

Krister Lindén
Research Director
National Coordinator of FIN-CLARIN
Supervisor for doctoral programme



Language Technology
Department of Digital Humanities
Centre of Excellence in Ancient Near Eastern Empires (ANEE)

Postal address:
PL 24 (Unioninkatu 40)
00014
Finland

Email: krister.linden@helsinki.fi
Mobile: +358504154717
Phone: +358294129303

Qualifications

Principal Investigator

Period : 01.06.2010 - 31.12.2017 in

Principal Investigator

Period : 15.05.2018 - 31.12.2025 in Department of Digital Humanities

Employment

Research Director

Department of Digital Humanities
University of Helsinki
Finland
1 Jan 2018 → present

Title of Docent

Faculty Common Matters (Faculty of Arts)
University of Helsinki
Finland
1 Jan 2010 → present

Deputy Team Supervisor

Centre of Excellence in Ancient Near Eastern Empires (ANEE)
University of Helsinki
Finland
1 Jan 2018 → 31 Dec 2025

Research outputs

2022

Italian Language and Dialect Identification and Regional French Variety Detection using Adaptive Naive Bayes

Jauhiainen, T., Jauhiainen, H. & Lindén, K., 12 Oct 2022, *Proceedings of the Ninth Workshop on NLP for Similar Languages, Varieties and Dialects: The 29th International Conference on Computational Linguistics*. Scherrer, Y., Jauhiainen, T., Ljubešić, N., Nakov, P., Tiedemann, J. & Zampieri, M. (eds.). Gyeongju: COLING , p. 119-129 11 p. (International conference on computational linguistics; vol. 29, no. 19).

Language Identification as part of the Text Corpus Creation Pipeline at the Language Bank of Finland

Jauhiainen, T., Piitulainen, J., Axelson, E. & Linden, K., 6 Oct 2022, *Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022)*. Berglund, K., La Mela, M. & Zwart, I. (eds.). Aachen: CEUR, p. 251-259 9 p. (CEUR Workshop Proceedings; vol. 3232).

Donate Speech: Collecting and Sharing a Large-Scale Speech Database for Social Sciences, Humanities and Artificial Intelligence Research and Innovation

Lindén, K., Jauhiainen, T., Lennes, M., Kurimo, M., Rossi, A., Kurki, T. & Pitkänen, O., Oct 2022, *CLARIN: The Infrastructure for Language Resources*. Fišer, D. & Witt, A. (eds.). Berlin: de Gruyter, p. 481-510 30 p. (Digital Linguistics; vol. 1).

The CLARIN Committee for Legal and Ethical Issues and the Normative Layer of the CLARIN Infrastructure

Kamocki, P., Tartu, U. & Lindén, K., Oct 2022, *CLARIN: The Infrastructure for Language Resources*. Fišer, D. & Witt, A. (eds.). Berlin: de Gruyter, p. 457-479 23 p. (Digital Linguistics; vol. 1).

Digital Approaches to Analyzing and Translating Emotion: What Is Love?

Alstola, T., Jauhiainen, H., Svärd, S., Sahala, A. & Linden, K., 30 Aug 2022, *The Routledge Handbook of Emotions in the Ancient Near East*. Sonik, K. & Steinert, U. (eds.). London: Routledge, p. 88–116 29 p.

Donera prat för AI-utveckling till Språkbanken i Finland

Linden, K., Lennes, M., Jauhiainen, T., Kurimo, M. & Rossi, A., 18 Aug 2022, In: *Språk i Norden*. 52, 1, p. 35-50 16 p., 132412.

Lahjoita puhetta: a large-scale corpus of spoken Finnish with some benchmarks

Moisio, A., Porjazovski, D., Rouhe, A., Getman, Y., Virkkunen, A., AlGhezi, R., Lennes, M., Grósz, T., Lindén, K. & Kurimo, M., 9 Aug 2022, (E-pub ahead of print) In: *Language Resources and Evaluation*. 33 p.

Legal Issues Related to the Use of Twitter Data in Language Research

Kamocki, P., Hanneschläger, V., Hoorn, E., Kelli, A., Kupietz, M., Linden, K. & Puksas, A., 8 Jul 2022, *Selected papers from the CLARIN Annual Conference 2021*. Monachini, M. & Eskevich, M. (eds.). Linköping: Linköping University Electronic Press, p. 68-75 8 p. (Linköping Electronic Conference Proceedings ; vol. 89).

The Interaction of Personal Data, Intellectual Property and Freedom of Expression in the Context of Language Research

Kelli, A., Linden, K., Kamocki, P., Vider, K., Labropoulou, P., Birštonas, R., Mantrov, V., Hanneschläger, V., Del Gratta, R., Värvi, A., Tavits, G., Vutt, A., Hoorn, E., Hajic, J. & Language, I., 8 Jul 2022, *Selected papers from the CLARIN Annual Conference 2021*. Monachini, M. & Eskevich, M. (eds.). Linköping : Linköping University Electronic Press, p. 76-87 12 p. (Linköping Electronic Conference Proceedings; vol. 189).

Building a Chatbot: Challenges under Copyright and Data Protection Law

Kelli, A., Tavast, A. & Lindén, K., 30 Jun 2022, *Contracting and Contract Law in the Age of Artificial Intelligence*. Ebers, M., Poncibò, C. & Zou, M. (eds.). Oxford: Hart publishing, p. 115-134 20 p.

Report on the Finnish Language

Linden, K. & Dyster, W., 28 Feb 2022, *European Language Equality: Language Technology Support of Europe's Languages in 2020/2021*. Giagkou, M., Piperidis, S., Rehm, G. & Dunne, J. (eds.). Dublin: ELE Consortium, 24 p.

BabyLemmatizer: A Lemmatizer and POS-tagger for Akkadian

Sahala, A., Alstola, T., Valk, J. & Linden, K., 2022, *CLARIN Annual Conference Proceedings, 2022*. Erjavec, T. & Eskevich, M. (eds.). Utrecht: CLARIN ERIC, p. 14–18 5 p. (CLARIN Annual Conference Proceedings).

EU Data Governance Act: New Opportunities and New Challenges for CLARIN

Kamocki, P. & Linden, K., 2022, *CLARIN Annual Conference Proceedings, 2022*. Erjavec, T. & Eskevich, M. (eds.). Utrecht: CLARIN ERIC, p. 44-47 4 p. (CLARIN Annual Conference Proceedings).

HeLI-OTS, Off-the-shelf Language Identifier for Text

Jauhiainen, T., Jauhiainen, H. & Lindén, K., 2022, *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*. Calzolari, N., Béchet, F. & Blache, et al., P. (eds.). Paris: European Language Resources Association (ELRA), p. 3912-3922 11 p.

The Resource Publishing Pipeline of the Language Bank of Finland

Dieckmann, U., Lennes, M., Piitulainen, J., Niemi, J., Axelson, E., Jauhiainen, T. & Linden, K., 2022, *CLARIN Annual Conference Proceedings, 2022*. Erjavec, T. & Eskevich, M. (eds.). Utrecht: CLARIN ERIC, p. 53-56 4 p. (CLARIN Annual Conference Proceedings).

2021

A multilingual net of lexical resources

Lindén, K., Niemi, J., Borin, G., Forsberg, M., Sanford Pedersen, B., Nimb, S., Orav, H., Kahusk, N. & Vider, K., 15 Nov 2021, *The Swedish FrameNet++: Harmonization, integration, method development and practical language technology applications*. Dannélls, D., Borin, L. & Heppin, K. F. (eds.). Amsterdam: John Benjamins, p. 123-137 15 p. (Natural Language Processing; vol. 14).

Legal issues related to the use of Twitter data in language research

Kamocki, P., Hanneschläger, V., Hoorn, E., Kelli, A., Kupietz, M., Lindén, K. & Puksas, A., 27 Sep 2021, *CLARIN Annual Conference Proceedings 2021*. Monachini, M. & Eskevich, M. (eds.). Utrecht: CLARIN ERIC, p. 150-153 4 p. (CLARIN Annual Conference Proceedings).

The interplay of legal regimes of personal data, intellectual property and freedom of expression in language research

Kelli, A., Lindén, K., Kamocki, P., Vider, K., Labropoulou, P., Birštonas, R., Mantrov, V., Hanneschläger, V., Del Gratta, R., Värvi, A., Tavits, G. & Vutt, A., 27 Sep 2021, *Proceedings of CLARIN Annual Conference 2021*. Monachini, M. & Eskevich, M. (eds.). Utrecht: CLARIN ERIC, p. 154-159 6 p. (CLARIN Annual Conference).

Sharing is caring: a legal perspective on sharing language data containing personal data and the division of liability between researchers and research organisations

Kelli, A., Lindén, K., Vider, K., Kamocki, P., Tavast, A., Birštonas, R., Tavits, G., Keskküla, M., Labropoulou, P., Kull, I., Värvi, A., Erikson, M., Vutt, A. & Calamai, S., 22 Jun 2021, *Selected Papers from the CLARIN Annual Conference 2020: Virtual Event, 2020, 5-7 October*. Navarretta, C. & Eskevich, M. (eds.). Utrecht: CLARIN ERIC, p. 129-147 19 p. (Linköping Electronic Conference Proceedings; no. 180).

Findings of the VarDial Evaluation Campaign 2021

Chakravarthi, B. R., Gaman, M., Ionescu, R. T., Jauhiainen, H., Jauhiainen, T., Linden, K., Ljubešić, N., Partanen, N., Priyadharshini, R., Purschke, C., Rajagopal, E., Scherrer, Y. & Zampieri, M., Apr 2021, *Proceedings of the 8th VarDial Workshop on NLP for Similar Languages, Varieties and Dialects*. Zampieri, M., Nakov, P., Ljubešić, N., Tiedemann, J., Scherrer, Y. & Jauhiainen, T. (eds.). Stroudsburg: The Association for Computational Linguistics, p. 1-11 11 p.

Fear in Akkadian Texts: New Digital Perspectives on Lexical Semantics

Svärd, S., Alstola, T., Jauhiainen, H., Sahala, A. & Linden, K., 2021, *The Expression of Emotions in Ancient Egypt and Mesopotamia*. Hsu, S-W. & Llop-Radua, J. (eds.). Leiden: Brill, p. 470-502 33 p. (Culture and History of the Ancient Near East; vol. 116).

Isikuandmeid sisaldavate keeleandmete jagamisega seonduv õiguslik raamistik: Teadlase ja teadusasutuse kohustused ning vastutus

Kelli, A., Vider, K., Tavast, A., Lindén, K., Kamocki, P., Birštonas, R., Labropoulou, P., Calamai, S., Tavits, G., Keskküla, M., Kull, I., Erikson, M., Ginter, C., Värvi, A. & Vutt, A., 2021, In: *Eesti Rakenduslingvistika Uingu Aastaraamat*. 17, p. 99-121 23 p.

Naive Bayes-based Experiments in Romanian Dialect Identification

Jauhiainen, T., Jauhiainen, H. & Linden, K., 2021, *Proceedings of the 8th VarDial Workshop on NLP for Similar Languages, Varieties and Dialects*. Zampieri, M., Nakov, P., Ljubešić, N., Tiedemann, J., Scherrer, Y. & Jauhiainen, T. (eds.). Stroudsburg: ACL Anthology, p. 76-83 8 p.

Naive Bayes-based Experiments in Romanian Dialect Identification

Jauhiainen, T., Jauhiainen, H. & Linden, K., 2021.

Suomalais-ugrilaiset kielet ja internet -projekti 2013-2019

Jauhiainen, T., Jauhiainen, H. & Linden, K., 2021, *Multilingual Facilitation*. Hämäläinen, M., Partanen, N. & Alnajjar, K. (eds.). Helsinki: University of Helsinki

2020

A Report on the VarDial Evaluation Campaign 2020

Gaman, M., Hovy, D., Ionescu, R. T., Jauhiainen, H., Jauhiainen, T., Linden, K., Ljubešić, N., Partanen, N., Purschke, C., Scherrer, Y. & Zampieri, M., 13 Dec 2020, *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*. Zampieri, M., Nakov, P., Ljubešić, N., Tiedemann, J. & Scherrer, Y. (eds.). International Committee on Computational Linguistics, p. 1-14

Optical character recognition with neural networks and post-correction with finite state methods

Drobac, S. & Linden, K., Dec 2020, In: *International Journal on Document Analysis and Recognition*. 23, 4, p. 279-295 17 p.

Improving Word Association Measures in Repetitive Corpora with Context Similarity Weighting

Sahala, A. & Lindén, K., Nov 2020, *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2020, Volume 1: KDIR, Budapest, Hungary, November 2-4, 2020*. Fred, A. L. N. & Filipe, J. (eds.). Setúbal: SCITEPRESS Science And Technology Publications, p. 48-58 11 p.

Akkadian Treebank for early Neo-Assyrian Royal Inscriptions

Luukko, M., Sahala, A., Hardwick, S. & Lindén, K., 27 Oct 2020, *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*. Evang, K., Kallmeyer, L., Ehren, R., Petitjean, S., Seyffarth, E. & Seddah, D. (eds.). Stroudsburg: The Association for Computational Linguistics, p. 124-134 11 p.

CLARIN contractual framework for sharing language data: the perspective of personal data protection

Kelli, A., Linden, K., Vider, K., Kamocki, P., Tavast, A., Birštonas, R., Tavits, G., Keskküla, M. & Labropoulou, P., 5 Oct 2020, *Proceedings of CLARIN Annual Conference 2020. 05 – 07 October 2020, Online Edition*. Navarretta, C. & Eskevich, M. (eds.). Utrecht: CLARIN ERIC, p. 171-177 7 p.

A CLARIN Contractual Framework for Sharing Personal Data for Scientific Research

Linden, K., Kelli, A. & Nousias, A., 3 Jul 2020, *Selected Papers from the CLARIN Annual Conference 2019*. Simov, K. & Eskevich, M. (eds.). Linköping: Linköping University Electronic Press, p. 75–84 10 p. 10. (Linköping Electronic Conference Proceedings; no. 172).

The Impact of Copyright and Personal Data Laws on the Creation and Use of Models for Language Technologies

Kelli, A., Tavast, A., Lindén, K., Vider, K., Birštonas, R., Labropoulou, P., Kull, I., Tavits, G., Värvi, A., Stranák, P. & Hajic, J., 3 Jul 2020, *Selected Papers from the CLARIN Annual Conference 2019*. Simov, K. & Eskevich, M. (eds.). Linköping: Linköping University Electronic Press, p. 53-65 13 p. (Linköping Electronic Conference Proceedings; no. 172).

Automated Phonological Transcription of Akkadian Cuneiform Text

Sahala, A., Silfverberg, M., Arppe, A. & Linden, K., 17 May 2020, *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. Calzolari ... [et al.], N. (ed.). Paris: European Language Resources Association (ELRA), p. 3528-3534 7 p.

BabyFST: Towards a Finite-State Based Computational Model of Ancient Babylonian

Sahala, A., Silfverberg, M., Arppe, A. & Linden, K., 17 May 2020, *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. Calzolari ... [et al.], N. (ed.). Paris: European Language Resources Association (ELRA), p. 3886-3894 9 p.

The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe

Rehm, G., Marheinecke, K., Hegele, S., Piperidis, S., Bontcheva, K., Hajic, J., Choukri, K., Vasiljevs, A., Backfried, G., Prinz, C., Gomez-Perez, J. M., Meertens, L., Lukowicz, P., van Genabith, J., Löscher, A., Slusallek, P., Irgens, M., Gatellier, P., Köhler, J., Le Bars, L. & 27 others, Anastasiou, D., Auksoirüte, A., Bel, N., Branco, A., Budin, G., Daelemans, W., De

Smedt, K., Garabík, R., Gavriilidou, M., Gromann, D., Koeva, S., Krek, S., Krstev, C., Lindén, K., Magnini, B., Odijk, J., Ogrodniczuk, M., Rögnvaldsson, E., Rosner, M., Pedersen, B., Skadina, I., Tadić, M., Tufi, D., Váradi, T., Vider, K., Way, A. & Yvon, F., 1 May 2020, *Proceedings of The 12th Language Resources and Evaluation Conference*. Calzolari [et al.], N. (ed.). Paris: European Language Resources Association (ELRA), p. 3322-3332 11 p.

A Finnish news corpus for named entity recognition

Ruokolainen, T., Kauppinen, P., Silfverberg, M. & Lindén, K., Mar 2020, In: *Language Resources and Evaluation*. 54, 1, p. 247-272 26 p.

Building Web Corpora for Minority Languages

Jauhiainen, H., Jauhiainen, T. & Linden, K., 2020, *Proceedings of the 12th Web as Corpus Workshop*. Barbaresi, A., Bildhauer, F., Schäfer, R. & Stemle, E. (eds.). Stroudsburg: The Association for Computational Linguistics, p. 23-32 10 p.

Compatibility of business models of Estonian language technology entrepreneurs with regulatory framework on the use of digital content

Kull, I., Vider, K., Mets, T., Kelli, A. & Linden, K., 2020, In: *International Comparative Jurisprudence*. 6, 1, p. 70–83 14 p.

Experiments in Language Variety Geolocation and Dialect Identification

Jauhiainen, T., Jauhiainen, H. & Linden, K., 2020, *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*. Zampieri, M., Nakov, P., Ljubešić, N., Tiedemann, J. & Scherrer, Y. (eds.). International Committee on Computational Linguistics (ICCL), p. 220-231 12 p.

Impact of Legal Status of Data on Development of Data-Intensive Products: Example of Language Technologies

Kelli, A., Tavast, A., Linden, K., Bristonas, R., Labropoulou, P., Vider, K., Kull, I., Tavits, G., Värvi, A. & Mantrov, V., 2020, *Legal Science: Functions, Significance and Future in Legal Systems II : Collection of Research Papers in Conjunction with the 7th International Scientific Conference of the Faculty of Law of the University of Latvia (16–18 October 2019, Riga)*. Damberg, A. (ed.). Riga: University of Latvia press, p. 383–400 18 p.

Keeleandmete õigusliku režiimi mõju nende abil loodud keelemudelitele

Kelli, A., Vider, K., Tavast, A., Lindén, K., Birštonas, R., Labropoulou, P., Värvi, A., Kull, I., Tavits, G. & Ginter, C., 2020, In: *Eesti rakenduslingvistika ühingu aastaraamat*. 16, p. 59-76 18 p.

Uralic Language Identification (ULI) 2020 shared task dataset and the Wanca 2017 corpora

Jauhiainen, T., Jauhiainen, H., Partanen, N. & Linden, K., 2020, *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*. Zampieri, M., Nakov, P., Ljubešić, N., Tiedemann, J. & Scherrer, Y. (eds.). International Committee on Computational Linguistics (ICCL), p. 173–185 13 p.

Vestlusrobotid ja autoriõigus

Kelli, A., Tavast, A. & Linden, K., 2020, In: *Juridica*. 5, p. 345–354 10 p.

2019

The Extent of Legal Control over Language Data: the Case of Language Technologies

Kelli, A., Tavast, A., Linden, K., Vider, K., Birštonas, R., Labropoulou, P., Kull, I., Tavits, G. & Värvi, A., 30 Sep 2019, *Proceedings of CLARIN Annual Conference 2019*. Simov, K. & Eskevich, M. (eds.). Utrecht: CLARIN ERIC, p. 69-74 6 p.

To Ask or not to Ask: Informed Consent to Participate and Using Data in the Public Interest

Linden, K., Kelli, A. & Nousias, A., 30 Sep 2019, *Proceedings of CLARIN Annual Conference 2019*. Simov, K. & Eskevich, M. (eds.). Leipzig: CLARIN ERIC, p. 56-60 5 p.

Language Model Adaptation for Language and Dialect Identification of Text

Jauhiainen, T. S., Linden, B. K. J. & Jauhiainen, H. A., Sep 2019, In: *Natural Language Engineering*. 25, 5, p. 561-583 23 p., 135132491900038.

Automatic Language Identification in Texts: A Survey

Jauhiainen, T., Lui, M., Zampieri, M., Baldwin, T. & Lindén, K., 25 Aug 2019, In: Journal of Artificial Intelligence Research. 65, p. 675-782 108 p.

Processing personal data without the consent of the data subject for the development and use of language resources

Kelli, A., Lindén, K., Vider, K., Kamocki, P., Birštonas, R., Calamai, S., Labropoulou, P., Gavriilidou, M. & Stranák, P., 28 May 2019, *Selected papers from the CLARIN Annual Conference 2018, Pisa, 8-10 October 2018*. Skadina, I. & Eskevich, M. (eds.). Linköping: Linköping University Electronic Press, p. 72-82 11 p. (Linköping Electronic Conference Proceedings; no. 159).

Discriminating between Mandarin Chinese and Swiss-German varieties using adaptive language models

Jauhiainen, T. S., Jauhiainen, H. A. & Linden, B. K. J., 30 Apr 2019, *Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2019)*. Stroudsburg: The Association for Computational Linguistics, p. 178-187 10 p.

Language and Dialect Identification of Cuneiform Texts

Jauhiainen, T. S., Jauhiainen, H. A., Alstola, T. & Linden, B. K. J., 30 Apr 2019, *Proceedings of the Sixth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2019)*. Stroudsburg: The Association for Computational Linguistics, p. 89-98 10 p.

Aššur and His Friends: A Statistical Analysis of Neo-Assyrian Texts

Alstola, T., Zaia, S., Sahala, A., Jauhiainen, H., Svärd, S. & Linden, K., 2019, In: Journal of Cuneiform Studies. 71, p. 159-180 22 p.

FinnTransFrame: translating frames in the FinnFrameNet project

Lindén, K., Haltia, H., Laine, A., Luukkonen, J., Piitulainen, J. & Väisänen, N., 2019, In: Language Resources and Evaluation. 53, 1, p. 141-171 31 p.

Improving OCR of historical newspapers and journals published in Finland

Drobac, S., Kauppinen, P. & Linden, K., 2019, *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*. New York: ACM, p. 97-102 6 p.

Optical Font Family Recognition using a Neural Network

Drobac, S. & Lindén, K., 2019, *Proceedings of the Research Data And Humanities (Rdhum) 2019 Conference: Data, Methods And Tools*. Jantunen, J. H., Brunni, S., Kunnas, N., Palviainen, S. & Västi, K. (eds.). Oulu: Studia Humaniora Ouluensia, p. 115-123 9 p. (Studia humaniora ouluensia; no. 17).

Wanca in Korp: Text corpora for underresourced Uralic languages

Jauhiainen, H., Jauhiainen, T. & Linden, K., 2019, *Proceedings of the Research data and humanities (RDHUM) 2019 conference : data, methods and tools*. Jantunen, J. H., Brunni, S., Kunnas, N., Palviainen, S. & Västi, K. (eds.). Oulu: University of Oulu, p. 21-40 20 p. (Studia Humaniora Ouluensia; no. 17).

2018

Vad får AI för bild av mänskligheten?

Linden, B. K. J., 26 Nov 2018, In: Yliopistolainen : Helsingin yliopiston henkilöstölehti. 4/2018, p. 9 1 p.

Challenges of transformation of research data into open data: The perspective of social sciences and humanities

Kelli, A., Mets, T., Vider, K., Värvi, A., Jonsson, L., Lindén, K. & Birštonas, R., Sep 2018, In: International Journal of Technology Management & Sustainable Development. 17, 3, p. 227-251 25 p.

Semantic Domains in Akkadian Text

Svärd, S. S., Jauhiainen, H. A., Linden, B. K. J. & Sahala, A. J. A., 7 Aug 2018, *CyberResearch on the Ancient Near East and Neighboring Regions: Case Studies on Archaeological Data, Objects, Texts, and Digital Archiving*. Juloux, V. B., Gansell, A. R. & di Ludovico, A. (eds.). Leiden: Brill, p. 224-256 33 p. (Digital Biblical Studies; no. 2).

HeLI-based Experiments in Discriminating Between Dutch and Flemish Subtitles

Jauhiainen, T. S., Jauhiainen, H. A. & Linden, B. K. J., Aug 2018, *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*. Zampieri, M., Nakov, P., Ljubešić, N., Tiedemann, J., Malmasi, S. & Ali, A. (eds.). Santa Fe: The Association for Computational Linguistics, p. 137-144 8 p.

HeLI-based Experiments in Swiss German Dialect Identification

Jauhiainen, T. S., Jauhiainen, H. A. & Linden, B. K. J., Aug 2018, *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*. Zampieri, M., Nakov, P., Ljubešić, N., Tiedemann, J., Malmasi, S. & Ali, A. (eds.). Santa Fe: The Association for Computational Linguistics, p. 254-262 9 p.

Iterative Language Model Adaptation for Indo-Aryan Language Identification

Jauhiainen, T. S., Jauhiainen, H. A. & Linden, B. K. J., Aug 2018, *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*. Zampieri, M., Nakov, P., Ljubešić, N., Tiedemann, J., Malmasi, S. & Ali, A. (eds.). Santa Fe: The Association for Computational Linguistics, p. 66-75 10 p.

The Dagstuhl Perspectives Workshop on Performance Modeling and Prediction

Ferro, N., Fuhr, N., Grefenstette, G., Konstan, J. A., Castells, P., Daly, E. M., Declerck, T., Ekstrand, M. D., Geyer, W., Gonzalo, J., Kuflik, T., Lindén, K., Magnini, B., Nie, J.-Y., Perego, R., Shapira, B., Soboroff, I., Tintarev, N., Verspoor, K., Willemsen, M. C. & 1 others, Zobel, J., Jun 2018, In: *SIGIR Forum*. 52, 1, p. 91-101 11 p.

Implementation of an Open Science Policy in the context of management of CLARIN language resources: a need for changes?

Kelli, A., Lindén, K., Vider, K., Labropoulou, P., Ketzan, E., Kamocki, P. & Stranák, P., 16 May 2018, *Selected papers from the CLARIN Annual Conference 2017, Budapest, 18–20 September 2017*. Piasecki, M. (ed.). Linköping: Linköping University Electronic Press, Vol. 147. p. 102-111 10 p. 009. (Linköping Electronic Conference Proceedings; no. 147).

Tieteen kansallinen termipankki (TTP) ja tiedonlouhinnan apuneuvot

Hardwick, S., Enqvist, E. J., Onikki-Rantajääskö, T. A. & Linden, B. K. J., 7 May 2018. 1 p.

Rethinking Summarization and Storytelling for Modern Social Multimedia

Rudinac, S., Chua, T.-S., Diaz-Ferreyra, N., Friedland, G., Gornostaja, T., Huet, B., Kaptein, R., Lindén, K., Moens, M.-F., Peltonen, J., Redi, M., Schedl, M., Shamma, D. A., Smeaton, A. & Xie, L., 28 Jan 2018, *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I*. Schoeffmann, K., Chalidabhongse, T. H., Ngo, C. W., Aramvith, S., O'Connor, N. E., Ho, Y.-S., Gabbouj, M. & Elgammal, A. (eds.). Cham: Springer International Publishing AG, Vol. 1. p. 632-644 13 p. (LNCS; no. 10704).

From Evaluating to Forecasting Performance: How to Turn Information Retrieval, Natural Language Processing and Recommender Systems into Predictive Sciences

Ferro, N., Fuhr, N., Grefenstette, G., Konstan, J. A., Castells, P., Daly, E. M., Declerck, T., Ekstrand, M. D., Geyer, W., Gonzalo, J., Kuflik, T., Lindén, K., Magnini, B., Nie, J.-Y., Perego, R., Shapira, B., Soboroff, I., Tintarev, N., Verspoor, K., Willemsen, M. C. & 1 others, Zobel, J., 2018, In: *Dagstuhl manifestos*. 7, 1, p. 96-139 44 p.

Keeleressursside loomise ja kasutamise seonduvaid isikuandmete kaitse küsimusi

Kelli, A., Vider, K., Kull, I., Siil, T., Lindén, K., Tavast, A., Värvi, A., Ginter, C. & Meister, E., 2018, In: *Eesti rakenduslingvistika ühingu aastaraamat*. 14, p. 77-94 18 p.

Rethinking Summarization and Storytelling for Modern Social Multimedia

Rudinac, S., Chua, T.-S., Diaz-Ferreyra, N., Friedland, G., Gornostaja, T., Huet, B., Kaptein, R., Linden, K., Moens, M.-F., Peltonen, J., Redi, M., Schedl, M., Shamma, D. A., Smeaton, A. & Xie, L., 2018, p. 141-153. 13 p.

2017

FinnFN 1.0: The Finnish frame semantic database

Lindén, K., Haltia, H., Luukkonen, J., Laine, A., Roivainen, H. & Väisänen, N., 14 Aug 2017, In: *Nordic journal of linguistics*. 40, 3, p. 1-25 25 p.

Evaluating HeLI with non-linear mappings

Jauhiainen, T. S., Linden, B. K. J. & Jauhiainen, H. A., 2017, *Fourth Workshop on NLP for Similar Languages, Varieties and Dialects - Proceedings of the Workshop*. Stroudsburg: The Association for Computational Linguistics, p. 102-108 7 p.

Evaluation of language identification methods using 285 languages

Jauhiainen, T. S., Linden, B. K. J. & Jauhiainen, H. A., 2017, *21st Nordic Conference of Computational Linguistics: Proceedings of the Conference*. Tiedemann, J. (ed.). Linköping: Linköping University Electronic Press, p. 183-191 9 p. (Linköping Electronic Conference Proceedings; no. 31).

OCR and post-correction of historical Finnish texts

Drobac, S., Kauppinen, P. S. & Linden, B. K. J., 2017, *Proceedings of the 21st Nordic Conference on Computational Linguistics, NoDaLiDa, 22-24 May 2017, Gothenburg, Sweden*. Tiedemann, J. (ed.). Linköping: Linköping University Electronic Press, p. 70-76 7 p. (Linköping Electronic Conference Proceedings; no. 131).

2016

FinnPos: an open-source morphological tagging and lemmatization toolkit for Finnish

Silfverberg, M., Ruokolainen, T., Linden, K. & Kurimo, M., Dec 2016, In: *Language Resources and Evaluation*. 50, 4, p. 863-878 16 p.

In-Document Adaptation for a Human Guided Automatic Transcription Service

Mansikkaniemi, A., Kurimo, M. & Linden, B. K. J., 13 Aug 2016, *Speech and Computer : 18th International Conference, SPECOM 2016, Budapest, Hungary, August 23-27, 2016, Proceedings*. Ronzhin, A., Potapova, R. & Németh, G. (eds.). Cham: Springer International Publishing AG, p. 395-402 8 p. (Lecture Notes in Computer Science; vol. 9811).

Data-Driven Spelling Correction using Weighted Finite-State Methods

Silfverberg, M., Kauppinen, P. & Linden, K., 12 Aug 2016, *The 54th Annual Meeting of the Association for Computational Linguistics: Proceedings of the SIGFSM Workshop on Statistical NLP and Weighted Automata*. Stroudsburg, PA: ACL, p. 51-59 9 p.

The strategic impact of META-NET on the regional, national and international level

Rehm, G., Uszkoreit, H., Ananiadou, S., Bel, N., Bieleveciene, A., Borin, L., Branco, A., Budin, G., Calzolari, N., Daelemans, W., Garabik, R., Grobelnik, M., Garcia-Mateo, C., van Genabith, J., Hajic, J., Hernaez, I., Judge, J., Koeva, S., Krek, S., Krstev, C. & 24 others, Linden, K., Magnini, B., Mariani, J., McNaught, J., Melero, M., Monachini, M., Moreno, A., Odijk, J., Ogrodniczuk, M., Pezik, P., Piperidis, S., Przepiorkowski, A., Rognvaldsson, E., Rosner, M., Pedersen, B. S., Skadina, I., De Smedt, K., Tadic, M., Thompson, P., Tufis, D., Varadi, T., Vasiljevs, A., Vider, K. & Zabarskaite, J., Jun 2016, In: *Language Resources and Evaluation*. 50, 2, p. 351-374 24 p.

Eduskunnan täysistuntovideoiden ja keskustelupöytäkirjojen automaattinen kohdistaminen AaltoASR-työkalulla

Mansikkaniemi, A., Kurimo, M., Lennes, M. E. & Linden, B. K. J., 23 May 2016.

The Regulatory and Contractual Framework as an Integral Part of the CLARIN Infrastructure

Kelli, A., Vider, K. & Lindén, K., 11 Apr 2016, *Selected Papers from the CLARIN Annual Conference 2015, October 14–16, 2015, Wrocław, Poland: NEALT Proceedings Series*. DeSmedt, K. (ed.). Linköping: Linköping University Electronic Press, Vol. 123. p. 13-24 12 p. 002. (Linköping Electronic Conference Proceedings; vol. 123).

Digitaalsete keeleressursside loomist ja kasutamist määrav õiguslik raamistik Eestis ja selle ühildumine CLARIN-i infrastruktuuriga

Kelli, A., Vider, K., Pisuke, H. & Linden, B. K. J., 2016, In: *Eesti rakenduslingvistika ühingu aastaraamat*. 12, p. 81-98 18 p.

HeLI, a Word-Based Backoff Method for Language Identification

Jauhiainen, T. S., Linden, B. K. J. & Jauhiainen, H. A., 2016, *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects: VarDial3, Osaka, Japan, December 12 2016*. p. 153-162 10 p.

2015

Using HFST—Helsinki Finite-State Technology for Recognizing Semantic Frames

Lindén, K., Hardwick, S., Silfverberg, M. & Axelson, E., 9 Dec 2015, *Systems and Frameworks for Computational Morphology*. Springer International Publishing AG, p. 124-136 13 p. (Communications in Computer and Information Science; vol. 537).

Regulatory and Contractual Framework as an Integral Part of CLARIN Infrastructure: the Estonian and Finnish Perspectives

Kelli, A., Vider, K. & Linden, K., Oct 2015, *CLARIN Annual Conference 2015: Book of Abstracts*. De Smedt, K. (ed.). Wrocław, Poland: CLARIN ERIC, p. 32-36 5 p.

Extracting Semantic Frames using hfst-pmatch

Hardwick, S., Silfverberg, M. & Linden, K., May 2015, *Proceedings of the 20th Nordic Conference of Computational Linguistics: NODALIDA 2015*. Magyesi, B. (ed.). Linköping: Linköping University Electronic Press, p. 305-308 4 p. 109:042. (Linköping Electronic Conference Proceedings).

Automated Lossless Hyper-Minimization for Morphological Analyzers

Drobac, S., Silfverberg, M. & Linden, K., 22 Apr 2015, *Proceedings of the 12th International Conference on Finite-State Methods and Natural Language Processing 2015: Collected papers*. Hanneforth, T. & Wurm, C. (eds.). Stroudsburg: ACL, 5 p.

The Finno-Ugric Languages and the Internet project

Jauhiainen, H., Jauhiainen, T. & Linden, K., 15 Jan 2015, *First International Workshop on Computational Linguistics for Uralic Languages: Proceedings of the Workshop*. Pirinen, T., Tyers, F. & Trosterud, T. (eds.). Tromsø: Septentrio Academic Publishing, Vol. 2. p. 87–98 12 p. (Septentrio Conference Series; vol. 2015, no. 2).

Discriminating similar languages with token-based backoff

Jauhiainen, T., Jauhiainen, H. & Linden, K., 2015, *Proceedings of the Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects*. The Association for Computational Linguistics, p. 44-51 8 p.

Language Set Identification in Noisy Synthetic Multilingual Documents

Jauhiainen, T. S., Linden, K. & Jauhiainen, H. A., 2015, *Computational Linguistics and Intelligent Text Processing*. Gelbukh, A. (ed.). Springer International Publishing AG, Vol. Part I. p. 633-643 11 p. (Lecture Notes in Computer Science; vol. 9041).

Tutkimuksen haasteet digitaalisessa ajassa

Linden, K., 2015, In: *Kansalliskirjasto*. 57, p. 41-42 2 p.

2014

Analyzing and Improving the Quality of a Historical News Collection using Language Technology and Statistical Machine Learning Methods

Kettunen, K., Honkela, T., Linden, K., Kauppinen, P., Pääkkönen, T. & Kervinen, J., 16 Aug 2014, *IFLA World Library and Information Congress Proceedings: 80th IFLA General Conference and Assembly*. Lyon, France: IFLA, 23 p.

FIN-CLARIN – en humanistisk forskningsinfrastruktur med betoning på språk

Linden, B. K. J., Jul 2014, In: *Språk i Norden = Sprog i Norden : årsskrift for Nordisk språkråd og språknemdene i Norden*. p. 126-132 7 p.

Part-of-Speech Tagging using Conditional Random Fields: Exploiting Sub-Label Dependencies for Improved Accuracy

Silfverberg, M., Ruokolainen, T., Linden, K. & Kurimo, M., 22 Jun 2014, *Unknown host publication*.

CLARA: A New Generation of Researchers in Common Language Resources and Their Applications

Koenraad De Smedt, Erhard Hinrichs, Detmar Meurers, Inguna Skadina, Bolette Pedersen, Costanza Navarretta, Núria Bel, Krister Linden, Marketa Lopatkova, Jan Hajic, Gisle andersen and Przemyslaw Lenkiewicz, 26 May 2014, *Proceedings of LREC 2014*. Reykjavik, Iceland: European Language Resources Association (ELRA), 9 p. #410

HFST-SweNER – A New NER Resource for Swedish

Kokkinakis, D., Niemi, J., Hardwick, S., Linden, K. & Borin, L., 26 May 2014, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J. & Piperidis, S. (eds.). Reykjavik, Iceland: European Language Resources Association (ELRA), 7 p. #391

Heuristic Hyper-minimization of Finite State Lexicons

Drobac, S., Linden, K., Pirinen, T. & Silfverberg, M., 26 May 2014, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA), Vol. 9. 6 p. #784

The Strategic Impact of META-NET on the Regional, National and International Level

Georg Rehm, Hans Uszkoreit, Sophia Ananiadou, Núria Bel, Audrone Bielevičienė, 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 Tufi, Tamás Váradi, Andrejs Vasiljevs, Kadri Vider, Jolanta Zabarskaite, 26 May 2014, *Proceedings of LREC 2014*. Reykjavik, Iceland: European Language Resources Association (ELRA), p. 1517-1524 8 p. #405

Accelerated Estimation of Conditional Random Fields using a Pseudo-Likelihood-inspired Perceptron Variant

Ruokolainen, T., Silfverberg, M., Kurimo, M. & Linden, K., 26 Apr 2014, *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics: EACL 2014*. ACL, 5 p.

State-of-the-Art in Weighted Finite-State Spell-Checking

Pirinen, T. & Linden, K., 6 Apr 2014, *Computational Linguistics and Intelligent Text Processing: 15th International Conference, CICLing 2014, Kathmandu, Nepal, April 6-12, 2014, Proceedings, Part II*. Gelbukh, A. (ed.). Berlin Heidelberg: Springer-Verlag, Vol. 2. p. 519-532 14 p. (Lecture Notes in Computer Science; vol. 8404).

2013

Software Newsroom – an approach to automation of news search and editing

Huovelin, J., Gross, O., Solin, O., Linden, K., Maisala, S. P. T., Oittinen, T., Toivonen, H., Niemi, J. & Silfverberg, M., 7 Nov 2013, In: *Journal of Print Media Technology research*. 2, 3, p. 141-156 15 p.

HFST—a System for Creating NLP Tools

Linden, K., Axelson, E., Drobac, S., Hardwick, S., Kuokkala, J., Niemi, J., Pirinen, T. & Silfverberg, M., Sep 2013, *Systems and Frameworks for Computational Morphology: Communications in Computer and Information Science*. Mahlow, C. & Piotrowski, M. (eds.). Springer-Verlag, p. 53-71 20 p. (Communications in Computer and Information Science).

Baltic and Nordic Parts of the European Linguistic Infrastructure

Skadina, I., Vasiljevs, A., Borin, L., Linden, K., Losnegaard, G., Olsen, S., Pedersen, B., Rozis, R. & De Smedt, K., 20 May 2013, *Proceedings of NODALIDA 2013*. Linköping University Electronic Press, 16 p.

Nordic and Baltic wordnets aligned and compared through “WordTies”

Pedersen, B., Linden, K., Vider, K., Forsberg, M., Kahusk, N., Niemi, J., Nygaard, L., Seaton, M., Orav, H., Borin, L., Voionmaa, K., Nisbeth, N. & Rögnvaldsson, E., 20 May 2013, *Proceedings of NODALIDA 2013*. Linköping University Press, 16 p. (Linköping Electronic Conference Proceedings; no. 85).

Is It Possible to Create a Very Large WordNet in 100 days? – an Evaluation

Linden, K. & Niemi, J., 2013, In: *Language Resources and Evaluation*. 48, 2, p. 191-201 10 p.

2012

Building shared language research environments inside the European Union: How to optimize the system based on experiences from real life

Oksanen, V. & Linden, K., 12 Sep 2012, *Unknown host publication*. 12 p.

Extending and Updating the Finnish Wordnet

Linden, K., Niemi, J. & Hyvärinen, M., 14 Aug 2012, *Shall We Play the Festschrift Game? Essays on the Occasion of Lauri Carlson's 60th Birthday*. Springer-Verlag, p. 67-98 32 p.

Shall We Play the Festschrift Game? Essays on the Occasion of Lauri Carlson's 60th Birthday

Sousa, D., Linden, K. & Nganga, W., 14 Aug 2012, Springer-Verlag. 241 p.

Creation of Shared Language Resource Repository in the Nordic and Baltic Countries

Vasiljevs, A., Forsberg, M., Goronstay, T., Hansen, D., Jóhannsdóttir, K., Linden, K., Lyse, G., Offersgaard, L., Olsen, S., Oksanen, V., Pedersen, B., Rögnvaldsson, E., Skadina, I. & de Smedt, K., 23 May 2012, *The eighth international conference on Language Resources and Evaluation (LREC): LREC 2012*. Istanbul, Vol. 8. 8 p.

Representing the Translation Relation in a Bilingual Wordnet

Niemi, J. & Linden, K., 23 May 2012, *The eighth international conference on Language Resources and Evaluation (LREC): LREC 2012*. Istanbul: European Language Resources Association (ELRA), Vol. 8. 8 p.

Specifying Treebanks, Outsourcing Parsebanks: FinnTreeBank 3

Voutilainen, A., Muhonen, K., Purtonen, T. K. & Linden, K., 23 May 2012, *Proceedings of LREC 2012: 8th ELRA Conference on Language Resources and Evaluation*. Calzolari, N., Choukri, K., Declerck, T., Dogan, M. U., Maegaard, B., Mariani, J., Odijk, J. & Piperidis, S. (eds.). Paris: European Language Resources Association (ELRA), p. 1927-1931 5 p.

Improving Finite-State Spell-Checker Suggestions with Part of Speech N-Grams

Pirinen, T., Silfverberg, M. & Linden, K., 9 Mar 2012, *Computational Linguistics and Intelligent Text Processing: 13th International Conference, CICLing 2012*. Gelbukh, A. (ed.). Delhi, India, 11 p.

Predictive Text Entry for Agglutinative Languages Using Unsupervised Morphological Segmentation

Silfverberg, M., Linden, K. & Hyvärinen, M., 9 Mar 2012, In: *Lecture Notes in Computer Science*. 12 p.

FinnTreeBank 2

Voutilainen, A., Purtonen, T. K., Muhonen, K., Hardwick, S. & Linden, K., Feb 2012

Finding a Location for a New Word in WordNet

Pääkkö, P. & Linden, K., 9 Jan 2012, *Proceedings of the Global Wordnet Conference*. Matsue, Japan, 8 p.

Linking and Validating Nordic and Baltic Wordnets - A Multilingual Action in META-NORD

Sanford Pedersen, B., Borin, L., Forsberg, M., Linden, K., Orav, H. & Rögnvaldsson, E., 9 Jan 2012, *Proceedings of the Global Wordnet Conference*. Matsue, Japan, 5 p.

Using a Bilingual Resource to Add Synonyms to a Wordnet: FinnWordNet and Wikipedia as an Example

Niemi, J., Linden, K. & Hyvärinen, M., 9 Jan 2012, *Proceedings of the Global Wordnet Conference*. Matsue, Japan, p. 227-231 5 p.

HFST - an Environment for Creating Language Technology Applications

Linden, K., Silfverberg, M., Pirinen, T., Hardwick, S., Drobac, S. & Axelson, E., 2012, *COMPUTATIONAL LINGUISTICS – APPLICATIONS*. Piasecki, M. & Przepiórkowski, A. (eds.). Springer-Verlag, 20 p. (Studies in Computational Intelligence).

Suomen kieli digitaalisella aikakaudella

Koskenniemi, K., Linden, K., Carlson, L., Vainio, M., Arppe, A., Lennes, M., Westerlund, H., Hyvärinen, M., Bartis, I., Nuolijärvi, P. & Piehl, A., 2012, Berlin: Springer-Verlag. 85 p.

2011

META-NORD: Towards Sharing of Language Resources in Nordic and Baltic Countries

Skadina, I., Vasiljevs, A., Borin, L., De Smedt, K., Linden, K. & Rögnvaldsson, E., 12 Nov 2011, *The 5th International Joint Conference on Natural Language Processing*. Chiang Mai, Thailand, 9 p.

HFST—Framework for Compiling and Applying Morphologies

Linden, K., Silfverberg, M., Axelson, E., Hardwick, S. & Pirinen, T., 26 Aug 2011, *Systems and Frameworks for Computational Morphology*. Mahlow, C. & Pietrowski, M. (eds.). Springer, Vol. Vol. 100. p. 67-85 18 p. (Communications in Computer and Information Science; vol. 100).

Do wordnets also improve human performance on NLP tasks?

Muhonen, K. & Linden, K., 12 May 2011, *NEALT Proceedings Series Vol. 11*. Riga, Latvia: Northern European Association for Language Technology , Vol. 11. 7 p. (NEALT Proceedings Series).

FIN-CLARIN: A Framework for Depositing and Disseminating Language Resources for R&D

Voutilainen, A. & Linden, K., 11 May 2011, *NEALT Proceedings Series*. Riga, Latvia, 4 p. (NEALT Proceedings Series).

Open Content Licenses: How to choose the right one

Oksanen, V. & Linden, K., 11 May 2011, *NEALT Proceedings Series*. Riga, Latvia, Vol. 13. 7 p. (NEALT Proceedings Series).

Combining Statistical Models for POS Tagging using Finite-State Calculus

Silfverberg, M. & Linden, K., 9 May 2011, *Proceedings of the 18th Conference of Computational Linguistics NODALIDA 2011*. Pedersen, B. & Skadiņa, I. (eds.). Northern European Association for Language Technology , p. 183–190 8 p. (NEALT Proceedings Series; vol. 11).

Designing a Dependency Representation and Grammar Definition Corpus for Finnish

Voutilainen, A., Linden, K. & Purtonen, T. K., 8 May 2011, *Las tecnologías de la información y las comunicaciones: Presente y futuro en el análisis de corpora: Actas del III Congreso Internacional de Lingüística de Corpus..* Pastor, C. & Mora, C. (eds.). Valencia, p. 151-158 9 p.

HFST-Framework for Compiling and Applying Morphologies

Linden, K., Axelson, E., Hardwick, S., Pirinen, T. A. & Silfverberg, M., 2011, *SYSTEMS AND FRAMEWORKS FOR COMPUTATIONAL MORPHOLOGY*. Mahlow, C. & Piotrowski, M. (eds.). Springer-Verlag, p. 67-85 19 p. (Communications in Computer and Information Science; vol. 100).

2010

FinnTreeBank 1

Voutilainen, A., Purtonen, T. K., Leisko-Järvinen, S., Kumlander, M., Linden, K., Nissinen, M. & Hardwick, S., 29 Dec 2010

FinnWordNet - a Finnish WordNet database

Linden, K., Carlson, L. & Niemi, J. A., Dec 2010

Creating and Weighting Hunspell Dictionaries as Finite-State Automata

Pirinen, T. & Linden, K., Nov 2010, In: *Investigationes Linguisticae* (Online Edition). 19

FinnWordNet - WordNet på finska via översättning

Linden, K. & Carlson, L., Nov 2010, In: *LexicoNordica*. 17, 16 p.

Building and Using Existing Hunspell Dictionaries and TEX Hyphenators as Finite-State Automata

Pirinen, T. & Linden, K., Oct 2010, *Proceedings of International Multiconference on Computer Science and Information Technology: Computational Linguistics—Applications (CLA'10)*. Ganzha, M. & Paprzycki, M. (eds.). Wisla, Poland, Vol. 5. p. 477–484 8 p. (Proceedings of the International Multiconference on Computer Science and Information Technology).

Part-of-Speech Tagging using Parallel Weighted Finite-State Transducers

Silfverberg, M. & Linden, K., Aug 2010, *Proceedings of IceTAL 2010: 7th International Conference on Natural Language Processing*. Reykjavik, 12 p.

Finite-State Spell-Checking with Weighted Language and Error Models: Building and Evaluating Spell-Checkers with Wikipedia as Corpus

Pirinen, T. & Linden, K., May 2010, *Proceedings of LREC 2010: Workshop on Creation and use of basic lexical resources for less-resourced languages*.

Laundry Symbols and License Management: Practical Considerations for the Distribution of LRs based on experiences from CLARIN

Oksanen, V., Linden, K. & Westerlund, H., May 2010, *Proceedings of LREC 2010 : Workshop on Language Resources: From Storyboard to Sustainability and LR Lifecycle Management*.

2009

HFST Library

Silfverberg, M., Linden, K., Pirinen, T. & Axelson, E., Jun 2009

HFST-optimized-lookup

Silfverberg, M., Hardwick, S. & Linden, K., Jun 2009

Conflict resolution using weighted rules in HFST-TWOLC

Silfverberg, M. & Linden, K., May 2009, In: *NEALT Proceedings Series*. 4, p. 174-181 8 p.

Guessers for Finite-State Transducer Lexicons

Lindén, K., Mar 2009, *Computational Linguistics and Intelligent Text Processing: 10th International Conference, CICLing 2009*. Gelbukh, A. (ed.). Springer-Verlag, p. 158-169 12 p. (Lecture Notes in Computer Science).

Corpus-based lexeme ranking for morphological guessers

Linden, K. & Tuovila, J., 2009, *State of the Art in Computational Morphology edited by Cerstin Mahlow, Michael Piotrowski*. Berlin, Heidelberg: Springer-Verlag, p. 118-135 18 p. (Communications in computer and information science).

Corpus-based paradigm Selection for morphological entries

Linden, K. & Tuovila, J., 2009, In: *NEALT Proceedings Series*. 4, p. 96-102 7 p.

Entry Generation for New Words by Analogy for Morphological Lexicons

Linden, K., 2009, In: *Northern European Journal of Language Technology*. 1, p. 1-25 25 p.

HFST runtime format: A compacted transducer format allowing for fast lookup

Silfverberg, M. & Linden, K., 2009, *Finite-State Methods and Natural Language Processing - FSMNLP 2009 Eight International Workshop*.

HFST tool for morphology: An efficient open-source package for construction of morphological analyzers

Linden, K., Silfverberg, M. & Pirinen, T., 2009, *State of the Art in Computational Morphology edited by Cerstin Mahlow, Michael Piotrowski*. Berlin, Heidelberg: Springer-Verlag, p. 28-47 20 p. (Communications in computer and information science).

Weighted finite-state morphological analysis of finnish compounding with HFST-LEXC

Linden, K. & Pirinen, T., 2009, In: NEALT Proceedings Series. 4, p. 89-95 7 p.

Weighting finite-state morphological analyzers using HFST tools

Linden, K. & Pirinen, T., 2009, *Finite-State Methods and Natural Language Processing - FSMNLP 2009 Eight International Workshop*.

2008

A probabilistic model for guessing base forms of new words by analogy

Linden, K., 2008, In: Lecture Notes in Computer Science. 4919, p. 106-116 11 p.

Assigning an Inflectional Paradigm using the Longest Matching Affix

Linden, K., 2008, *Ei mitään ongelmia: Juhlakirja Juhani Reimanille 50-vuotispäiväksi 23.1.2008*. Wiberg, M. & Koura, A. (eds.). Turku

2007

Expert panel report: the Nordic countries : a leading region in language technology

Koskenniemi, K. (ed.), Linden, K. (ed.) & Nordgård, T. (ed.), 2007, Helsinki: University of Helsinki. Department of General Linguistics. 76 p. (Publications; vol. 40)

Opiskelijapalautteen keräämisestä

Kauhanen, I. & Linden, K., 2007, Helsinki: Helsingin yliopisto, yleisen kielitieteen laitos. 18 p. (Opintomonisteen / Yleisen kielitieteen laitos. Helsingin yliopisto; no. No. 4)

SpråkVis - Språkteknologisk vismansrapport: En utgivad sammanfattning av The Nordic Countries - A leading Region in language Technology

Koskenniemi, K. (ed.), Linden, K. (ed.) & Nordgård, T. (ed.), 2007, Helsinki: Helsingfors universitet. Institutionen för allmän språkvetenskap. 12 p.

2006

Common Infrastructure for Finite-State Based Methods and Linguistics Descriptions

Yli-Jyrä, A. M., Koskenniemi, K. & Linden, K., 2006, *International Workshop Towards a Research Infrastructure for Language Resources*. 2 p.

Multilingual modelling of cross-lingual spelling variants

Linden, K., 2006, In: Information Retrieval Journal. 9, 3, p. 295-310 16 p.

2005

Inquiries into words, constraints and contexts: Festschrift in the honour of Kimmo Koskenniemi on his 60th birthday

Arppe, A., Carlson, L., Linden, K., Piitulainen, J. O., Suominen, M., Vainio, M., Westerlund, H. & Yli-Jyrä, A. M., 2005, [S.l.]: CSLI publications. 288 p. (CSLI Studies in Computational Linguistics ONLINE)

Morfessor and Hutmegs: Unsupervised Morpheme Segmentation for Highly-Inflecting and Compounding Languages
Creutz, M., Lagus, K., Linden, K. & Virpioja, S. P., 2005, *The Second Baltic Conference on Human Language Technologies: Proceedings*. Langemets, M. & Penjam, P. (eds.). Tallinn: Institute of Cybernetics; Institute of the Estonian Language

Word sense discovery and disambiguation

Linden, K., 2005, Helsinki: University of Helsinki.

Word senses

Linden, K., 2005, *Inquiries into words, constraints and contexts: Festschrift in the honour of Kimmo Koskenniemi on his 60th birthday*. [S.l.]: [s.n.], p. 249-258 10 p.

2004

Evaluation of Linguistic Features for Word Sense Disambiguation with Self-Organized Document Maps

Linden, K., Nov 2004, In: *Language Resources and Evaluation*. 38, 4, p. 417-435 19 p.

Discovering Synonyms and Other Related Words

Linden, K. & Piitulainen, J. O., Aug 2004, *Proceedings of COLING 2004: CompuTerm 2004: 3rd International Workshop on Computational Terminology*. Ananadiou, S. & Zweigenbaum, P. (eds.). p. 63-70

Morpheme Segmentation Gold Standards for Finnish and English

Creutz, M. J. P. & Linden, B. K. J., 2004, *Publications in Computer and Information Science: Report A77*. Helsinki University of Technology, 33 p.

2003

Word Sense Disambiguation with THESSOM

Linden, K., Sep 2003, *Proceedings of WSOM'03 - Intelligent Systems and Innovational Computing*.

2002

Word Sense Disambiguation in Document Space

Linden, K. & Lagus, K., Oct 2002, *Proceedings of 2002 IEEE International Conference on Systems, Man and Cybernetics*.

1997

Language Applications with Finite-State Technology

Linden, K., 1997, In: *International Journal of Corpus Linguistics*. p. 281-288

1987

Unification as a grammatical tool

Carlson, L. & Linden, K., 1987, In: *Nordic journal of linguistics*. 10, p. 111-136

Projects

1. **FIN-CLARIN - Kielipankki**
Linden, K., Piitulainen, J., Niemi, J., Lennes, M., Bartis, I., Westerlund, H., Drobac, S., Axelson, E. & Kauppinen, P.
05/06/2006 → ...
2. **FinnWordNet - A Finnish WordNet**
Linden, K., Niemi, J., Hyvärinen, M., Muhonen, K. & Pääkkö, P.
01/01/2010 → ...

3. **HFST - Helsinki Finite-State Technology**
Linden, K., Koskenniemi, K., Yli-Jyrä, A., Hulden, M., Silfverberg, M., Pirinen, T., Axelson, E., Hardwick, S., Niemi, J. & Hurskainen, A.
01/01/2005 → ...
4. **Lahjoita puhetta -kampanjan suunnittelu ja toteutus**
Jauhiainen, T., Lennes, M. & Linden, K.
01/11/2019 → ...
5. **ANEE: ANEE Team 1: Digital Humanities Approaches**
Svärd, S., Linden, K., Bach, J., Jauhiainen, T., Jauhiainen, H., Sahala, A., Konstantopoulos, G., Fink, S., Zaia, S., Luukko, M. I., Bennett, E. R., Alstola, T., Vanderstraeten, E., Autere, K., Giessler, J., Debourse, C. H. I., Tams, L., Valk, J., Uotila, R. & Jones, C. W.
01/01/2018 → 31/12/2025
6. **ANEE: Centre of Excellence in Ancient Near Eastern Empires**
Silverman, J. M., Svärd, S., Lahelma, A., Nikki, N., Töyräänvuori, J., Jokiranta, J., Laine, L., Kletter, R., Bonnie, R., Thomas, S., Holmqvist-Sipilä, E., Kouki, P., Wasmuth, M., Wallis, C., Lorenzon, M., Jauhiainen, H., Jauhiainen, T., Alstola, T., Bach, J., Fink, S., Konstantopoulos, G., Zaia, S., Sahala, A., Mieltunen, P. H., Avneri Meir, R., Aissaoui, A. I., Arpe, O. P., Luukko, M. I., Kovanen, H., Linden, K., Bennett, E. R., Saari, S., Vanderstraeten, E., Autere, K., Debourse, C. H. I., Kailamäki, U., Kautonen, S. K., Spunaugle, A., Tams, L., Jones, C. W., Wijnsma, U. Z., Valk, J., Gheorghide, P., Autere, K., Uotila, R., Reinikainen, S. J. J., Cutillas Victoria, B., Holappa, M., Kletter, R. & Smith, S. L.
01/01/2018 → 31/12/2025
7. **LIST: Language Identification of Speech and Text**
Jauhiainen, T. & Linden, K.
01/01/2021 → 31/12/2022
8. **Semantic Domains in Akkadian Texts**
Linden, K., Alstola, T., Jauhiainen, H., Jauhiainen, T., Sahala, A., Svärd, S. & Zaia, S.
01/09/2016 → 31/08/2020
9. **FIN-CLARIN UPGRADE - CLARINin vahvistaminen Suomessa FIN-CLARIN UPGRADE - Förstärkning av CLARIN i Finland**
Linden, K., Lennes, M., Westerlund, H., Piitulainen, J., Niemi, J., Axelson, E., Jauhiainen, T., Drobac, S., Hardwick, S., Kauppinen, P., Voutilainen, A., Dieckmann, U., Palolahti, M., Matthiesen, M. & Aalto, T.
01/01/2017 → 31/12/2019
10. **Finno-Ugric Languages and the Internet**
Linden, K., Jauhiainen, H. & Jauhiainen, T.
01/01/2013 → 31/12/2018
11. **FIN-CLARIN-INVEST2 - Yhteiskäyttöisten kielivarojen ja -teknologian kohentaminen CLARINia varten Suomessa FIN-CLARIN-INVEST2 - Uppgradering av de gemensamma språkresurserna och -teknologin för CLARIN i Finland**
Linden, K.
01/01/2015 → 28/02/2017
12. **FINNCLARIN CONTENT**
Linden, K. & Koskenniemi, K.
01/01/2011 → 31/12/2015
13. **FIN-CLARIN-INVEST - Yhteiskäyttöisten kielivarojen ja -teknologian kohentaminen CLARINia varten Suomessa FIN-CLARIN-INVEST - Uppgradering av de gemensamma språkresurserna och -teknologin för CLARIN i Finland**
Linden, K.
01/01/2014 → 31/12/2014
14. **CLARA - Common Language Resources and their Applications**
Koskenniemi, K. & Linden, K.
Unknown funder
01/12/2009 → 31/03/2014
15. **Respeaking**
Tiittula, L., Linden, K., Bacon, H., Carlson, L., Eskelinen, J., Hartama-Heinonen, R., Koponen, M., Kukkonen, P., Mikkonen, K. & Pekkilä, E. O.
Humanistinen tdk
01/02/2013 → 31/12/2013
16. **FIN-CLARIN-BUILD - CLARINin vahvistaminen Suomessa FIN-CLARIN-BUILD - etablering av CLARIN i Finland**
Linden, K.
14/12/2012 → 31/12/2013
17. **META-NORD: Baltic and Nordic Parts of the European Open Linguistic Infrastructure**
Linden, K., Koskenniemi, K. & Bartis, I.
EU FP7
01/02/2011 → 31/01/2013

18. **FIN-CLARIN: Finnish Language Resource Infrastructure**
Koskenniemi, K. & Linden, K.
01/01/2007 → 31/12/2012
19. **Common Language Resources and Technology Infrastructure (CLARIN)**
Koskenniemi, K., Linden, K., Oksanen, V. & Yangarber, R.
01/01/2008 → 30/06/2011

Activities

Post-doctoral mentor / Måns Huldén

Krister Linden (Post doc mentor)

12 Jan 2009 → 11 Dec 2011

Activity: Examination types › Post doctoral mentor

Examination of doctoral dissertation / Grégoire Détrez

Krister Linden (Opponent) & Bo Krister Johan Linden (Pre-examiner)

2016

Activity: Examination types › Opponent of doctoral dissertation

Exmination of doctoral dissertation / Sami Virpioja

Krister Linden (Opponent) & Bo Krister Johan Linden (Pre-examiner)

2012

Activity: Examination types › Opponent of doctoral dissertation

Examination of doctoral dissertation / Tuomo Kakkonen

Krister Linden (Pre-examiner)

1 Sep 2007 → 15 Sep 2007

Activity: Examination types › Pre-examiner of doctoral thesis

Examination of doctoral dissertation / Vesa Siivola

Krister Linden (Pre-examiner)

15 May 2007 → 31 May 2007

Activity: Examination types › Pre-examiner of doctoral thesis

Contributions to Computational Assyriology / Aleksu Sahala

Krister Linden (Supervisor) & Saana Svärd (Supervisor)

1 Jan 2017 → 31 Aug 2021

Activity: Examination types › Supervisor or co-supervisor of doctoral thesis

PhD supervision of Rosa Suviranta

Tuomo Hiippala (Supervisor) & Krister Linden (Co-supervisor)

2022 → ...

Activity: Examination types › Supervisor or co-supervisor of doctoral thesis

Supervisor of doctoral thesis in Language Technology / Emily Öhman

Krister Linden (Co-supervisor)

1 Aug 2016 → ...

Activity: Examination types › Supervisor or co-supervisor of doctoral thesis

Supervisor of doctoral thesis in Language Technology / Miikka Silfverberg

Bo Krister Johan Linden (Supervisor) & Anssi Yli-Jyrä (Supervisor)

1 Jan 2010 → 22 Oct 2016

Activity: Examination types › Supervisor or co-supervisor of doctoral thesis

Supervisor of doctoral thesis in Language Technology / Senka Drobac

Krister Linden (Supervisor)

1 Jan 2012 → ...

Activity: Examination types › Supervisor or co-supervisor of doctoral thesis

Supervisor of doctoral thesis in Language Technology / Tommi Jauhiainen

Krister Linden (Supervisor)

1 Jun 2011 → ...

Activity: Examination types › Supervisor or co-supervisor of doctoral thesis

Supervisor of doctoral thesis in Language Technology / Tommi Pirinen

Krister Linden (Supervisor)

1 Jan 2010 → 2 Dec 2014

Activity: Examination types › Supervisor or co-supervisor of doctoral thesis

Cabinet of Commissioner Leonard Orban at the European Commission (External organisation)

Krister Linden (Expert member)

1 Apr 2008 → 31 Mar 2009

Activity: Membership types › Membership or other role in public Finnish or international organization

Nordiska Ministerrådet (External organisation)

Krister Linden (Expert member)

1 Jan 2006 → 31 Dec 2006

Activity: Membership types › Membership or other role in public Finnish or international organization

CLARIN - Common Language Resources and Technology Infrastructure (External organisation)

Krister Linden (Board member)

1 Jan 2010 → 31 Dec 2011

Activity: Membership types › Membership or other role in national/international committee, council, board

ESF - European Science Foundation (External organisation)

Krister Linden (Expert member)

24 May 2011 → 23 May 2015

Activity: Membership types › Membership or other role in national/international committee, council, board

HAKA - Identity federation of the Finnish universities, polytechnics and research institutions (External organisation)

Krister Linden (Board member)

2012 → 2013

Activity: Membership types › Membership or other role in national/international committee, council, board

KIT Language Technology Graduate School of Finland (External organisation)

Krister Linden (Expert member)

1 Jan 2001 → 31 Dec 2005

Activity: Membership types › Membership or other role in national/international committee, council, board

Langnet (External organisation)

Krister Linden (Board member)

1 Jan 2012 → 31 Dec 2015

Activity: Membership types › Membership or other role in national/international committee, council, board

Lantern - Language Technology Research Network (External organisation)

Krister Linden (Secretary)

1 Jan 2005 → 31 Dec 2007

Activity: Membership types › Membership or other role in national/international committee, council, board

META-TRUST (External organisation)

Krister Linden (Board member)

12 Sep 2012 → 31 Jan 2017

Activity: Membership types › Membership or other role in national/international committee, council, board

NYKYKIELET 2010-2017 (Organisational unit)

Krister Linden (Board member)

1 Jan 2014 → 31 Dec 2017

Activity: Membership types › Membership or other role in national/international committee, council, board

Northen European Association for Language Technology (External organisation)

Krister Linden (Chair)

1 Jan 2010 → 31 Dec 2013

Activity: Membership types › Membership or other role in national/international committee, council, board

Northen European Association for Language Technology (External organisation)

Krister Linden (Expert member)

1 Jan 2014 → 31 Dec 2017

Activity: Membership types › Membership or other role in national/international committee, council, board

Global Wordnet Association (External organisation)

Krister Linden (Member)

2012 → ...

Activity: Membership types › Membership or other role in research network

NEALT - North European Association for Language Technology (External organisation)

Krister Linden (Member)

2009 → ...

Activity: Membership types › Membership or other role in research network

Department of Modern Languages (Organisational unit)

Krister Linden (Assessment)

2014

Activity: Membership types › Assessment of candidates for academic posts