

CURRICULUM VITAE

FULL NAME: Antti Veikko Gabriel **ARPPE**

DATE AND PLACE OF BIRTH: Helsinki, Finland, 20 December 1967

DEGREES AWARDED

- 27 Jan 2009 **Doctor of Philosophy** (General Linguistics), awarded by the Faculty of Arts, University of Helsinki [Helsingin yliopisto]
- 29 Aug 1995 **Master of Science** (Industrial Management), incorporating a B.Sc. degree, Helsinki University of Technology [Teknillinen korkeakoulu, now part of Aalto University]; Major: Corporate strategy and international marketing; I Minor: Computational linguistics (at University of Helsinki, Department of General Linguistics); II Minor: Knowledge engineering
-

ACADEMIC EDUCATION

- 1999-2009 University of Helsinki, Faculty of Arts, Department of General Linguistics: graduate and postgraduate studies
Supervisors: Prof. Fred Karlsson, Prof. Lauri Carlson (University of Helsinki), Adj. Prof., Univ. Lecturer Urho Määttä (University of Tampere) & Adj. Prof., Senior Researcher Juhani Järvikivi (Max Planck Institute of Psycholinguistics)
- 1987-1995 Helsinki University of Technology, Faculty of Mechanical Engineering, Department of Industrial Management: combined undergraduate and graduate studies (for M.Sc.)

Other graduate and postgraduate studies:

- 2 May 2012 Basic Course in University-level Pedagogy (10 credits), Faculty of Arts, University of Helsinki
- 4 May 2005 Basic Studies in Assyriology (40 credits), Department of Asian and African Studies
- 30 Sept 2004 Basic Studies in Exact Methods, Half-Module (26 credits), Department of General Linguistics
- Sept-Dec 1999 University of Helsinki, Department of General Linguistics: Student in the European Master of Linguistics program (34 credits, incorporated in PhD degree)
- Jan-Jun 1999 University of Lund, Department of Linguistics: Visiting postgraduate student in the European Master of Linguistics Program (50 credits)

CURRICULUM VITAE

CURRENT PRIMARY POSITION

1 July 2018- **Associate Professor** (promoted to Full Professor 1 July 2024-) in Quantitative Linguistics, Department of Linguistics, University of Alberta

CURRENT SECONDARY ROLES

1 Apr 2019- **Project Director**, *21st Century Tools for Indigenous Languages* (SSHRC-funded Partnership project: <https://21c.tools/>)

1 Sept 2013- **Founding Director**, Alberta Language Technology Lab (ALTLab), University of Alberta (URL: <http://altlab.ualberta.ca/>)

PREVIOUS PROFESSIONAL POSITIONS, ROLES, AND APPOINTMENTS

1 Aug 2012 – 31 Jun 2018 **Assistant Professor** (tenure-track) in Quantitative Linguistics, Department of Linguistics, University of Alberta

1 Jan 2010 – 31 Jul 2012 **Post-Doctoral Research Fellow** [tutkijatohtori], Department of Modern Languages, University of Helsinki

1 Jul 2010 – 14 Sept 2021 **Visiting Scholar** (1 Jul 2010 – 31 Aug 2011); **Post-doctoral Research Fellow** (1 Sept 2011 – 30 Jul 2012), **Instructor** (part-time: 1 Sept – 31 Dec 2011), Department of Linguistics, University of Alberta

1 Aug 2008 – 31 Dec 2009 **Research Co-Ordinator** [tutkimuskoordinaattori] (ESFRI/CLARIN project [Common Language Resources and technology Infrastructure), Department of General Linguistics (Language Technology), University of Helsinki

1 Jan 2003 – 31 Dec 2007 **Post-Graduate Student** [tutkijakoulutettava] in LANGNET, the Finnish graduate school in language studies

1 Sept 2000 – 31 Dec 2009 **Research & Business Associate** (–31 May 2004); **External Consultant** (1 Jun 2004–) (part-time, contract-based), Lingsoft Incorporated

1 Mar 2000 – 31 Dec 2002 **Researcher** [tutkija], GILTA project – Managing large text masses (Finnish Funding Agency for Technology and Innovation, USIX program), Department of General Linguistics, University of Helsinki

1 Aug 1998 – 31 Aug 2000 **Deputy Managing Director**, Lingsoft Incorporated (on study leave January-December 1999; part-time January – August 2000)

CURRICULUM VITAE

| | |
|------------------------------|---|
| 1 Dec 1995 – 31 July 1998 | Product Manager (proofing tools and electronic dictionaries for the Nordic languages), Lingsoft Incorporated |
| 1 Oct 1994 – 30 Nov 1995 | Diploma Thesis Worker [diplomityöntekijä] (–31 July 1995); Business Analyst (1 Aug 1995–), Lingsoft Incorporated |
| Feb 1994 – Jun 1994 | Research Assistant (of Prof. Lauri Carlson), “User profiles of Translators' Aids”, EAGLES-project, University of Helsinki |
| May 1993 – Jan 1994 | Course Co-Ordinator , Cross-cultural Business, Institute of Industrial Management, Helsinki University of Technology |
| Sept 1993 – Feb 1994 | Intern , Special study in corporate strategy and international marketing on the markets of machine translation in Finland for Kielikone Ltd. |
| Aug 1990 – Jan 1991 | Teaching Assistant , EuroMBA program, Institute of Industrial Management, Helsinki University of Technology |

RESEARCH / PUBLICATIONS

BOOKS – Scientific monographs

Arppe, Antti (2008). *Univariate, bivariate and multivariate methods in corpus- based lexicography – a study of synonymy*. [PhD Dissertation] Publications of the Department of General Linguistics, University of Helsinki, No. 44. [300pp] (Electronic version with appendices: URN: <http://urn.fi/URN:ISBN:978-952-10-5175-3>. [614pp])

Arppe, Antti (1995). *The Strategic Opportunities of a Small High-Technology Company on Emerging Markets*. [Unpublished Master’s thesis]. Institute of Strategy and International Marketing, Department of Industrial Management, Helsinki University of Technology. URL: <http://urn.fi/URN:NBN:fi:aalto-2020120342343> [136pp]

BOOKS – EDITED COLLECTIONS

Harrigan, Atticus Aditi Chaudhary, Shruti Rijhwani, Sarah Moeller, **Antti Arppe**, Alexis Palmer, Ryan Henke, Daisy Rosenblum (eds.) (2023). *Proceedings of the Sixth Workshop on the Use of Computational Methods in the Study of Endangered Languages*. URL: <https://aclanthology.org/2023.computel-1.0/>

Moeller, Sarah, Antonios Anastasopoulos, **Antti Arppe**, Aditi Chaudhary, Atticus Harrigan, Josh Holden, Jordan Lachler, Alexis Palmer, Shruti Rijhwani & Lane Schwartz (eds.) (2022). *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*. [200pp] Dublin, Ireland: Association of Computational Linguistics. URL: <https://aclanthology.org/volumes/2022.computel-1/>

Arppe, Antti, Jeff Good, Atticus Harrigan, Atticus, Jordan Lachler, Sarah Moeller, Mans Hulden, Alexis Palmer, Miikka Silfverberg & Lane Schwartz (eds.) (2021). *Proceedings of the 4th Workshop on the Use of Computational Methods in the Study of Endangered*

CURRICULUM VITAE

Languages (ComputEL-4). [153 pages] Association of Computational Linguistics. URL: <https://aclanthology.org/volumes/2021.computel-1/>

Arppe, Antti, Jeff Good, Mans Hulden, Jordan Lachler, Alexis Palmer, Lane Schwartz, Lane, Miikka Silfverberg (eds.) (2019). *ComputEL-3: Proceedings of the 3rd Workshop on Computational Methods in the Study of Endangered Languages*, Volumes 1-2. URL: <https://journals.colorado.edu/index.php/computel/index>

Arppe, Antti, Jeff Good, Mans Hulden, Jordan Lachler, Alexis Palmer, Lane Schwartz, Lane (eds.) (2017). *ComputEL-2: Proceedings of the 2nd Workshop on Computational Methods in the Study of Endangered Languages*. URL: <https://aclanthology.org/volumes/W17-01/>

Suominen, Mickael, **Antti Arppe**, Anu Airola, Orvokki Heinämäki, Matti Miestamo, Urho Määttä, Jussi Niemi, Kari K. Pitkänen & Kaius Sinnemäki (eds.). (2006) *A Man of Measure. Festschrift in Honour of Fred Karlsson in his 60th Birthday*. Special Supplement to SKY Journal of Linguistics, Volume 19/2006, Turku: Linguistic Association of Finland. URL: <https://www.linguistics.fi/julkaisut/sky2006special.shtml> [519pp]

Arppe, Antti; Lauri Carlson, Krister Lindén, Jussi Piitulainen, Mickael Suominen, Martti Vainio, Hanna Westerlund & Anssi Yli-Jyrä (eds.) (2005) *Inquiries into Words, Constraints, and Contexts. Festschrift for Kimmo Koskenniemi on his 60th Birthday*. CSLI Studies in Computational Linguistics ONLINE, Stanford: CSLI Publications. URL: <https://web.stanford.edu/group/cslipublications/cslipublications/koskenniemi-festschrift/> [330pp]

Aho, Jaakko, **Antti Arppe**, Markus Henrikson, Juha Hintsa, Ari Hyppönen, Juha Jouppi, Lassi Linnanen, Marko Luhtala, Mika Läärä, Sari Matero, Juha Pankakoski, Virpi Pulli, Satu Rahikainen & Tuomas Sandholm (eds.) (1991) *The Innovative Enterprise*. Prodeko Publications, Espoo, Finland.

BOOKS – TRANSLATIONS

Naumanen, Mika (2002) *Nuorten teknologiayritysten menestystekijät* [The success factors of young technology companies] (Sitra reports series, ISSN 1457- 571X; 28). Edita Publishing Oy, Helsinki. ISBN 951-37-3819-1 (translated from English to Finnish by **Antti Arppe**).

ARTICLES – REFEREED IN SCIENTIFIC JOURNALS

Arppe, Antti, Atticus Harrigan, Katherine Schmirler, Daniel Dacanay & Rose Mackinaw (forthcoming). nêhiyawî-pîkiskwêwina maskwacîsihk: Spoken Dictionary of Maskwacîs Cree. *Dictionaries (Special Issue on Indigenous Lexicography)*.

Schmirler, Katherine, Inge Genee & **Antti Arppe**. (forthcoming). Morphophonological rule development and real-time rule testing with XFST: A model for Blackfoot. *Papers of the Fifty-Third Algonquian Conference (PAC53)*. Michigan State University Press, 53. [20pp]

CURRICULUM VITAE

Schmirler, Katherine & **Antti Arppe**. (forthcoming). Plains Cree textual analysis with PCA: Across the Bloomfield and Ahenakew-Wolfart subcorpora. *Papers of the Fifty-Third Algonquian Conference (PAC-53)*. Michigan State University Press, 53. [21pp]

Dacanay, Dacanay, Jolene Poulin & **Antti Arppe** (forthcoming) kwêyask kotahâskwâtam: [The Effects of Altering Specificity in WordNet on the Accuracy of Computational Semantic Classifications of Plains Cree \(nêhiyawêwin\)](#). *Papers of the Fifty-Third Algonquian Conference (PAC-53)*. Michigan State University Press, 53. [21pp]

Harrigan, Atticus & **Antti Arppe** (2022). [Leveraging Majority Language Resources for Plains Cree Semantic Classification](#). *Papers of the 52nd Algonquian Conference (PAC-52)*, pp. 1-20, East Lansing, Michigan, Michigan State University (MSU) Press.

Pankratz, Elizabeth, **Antti Arppe** & Jordan Lachler. (2022). [Low hanging fruit and the Boasian trilogy in digital lexicography of morphologically rich languages: Lessons from a survey of Indigenous language resources in Canada](#). *Nordlyd*, 46:1. 193-204, URL: <https://septentrio.uit.no/index.php/nordlyd/article/view/6441/6653>

Arppe, Antti, Katherine Schmirler, Atticus Harrigan & Arok Wolvengrey (2020). [A Morphosyntactically Tagged Corpus for Plains Cree](#). Macauley, Monica, & Noodin, Margaret (Eds.). *Papers of the 49th Algonquian Conference (PAC49)*) 49: 1-16. East Lansing, Michigan: Michigan State University (MSU) Press. URL: <https://doi.org/10.14321/j.ctvv417gp.5>

Antti Arppe, Katherine Schmirler, Miikka Silfverberg, Mans Hulden & Arok Wolvengrey (2019). [Insights from computational modelling of the derivational structure of Plains Cree stems](#). In: M. Macaulay & M. Noodin (eds), *Papers of the Forty-Eighth Algonquian Conference (PAC48)*, 48, 1-18. East Lansing, Michigan: MSU Press. URL: <http://dx.doi.org/10.14321/j.ctvc5pc5j.4>

Schmirler, Katherine, Atticus Harrigan, **Antti Arppe** & Arok Wolvengrey (2019). [Plains Cree verbal derivational morphology: a corpus investigation](#). In: M. Macaulay & M. Noodin (eds), *Papers of the Forty-Eighth Algonquian Conference (PAC48)*, 48, 239-254. East Lansing, Michigan: MSU Press. URL: <https://doi.org/10.14321/j.ctvc5pc5j.18>

Arppe, Antti, Chris Harvey, Marie-Odile Junker & J. Randolph Valentine (2018). [Algonquian Verb Paradigms: A Case for Systematicity and Consistency](#). In: M. MacCauley, and M. Noodin (eds), *Papers of the 47th Algonquian Conference (PAC47)*, 47, 1-22. East Lansing, Michigan: MSU Press. URL: <https://doi.org/10.14321/j.ctt1x76dj5.4>

Harrigan, Atticus, **Antti Arppe** & Arok Wolvengrey (2018). [Toward a Detailed Plains Cree VAI Paradigm](#). In: M. MacCauley, and M. Noodin (eds), *Papers of the Forty-Seventh Algonquian Conference (PAC47)*, 47, 129-146. East Lansing, Michigan: MSU Press. URL: <https://doi.org/10.14321/j.ctt1x76dj5.10>

Bontogon, Megan, **Antti Arppe**, Lene Antonsen, Dorothy Thunder & Jordan Lachler (2018). [Intelligent Computer Assisted Language Learning \(ICALL\) for Plains Cree: An In-Depth User Experience Evaluation](#). *Canadian Languages Review (Special Issue on Indigenous Languages)*. DOI: <https://doi.org/10.3138/cmlr.4054>

CURRICULUM VITAE

Harrigan, Atticus, Katherine Schmirler, **Antti Arppe**, Lene Antonsen, Sjur N. Moshagen, Trond Trosterud & Arok Wolvengrey (2017). [Learning from the Computational Modeling of Plains Cree Verbs](#). *Morphology* 27, pages 565–598. URL: <https://link.springer.com/article/10.1007%2Fs11525-017-9315-x> [34 pages: 15,062 words]

Paradis, Johanne, Ruiting Jia & **Antti Arppe** (2017). The acquisition of tense morphology over time by English second language children with specific language impairment: Testing the cumulative effects hypothesis. *Applied Psycholinguistics*, 38: 1-28, doi:10.1017/S0142716416000485. [28 pages: 13,185 words]

Paradis, Johanne, Yasemin Tulpar, & **Antti Arppe** (2016). [Chinese L1 children's English L2 verb morphology over time: Individual variation in long-term outcomes](#). *Journal of Child Language*, 43(3): 553-80. DOI: [10.1017/S0305000915000562](https://doi.org/10.1017/S0305000915000562) [27 pages: 12,202 words]

Miestamo, Matti; Dik Bakker & **Antti Arppe** (2016). [Sampling for variety](#). *Linguistic Typology*, 2: 233–296. URL: <https://doi.org/10.1515/lingty-2016-0006> [67 pages: 21,030 words]

Divjak, Dagmar, Ewa Dabrowska & **Antti Arppe** (2016). Machine Meets Man: evaluating the psychological reality of corpus-based probabilistic models. *Cognitive Linguistics*, 27:1, 1-34 [34 pages: 12,810 words]

Divjak, Dagmar & **Antti Arppe** (2013) [Extracting prototypes from exemplars. What evidence do corpora contain for the representation of linguistic categories?](#) *Cognitive Linguistics* 24(2): 221-274. DOI: <https://doi.org/10.1515/cog-2013-0008> [50 pages: 17,168 words]

Han, Weifeng, **Antti Arppe** & John Newman (accepted 2013, in press). [Topic marking in Shanghaiese](#). *Corpus Linguistics and Linguistic theory*. DOI: <https://doi.org/10.1515/cllt-2013-0014> [30 pages: 11,530 words]

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert & Arne Zeschel (2010). Cognitive Corpus Linguistics: Five points of debate on current theory and methodology. *Corpora* 5:1, pp. 1-27. [27pp]

Antti Arppe & Juhani Järviö (2007). [Every method counts – Combining corpus-based and experimental evidence in the study of synonymy](#). *Corpus Linguistics and Linguistic Theory*, 3:2, 131-159. DOI: <https://doi.org/10.1515/CLLT.2007.009> [29pp]

Magnusson, Camilla; **Antti Arppe**, Tomas Eklund, Barbro Back, Hannu Vanharanta, & Ari Visa (2004). [The language of quarterly reports as an indicator of change in the company's financial status](#). *Information & Management*, 42:4, 561-574. URL: <https://doi.org/10.1016/j.im.2004.02.008> [14pp]

ARTICLES – NON-REFEREED ARTICLES, COMMENTARIES AND REVIEWS IN SCIENTIFIC JOURNALS

Arppe, Arppe (2009). [Monta tapaa ajatella – tilastollisten menetelmien hyödyntäminen aineistolähtöisessä sanastontutkimuksessa](#). [Many ways of *thinking* – applying statistical

CURRICULUM VITAE

methods in corpus-based lexicography]. *Virittäjä*, 113:2. URL: <https://journal.fi/virittaja/article/view/4185> [8pp]

Antti Arppe & Juhani Järvikivi (2007). [Take empiricism seriously! – In support of methodological diversity in linguistics](#) [Commentary of Geoffrey Sampson 2007. Grammar without Grammaticality.] *Corpus Linguistics and Linguistic Theory*, 3:1, 99-109. URL: <https://doi.org/10.1515/CLLT.2006.007> [11pp]

Arppe, Arppe (2002). [Ei yhtä ainoaa polkua – Suomalaisia kokemuksia matkalla kieliteknologisesta tutkimuksesta liiketoimintaan](#) [No single path – Finnish lessons in the commercialisation of language engineering research]. *Puhe ja kieli* 22:1, 37-44. URL: <https://journal.fi/pk/article/view/8058> [8pp]

ARTICLES – REFEREED ARTICLES IN BOOKS AND CONFERENCE PROCEEDINGS

Antti Arppe, Andrew Neitsch, Daniel Dacanay, Jolene Poulin, Daniel W. Hieber & Atticus G. Harrigan (2023). Finding words that aren't there: Using word embeddings to improve dictionary search for low-resource languages. [Third Workshop on NLP for Indigenous Languages of the Americas \(AmericasNLP-2023\)](#), Toronto, 14 July 2023. URL: <https://aclanthology.org/2023.americasnlp-1.15/>

Poulin, Jolene, Daniel Dacanay & **Antti Arppe** (2023). Speech Database (Speech-DB) – An on-line platform for storing, validating, searching, and recording spoken language data. [2nd Workshop on NLP Applications to Field Linguistics \(Field Matters\)](#), 30-39, EACL, Dubrovnik, 6 May 2023. URL: <https://aclanthology.org/2023.fieldmatters-1.4/>

Holden, Joshua, Christopher Cox & **Antti Arppe** (2022). [An Expanded Finite-State Transducer for Tsuut'ina Verbs](#). *Proceedings of the 13th Language Resources and Evaluation Conference (LREC-2022)*, 13, pp. 5143-5152, URL: <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.551.pdf>

Lane, William, Atticus Harrigan & **Antti Arppe**. (2022). [Interactive Word Completion for Plains Cree](#). *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 1, pp. 3284-3294 [6338 words], URL: <https://aclanthology.org/2022.acl-long.232/>

Harrigan, Atticus & **Antti Arppe**. (2021). [Leveraging English Word Embeddings for Semi-Automatic Semantic Classification in Nêhiyawêwin \(Plains Cree\)](#). *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP-1)*, pp. 113-121. URL: <https://aclanthology.org/2021.americasnlp-1.12/>

Dacanay, Daniel, Atticus Harrigan, Arok Wolvengrey & **Antti Arppe** (2021). [The More Detail, the Better? Investigating the Effects of Semantic Ontology Specificity on Vector Semantic Classification with a Plains Cree / nêhiyawêwin Dictionary](#). *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP-1)*, pp. 143-152. URL: <https://aclanthology.org/2021.americasnlp-1.15.pdf>

CURRICULUM VITAE

Dacanay, Daniel, Atticus Harrigan & **Antti Arppe**. (2021). [Computational Analysis versus Human Intuition: A Critical Comparison of Vector Semantics with Manual Semantic Classification in the Context of Plains Cree](#). *Proceedings of the 4th Workshop on the Use of Computational Methods in the Study of Endangered Languages (ComputEL-4)*, pp. 33-43, March 2021. URL: <https://aclanthology.org/2021.computel-1.5/>

Sahala, Aleks, Miikka Silfverberg, **Antti Arppe** & Krister Lindén (2020). [BabyFST – Towards a Finite-State Based Computational Model of Ancient Babylonian](#). *Language Resources and Evaluation (LREC)*, Marseille, France. URL: <https://aclanthology.org/2020.lrec-1.479/>

Sahala, Aleks, Miikka Silfverberg, **Antti Arppe** & Krister Lindén (2020). [Automated Phonological Transcription of Akkadian Cuneiform Text](#). *Language Resources and Evaluation (LREC)*, Marseille, France. URL: <https://aclanthology.org/2020.lrec-1.433.pdf>

Schmirler, Katherine & **Antti Arppe**. (2019). [Modelling Plains Cree Negation with Constraint Grammar](#). *Proceedings of the NoDaLiDa 2019 Workshop on Constraint Grammar - Methods, Tools and Applications*, September 2019, Turku, Finland, NEALT Proceedings Series 43, pp. 27-34. URL: <https://ep.liu.se/ecp/168/007/ecp19168007.pdf> [5369 words]

Harrigan, Atticus, **Antti Arppe** & Timothy Mills (2019). [Preliminary Plains Cree Speech Synthesizer](#). In: Arppe, Antti, Good, Jeff, Hulden, Mans, Lachler, Jordan, Palmer, Alexis, Schwartz, Lane, & Silfverberg, Miikka (Eds.). *Proceedings of the 3rd Workshop on the Use of Computational Methods in the Study of Endangered Languages (ComputEL-3), 1*, 64-73. Stroudsburg, Pennsylvania: Association of Computational Linguistics. URL: <https://journals.colorado.edu/index.php/computel/issue/view/43>

Antti Arppe, Katherine Schmirler, Miikka Silfverberg, Mans Hulden & Arok Wolvengrey (2019). [Insights from computational modelling of the derivational structure of Plains Cree stems](#). In: M. Macaulay & M. Noodin (eds), *Papers of the Forty-Eighth Algonquian Conference (PAC48)*, 48, 1-18. East Lansing, Michigan: MSU Press. URL: <http://dx.doi.org/10.14321/j.ctvc5pc5j.4>

Schmirler, Katherine, Atticus Harrigan, **Antti Arppe** & Arok Wolvengrey (2019). [Plains Cree verbal derivational morphology: a corpus investigation](#). In: M. Macaulay & M. Noodin (eds), *Papers of the Forty-Eighth Algonquian Conference (PAC48)*, 48, 239-254. East Lansing, Michigan: MSU Press. URL: <https://doi.org/10.14321/j.ctvc5pc5j.18>

Littell, Patrick, Anna Kazantseva, Roland Kuhn, Aidan Pine, **Antti Arppe**, Christopher Cox, & Marie-Odile Junker (2018). [Indigenous language technologies in Canada: Assessment, challenges, and successes](#). In: Emily Bender, Leon Derczynski, and Pierre Isabelle (eds.), *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, 2620–2632. Santa Fe, New Mexico: Association of Computational Linguistics. URL: <https://www.aclweb.org/anthology/C18-1222>

Lachler, Jordan, Lene Antonsen, Trond Trosterud, Sjur N. Moshagen, & **Antti Arppe** (2018). [Modeling Northern Haida Morphology](#). *Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018)*, 2326-2330, Miyazaki, Japan:

CURRICULUM VITAE

European Language Resource Association. URL: <https://www.aclweb.org/anthology/L18-1368.pdf>

Lovick, Olga, Christopher Cox, Miikka Silfverberg, **Antti Arppe** & Mans Hulden (2018). [A computational architecture for the morphology of Upper Tanana](#). *Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018)*, 1874-1879. Miyazaki, Japan: European Language Resource Association. URL: <http://www.lrec-conf.org/proceedings/lrec2018/pdf/816.pdf>.

Schmirler, Katherine, **Antti Arppe**, Trond Trosterud & Lene Antonsen (2018). [Building a Constraint Grammar Parser for Plains Cree Verbs and Arguments](#). *Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018)*, 2981-2988. Miyazaki, Japan: European Language Resource Association. URL: <http://www.lrec-conf.org/proceedings/lrec2018/pdf/873.pdf>

Arppe, Antti, Chris Harvey, Marie-Odile Junker & J. Randolph Valentine (2018). [Algonquian Verb Paradigms: A Case for Systematicity and Consistency](#). In: M. MacCauley, and M. Noodin (eds), *Papers of the 47th Algonquian Conference (PAC47)*, 47, 1-22. East Lansing, Michigan: MSU Press. <https://doi.org/10.14321/j.ctt1x76dj5.4>

Harrigan, Atticus, **Antti Arppe** & Arok Wolvengrey (2018). [Toward a Detailed Plains Cree VAI Paradigm](#). In: M. MacCauley, and M. Noodin (eds), *Papers of the Forty-Seventh Algonquian Conference (PAC47)*, 47, 129-146. East Lansing, Michigan: MSU Press. <https://doi.org/10.14321/j.ctt1x76dj5.10>

Antti Arppe, Marie-Odile Junker & Delasie Torkornoo (2017). [Converting a comprehensive lexical database into a computational model: The case of East Cree verb inflection](#). *Proceedings of the 2nd Workshop on Computational Methods for Endangered Languages (ComputEL-2)*, 43-47, University of Hawai'i, Manoa, 6-7 March 2017. ACL Anthology. URL: <https://aclanthology.org/W17-0108.pdf>

Bowers, Dustin, **Antti Arppe**, Jordan Lachler, Sjur N. Moshagen & Trond Trosterud (2017). [A Morphological Parser for Odawa](#). *Proceedings of the 2nd Workshop on Computational Methods for Endangered Languages (ComputEL-2)*, 1-9, University of Hawai'i, Manoa, 6-7 March 2017. ACL Anthology. URL: <https://aclanthology.org/W17-0101.pdf>

Divjak, Dagmar, **Antti Arppe** & R. Harald Baayen. Does language-as-used fit a self-paced reading paradigm? A case study using data from Russian (2016). In: Anstatt, Tanja, Anja Gattnar & Christina Clasmeier (eds.), *Slavic Languages in a Black Box. Changes and Challenges for Empirical and Experimental Research*, Narr: Tübinger Beiträge zur Linguistik 554, 52-83

Hubert, Isabell, **Antti Arppe**, Jordan Lachler & Eddie A. Santos. (2016). [Training & Quality Assessment of an Optical Character Recognition Model for Northern Haida](#). In: 10th International Conference on Language Resources and Evaluation (LREC). Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asunción Moreno, Jan Odijk, Stelios Piperidis (eds). European Language Resources Association, ISBN 978-2-9517408-9-1, 3227-3234. URL: <https://aclanthology.org/L16-1514/> [8 pages]

CURRICULUM VITAE

Arppe, Antti, Jordan Lachler, Trond Trosterud, Lene Antonsen & Sjur N. Moshagen (2016). Basic Language Resource Kits for Endangered Languages: A Case Study of Plains Cree. In: Claudia Soria, Laurette Pretorius, Thierry Declerck, Joseph Mariani, Kevin Scannell, Eveline Wandl-Vogt (editors). *CCURL 2016 - Collaboration and Computing for Under-Resourced Languages – Towards an Alliance for Digital Language Diversity (LREC 2016 Workshop)*, Portorož, Slovenia, 23 May 2016, 1-8. URL: http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-CCURL2016_Proceedings.pdf [8 pages].

Evert, Stefan, and **Antti Arppe** (2015). Some theoretical and experimental observations on naive discriminative learning. *6th Conference on Quantitative Investigations in Theoretical Linguistics*, 3-5 November 2015, University of Tübingen, Germany. URL: <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/64864>

Conor Snoek, Dorothy Thunder, Kaidi Lõo, **Antti Arppe**, Jordan Lachler, Sjur Moshagen, and Trond Trosterud (2014). [Modeling the Noun Morphology of Plains Cree](#). In: Proceedings of *ComputEL: Workshop on the use of computational methods in the study of endangered languages*, 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland, 26 June 2014, 34-42. URL: <https://aclanthology.org/W14-2205/>

Paukkeri, Mari-Sanna, Jaakko Väyrynen & **Antti Arppe** (2012). Exploring Extensive Linguistic Feature Sets in Near-Synonym Lexical Choice. *Proceedings, Part II. Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, 11-17 March 2012, New Delhi, India, 1-12. URL: https://link.springer.com/chapter/10.1007/978-3-642-28601-8_1 [12pp]

Arppe, Antti (2009). Linguistic choices vs. probabilities – how much and what can linguistic theory explain? In: Featherston, Sam & Susanne Winkler (eds.) *The Fruits of Empirical Linguistics. Volume 1: Process*. Berlin: de Gruyter, 1-24. [24pp]

Arppe, Antti (2002). The usage patterns and selectional preferences of synonyms in a morphologically rich language. In: Morin, Annie & Pascale Sébillot (eds.) 2002. *JADT-2002. 6th International Conference on Textual Data Statistical Analysis*, 13-15 March 2002, Vol. 1, 21-32. Rennes: INRIA. [12pp]

Arppe, Antti (1995). Term Extraction from Unrestricted Text. *10th Nordic Conference of Computational Linguistics*, 30-31 May 1995. Department of General Linguistics, University of Helsinki. [8pp]

ARTICLES – “LIGHTLY” OR NON-REFEREED ARTICLES IN BOOKS AND CONFERENCE PROCEEDINGS

Arppe, Antti, Christopher Cox, Mans Hulden, Jordan Lachler, Sjur N. Moshagen, Miikka Silfverberg & Trond Trosterud (2017). [Computational Modeling of Verbs in Dene Languages: The Case of Tsuut’ina](#). In Alessandro Jaker (ed.), *Working Papers in Athabaskan (Dene) Languages (Alaska Native Language Center Working Papers)* 13, 51-69. Fairbanks, Alaska: Alaska Native Language Center.

CURRICULUM VITAE

Arppe, Antti & Mikko Lounela (2012). Aineiston edustavuudesta ja luotettavuudesta [On the representativeness and reliability of texts]. In: Heikkinen, Vesa et al. (Editors). *Genreanalyysi – tekstilajitutkimuksen käsikirja [Genre analysis – handbook of text genre research]*, [5pp]

Arppe, Antti (2007). [Multivariate methods in corpus-based lexicography: A study of synonymy in Finnish](#). *The Fourth Biennial Corpus Linguistics 2007 Conference*, July 28-30, 2007, Birmingham, UK. URL: <https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/107Paper.pdf> [18pp]

Arppe, Antti (2006). [Frequency Considerations in Morphology, Revisited – Finnish Verbs Differ, Too](#). In: Suominen, Mickael et al. (eds.), 175-189. URL: https://www.linguistics.fi/julkaisut/SKY2006_1/1.3.1.ARPPE.pdf [25pp]

Arppe, Antti (2005). The role of morphological features in distinguishing semantically similar words in Finnish. *Proceedings from the Corpus Linguistics Conference Series*, Vol. 1, no. 1, ISSN 1747-9398. Third Biennial Corpus Linguistics 2005 Conference, July 14-17, 2005, Birmingham, UK. [15pp]

Arppe, Antti (2005). [The Very Long Way from Basic Linguistic Research to Commercially Successful Language Technology: the Case of Two-Level Morphology](#). In: **Arppe, Antti** et al. (eds.), 2-17. URL: <https://web.stanford.edu/group/cslipublications/cslipublications/koskenniemi-festschrift/1-arppe.pdf> [16pp]

Arppe, Antti (2001). Focal points in frequency profiles – how some word forms in a paradigm are more significant than others in Finnish. *Proceedings of COMPLEX 2001 – 6th Conference on Computational Lexicography and Corpus Research “Computational Lexicography and New EU Languages”*, 28-30 June 2001, Centre for Corpus Linguistics, Department of English, University of Birmingham, UK, 1-8. [8pp]

Arppe, Antti (2001). Corpus-Based Observations on the Use and Occurrence of Inflected Forms in Finnish – the Case of Synonyms. Niemi, Jussi & Janne Heikkinen (eds.) *Nordic and Baltic Morphology. Papers from a NorFA Course, Tartu, June 2000. Studies in Languages 36*, University of Joensuu, Faculty of Humanities.

Arppe, Antti (2001). Lärdomar från utveckling av inflekterande synonymordböcker. Gellerstam, Martin et al. (eds.) *Nordiska studier i Lexikografi 5. Rapport från 'Konferens om lexikografi i Norden'*, Gothenburg May, 26-29, 1999. Skrifter utgivna av Nordiska föreningen för lexikografi (6) i samarbete med Nordiska språkrådet och Meijerbergs institut, Gothenburg, Sweden, 31-43 [13pp]

Arppe, Antti, Mari Voipio & Malene Würtz (2000). Creating Inflecting Electronic Dictionaries. Lindberg, Carl-Erik & Lund, Steffen Nordahl (eds.) *17th Scandinavian Conference of Linguistics*, Nyborg, 20-22 August 1998. Odense Working Papers in Language and Communication No 19, April 2000, Vol 1. University of Southern Denmark, Odense, Denmark, 1-10 [10pp]

Arppe, Antti (2000). Developing a Grammar Checker for Swedish. Nordgård, Torbjørn (ed.) *Proceedings from the 12th Nordiske datalingvistikkdager, Trondheim*, 9-10 December 1999. Department of Linguistics, Norwegian University of Science and Technology, Trondheim, 13-27. [15pp]

CURRICULUM VITAE

Ahmad, Khurshid, Antoine Ogonowski, Eva Dauphin, Jean-David Sta & **Antti Arppe** (1996) Engineering Terminology – A Case for a Linguistically Informed Database. *Proceedings of the 4th Congress on Terminology and Knowledge Engineering (TKE)*, Vienna, 26-28 August 1996. Frankfurt: Index Verlag. 166-178.

Arppe, Antti (1996). Information Explosion and the Use of Linguistic Tools in Finland. Harakka, Terttu & Merja Koskela (eds.) *Kieli ja tietokone. Association Finlandais de Linguistique Appliquée, Yearbook 54*. Jyväskylä, Finland.

OTHER ARTICLES AND PUBLICATIONS FOR THE SCIENTIFIC COMMUNITY

Arppe, Antti, Sari Bruun, Kimmo Koskenniemi, Krister Lindén, Mikael Lindén, Ville Oksanen & Hanna Westerlund (eds.) (2011). A report including Model Licensing Templates and Authorization and Authentication Scheme (Deliverable D7S-2.1). Common Language Resources and Technology Infrastructure (CLARIN: EC FP7 project no. 212230).

Arppe, Antti, Sari Bruun, Krister Lindén, & Hanna Westerlund (eds.) (2011). Set of Federation Agreements for CLARIN Centres (Deliverable D7S-4.1). Common Language Resources and Technology Infrastructure (CLARIN: EC FP7 project no. 212230).

Arppe, Antti (1994). Translators' Aids: User Profiles (main text chapter + appendix chapter). In: McNought John et al. (eds.). LRE EAGLES report. Commission of the European Union, Luxembourg.

ARTICLES IN NON-SCIENTIFIC JOURNALS OR FORA

Arppe, Antti (2002). Forward with Feet on the Ground – Speech Technology the Finnish Way. [Formerly: Crossing the Chasm – Speech Technology the Finnish Way]. Euromap project. Available originally at: URL: <http://www.hltcentral.org/page-1054.0.shtml>

Arppe, Antti (2002). Kärki on kapea mutta kärjen tuntumassa ollaan – Puheteknologiaa suomalaisittain. Euromap project. URL: <http://www.csc.fi/euromap/artikkelit/puheteknologia.phtml.fi>

Arppe, Antti (2002). No Single Path – Finnish Lessons in the Commercialization of Language Technology Research. [Formerly: Finnish Lessons: No Single Path to Market]. Euromap project. Available originally at: URL: <http://www.hltcentral.org/page-969.shtml>

Arppe, Arppe (2002). Ei yhtä ainoaa polkua - Suomalaisia kokemuksia matkalla kieliteknologisesta tutkimuksesta liiketoimintaan. Euromap project. URL: <http://www.csc.fi/euromap/artikkelit/arppe.phtml.fi>

ARTICLES – UNDER PREPARATION

Antti Arppe, Miikka Silfverberg & Chris Hammerly (submitted in April 2023). Flattening complex finite-state morphological models: A case study of Odawa. *SIGMORPHON-2023*.

Dacanay, Daniel & **Antti Arppe** (accepted in May 2023). Digitizing, Translating, and Standardizing Pr. Albert Lacombe's *Dictionnaire de la langue des Cris*. *Papers of the 54th Algonquian Conference (PAC-54)*.

Weber, Natalie, **Antti Arppe**, Ksenia Bogomolets, Andrew Cowell, Rose-Marie Déchaine, Christopher Hammerly, Sarah Murray, Katherine Schmirler & Rachel Vogel (Accepted in May 2023). Microparametric variation in prosodic structure: case studies from Algonquian. *Papers of the 54th Algonquian Conference (PAC-54)*.

ABSTRACTS – REFEREED ABSTRACTS/PRESENTATIONS IN CONFERENCE AND WORKSHOP PROCEEDINGS

Weber, Natalie, **Antti Arppe**, et al. et al. (2022). Special session on Microparametric variation in prosodic structure: Case studies from Algonquian. *98th Annual Meeting of the Linguistic Society of America*, Denver, CO, 5–8 January 2023

Arppe, Antti, Jolene Poulin, Eddie Antonio Santos, Andrew Neitsch, Atticus Harrigan, Katherine Schmirler, Daniel Hieber, Ansh Dubey & Arok Wolvengrey (2022). [Towards a morphologically intelligent and user-friendly on-line dictionary of Plains Cree – next next round](https://altlab.ualberta.ca/wp-content/uploads/2023/03/itwewina_AC54_Oct2022.pptx.pdf). *54th Algonquian Conference*, University of Colorado, Boulder, 20-23 October 2022. URL: https://altlab.ualberta.ca/wp-content/uploads/2023/03/itwewina_AC54_Oct2022.pptx.pdf

Arppe, Antti, Atticus Harrigan, Katherine Schmirler, Daniel Dacanay, Jolene Poulin & Rose Makinaw. (2022). Towards the Spoken Dictionary of Maskwacîs Cree. *54th Algonquian Conference*, University of Colorado, Boulder, 20-23 October 2022.

Dacanay, Daniel & **Antti Arppe** (2023). Digitizing, Translating, and Standardizing Pr. Albert Lacombe's *Dictionnaire de la langue des Cris*. *54th Algonquian Conference*, University of Colorado, Boulder, 20-23 October 2022.

Arppe, Antti, Atticus Harrigan, Katherine Schmirler, Daniel Dacanay, Jolene Poulin & Rose Makinaw. (2022). *ékosi ê-nêhiyawî-pîkiskwêcik maskwacîsihk Towards a Spoken Dictionary of Maskwacîs Cree*. Special Session on **Partnership in Practice** at ComputEL-5, at the ACL-2022 Conference, Dublin, Ireland. 2022-05-27.

Schmirler, Katherine & **Antti Arppe** (2019). Plains Cree actors and goals: Across time periods and genres. *51st Algonquian Conference*, Université de Montréal, Quebec, 27 October 2019.

Harrigan, Atticus & **Antti Arppe**. (2019). Automatic semantic classification of Plains Cree verbs. *51st Algonquian Conference*, Université de Montréal, Quebec, 26 October 2019.

CURRICULUM VITAE

Arppe, Antti, Katherine Schmirler, Eddie Antonio Santos, Arok Wolvengrey (2019). Towards a Systematic and Quantitative Identification of Sound Change Rules – Algonquian Historical Linguistics Meets Optical Character Recognition. *51st Algonquian Conference*, Université de Montréal, Quebec, 26 October 2019.

Arppe, Antti, Eddie Antonio Santos, Atticus Harrigan, Katherine Schmirler, Arok Wolvengrey, Arok (2019). Towards a morphologically intelligent on-line dictionary of Plains Cree. *51st Algonquian Conference*, Université de Montréal, Quebec, 26 October 2019.

Junker, Marie-Odile, Jérémie Ambroise, **Antti Arppe**, Bill Jankewicz, Mimie Neacappo, Claire Owen (2019). A Digital Infrastructure to Sustain Algonquian Languages: Dictionaries and Linguistic Atlas. Progress Report 3. *51st Algonquian Conference*, Université de Montréal, Quebec, 25 October 2019.

Arppe, Antti, Katherine Schmirler, Atticus Harrigan & Arok Wolvengrey (2019). Small Can Be Beautiful - A Morpho-Syntactically Tagged Text Collection for Plains Cree. *Corpus Linguistics Conference*, 23 July 2019, Cardiff University, Wales, UK.

Pankratz, Elizabeth, Jordan Lachler, **Antti Arppe** (2019). A survey of Indigenous language tools: The emergence of two lexicographical approaches among Indigenous languages of Canada. *International Conference of Language Documentation and Conservation (ICLDC6)*, University of Hawai'i.

Heinrich, Claudia & **Antti Arppe** (2018). An exploratory corpus-based investigation of the nominal diminutive in Plains Cree. *50th Algonquian Conference*, October 2019.

Dustin Bowers, **Antti Arppe**, Jordan Lachler, Sjur Moshagen, and Trond Trosterud (2017). A Morphological Parser for Odawa. *2nd Workshop on Computational Methods for Endangered Languages (ComputEL-2)*, University of Hawai'i, Manoa, United States, 6-7 March 2017.

Arppe, Antti, Marie-Odile Junker & Delasie Torkornoo (2017). Converting a comprehensive lexical database into a computational model: The case of East Cree verb inflection. *2nd Workshop on Computational Methods for Endangered Languages (ComputEL-2)*, University of Hawai'i, Manoa, United States, 6-7 March 2017.

Arppe, Antti, Katherine Schmirler, Miikka Silfverberg & Arok Wolvengrey (2016). Computational modeling of the derivational morphology of Plains Cree verbs. *48th Algonquian Conference*, Milwaukee, WI, United States, 14-16 October 2016.

Schmirler, Katherine, Atticus Harrigan, **Antti Arppe** & Arok Wolvengrey (2016). Plains Cree Verbal Derivational Morphology: A Corpus Investigation. *48th Algonquian Conference*, Milwaukee, WI, United States, 14-16 October 2016.

Harrigan, Atticus, Katherine Schmirler, **Antti Arppe** & Arok Wolvengrey (2016). Lexical Subtypes in a Plains Cree Corpus. *48th Algonquian Conference*, Milwaukee, WI, United States, 14-16 October 2016.

Dustin Bowers, **Antti Arppe**, Jordan Lachler, Sjur Moshagen, and Trond Trosterud (2017). A Morphological Parser for Odawa. *2nd Workshop on Computational Methods for Endangered Languages (ComputEL-2)*, University of Hawai'i, Manoa, United States, 6-7 March 2017.

CURRICULUM VITAE

Christopher Cox, Måns Huldén, Miikka Silfverberg, Jordan Lachler, Sally Rice, Sjur N. Moshagen, Trond Trosterud & **Antti Arppe** (2016) Computational modeling of the verb in Dene languages. The case of Tsuut'ina. Dene Languages Conference, Yellowknife, North-West Territories, Canada, 6-7 June 2016.

Arppe, Antti, Atticus G. Harrigan & Katherine Schmirler (2016) So similar in principle, but so different in practice. Mixing texts, elicitation, and experimentation in the study of the Plains Cree independent and conjunct verb constructions. 2nd Workshop on *New Ways of Analyzing Syntactic Variation* (Nwasv2), 19-20 May 2016, Ghent, Belgium.

Harrigan, Atticus G., Lene Antonsen, **Antti Arppe**, Dustin Bowers, Katherine Schmirler, Trond Trosterud, & Arok Wolvengrey (2016) Learning from the Computational Modeling of Plains Cree Verbs. Workshop on *Computational methods for descriptive and theoretical morphology*, 17th International Morphology Meeting, 18 February 2016, Vienna, Austria.

Evert, Stefan & **Antti Arppe** (2015) Some Theoretical and Experimental Observations of Naïve Discriminative Learning. 6th Workshop on *Quantitative Investigations in Theoretical Linguistics*, University of Tübingen, 4-6 November, 2015.

Arppe, Antti, Chris Harvey, Marie-Odile Junker & Rand Valentine (2015) Verb paradigms. *47th Algonquian Conference*, 22-25 October 2015, Winnipeg, Manitoba.

Atticus G. Harrigan & **Antti Arppe** (2015) miswêw êkwa ê-tipiskâk, mâka mâcîw âhpô ê-mâcî? (2015) *47th Algonquian Conference*, 22-25 October 2015, Winnipeg, Manitoba.

Atticus G. Harrigan, Lene Antonsen, **Antti Arppe**, Trond Trosterud & Arok Wolvengrey (2015) Modeling the Plains Cree verb. *47th Algonquian Conference*, 22-25 October 2015, Winnipeg, Manitoba.

Marrville, Caelan & **Antti Arppe** (2015). Linguistic development of the Alberta Bituminous Sands. *Corpus Linguistics 2015*, 21-25 July 2015, University of Lancaster, UK. [AA co-responsible for general research design and interpretation of results.] Presented 21-Jul-2015].

Arppe, Antti, Lene Antonsen, Trond Trosterud, Sjur Moshagen, Dorothy Thunder, Conor Snoek, Timothy Mills, Juhani Järvi-kivi, and Jordan Lachler. (2015) Turning language documentation into reader's and writer's software tools. *4th International Conference on Language Documentation and Conservation (ICDLC)*. 26 February – 1 March 2015, Honolulu, HI. (500 words; 1 page); [Presented: 28-Feb-2015]

Arppe, Antti, Lene Antonsen, Trond Trosterud, Conor Snoek, Dorothy Thunder, Jordan Lachler, Jean Okimâsis, and Arok Wolvengrey. (2014). Linguistic insights from computational modeling of Plains Cree morphology. *46th Algonquian Conference*, October 23-26, 201, Mohegan Sun Convention Center, Uncasville, Connecticut. (500 words; 1 pages); [Presented: 25-Oct-2014]

Conor Snoek, Dorothy Thunder, Kaidi Lõo, **Antti Arppe**, Jordan Lachler, Juhani Järvi-kivi, Timothy Mills, Sjur N. Moshagen & Trond Trosterud (2014). Literacy and language learning tools for Plains Cree. *Prairies Workshop on Languages and Linguistics*, Brandon, Manitoba, 1 March 2014.[500 words; 1 pages]

CURRICULUM VITAE

Mildred Lau & **Antti Arppe** (2013). We don't all 'think' exactly alike: empirical evidence for cross-linguistic lexical contrast. *5th Conference on Quantitative Investigations in Theoretical Linguistics*, September 12-14 2013, Leuven, Belgium. [500 words; 1 pages]

Conor Snoek, Dorothy Thunder, Kaidi Lõo, **Antti Arppe**, Jordan Lachler, Sjur Moshagen & Trond Trosterud (2014). Modeling the Nominal Morphology of Plains Cree. *ComputEL: Workshop on the use of computational methods in the study of endangered languages*, 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland, 26 June 2014.

Dowbor, Dowbor, **Antti Arppe** & Sally Rice (2013). Bridging Lexical and Constructional Synonymy - the English Passive and its Auxiliary Verbs. *American Association for Corpus Linguistics Conference (AACL2013)*, SDSU, San Diego, CA, 20 January 2013

Abdulrahim, Dana & **Antti Arppe** (2013). Converging Linguistic Evidence: The synonymy of Arabic COME verbs. *27th Annual Arabic Linguistics Symposium*, Indiana University at Bloomington, 28 February 2013.

Dagmar Divjak, **Antti Arppe** & Ewa Dąbrowska (2013). Man meets machine. Predicting lexical preferences using conditional probabilities. *12th International Cognitive Linguistics Conference*, University of Alberta, Edmonton, Canada, 24 June 2013.

Dagmara Dowbor, **Antti Arppe** & Sally Rice (2013). English passive constructions: Many variations on an elusive schema. *12th International Cognitive Linguistics Conference*, University of Alberta, Edmonton, Canada, 27 June 2013.

Walton, Linda & **Antti Arppe** (2013) Semantically driven Tagalog morphology. *12th International Cognitive Linguistics Conference (ICLC-12)*, 23-28 June 2013, Edmonton, Alberta, 28 June 2013.

Arppe, Antti, Patrick Bolger & Dagmara Dowbor (2012). Grammar quadrangulated — contrasting frequency, selection, acceptability and processing. *8th International Conference on the Mental Lexicon*, McGill University, Montreal, Quebec, 24 October 2012.

Arppe, Antti, Patrick Bolger & Dagmara Dowbor (2012). The more evidential diversity, the merrier — contrasting linguistic data on frequency, selection, acceptability and processing. *New Ways of Analyzing Syntactic Variation (NWASV)*, Radboud University, Nijmegen, the Netherlands, 15 November 2012.

Tucker, Benjamin V. & **Antti Arppe** (2012). Modeling the occurrence of the allophones of /t/ in a spontaneous speech corpus using a discriminative learning approach. *The 13th Conference on Laboratory Phonology (LabPhon 13)*, Stuttgart, Germany, July 27-29, 2012.

Dowbor, Dagmara, **Antti Arppe** & Sally Rice (2012). The English Passive Under Corpus Scrutiny: Some Cognitive Models of Transitivity Revisited. *Conceptual Structure, Discourse, and Language (CSDL-11)*, 17-20 May 2012, Vancouver, Canada.

Arppe, Antti & Patrick Bolger (2012). Grammar triangulated – contrasting frequency, acceptability and processing. *Workshop on Grammar between gradience and frequency*,

CURRICULUM VITAE

Annual Conference of DGfS (Deutsche Gesellschaft für Sprachwissenschaft), Frankfurt am Main, Germany, 6-9 March 2012.

Arppe, Antti & Benjamin V. Tucker (2012). You should model what you observe – the case of the allophonic realizations of the English /t/ in a spontaneous speech corpus. Pre-proceedings of the *International Conference on Linguistic Evidence*. Tübingen, Germany, 9-11 February 2012.

Benjamin V. Tucker & **Antti Arppe** (2011). Allophonic realizations of the phoneme /t/ in an English spontaneous speech corpus. Workshop on *Production and Comprehension of Conversational Speech*, 12-13 December 2011, the Netherlands.

Miestamo, Matti, Dik Bakker & **Antti Arppe** (2011). Sampling for Variety. *Association for Linguistic Typology 9th Biennial Conference*, University of Hong Kong, 21-24 July 2011.

Arppe, Antti (2011). From modeling lexical synonyms to constructional alternations. *Corpus Linguistics 2011 (CL2011)*, University of Birmingham, England, 20-22 July 2011.

Antti Arppe & R. Harald Baayen (2011). Statistical classification and principles of human learning. *4th Conference on Quantitative Investigations in Theoretical Linguistics (QITL4)*, Humboldt-Universität zu Berlin, Germany, 28-31 March 2011.

Arppe, Antti & Benjamin V. Tucker (2011). Allophonic variation is regular. *New Tools and Methods for Very-Large-Scale Phonetics Research (VLSP-11)*, University of Pennsylvania, USA, 8-31 January 2011.

Divjak, Dagmar & **Antti Arppe** (2010). Extracting prototypes from corpus data: a distributional account of representing near-synonymous verbs. *Interdisciplinary Workshop on Verbs. The Identification and Representation of Verb Features (Verb 2010)*, Scuola Normale Superiore & Università di Pisa, Italy, 4-5 November 2010.

Arppe, Antti & Dagmar Divjak (2010). Synonymy is both Gradient and Context-Dependent. Workshop on *Computational Approaches to Synonymy* at the Symposium on *Re-thinking synonymy: semantic sameness and similarity in languages and their description*, Linguistic Association of Finland (SKY), Helsinki, 28-30 October 2010.

Arppe, Antti & Dagmar Divjak (2010). Exemplars and prototypes as two faces of the same coin – what evidence do corpora contain for the representation of linguistic categories? *Seventh International Conference on the Mental Lexicon*, University of Windsor, Ontario, Canada, 30 June – 3 July 2010.

Arppe, Antti & Dagmar Divjak (2010). Exemplars, prototypes, or both – what evidence do corpora contain for the representation of linguistic categories? *Linguistic Evidence 2010*, Tübingen, Germany, 11-13 February 2010

Divjak, Dagmar & **Antti Arppe** (2009). Near-synonymy or the art of expressing roughly the same meaning. *Slavic Conference of Cognitive Linguistics (SCLC- 2009)*, Prague, Czech Republic, 15-17 October 2009.

CURRICULUM VITAE

Arppe, Antti (2008) Linguistic choices vs. probabilities – how much and what can linguistic theory explain? Pre-proceedings of the *International Conference on Linguistic Evidence*. Tübingen, Germany, 31 January – 2 February 2008.

Arppe, Antti (2006) Complex phenomena deserve complex explanations. *Quantitative Investigations in Theoretical Linguistics Conference* (QITL2/2006), Osnabrück, Germany, 1-2 June 2006. Also available on-line at: <http://www.cogsci.uni-osnabrueck.de/~qitl/>

Arppe, Antti (2005) On the limits of generalizing from quantitative, corpus-based evidence in a morphologically rich language. Pre-proceedings of the *International Conference on Linguistic Evidence*. Tübingen, Germany, 2-4 June 2006. URL: <http://www.sfb441.uni-tuebingen.de/LingEvid2006/abstracts/arppe.pdf>

Arppe, Antti (2004) Every method makes a difference - describing the use of a Finnish synonym pair. Pre-proceedings of *International Conference on Linguistic Evidence*, 29-31 January 2004, Tübingen, Germany.

Arppe, Antti & Järvikivi, Juhani (2002) Verbal Synonymy in Practice: Combining Corpus-Based and Psycholinguistic Evidence. Workshop on *Quantitative Investigations in Linguistics* (QITL1/QITL-2002), Osnabrück, Germany, 3-5 October 2002. URL: <http://www.cogsci.uni-osnabrueck.de/~qitl/>

PUBLISHED RESEARCH MATERIALS, SOFTWARE AND WEB-SITES

MATERIALS (LINGUISTIC):

Arppe, Antti (2013). *Extracting exemplars and prototypes*. R vignette to accompany Divjak & Arppe (2013). [9 pages]

Arppe, Antti, Weifeng Han & John Newman (2013) *Polytomous logistic regression with Shanghainese topic markers*. R vignette to accompany Han, Newman & Arppe (in press). [11 pages]

amph (2008). A micro-corpus of 3404 occurrences of the four most common Finnish THINK lexemes, *ajatella*, *mieltä*, *pohtia*, and *harkita*, in Finnish newspaper and Internet newsgroup discussion texts, containing extracts and linguistic analysis of the relevant context in the original corpus data, scripts for processing this data, R functions for its statistical analysis, as well as a comprehensive set of ensuing results as R data tables. Compiled and analyzed by **Antti Arppe**. URL: <http://urn.fi/urn:nbn:fi:lb-2015021301>

SFNET (2002-2003). ~100 million words of Finnish internet newsgroup discussions posted during October 2002 – April 2003. Compiled by Tuuli Tuominen, Panu Kalliokoski, Computing Centre, University of Helsinki, and **Antti Arppe**, Department of General Linguistics, University of Helsinki, and CSC – IT Center for Science, Finland. URL: <http://urn.fi/urn:nbn:fi:lb-20150126>

Spoken Dictionary of Maskwacis Cree (on-going since 2014). Maskwacis Cree speakers and Elders: Mary Jean Littlechild, Louise Wildcat, Jerry Roasting, Harley Simon, Annette Lee, Arlene Makinaw, Rosie Rowan, Rose Makinaw, Kisikaw, Betty Simon, Brian Lightning, Brian Lee, Linda Oldpan, Miriam Buffalo, Debora Young, Ivy

CURRICULUM VITAE

Raine, Paula Mackinaw, Norma Linda Saddleback, and Renee Makinaw, under the management of Rose Makinaw, Patricia Johnson and Brian Wildcat, and with the help of a number of UofA undergraduate and graduate research assistants and volunteers. Edited by **Arppe, Antti**, Atticus Harrigan, Katherine Schmirler, Dustin Bowers, Megan Bontogon, Sarah Giesbrecht, Jolene Poulin, Daniel Dacanay, Conor Snoek, Timothy Mills & Jordan Lachler. Currently ~20,000 words and ~2,000 sentences. URL: <https://speech-db.altlab.app/maskwacis/>

SOFTWARE:

Computational models and applications

Arppe, Antti, Marie-Odile Junker & Delasie Torkornoo (under development). *Finite-state transducer-based computational model of East Cree morphology* (under development). URL: <https://github.com/giellalt/lang-crl>

Arppe, Antti, Atticus Harrigan, Katherine Schmirler, Lene Antonsen, Trond Trosterud, Sjur Moshagen, Arok Wolvengrey, Jean Okimâsis, Dorothy Thunder, Jordan Lachler & Conor Snoek. *Finite-state transducer-based computational model of Plains Cree morphology* (under development). URL: <https://github.com/giellalt/lang-crk>

Arppe, Antti, Arok Wolvengrey, Jolene Poulin, Andrew Neitsch, Eddie A. Santos, Ryan Johnson, Atticus Harrigan, Katherine Schmirler, Dorothy Thunder, Trond Trosterud & Lene Antonsen (under development) *itwêwina: nêhiyawêwin – âkayâsimowin / English – Plains Cree intelligent on-line dictionary*. URL: <https://itwewina.altlab.app> – Code: <https://github.com/UAlbertaALTLab/morphodict>

Bowers, Dustin, **Antti Arppe** & Jordan Lachler. *Finite-state transducer-based computational model of Odawa morphology* (under development). URL: <https://github.com/giellalt/lang-otw>

Cox, Christopher, Josh Holden, Mans Hulden, Miikka Silfverberg, Jordan Lachler, Sjur N Moshagen, Trond Trosterud, Karoline Antonsen & **Antti Arppe** (under development). *Finite-state transducer-based computational model of Tsuut'ina morphology*. URL: <https://github.com/giellalt/lang-srs>

Lachler, Jordan, Lene Antonsen, Trond Trosterud & **Antti Arppe**. *Finite-state transducer-based computational model of Northern Haida morphology* (under development). URL: <https://github.com/giellalt/lang-hdn>

Statistical analysis:

Antti Arppe, Peter Hendrix, Petar Milin, R. Harald Baayen, Tino Sering & Cyrus Shaoul. (2015) *ndl v.0.2.17*. URL: <https://cran.r-project.org/web/packages/ndl/>

Antti Arppe. (2014) *polytomous v.0.1.6*. URL: <https://cran.r-project.org/web/packages/polytomous/>

Shaoul, Cyrus, **Antti Arppe**, Peter Hendrix, Petar Milin & R. Harald Baayen. (2014) *ndl v.0.2.16*. URL: <https://cran.r-project.org/web/packages/ndl/>

CURRICULUM VITAE

Antti Arppe (2012). *polytomous: Polytomous logistic regression for fixed and mixed effects*. R package version 0.1.4. URL: <http://CRAN.R-project.org/package=polytomous>

Antti Arppe, Petar Milin, R. Harald Baayen and with contributions from Peter Hendrix (2011). *ndl: Naive Discriminative Learning*. R package version 0.1.1. URL: <http://CRAN.R-project.org/package=ndl>

OTHER PUBLICATIONS TO THE GENERAL PUBLIC

ARTICLES AND LETTERS-TO-THE-EDITOR

Niukkuus ja epävarmuus kannustavat? [Do scarcity and insecurity motivate?] *Helsingin Sanomat*, 7 April 2005.

Onko makkaratehdas paras esikuva yliopistoille? [Is the sausage factory the best model for universities?] *Kanava*, 4-5/2005, 2 June 2005.

Sopiiko markkinatalouden malli julkiselle sektorille? [Does the market model suit the public sector?] *Helsingin Sanomat*, 16 October 2006. [together with Tiina Arppe, Jussi Pakkasvirta and Martti Vainio].

Tieteellinen omavaraisuus turvattava [Scientific self-sufficiency must be safe-guarded], *Suomen Kuvalehti*, 3 November 2006. [together with Tiina Arppe, Jussi Pakkasvirta & Martti Vainio].

INVITED TALKS AND OTHER PRESENTATIONS

- 9 May 2023 Finding words that aren't there – Using word embeddings to improve dictionary search for low-resource languages. *TÜling Lecture Series*, Department of Linguistics, Tartu University, Estonia.
- 7 December 2022 Finding words that aren't there – Using word embeddings to improve dictionary search for low-resource languages. Department of Linguistics, University of British Columbia.
- 21 October 2022 Language Technologies Panel, *54th Algonquian Conference*, University of Colorado, Boulder, 21 October 2022.
- 7 October 2022 Finding words that aren't there – Using word embeddings to improve dictionary search for low-resource languages. *Linguistics Colloquium*, Department of Linguistics, University of Alberta.
- 14 September 2022 Finding words that aren't there – Using word embeddings to improve dictionary search for low-resource languages. *OFAI 2022 Lecture Series*, Austrian Research Institute for Artificial Intelligence (ÖFAI).

CURRICULUM VITAE

- 28 April 2022 Finding words that aren't there. *Helsinki Digital Humanities Research Seminar*, Department of Modern Languages, University of Helsinki.
- 6 December 2019 21st century language technology and tools meet 21st century challenges and opportunities, *Language Technology for All (LT4ALL)*, UNESCO Headquarters, Paris, France.
- April 2019 Language Technology – A Canadian Perspective. *Is it possible to create technical tools for Indigenous languages? Special Event at the 18th Session of the United Nations Permanent Forum on Indigenous Issues (UNPFII)*, 24 April 2019, UN Headquarters, New York, United States.
- 31 May 2016 Turning language documentation into reader's and writer's software tools. *National Research Council*, Ottawa, Canada.
- 26 February 2014 On the original inspirations of Tolkien's languages – Quenya and Finnish. Guest lecture for *Last Alliance*, the UofA's Tolkien Society.
- 5 February 2014 Tolkien's childhood haunts in the Birmingham area, UK. Guest Lecture for *Last Alliance*, UofA's Tolkien Society.
- 13 March 2010 Towards polytomous mixed effects regression analysis and modeling. *Center for Comparative Psycholinguistics (CCP)*, Department of Linguistics, University of Alberta, Canada.
- 26 February 2010 Polytomous logistic regression analysis and modeling of linguistic alternations. *Center for Comparative Psycholinguistics (CCP)*, Department of Linguistics, University of Alberta, Canada.
- 10 February 2010 Exemplars, prototypes, or both – what evidence do corpora contain for the representation of linguistic categories?, *Linguistic Evidence 2010*, University of Tübingen, Germany.
- 21 April 2005 Produktutveckling av språkgranskningsprogram. *Seminarium of språkkontrollprogram* [Seminar on Proofing tools], April 21-22, 2005. *Nordisk språkråd*, Hotell Kalkstrand, Pargas, Finland.
- 18 October, 2003 Varför är språkkontrollprogram sådana som de är? Hur kunde de bli bättre eller nyttigare? Seminar of Nordic language councils on language technology, October 17-19, 2003. *Nordisk språkråd*, Schæffersgården, Copenhagen, Denmark.
- 27 March 2003 What has worked and what hasn't in commercializing language technology – a Finnish experience. Information Day: "Language Technology: Technological Innovations and Business Solutions" (Wrap-up seminar of the *Greek national Euromap project*). Amphitheatre "Zerva" National Documentation Centre, Vas. Constantinou 48 av., Athens, Greece. URL: http://www.ilsp.gr/hope_files/en/imerida27_03_03.htm

CURRICULUM VITAE

- 6 February 2003 Bruket av finska verbsynonymer i praktiken enligt korpuslingvistiska och psykolingvistiska metoder. Doctoral seminar, *Department of Nordic languages*, University of Helsinki.
- 27 January 2000 Humanistisesta perustutkimuksesta kaupallisesti menestyksekkääseen liiketoimintaan. Ohjelmistojen ja add-in-komponenttien lisensointi kansainvälisille ohjelmistotaloille. Yritysesimerkki: Lingsoft Oy [From basic research in the humanities to commercially successful business. The licensing of software and add-on components to international software companies. Corporate Example: Lingsoft, Inc.]. Ohjelmistojen tuotteistaminen -kurssi [Commercialization of software course], *Department of Computer Science*, University of Helsinki.
- 7 May 1999 Den långa vägen från lingvistisk teori till kommersiell verksamhet (The long road ...), *Language technology day*, University of Gothenburg, Department of General Linguistics.
- 29 April 1999 Den långa vägen från lingvistisk teori till kommersiell verksamhet (The long road ...), *Department of General Linguistics*, University of Lund.
- 24 March 1999 Den långa vägen från lingvistisk teori till kommersiell verksamhet (The long road from linguistic theory to commercial activity), *Department of General Linguistics*, University of Uppsala.
- 22 March 1999 Final panel, *Temadag om datorstödd språkgranskning* (Theme day of computer-aided language checking), *Royal Institute of Technology*, Stockholm, NADA (Numerisk analys och datalogi).

RESEARCH AWARDS, ALLOCATIONS, GRANTS AND ROLES

- 2023 Social Sciences and Humanities Research Council (SSHRC); Partnership Grant (435-2023-0474); *Mapping and modelling grammatical person and voice in Ojibwe*; **Co-Investigator** (Principal Investigator: Ass't Professor Christopher Hammerly, University of British Columbia): \$263,358 (2023-2026).
- 2019 University of Alberta, Office of the Vice-President (Research), Faculty of Arts, and Department of Linguistics; Contributions to SSHRC Partnership Grant (895-2019-1012): *21st Century Tools for Indigenous Languages*; **Principal Investigator**: \$755,555.22 (\$466,094.72 CAD in cash, and \$289,460.50 in kind) (2019-2026)
- 2019 Social Sciences and Humanities Research Council (SSHRC); Partnership Grant (895-2019-1012); *21st Century Tools for Indigenous Languages*; **Project Director/Principal Investigator** (31 co-investigators and collaborators): \$2,500,000 (2019-2026/7).

CURRICULUM VITAE

- 2018 Social Sciences and Humanities Research Council (SSHRC); Partnership Grant Letter-of-Intent/Stage 1 (895-2018-0026); *21st Century Tools for Indigenous Languages*; **Principal Investigator** (multiple co-investigators and collaborators): \$20,000 (2018-2019).
- 2017 Kule Institute for Advanced Study (KIAS); Research Cluster Grant: *Deep Learning for Sound Recognition*; **Co-Investigator** (Principal Investigator: Michael Frishkopf): \$55,000 (2017-2020).
- 2016 Social Sciences and Humanities Research Council (SSHRC); Connections Outreach Grant (611-2016-0207): *21st Century Tools for Indigenous Languages*; **Principal Investigator** (co-investigators: Jordan Lachler, Christopher Cox [Carleton University], Mans Hulden [University of Colorado, Boulder], and a number of collaborators): \$50,000 (2016-2018)
- 2016 University of Alberta; Indigenous Community Engagement, Research and Learning: *Responding to the Call: Supporting Indigenous Language Sustainability in Alberta*; **Co-Investigator** (Principal Investigator: Jordan Lachler): \$50,000 (2016-2018)
- 2015 University of Alberta, Support for the Advancement of Scholarship (SAS) fund: participation and travel grant to QITL-5, University of Tübingen, Germany: \$3000.
- 2015 Social Sciences and Humanities Research Council (SSHRC); Savoir (435-2014-1199): *Une infrastructure numérique pour que vivent les langues algonquiennes: Dictionnaires et Atlas linguistique*; **Collaborator** (Principal Investigator: Marie-Odile Junker [Carleton University]): \$500,000 (2014-2017)
- 2015 Kule Institute for Advanced Study (KIAS); Research Cluster Grant: *21st Century Tools for Indigenous Languages*; **Principal Investigator** (co-investigator: Jordan Lachler, and a number of collaborators): \$100,000 CAD (2015-2018)
- 2014 Social Sciences and Humanities Research Council (SSHRC); Partnership Development Grant (890-2013-0047): *21st Century Tools for Indigenous Languages*; **Principal Investigator** (co-investigators: Jordan Lachler, Juhani Järvikivi, Timothy Mills, Trond Trosterud [University of Tromsø], and a number of collaborators): \$198,702 CAD (March 2014 – March 2017)
- 2014 University of Alberta, Office of the Vice-President (Research); In-cash contribution to SSHRC PDG (890-2013-0047): *21st Century Tools for Indigenous Languages*; **Principal Investigator**: \$40,000 CAD (2014-2016)
- 2014 University of Alberta, Endowment for the Future (EFF) – CERF special: *Technological infrastructure for “21st Century Tools for Indigenous Languages”*; **Principal Investigator**: \$35,000 (April 2014 – June 2019)
- 2013 Killam Research Fund, University of Alberta; Cornerstones Grant: *21st Century Tools for Indigenous Languages*; **Principal Investigator** (co-

CURRICULUM VITAE

- investigators: Juhani Järvikivi, Jordan Lachler, Timothy Mills); \$50,000 CAD (September 2013 – January 2015)
- 2013 Distinguished Visitor Fund: Professor Trond Trosterud, University of Tromsø, **Principal Organizer**: \$7,750 CAD (October 2013)
- 2009 University of Helsinki, three-year full-time **Postdoctoral Fellowship** (overall salary of ~116,250€) + 3 x 3,000€/year = 9,000€ grant for research and travel costs (January 2010 – July 2012)
- 2009 Academy of Finland, *Grant for research and researcher training abroad* (for visiting University of Alberta, Canada, 2010-2011): 16,000€
- 2009 Finnish Ministry of Education: *FIN-CLARIN (Integrating the language resources in Finland)*; **Co-applicant** (Principal Applicant: Kimmo Koskenniemi): 900,000€ (2010-2013)
- 2009 University of Helsinki: *Development of a prototype for resource access and permission management*; **Co-applicant** (Principal applicant: Kimmo Koskenniemi) €100,000 (2009)
- 2007 University of Helsinki, Grant for finishing Ph.D. Dissertation, €3,600.
- 2007 E. A. Saarimaa fund award (€250) for an academic article concerning the Finnish language (Arppe, Antti, 2006. Frequency Considerations in Morphology, Revisited – Finnish Verbs Differ, Special Supplement to *SKY Journal of Linguistics*)
- 2004 *Claris Exemplis* diploma for initiatives in organizing courses on practical professional working skills (e.g., commercialization of technology; project management), Faculty of Arts, University of Helsinki
- 2003 LANGNET – Finnish graduate school in language studies, funded by the Finnish Ministry of Education: four-year full-time postgraduate/doctoral scholarship (overall salary of ~€90,000 during 2004-2007)
- 2000 – 2002 GILTA project, funded by TEKES – the Finnish Funding Agency for Technology and Innovation: three-year full-time postgraduate project researcher salary (€67,500)
- 2000 – 2012 Several travel grants from the Chancellor of the University of Helsinki (€900/2006; €1400/2005; €210/2004) and the LANGNET doctoral school (€1234/2007; €1400/2005).

PARTICIPATION IN AND MANAGEMENT OF RESEARCH AND DEVELOPMENT PROJECTS

Co-ordination of and participation in research projects:

CLARIN project (EU-ESFRI funded 2008-2010). Department of General Linguistics (Language Technology): **Research Co-ordinator** [tutkimuskoordinaattori] in Work Package 7 (WP7) – Intellectual Property Rights and Business Models; Deputy member of

CURRICULUM VITAE

Executive Board of CLARIN project 2008-2009; co-ordination of CLARIN activities in Finland (FIN-CLARIN).

GILTA project (USIX/TEKES). Department of General Linguistics, University of Helsinki (other participating institutions: Signal Processing Laboratory, and Pori School of Technology and Economics, Tampere University of Technology; and Laboratory of Information Systems, Åbo Akademi University): **Deputy member of the steering committee** 2000-2002

Management of and participation in product development projects:

Finnish speller, hyphenator and inflecting thesaurus (Lingsoft): participation in product specification and testing 1995; product management 1996-1998

Swedish speller, hyphenator and inflecting thesaurus (Lingsoft): head of product development 1996-1998.

Swedish grammar checker (Lingsoft): head of product development 1997-1998.

Norwegian (Bokmål & Nynorsk) speller, hyphenator and inflecting thesaurus (Lingsoft): head of product development 1996-1998

Danish speller, hyphenator and inflecting thesaurus (Lingsoft): head of product development 1996-1998

Finnish, Norwegian (Bokmål) and Danish grammar-checkers (Lingsoft): participation in product specification, project planning and budgeting 2000.

SUPERVISION AND TEACHING

SUPERVISION OF THESES, DISSERTATIONS, AND STUDENT RESEARCH

Advising/Hosting of Post-Doctoral Researchers:

2020- Ph.D **Daniel Hieber** (Advisor)

2020- Ph.D **Joshua Holden** (Advisor)

2015-2019 Ph.D **Miikka Silfverberg** (Local host during visits to UAlberta)

2015-2017 Ph.D **Dustin Bowers** (Advisor)

Supervision or Examination of Ph.D students in Linguistics:

2023- M.A. **Cameron Duval** (Supervisor)

2021 Ph.D **Antti Kanner**, Department of Finnish and Swedish Languages and Literature and Fenno-Ugric Studies, University of Helsinki (External Examiner)

CURRICULUM VITAE

- 2021 Ph.D Candidate **Rodrigo Becerra Parra** (Ph.D candidacy committee member, 19 Feb 2021).
- 2021- M.A. **Olga Kriukova**, University of Saskatchewan (Co-supervisor; Supervisor: Prof. Olga Lovick, USask)
- 2020-2023 Ph.D **William Abbott Lane**, Charles Darwin University, Darwin, Australia (Co-supervisor; Principal supervisor, Prof. Steven Bird)
- 2020- M.A. **Rae Anne Claxton** (Supervisor)
- 2019-2020 Ph.D **Hanna Westerlund**, Department of Modern Languages, University of Finland (External Examiner, on two occasions)
- 2017-2020 Ph.D **Brian Rusk** (Ph.D. candidacy committee member)
- 2017- Ph.D Candidate, B.Sc. (Honors) **Atticus Harrigan** (Supervisor)
- 2016-2020 Ph.D **Krithika Govindarajan** (Supervisory committee member)
- 2015-2022 Ph.D **Katherine Schmirler** (Supervisor)
- 2014- Ph.D Candidate, M.A. **Claudia Heinrich** (Supervisory committee member)
- 2014-2019 Ph.D **Isabell Hubert** (Supervisory committee member)
- 2012-2013 Ph.D. **Philip Dilts** (Co-supervisor)
- 2010-2013 Ph.D **Dana Abdulrahim** (Supervisory committee member)
- 2013 Ph.D **Hien Pham** (Ph.D candidacy committee member)
- 2012-2019 M.A. **Dagmara Dowbor** (Supervisory committee member)
- 2012-2014 Ph.D **Doug Inglis** (Supervisory committee member)
- 2012-2017 Ph.D **Audra Phillips** (Supervisory committee member)
- 2015 Ph.D **Vincent Porretta** (Supervisory committee member)
- 2012-2014 M.A. **Linda Walton** (Supervisor)
- 2012-2014 Ph.D **Guangting Wang** (Co-supervisor; Supervisor: R. Harald Baayen)
- 2013-2017 Ph.D **Caelan Marrville** (Co-supervisor; Supervisor: Juhani Järviäivi)
- 2013-2017 Ph.D **Kaidi Lõo** (Supervisory committee member)
- 2013-2019 Ph.D **Abdul-Aziz Alarifi** (Supervisory committee member)
- 2013-2014 Ph.D **Conor Snoek** (Research Assistantship supervisor)

Supervision of M.A./M.Sc. students in Linguistics:

- 2020-2023 M.A. **Karoline Antonsen** (Supervisor; Co-supervisor Dr. Joshua Holden)
- 2023 M.A. **Kirandeep Kaur**, University of Lethbridge (External examiner)
- 2022- B.A. (Honors) **Daniel Dacanay** (Supervisor)

CURRICULUM VITAE

- 2021-2023 M.A. **Dominik Kadlec**, University of Lethbridge (Co-supervisor; Supervisor: Prof. Inge Genee, ULethbridge)
- 2021-2022 B.A. **Ruben Mögel** (Supervisor)
- 2015-2016 B.Sc. (Honors) **Atticus Harrigan** (Supervisor; fast-tracked to our Ph.D program)
- 2001-2002 M.A. (General linguistics) **Camilla Magnusson**, Department of General Linguistics, University of Helsinki (practical supervisor)

Supervision of B.A./B.Sc. students in Linguistics:

- 2019-2022 B.A. (Honors) **Daniel Dacanay** (Supervisor, RA-ship supervisor)
- 2017-2018 B.A. (Honors) **Tanzi Reule** (Supervisor, RA-ship supervisor)
- 2015-2016 B.A. (Honors) **Megan Bontogon** (Supervisor, RA-ship supervisor)
- 2014-2015 B.A. (Honors) **Atticus Harrigan** (RA-ship supervisor)
- 2013 B.A. (Honors) **Mildred Lau** (Supervisor)

Supervision of graduate/undergraduate students and professionals in Computer Science:

- 2023 M.Sc. **Pranjal Naringrekar** (interim full-time software developer)
- 2023 CMPUT 401 capstone course, W2023: team of 6 CS students: **Saba Gul, Arakkal Sureshabu Sooraj, Jaden Huang, Ke Li, Yu Liu, Matthew Shocrylas** (Vocabulary Explorer)
- 2023- **Oleksandr Kasych** (software development intern)
- 2023 **Bayley Sapara** (software development intern)
- 2022-2023 **Ansh Dubey** (software development intern)
- 2022 CMPUT 401 capstone course, W2022: team of 6 CS students: **Zijie Tan, Fadi Bonilla, Cenab Bora, Uladzimir Bondarau, James Schaefer-Pham, Nicholas DeMarco** (paradigm panes)
- 2021 M.Sc. (Computer Science) **Arnob Mallik** (Thesis examination committee, 23 April 2021; Principal supervisor: Prof. Gregerz Kondrek, Computing Science, University of Alberta).
- 2020-2023 B.Sc. **Jolene Poulin** (full-time software developer)
- 2020-2021 M.Sc. **Andrew Neitsch** (full-time senior software developer)
- 2020 M.Sc. **Shüyu (Serena) Zhao**: (senior software development intern)
- 2020 **Aida Radu**: (junior software development intern)
- 2020 **Kobe Amoh**: (junior software development intern)
- 2019-2020 B.Sc. **Siyuan (Matt) Yan** (software development intern)

CURRICULUM VITAE

- 2019 CMPUT 401 capstone course, W2019: team of 3 CS students: **Ryan Pineo, Sarah Shimizu, Wilton (Weichen) Wang** (morphologically intelligent dictionary)
- 2018-2021 M.Sc. **Eddie Antonio Santos** (full-time software development lead; funded by the National Research Council)
- 2002-2005 M.Sc. (Engineering) **Heikki Vesalainen**, Laboratory of Computer and Information Science, Department of Computer Science and Engineering, Helsinki University of Technology (Instructor)
-

TEACHING

Teaching at university level: (S)pring – (F)all – (W)inter:

Finland:

Cross-Cultural Business (4 [ECTS] credits), Institute of Strategy and International Marketing, Helsinki University of Technology: F1993.

Teknologian kaupallistaminen (Kaupallinen kieliteknologia) [Commercialization of (language) technology] (2-4 credits), Language technology, Department of General Linguistics, University of Helsinki: F1999, F2000, F2001, F2002; University of Oulu: F2002 (1-day intensive package).

Projektityöskentely [Working in projects] (4 credits), Language technology, Department of General Linguistics, University of Helsinki / LANGNET graduate school of language studies: S2002, S2003, S2004, F2004, F2008.

UNIX:in perusteet ja tekstikorpusten tietojenkäsittely [Basic UNIX programming and text corpus processing (6 credits), Department of Informatics, University of Tampere/KIT network: F2002.

Korpuset ja kieli [Corpora and language] (4 credits), Language technology / General linguistics, Department of General Linguistics: S2003, S2004.

Kielitieteen lukemattomat klassikot -lukupiiri [Workshop on "Unread Classic Works in Linguistics"], Department of General Linguistics, University of Helsinki: F2003-S2004 (several guest presenters; co-ordination and arrangements together with Reetta Konstenius).

Alberta:

Introduction to Linguistic Analysis (LING101): W2017, F2017

Writing Systems (LING 325) (3 credits), Department of Linguistics, University of Alberta: W2013, W2014, W2015, F2015, F2016, F2017, F2019, F2021, F2023.

Practical Field Lexicography (LING 375), Department of Linguistics, University of Alberta (co-taught with Jordan Lachler and Timothy Mills): W2015.

CURRICULUM VITAE

Linguistics Directed Research I (LING 375): Peggy Jankovic, Megan Bontogon, Atticus Harrigan, Tanzi Reule, Daniel Dacanay, Xin He (Alyssa) Jiang.

Linguistics Directed Research II (LING 475): Megan Bontogon, Atticus Harrigan, Tanzi Reule, Daniel Dacanay.

Corpus Linguistics (LING 519), Department of Linguistics, University of Alberta: F2013, F2014.

Introduction to Statistics with Linguistic Data (LING 523, LING 399/599), Department of Linguistics, University of Alberta: S2013, F2013, F2014, F2015, F2016, F2017, F2019, F2023.

Linguistics of Tolkien (LING 399/599), Department of Linguistics, University of Alberta. (co-taught with Benjamin V. Tucker): F2014.

Quantitative Methods in Linguistics (LING 603) (3 credits), Department of Linguistics, University of Alberta: F2011, W2013, W2014, W2015, W2017, W2021, W2022.

Linguistics Generals Paper I/II (LING 683/694): Caelan Marrville, Jennifer Hinnell, Claudia Heinrich, Isabell Hubert, Katherine Schmirler, Atticus Harrigan.

Computational Linguistics (LING 520), Department of Linguistics, University of Alberta: W2018, F2020, F2021.

Computational Linguistics (LING 520), Department of Linguistics; (MDST 5990) Multidisciplinary Studies, University of Lethbridge: F2021.

Researcher training given in workshop form:

Exploratory univariate analysis. As part of: Introduction to Corpus Linguistics for Slavic Linguists organized by Dagmar Divjak and Steven Clancy, Slavic Cognitive Linguistics Conference, Prague, Czech Republic, October 15, 2009.

Introduction to Modeling Naive Discriminative Learning. 1-week-long 5x3h intensive course within the *STEP@CCP: Spring training in Experimental Psycholinguistics*, 4-8-May 2014, organized by the Centre for Comparative Psycholinguistics, University of Alberta.

A Short Overview and Introduction to Statistics and R, 3-day-long 2.5x3h intensive course within the *STEP@CCP: Spring training in Experimental Psycholinguistics* in May 19-23, 2015, co-taught with Dr. Vincent Porretta, organized by the Centre for Comparative Psycholinguistics at the UofA.

Individual lectures as part of university-level courses:

Kokemuksia [kieliteknologisten] ohjelmistotuotteiden sopimusneuvotteluista [Experiences from licensing negotiations of language-technology software]: Immateriaalioikeudet [Immaterial property rights], Language technology, Department of General Linguistics, University of Helsinki: S2003, S2004.

CURRICULUM VITAE

Monimenetelmäisyys kielitieteessä [Multimethodological perspectives in linguistics] - Kielitieteellistä tutkimusmetodiikkaa -kurssin moduuli II [Linguistic research methodology, module II], Department of General Linguistics, University of Helsinki / LANGNET graduate school: F2007.

Exploiting Multiple Methods in Linguistics: Corpora and Language Dynamics, Helsinki, 12-14 April 2010, LANGNET graduate school, Finland & Doctoral School of Humanities, Tallinn University, Estonia.

Theory vs. practice – some well as well as less known characteristics of a morphologically rich language – Finnish. 16 November 2010 & 17 March 2011. Introduction to Linguistic analysis (LING 101). Department of Linguistics, University of Alberta.

Finnish is written as it is spoken – and spoken as it is written – or is it? 17 March 2011, Writing systems (LING 325). Department of Linguistics, University of Alberta.

Mixed-effects logistic regression with polytomous outcomes, 2 December 2010. Mixed-effects modeling. Department of Linguistics, University of Alberta.

OTHER ACADEMIC AND PROFESSIONAL ACTIVITIES

Advisory/Editorial Board Memberships for Academic Book Series

Studies in Corpus Linguistics, John Benjamins Publishing Company: Advisory Board Member, 2016-

Reviews of manuscripts for scientific journals:

Applied Linguistics: 2014

Canadian Modern Languages Review: 2022, 2023

Cognitive Linguistics: 2012

Corpus Linguistics and Linguistic Theory: 2013, 2015

Frontiers in Linguistics: 2022

International Journal of American Studies: 2017

Language: 2011

Language Resources and Evaluation: 2019 x 3; 2020

Mental Lexicon: 2014, 2016

Natural Language Engineering: 2003-2004; 2011

Papers of the Algonquian Conference: 2021

SKY Journal of Linguistics: 2011

Software: Practice and Experience: 2010

CURRICULUM VITAE

TAL Journal (Special Issue on *NLP for under-resourced languages*): 2018

Virittäjä (Finland): 2008

Reviews of applications for research project grants:

National Research Foundation (United States): 2019

Estonian Research Council: 2013, 2019

Social Sciences and Humanities Research Council (SSHRC): Insight Grant application (2014); Partnership Grant Midterm Review Committee (2023-)

University of Leuven, Research Council: 2011

Icelandic Research Fund: 2010

Invited participation in research strategy/planning events:

2019 Language Technologies for All (LT4All) conference, 4-6 December 2019, UNESCO Headquarter (Paris, France).

2019 18th Session of the United Nations Permanent Forum on Indigenous Issues (UNPFII), 22 April – 3 May 2019, UN Headquarters, New York United States.

2015 Trans-Atlantic Platform during 1-2 June 2015 (hosted by SSHRC at Congress 2015 in Ottawa) in three short workshops: “Supporting Research into Social and Cultural Identities: Diversity, (In)equality and Differences”, “Supporting Transformational Research on the Environment”, and “Supporting Research to Build Resilient and Innovative Societies”.

Chair/Principal organizer of academic conferences and scientific events:

2023 55th Algonquian Conference (AC55), University of Alberta, 19-22 October 2023 (3-day conference with ~50 participants).

2022 Computational Workshop on Dene languages, University of Saskatchewan, SK, 8-11 November 2022. **Co-Organizer** (Principal organizer: Dr. Olga Lovick; 4-day workshop with altogether 5+ participants).

2022 Computational Workshop on Plains/Western Cree, First Nations University of Canada, Regina, SK, 26-30 September 2022. (4-day workshop with altogether 7 participants).

2018 50th Algonquian Conference (AC50), University of Alberta, 25-28 October 2018. (3-day conference with ~50 participants).

2018 Computational Workshop on Plains Cree, University of Alberta, 4-6 July 2018. (2-day workshop with 6 participants)

CURRICULUM VITAE

- 2017 Computational Workshop on Plains/Western Cree, University of Alberta, Edmonton, 15-17 May 2017. (3-day workshop with altogether 8 participants).
- 2017 3rd Computational Workshop on Indigenous Languages (III Dene CWIL), University of Colorado, Boulder, 28-30 April 2017. (3-day workshop with altogether 11 participants).
- 2016 2nd Computational Workshop on Indigenous Languages (CWIL-2016), University of Alberta, Edmonton, 2-14 May 2016. (2-week workshop with altogether 15 participants)
- 2016 Computational Workshop on Plains and Woods Cree, University of Alberta, Edmonton, March 2016. (4-day workshop with 8 participants).
- 2015 Computational Workshop on Ojibwe, University of Alberta, December 2015. (5-day workshop with 6 participants)
- 2015 Computational Workshop on Tsuut'ina, University of Alberta, Edmonton, October 2015. (3-day workshop with 6 participants)
- 2015 1st Computational Workshop on Indigenous Languages (CWIL-2015), University of Alberta, Edmonton, 6-20 June 2015. (2-week workshop with altogether 15 participants)
- 2014 SSHRC/Algonquian/Cree mini-summit, University of Alberta, June 2014. (1-week workshop with 15 participants)
- 2008 3rd Workshop on Quantitative Investigations in Theoretical Linguistics (QITL-3), Helsinki, 2-4 June 2008: **Co-Chair & Secretary of the Organizing Committee.**

Memberships of organizing committees for academic conferences and scientific events:

- 2023- 7th Workshop on Computational Methods for Endangered Languages (ComputEL-7). In: 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024), Malta, 21-22 March 2024.
- 2023 6th Workshop on Computational Methods for Endangered Languages (ComputEL-6). University of Hawai'i (virtual), 5-6 March 2023.
- 2021-2022 11th Language Resources and Evaluation Conference 2023: **Area Co-Chair** for Endangered Languages.
- 2022 5th Workshop on Computational Methods for Endangered Languages (ComputEL-5). 60th Conference of the Association of Computational Linguistics (ACL), Dublin, Ireland, 26-27 May 2022.
- 2021 4th Workshop on Computational Methods for Endangered Languages (ComputEL-4). University of Hawai'i, Manoa, Honolulu, Hawai'i, 2-3 March 2021.

CURRICULUM VITAE

- 2019 3rd Workshop on Computational Methods for Endangered Languages (ComputEL-3). University of Hawai'i, Manoa, Honolulu, Hawai'i, 26-27 February 2019.
- 2017 2nd Workshop on Computational Methods for Endangered Languages (ComputEL-2). University of Hawai'i, Manoa, Honolulu, Hawai'i, 6-7 March 2017.
- 2013 International Cognitive Linguistics Conference (ICLC), University of Alberta, Edmonton, June 2013.
- 2010 Re-thinking synonymy: semantic sameness and similarity in languages and their description (SKY Symposium 2010), Helsinki, 27-29 October 2010.
- 2009 1st Networking Event for European Research Infrastructures (NEERI-09), Helsinki 30 September – 2 October 2009.
- 2005 Finite State Methods in Natural Language Processing (FSMNLP-2005). Helsinki, 1-2 September 2005.
- 2004 20th Scandinavian Conference of Linguistics (SCL-20), Helsinki, 7-9 January 2004 (responsible for workshops).

Memberships of scientific/program committees/reviewing for academic conferences:

- 2023 6th Workshop on Computational Methods for Endangered Languages (ComputEL-6), March 2023.
- 2022 Language Resources and Evaluation (LREC), LREC, Marseille, France, June 2022
- 2022 SIGUL-2022 Workshop, Language Resources and Evaluation (LREC), Marseille, France, June 2022.
- 2022 5th Workshop on Computational Methods for Endangered Languages (ComputEL-4), May 2022.
- 2021 Empirical Methods in Natural Language Processing (EMNLP 2021), Punta Cana, Dominican Republic, November 2021.
- 2021 4th Workshop on Computational Methods for Endangered Languages (ComputEL-4), March 2021.
- 2020 28th International Conference on Computational Linguistics in 2020 (COLING 2020), Barcelona, Spain, 13-18 September 2020.
- 2020 1st Joint SLTU-CCURL Workshop (at LREC), Marseille, France, May 2020 (cancelled/virtualized)
- 2020 Language Resources and Evaluation (LREC), LREC, Marseille, France, May 2020 (cancelled/virtualized)
- 2019 Language Technologies for All (LT4All) conference, 4-6 December 2019, UNESCO Headquarters (Paris, France).

CURRICULUM VITAE

- 2019 3rd Workshop on Computational Methods for Endangered Languages (ComputEL-3), February 2019.
- 2018 All Together Now? Computational Modeling of Polysynthetic Languages (PolySynth 2018) (at COLING 2018), 25-26 August 2018, Santa Fe, New Mexico.
- 2018 27th International Conference on Computational Linguistics (COLING-2018), August 21-25, 2018, Santa Fe, New Mexico, USA.
- 2018 Language Resources and Evaluation (LREC), ELRA, Miyazaki, Japan, May 2018
- 2017 2nd Workshop on Computational Methods for Endangered Languages (ComputEL-2), March 2017.
- 2016 26th International Conference on Computational Linguistics (COLING-2016), 11-16 December 2016, Osaka, Japan.
- 2016 Language Resources and Evaluation (LREC), ELRA, Portoroz, Slovenia, May 2016.
- 2016 2nd New Ways of Analyzing Syntactic Variation (NWASV2), University of Ghent, May 2016.
- 2015 Sixth Conference of Quantitative Investigations in Theoretical Linguistics (QITL-6), University of Tübingen, Germany, November 2015.
- 2013 International Cognitive Linguistics Conference, University of Alberta, Edmonton, June 2013.
- 2013 Fifth Conference of Quantitative Investigations in Theoretical Linguistics (QITL-6), University of Leuven, Belgium, September 2013.
- 2011 Fourth Conference of Quantitative Investigations in Theoretical Linguistics (QITL-4), Humboldt University in Berlin, 28-31 March 2011.
- 2010 Re-thinking synonymy: semantic sameness and similarity in languages and their description (SKY Symposium 2010), Helsinki, 27-29 October 2010.
- 2009 1st Networking Event for European Research Infrastructures (NEERI-09), Helsinki, 30 September – 2 October 2009.
- 2009 17th Nordic Conference on Computational Linguistics (NoDaLiDa 2009), Odense, 14- 16 May 2009.

Internal service and administration:

- 2022- Undergraduate committee, Department of Linguistics, University of Alberta: Member
- 2020- Indigenous Language Sustainability (ILS) Committee, Department of Linguistics, University of Alberta: Member 2020-2022; Chair 2022-

CURRICULUM VITAE

- 2019-2020 Advisory Selection Committee, tenure-track Assistant Professor in Indigenous Language Sustainability (ILS), Department of Linguistics, University of Alberta.
- 2017-2022 Web Committee, Department of Linguistics, University of Alberta.
- 2016 President's *Ad hoc* Committee on Indigenous Language Sustainability, University of Alberta: Member 2016
- 2013-2015 Arts Council for Technology & Innovation (ACTI), Faculty of Arts, University of Alberta
- 2012-2015 Graduate Committee, Department of Linguistics, University of Alberta: preparation of the selection of departmental graduate students and graduate studies policy as well as screening of awards
- 2012-2018 Library Liaison
- 2001-2006 Executive Board, Department of General Linguistics, University of Helsinki: Deputy member 2001-2003; Member 2004-2006.

Positions of trust in the academic world

- 2019- ACL-SIGEL: Association of Computational Linguistics (ACL) Special Interest Group on Endangered Languages, Interim President 2019-2021, President 2021-2023; Past President 2023-.
- 2016- International Advisory Board of the ELRA [European Language Resources Association] Committee on Less-Resourced Languages: Member 2016-
- 2016- Interdisciplinary Committee on Applied Graduate Statistics, University of Alberta. Department Representative (Linguistics) 2016-
- 2013- Canadian Institute/Initiative of Nordic Studies: Member of the Board 2013-; Treasurer 2013-2016; Chair 2016-2020, Member of the Board 2020-
- 2001-2009 Linguistic Association of Finland (SKY ry.): Auditor 2001-2006; Deputy board member 2007; Treasurer 2008–2009
- 2007-2009 Finnish Association of Cognitive Linguistics (FiCLA ry.): Board member 2007-2009

Memberships of academic professional organizations:

- 2019- ACL-SIGEL
- 2014- Association of Computational Linguistics (ACL)
- 2001- Finnish Association of Cognitive Linguistics (FiCLA)
- 1995- Linguistic Association of Finland [Suomen kielitieteellinen yhdistys] (SKY)