjur Nørstebø Moshagen och Per Langgård, redaktörer, *Proceedings of the NODALIDA 2011 workshop on*

10 *Språkbanken*

- Visibility and Availability of LT Resources*, band 13 av *NEALT Proceedings Series*. Northern European Association for Language Technology (NEALT), 2011.
15. Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi och Sara Tonelli. Shallow Discourse Parsing with Conditional Random Fields. I: Haifeng Wang och David Yarowsky, redaktörer, *Proceedings of 5th International Joint Conference on Natural Language Processing*, ss 1071–1079, Chiang Mai, Thailand, november 2011.
  16. Sucheta Ghosh, Sara Tonelli, Giuseppe Riccardi och Richard Johansson. End-to-End Discourse Parser Evaluation. I: *Fifth IEEE International Conference on Semantic Computing (ICSC)*, ss 169–172, Palo Alto, United States, september 2011.
  17. Harald Hammarström och Lars Borin. Unsupervised learning of morphology. *Computational Linguistics*, 37(2):309–350, juni 2011.
  18. Dorte Haltrup Hansen, Bolette Pedersen, Martha D. Brandt och Markus Forsberg et al. First batch of language resources. Teknisk rapport, december 2011.
  19. Karin Heppin. MedEval – A Swedish medical test collection with doctors and patients user groups. *Journal of Biomedical Semantics*, 2(Suppl 3):S4, 2011.
  20. Qi Ju, Richard Johansson och Alessandro Moschitti. Towards Using Reranking in Hierarchical Classification. I: *Proceedings of the Joint ECML/PKDD-PASCAL Workshop on Large-Scale Hierarchical Classification*, Athens, Greece, september 2011.
  21. Dimitrios Kokkinakis. Evaluating the Coverage of three Controlled Health Vocabularies with Focus on Findings, Signs & Symptoms. I: *Workshop on Creation, Harmonization and Application of Terminology Resources Co-located with NODALIDA*, 2011.
  22. Dimitrios Kokkinakis. Health Portals and Clinical Phenotypes - Recognition using SNOMED CT. I: *Proceedings of the 9th Scandinavian Conference on Health Informatics*, Oslo, Norway, 2011. Tapir Akademiske Forlag.
  23. Dimitrios Kokkinakis. Medicinska terminologier - officiella standarder och verklighet. I: *Terminologisk workshop*, Karlstad, Sweden, 2011.
  24. Dimitrios Kokkinakis. Natural language processing of clinical data with a focus on diffuse symptoms. I: *Svenska Läkaresällskapets Riksstämman*, 2011.
  25. Dimitrios Kokkinakis. Reducing Complexity in Parsing Scientific Medi-

- cal Data, a Diabetes Case Study. I: *Proceedings of the Workshop on Bio-medical Natural Language Processing in conjunction with Recent Advances in Natural Language Processing (RANLP)*, Hissar, Bulgaria, 2011.
26. Dimitrios Kokkinakis. What is the Coverage of SNOMED CT(R) on Scientific Medical Corpora? I: *XXIII International Conference of the European Federation for Medical Informatics*, 2011.
  27. Dimitrios Kokkinakis och Mats Malm. Character Profiling in 19th Century Fiction. I: *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage in conjunction with the Recent Advances in Natural Language Processing (RANLP)*, Hissar, Bulgaria, 2011.
  28. Sofie Johansson Kokkinakis och Ulrika Magnusson. Computer based quantitative methods applied to first and second language student writing. I: *Young urban Swedish. Variation and change in multilingual settings*, ss 105–124. University of Gothenburg, Gothenburg, Sweden, 2011.
  29. Sofie Johansson Kokkinakis och Elena Volodina. Corpus-based approaches for the creation of a frequency based vocabulary list in the EU project KELLY – issues on reliability, validity and coverage. I: *Proceedings of Electronic lexicography in the 21st century (eLex 2011)*, Bled, Slovenia, november 2011.
  30. Sudheer Kolachina, Taraka Rama och Lakshmi Bai. Maximum parsimony method in the subgrouping of Dravidian languages. I: Amir Zeldes och Anke Lüdeling, redaktörer, *Proceedings of Quantitative Investigations in Theoretical Linguistics 4 (QITL-4)*, Berlin, Humboldt-Universität, mars 2011.
  31. Gyri Smørdal Losnegaard, Anje Müller Gjesdal och Martha D. Brandt et al. Report on resources (actually or potentially) available to the META-NORD consortium. Teknisk rapport, juli 2011.
  32. Kaili Müürisep, Martha D. Brandt och Markus Forsberg et al. Report on methodology and criteria followed for the selection of resources. Teknisk rapport, oktober 2011.
  33. Taraka Rama och Lars Borin. Estimating Language Relationships from a Parallel Corpus. A Study of the Europarl Corpus. I: Bolette Sandford Pedersen, Gunta Nešpore och Inguna Skadina, redaktörer, *Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011*, band 11 av *NEALT Proceedings Series*, ss 161–167. Northern European Association for Language Technology (NEALT), 2011.
  34. Carola Ribeck. Identifying Lexical Bundles in Secondary School Textbooks. I: Xygmunt Vetualani, redaktör, *Human Language Technologies*

- as a Challenge for Computer Science and Linguists, Proceedings of the 5th Language & Technology Conference*, ss 202–206, Poznan, Poland, november 2011.
35. Eiríkur Rögnvaldsson, Lars Borin och Jonas Lindh et al. Description of third parties networks for each language covered in the META-NORD project. Teknisk rapport, juli 2011.
  36. Eiríkur Rögnvaldsson, Lars Borin och Martha D. Brandt et al. Long-time viability and sustainability strategy. Teknisk rapport, juli 2011.
  37. Eiríkur Rögnvaldsson, Lars Borin, Jonas Lindh och Martha D. Brandt et al. META-NORD action plan. Teknisk rapport, juli 2011.
  38. Anju Saxena och Lars Borin. Dialect Classification in the Himalayas: a Computational Approach. I: Bolette Sandford Pedersen, Gunta Nešpore och Inguna Skadina, redaktörer, *Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011*, band 11 av *NEALT Proceedings Series*, ss 307–310. Northern European Association for Language Technology (NEALT), 2011.
  39. Inguna Skadina, Andrejs Vasiljevs, Lars Borin, Koenraad De Smedt, Krister Lindén och Eiríkur Rögnvaldsson. META-NORD: Towards Sharing of Language Resources in Nordic and Baltic Countries. I: *Proceedings of the Workshop on Language Resources, Technology and Services in the Sharing Paradigm*, ss 107–114, Chiang Mai, Thailand, november 2011. Asian Federation of Natural Language Processing.
  40. Frida Smith, Eva Carlsson, Febe Friberg, Dimitrios Kokkinakis, Markus Forsberg, Matilda Öhrn och Joakim Öhlén. Developing a toolkit for written information materials for patients with colorectal cancer undergoing elective surgery. I: *Svenska Läkaresällskapets Riksstämman*, 2011.
  41. Jean Vancoppenolle, Eric Tabbert, Gerlof Bouma och Manfred Stede. A German Grammar for Generation in OpenCCG. I: Hedeland, Schmidt och Wörner, redaktörer, *Multilingual Resources and Multilingual Applications, Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2011*, ss 145–151, Hamburg, Germany, 2011.
  42. Andrejs Vasljev, Bolette Sandford Pedersen, Koenraad De Smedt, Lars Borin och Inguna Skadina. META-NORD: Baltic and Nordic Branch of the European Open Linguistic Infrastructure. I: Sjur Nørstebø Moshagen och Per Langgård, redaktörer, *Proceedings of the NODALIDA 2011 Workshop Visibility and Availability of LT Resources*, band 13 av *NEALT Proceedings Series*, s 18–22. Northern European Association for Language Technology (NEALT), 2011.

43. Kadri Vider, Krista Liin, Lars Borin, Martha D. Brandt och Markus Forsberg et al. Selection of resources, agreements, detailed work plan. Teknisk rapport, december 2011.
44. Hanna Westerlund, Ville Oksanen, Krister Lindén, Leif-Jöran Olsson och Jonas Lindh et al. IPR principles and provisions as agreed with META-NET and partner projects. Teknisk rapport, maj 2011.
45. Søren Wichmann, Taraka Rama och Eric Holman. Phonological diversity, word length, and population sizes across languages: The ASJP evidence. *Linguistic Typology*, 15(2):177–197, november 2011.