
Språkbanken

Årsrapport 2015

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 20th Nordic Conference of Computational Linguistics (NODALIDA 2015), Vilnius, Litauen/maj (5)
- Workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities, (NODALIDA 2015), Vilnius, Litauen/maj (2)
- The 2015 Annual Conference of the North American Chapter of the ACL: Human Language Technologies, Denver, Colorado/maj (3)
- The 9th Workshop on Innovative Use of NLP for Building Educational Applications, Denver, Colorado/juni (1)
- The 4th Workshop on Computational Linguistics for Literature, Denver, Colorado/juni (1)
- The 1st Workshop on Vector Space Modeling for Natural Language Processing, Denver, Colorado/juni (1)
- International Conference: Digital Literary Studies, Coimbra, Portugal/maj (1)
- International Conference in Recent Advances in Natural Language Processing (RANLP), Chisarja, Bulgarien/september (2)
- Workshop on Grammar Engineering Across Frameworks (GEAF), 53rd Annual Meeting of the ACL and 7th IJCNLP, Peking, Kina/juli (1)
- The 7th International Conference on Computer Supported Education (CSED), Lissabon, Portugal/maj (1)
- The 14th International Workshop on Treebanks and Linguistic Theories (TLT14), Warszawa, Polen/december (1)
- Workshop on Corpus based Research in the Humanities (CRH), Warszawa, Polen/december (1)
- Nordiske forskningskonferansen om nordiske språk som andre- og fremmedspråk (NordAnd-12), Norge/juni (1)
- Digital Humanities (DH2015), Sydney/juni (1)

Andra presentationer, plats/månad

- Centre for Language Technology (CLT) seminarium, Göteborg/under året (4 seminarier)
- The 6th CLT workshop, Nääs Fabriker, Tollerod/oktober (13 presentationer)
- Workshop of Working Group 3 of the European Network of e-Lexicography on the Automatic Extraction of Good Dictionary Examples, Wien/februari (1 presentation)
- Summer Datathon on Linguistic Linked Open Data (SD-LLOD), Cerdilla/juni (1 presentation)

- CLARIN annual conference, Wrocław/oktober (2 presentationer)
- CLARIN Workshop on Named Entity Recognition in Digital Humanities, Helsinki/juni (3 presentationer)
- Digisams Swe-Clarín-verkstad “Digitala vägar till samlingarna”, KB Stockholm/oktober (2 presentationer)
- eXist-db-förkonferens XML-Prag, Prag/februari (3 presentationer)
- The 5th annual Språkbanken autumn workshop (höstworkshop), Göteborg/oktober (3 presentationer)
- UFAL Karlsuniversitetet, Prag/november (1 presentation)
- “Text reuse of biblical text/quotes in Swedish 19th century fiction”. NorLit 2015: THE BOOK TO COME – The 13th biannual conference of the Nordic Association for Comparative Literature, Göteborg/augusti
- “Relations, co-ordinations and rhetorics”. Swe-Clarins första HSS workshop i Göteborg/april
- “Karp & Korp”. NEL meeting (European Network for e-Lexicography), Wien/februari
- “Sentence selection in the context of Swedish as a second language: potential for GDEX”. ENEL workshop (European Network for e-Lexicography), Wien/februari
- “4th NLP for CALL”. Introduction for the 20th Nordic Conference of Computational Linguistics (NoDaLiDa) workshop, Vilnius/maj
- “Introducing SVALex: a corpus-based lexical resource for second language learning”. ENEL meeting (European Network for e-Lexicography), Herstmonceux slott/augusti
- “Embedding a Semantic Network in a Word Space”. Internationell datavetenskap institution/maj
- “Embedding a Semantic Network in a Word Space”. Google, Mountain View, Kalifornien/juni
- “Introducing index lists as a tool for identifying typical linguistic features of specialized registers, with examples from Swedish textbooks in natural sciences”. The International Association for Research on Textbooks and Educational Media (IARTEM), Berlin/december
- “Didaktiskt baserat läsförståelsetest och pedagogiska insatser”. Skriv!Les! Stavanger/maj
- “Exploiting FrameNet for Swedish – Can it be done?” Seminarium Institutionen för Lingvistik och filologi, Uppsala universitet/mars

I media

Utöver regelrätta publikationer och presentationer har Språkbanken och språkbankens resurser omnämnts i media i flera fall.

4 *Språkbanken*

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

- Ko|arp-hackathon workshop, Göteborg/november (ca. 11)
- Höstworkshop, Göteborg/oktober (ca. 60)
- Swe-Clarins första workshop för humaniora och samhällsvetenskap (HS), Göteborg/april (ca. 60)
- Swe-Clarins och det talade språket I, SWE-CLARIN workshop, Stockholm/november (ca. 50)
- 4th workshop on NLP for computer-assisted language learning (NLP4CALL), NoDaLiDa, Vilnius/maj (ca. 20)
- eXist-db-förkonferens XML-Prag, Prag/februari (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 5); finansiär ISA
- Automatic identification of language change; egen finansiering
- CLT:s textteknologilaboratorium (år 7 av 7); finansiär CLT (Göteborgs universitet)
- Distributional Methods to Represent the Meaning of Frames and Constructions (år 2 av 5); finansiär VR
- Diabase – Towards a diachronic BLARK; egen finansiering
- Digital areallingvistik (år 6); finansiär VR
- Dramawebben (år 3 av 3); finansiär VR
- GRUS – Gymnasie- och grundskoleelevers skrivande i ”en dator till en elev”-satsning (år 5); finansiär ISA
- Koala – Korps lingvistiska annotationer (år 2 av 3), finansiär RJ
- Kulturomik: Mot kunskapsbaserad storskalig kunskapsutvinning ur svensk text (år 4 av 5); finansiär VR
- Litteraturbanken (permanent); finansiär Svenska Akademien, Kungl. Vetternhetsakademien
- MApiR – Metoder för automatisk Analys av Text i digitala Historiska Resurser (år 2 av 3), finansiär MAW
- Science and literacy teaching (år 1 av 4); finansiär VR
- SweCxn – ett svenskt konstruktikon (år 3 av 3); finansiär RJ
- Swe-Clarins (år 2 av 5); finansiär VR
- SweFN++ (år 5); finansiär CLT (Göteborgs universitet), VR

Flera projektansökningar har inlämnats för pågående och kommande projekt.

- Språkliga och extra-lingvistiska parametrar för tidig upptäckt av kognitiv svikt. Ansökt till Riksbankens Jubileumsfond (beviljad)
- Perspektiv på det emotionella landskapet i svensk fiktion. Ansökt till Marcus och Amalia Wallenberg (avslagen)
- Översättningens kulturer. Ansökt till Vetenskapsrådet (avslagen)
- Bootstrapping Language Learning Interactive Symbiotic Suite (BLLISS). EU application Horizon 2020, ansökt till Call H2020-ICT-2015 (avslagen)
- Profiling Swedish as a Second Language: corpus-driven text studies based on the Common European Framework of Reference. Ansökt till Vetenskapsrådet (avslagen)
- Infrastructure to support L2 research, teaching and assessment. Ansökt till Vetenskapsrådet Inventory of research infrastructure needs 2015 (okänt)
- Classification of Swedish learner essays by CEFR levels. Ansökt till institutionen för svenska språket (beviljad)
- Science and literacy teaching. Ansökt till Vetenskapsrådet (beviljad)

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.

6 *Språkbanken*

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

- CDH – Centrum för Digital Humaniora, Göteborgs universitet
- DART kommunikations- och dataresurscenter
- Dialekt- och ortnamnsarkivet, Institutet för språk och folkminnen, Göteborg
- 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
- IMPACT Centre of Competence
- 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
- PARSEME (PARSing and Multi-word Expressions)
- Tekstlaboratoriet, UiO – Universitetet i Oslo

BESÖK

Ett flertal gästforskare och seminariegäster har besökt oss under året.

- Eckhart Bick, Syddansk universitet
- Chris Biemann, TU Darmstadt
- Marco Büchler, Georg August Göttingen universitet
- Christian Chiarcos, Universität Potsdam
- Paul Cook, New Brunswick universitet
- Thomas François, CENTAL (Université catholique de Louvain)
- Olga Gerassimenko, Tartu universitet
- Jey Han Lau, King's College
- Gerhard Jäger, Tübingen universitet
- Neeme Kahusk, Tartu universitet
- Christos Koniaris, KTH
- Beáta Megyesi, Uppsala universitet
- Roberto Navigli, Sapienza universitet
- Jyrki Niemi, Helsinki universitet
- Mattias Nilsson Benfatto, Karolinska Institutet
- Kristina Nilsson Björkenstam, Stockholm universitet
- Barbara Plank, Københavns universitet
- Josef Ruppenhofer, Universität Hildesheim
- Magnus Sahlgren, Gavagai
- Pavel Straňák, Karlsuniversitetet i Prag
- Oscar Täckström, Google
- Jörg Tiedemann, Uppsala universitet
- Tiago Timponi Torrent, Juiz de Fora Federal universitet

8 *Språkbanken*

- Kadri Vider, Tartu universitet
- Torsten Zesch, Universität Duisburg-Essen
- Gustaf Öqvist Seimyr, Karolinska Institutet
- Lilja Øvrelid, Oslo universitet

Dessutom har vi haft ett större studiebesök från Centrum för Språkinledning med en omfattning om 38 personer: 36 inlärare av svenska som andraspråk och 2 lärare. Syftet med besöket har varit att utöka kontakter med skolan och experimentera med meningsurvalsalgoritmer.

BEDÖMNINGSUPPDRAG

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

- Tidskriftsartiklar
 - The Semantic Web Journal (SWJ)
 - Medical Informatics – candidate articles for Yearbook of Med. Informatics
 - Computer-Assisted Language Instruction Consortium (CALICO) journal
 - Computational Linguistics
 - Lingua
- Konferensbidrag
 - Conference on Building and Exploring Web Based Environments (WEB)
 - Conference on Computational Natural Language Learning (CoNLL)
 - Conference of the Association for Computational Linguistics conference (ACL)
 - European Conference on Information Retrieval (ECIR)
 - Empirical Methods on Natural Language Processing (EMNLP)
 - International Joint Conference on Natural Language Processing (IJCNLP)
 - International Conference on Parsing Technologies (IWPT)
 - International Conference on Applications of Natural Language to Information Systems (NLDB)
 - Language Resources and Evaluation Conference (LREC)
 - Lexical and Computational Semantics (*SEM)
 - Workshop on Building Education Applications (NAACL)

- Workshop on Natural Language Processing Techniques for Educational Applications (ACL & IJCNLP)
- Workshop on NLP for computer-assisted language learning (NoDaLiDa)
- Workshop on Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities (NODALIDA)
- Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)
- Övrigt
 - Betygskommitté, licentiatseminarium, Göteborgs universitet
 - Betygskommitté, disputation, Göteborgs universitet
 - Betygskommitté, disputation, Stockholm universitet

NY PERSONAL

Under året har Språkbanken anställt två systemutvecklare/forskningsingenjörer, Tomasz Kosiński och Maria Öhrman samt postdoktor, Shafqat Mumtaz Virk.

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:
 - Computer-based analysis of school textbooks and normative texts (grundnivå/masternivå) ht2015
 - Empirisk skriftspråksforskning (forskarnivå) ht2015 (kursansvar)
 - Grundläggande språkteknologi (masternivå) ht2015
 - Informationssökning (kandidatnivå/masternivå) vt2015
 - Informationshantering och informationsstrukturer (masternivå) vt2015
 - Introduktion till programmering (masternivå) ht2015 (kursansvar)
 - Korpuslingvistik (forskarnivå) ht2015 (kursansvar)
 - Maskininlärning (masternivå) ht2015 (kursansvar)
 - Metodkurs inför MLT masterprojekt (masternivå) vt2015 (kursansvar)
 - Natural language processing (masternivå) ht2015
 - Semantik och pragmatik (grundnivå) vt2015
 - Skriva (grundnivå) ht2015
 - Språk, datorer och textbehandling (grundnivå) vt2015

10 *Språkbanken*

- Språkssystemet (grundnivå) vt2015
- Språkteknologiresurser (forskarnivå) ht2015 (kursansvar)
- Språkvetenskaplig problemlösning (grundnivå) vt2015
- Statistiska metoder (masternivå) vt2015 (kursansvar)
- Swedish language and culture (forskarnivå) ht2015
- Svenska 5 för gymnasielärare (forskarnivå) ht2015
- Svenska för studenter med finska som modersmål (masternivå) sommarkurs
- Svenska som andraspråk, 1, för gymnasielärare (grundnivå) vt2015
- Svenska som andraspråk 1 för lärare åk 7-9 (grundnivå) vt2015
- Svenska 5 för gymnasielärare (grundnivå) vt2015
- Textanalys. Språkteknologiskt moment (grundnivå) ht2015
- Grundstudenter:
 - Handledning av fyra grundstudenter, fristående kurser
- Masterstudenter:
 - Handledning av tre masterstudenter från MLT-programmet
- Doktorander:
 - Taraka Rama Kasicheyanula, Computational historical linguistics, började 2010, handledare Lars Borin och Søren Wichmann
 - Judy Ribeck, Syntactic analysis of textbooks used in Sweden when teaching mathematics and natural science, började 2010, handledare Emma Sköldberg och Lars Borin
 - Ildikó Pilán, Swedish second language readability from a semantic perspective, började 2013, handledare Lars Borin och Elena Volodina
 - Luis Nieto Piña, Distributional methods to represent the meaning of frames and constructions, började 2014, handledare Richard Johansson
 - Prasanth Kolachina, GU CSE, bihandledare Richard Johansson
 - Mikael Kågebäck, Chalmers CSE, bihandledare Richard Johansson
- Disputationer:
 - Taraka Rama, Studies in computational historical linguistics: Models and analyses, disputation november 2015.
 - Judy Ribeck, Steg för steg. Naturvetenskapligt ämnesspråk som räknas, disputation december 2015.

PUBLIKATIONER

1. Yvonne Adesam, Gerlof Bouma och Richard Johansson. Defining the Eukalyptus forest – the Koala treebank of Swedish. I: *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA) May 11-13, Vilnius, Lithuania. Edited by Beáta Megyesi*, ss 1–9, 2015.
2. Yvonne Adesam, Gerlof Bouma och Richard Johansson. Multiwords, Word Senses and Multiword Senses in the Eukalyptus Treebank of Written Swedish. I: *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14), 11–12 December, Warsaw, Poland*, ss 3–12, 2015.
3. Malin Ahlberg, Peter Andersson, Markus Forsberg och Nina Tahmassebi. A case study on supervised classification of Swedish pseudo-coordination. *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA) May 11-13, Vilnius, Lithuania*, ss 11–19, 2015.
4. Malin Ahlberg, Markus Forsberg och Måns Huldén. Paradigm classification in supervised learning of morphology. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, ss 1024–1029, 2015.
5. Christiane Andersen, Markus Forsberg, Martin Hammarstedt och Alexander Pankow. Sibirientyska kvinnor (Siberian German women), 2015.
6. Jim Barnett, Rahul Akolkar, RJ Auburn, Michael Bodell, Daniel C. Burnett, Jerry Carter, Scott McGlashan, Torbjörn Lager, Mark Helbing, Rafah Hosn, T.V. Raman, Klaus Reifenrath, No'am Rosenthal och Johan Roxendal. State Chart XML (SCXML): State Machine Notation for Control Abstraction. Teknisk rapport, Massachusetts, USA, 2015.
7. Jenny Bergenmar och Leif-Jöran Olsson. Tracing cultural transfer through multiple translation analysis. The case of the Swedish 19th-century Bourgeois novel in German and Czech. I: *Digital Literary Studies. International Conference*, 2015.
8. Lars Borin, Luis Nieto Piña och Richard Johansson. Here be dragons? the perils and promises of inter-resource lexical-semantic mapping. I: *Linköping Electronic Conference Proceedings. Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities. Workshop at NODALIDA, May 11, 13-18, Vilnius*, band 112, ss 1–11, 2015.
9. Karin Friberg Heppin och Dana Dannélls. Polysemy and questions of

12 Språkbanken

- lumping or splitting in the construction of Swedish FrameNet. I: *Proceedings of the Workshop on Semantic resources and Semantic Annotation for Natural Language Processing and the Digital Humanities at NODALIDA, Vilnius, 11th May*, ss 12–20, 2015.
10. Mehdi Ghanimifard och Richard Johansson. Enriching Word-sense Embeddings with TranslationalContext. I: Galia Angelova, Kalina Bontcheva, Ruslan Mitkov. International Conference, Hissar och Bulgaria 7–9 September, redaktörer, *Proceedings of Recent Advances in Natural Language Processing*, ss 208–215, 2015.
 11. Normunds Gruzitis och Dana Dannélls. A multilingual FrameNet-based grammar and lexicon for controlled natural language. *Language resources and evaluation*, 2015.
 12. Normunds Gruzitis, Dana Dannélls, Benjamin Lyngfelt och Aarne Ranta. Formalising the Swedish Constructicon in Grammatical Framework. I: *Proceedings of the Grammar Engineering Across Frameworks (GEAF) Workshop, 53rd Annual Meeting of the ACL and 7th IJCNLP, Beijing, China, July 26-31*, ss 49—56, 2015.
 13. Katarina Heimann Mühlenbock, Sofie Johansson Kokkinakis, Caroline Liberg, Åsa af Geijerstam, Jenny Wiksten Folkeryd, Arne Jönsson, Erik Kanebrant och Johan Falkenjack. A multivariate model for classifying texts' readability. I: *ACL Anthology - Proceedings of the 20th Nordic Conference of Computational Linguistics (NoDaLiDa). May 11–13, Vilnius, Lithuania.*, band 23, ss 257–261, 2015.
 14. Helge Holzmann, Nina Tahmasebi och Thomas Risse. Named entity evolution recognition on the Blogosphere. *International Journal on Digital Libraries*, 15(2-4):209–235, 2015.
 15. Richard Johansson och Luis Nieto Piña. Combining relational and distributional knowledge for word sense disambiguation. I: *Proceedings of the 20th Nordic Conference of Computational Linguistics, May 12-13, Vilnius, Lithuania. Linköping Electronic Conference Proceedings 109, Linköping University Electronic Press*, ss 69–78, 2015.
 16. Richard Johansson och Luis Nieto Piña. Embedding a Semantic Network in a Word Space. I: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Denver, United States, May 31 – June 5*, ss 1428–1433, 2015.
 17. Sofie Johansson Kokkinakis. Datorbaserad bedömning av en- och flerspråkiga elevers läsförståelse. I: *Nordiske forskningskonferansen om nordiske språk som andre- og fremmedspråk, Hamar*, 2015.

18. Erik Kanebrant, Katarina Heimann Mühlenbock, Sofie Johansson Kokkinakis, Arne Jönsson, Caroline Liberg, Åsa af Geijerstam, Jenny Wiksten Folkeryd och Johan Falkenjack. T-MASTER – A tool for assessing students' reading abilities. I: *Proceedings of the 7th International Conference on Computer Supported Education (CSED), Vol. 1, Ed. by Markus Helfert. May 23-25, Lisbon, Portugal*, ss 220–227, 2015.
19. Dimitrios Kokkinakis, Ann Ighe och Mats Malm. Gender-Based Vocation Identification in Swedish 19th Century Prose Fiction using Linguistic Patterns, NER and CRF Learning. I: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature (Clfl). Co-located with the NAACL/HLT. Denver, Colorado, USA*, s 9, 2015.
20. Dimitrios Kokkinakis och Mats Malm. Detecting Reuse of Biblical Quotes in Swedish 19th Century Fiction using Sequence Alignment. I: *Corpus-based Research in the Humanities workshop (CRH)*, s 8, 2015.
21. Mikael Kågebäck, Fredrik Johansson, Richard Johansson och Devdatt Dubhashi. Neural context embeddings for automatic discovery of word senses. I: *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing. Denver, United States*, ss 25–32, 2015.
22. Luis Nieto Piña och Richard Johansson. A Simple and Efficient Method to Generate Word Sense Representations. I: Galia Angelova, Kalina Bontcheva, Ruslan Mitkov och Bulgaria 7-9 September Hissar, redaktörer, *Proceedings of International Conference in Recent Advances in Natural Language Processing*, ss 465–472, 2015.
23. Ildikó Pilán. Helping Swedish words come to their senses: word-sense disambiguation based on sense associations from the SALDO lexicon. I: Beáta Megyesi, redaktör, *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA). May 11–13, Vilnius, Lithuania*, nummer 109, ss 275–279, 2015.
24. Ildikó Pilán, Sowmya Vajjala och Elena Volodina. A readable read: Automatic Assessment of Language Learning Materials based on Linguistic Complexity. *Research in Computing Science*, 2015.
25. Taraka Rama. Automatic cognate identification with gap-weighted string subsequences. I: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, May 31 – June 5, Denver, Colorado, USA*, ss 1227–1231, 2015.
26. Taraka Rama och Lars Borin. Comparative evaluation of string similarity measures for automatic language classification. I: *Sequences in Language and Text*. De Gruyter Mouton, 2015.

14 *Språkbanken*

27. Nina Tahmasebi, Lars Borin, Gabriele Capannini, Devdatt Dubhashi, Peter Exner, Markus Forsberg, Gerhard Gossen, Fredrik Johansson, Richard Johansson, Mikael Kågebäck, Olof Mogren, Pierre Nugues och Thomas Risse. Visions and open challenges for a knowledge-based culturomics. *International Journal on Digital Libraries*, 15(2-4):169–187, 2015.
28. Elena Volodina, Lars Borin och Ildikó Pilán, redaktörer. *Proceedings of the fourth workshop on NLP for computer-assisted language learning at Nodalida, May, 11, Vilnius, Lithuania*. Linköping University Press, Linköping, 2015.
29. Elena Volodina och Dijana Pijetlovic. Lark Trills for Language Drills: Text-to-speech technology for language learners. I: *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications, June 4, Denver, Colorado, USA*, ss 107–117, 2015.