CV (last updated March 2022)

prof.dr. Gertjan van Noord (1961)

Language Technology
Center for Language and Cognition Groningen
University of Groningen
PObox 716, NL 9700 AS Groningen
Netherlands
+31 (0)50-3637811

`G.J.M.van.Noord@rug.nl`
`http://www.let.rug.nl/~vannoord`

Education

Ph.D., 1993, Utrecht University. Diss: *Reversibility in Natural Language Processing*. Advisors: Jan Landsbergen and Jan van Eijck.

M.A., 1987, General Linguistics (major in Computational Linguistics), Utrecht University. *Cum Laude.*

Employment

2017–present University of Groningen. Director of the CLCG.

1/2011–present University of Groningen. Professor in Language Technology.

3/1999–1/2011 University of Groningen. Associate Professor of *Alfa-informatica* (Humanities Computing).

1/1992–3/1999 University of Groningen. Assistent Professor of *Alfa-informatica* (Humanities Computing).

1/1990–1/1991 University of the Saarland, Saarbrücken. Researcher at the Computational Linguistics dept. with prof. Hans Uszkoreit for SFB 314 project on Bidirectional Linguistic Deduction.

9/1987–1/1990 University of Utrecht. Researcher at the General Linguistics dept. for the EC-funded Eurotra project.

Project management and Ph.D.-student supervision

2019–present Supervision of PhD student Lukas Edman

2019–present Supervision of PhD student Prajit Dhar

2019 PaQuUd. Project funded by Clariah.

2019–present Co-supervision of PhD student Ahmet Üstün.

2017-2018 SPOD. Project funded by Clariah.

2015-2016 PaQu in Clariah. Project funded by Clariah for further development of the PaQu tool.

2014-2015 PaQu in Clarin. Project funded by Clarin for the development of a search tool for annotated treebanks.

9/2014-9/2018 Parsing Algorithms for Uncertain Input. Project funded by Nuance Foundation. Applicant and coordinatator. Supervision Ph.D.-student Rob van der Goot (per september 2014).

11/2013-11/2017 QTLeap. EU FP7 project. Co-applicant and coordinator University of Groningen. Supervision Ph.D.-student Dieke Oele.

09/2012-09/2016 Supervision Ph.D.-student Simon Suster.

10/2010-09/2012 Clarin TTNWW: TST-Tools voor het Nederlands als Webservices in een Workflow. Co-applicant and coordinator University of Groningen.

09/2009-09/2011 Supervision Ph.D.-student Gideon Kotzé.

09/2009-09/2011 STEVIN Paco-MT project. Coordinator University of Groningen.

09/2008-09/2012 Supervision Ph.D.-student Kostodin Cholakov.

05/2008-05/2011 STEVIN Duoman project. Co-applicant and coordinator University of Groningen.

05/2008-05/2011 STEVIN Daisy project. Co-applicant and coordinator University of Groningen. Ph.D.-student Daniël de Kok.

09/2007-09/2011 Supervision Ph.D.-student Yan Zhao.

09/2007-09/2011 Funding for continuation of NWO PIONIER. Ph.D.-student Barbara Plank, Postdoc Jörg Tiedemann.

11/2006-11/2009 STEVIN LASSY project. Main applicant and Principal investigator. Student-assistents (annotators) and Postdoc Erik Tjong Kim Sang.

09/2005-09/2009 Supervision Ph.D.-student Tim van de Cruys.

11/2004-11/2006 STEVIN IRME project. Co-applicant and coordinator University of Groningen. Postdoc Begona Villada.

11/2004-11/2006 STEVIN D-Coi project. Co-applicant and coordinator University of Groningen. Student-assistents (annotators) and scientific programmer Geert Kloosterman.

09/2003-12/2007 Supervision Ph.D.-student Francisco Borges.

11/1999-11/2005 NWO PIONIER 'Algorithms for Linguistic Processing'. Principal investigator. 4 Ph.D.-students (Tanja Gaustad, Begoña Villada, Robbert Prins, Leonoor van der Beek) and Post-docs (Mark-Jan Nederhof, Robert Malouf, Jan Daciuk, Tony Mullen).

1/1995-6/2000 Theme-group leader of the NWO Priority Programme on Language and Speech Technology. 1 Ph.D.-student (Rob Koeling) and 1 Post-doc (Mark-Jan Nederhof).

Promotor

- 2019. Rob van der Goot. Normalization and Parsing Algorithms for Uncertain Input.

- 2018. Dieke Oele. Automated Translation with Interlingual Word Representations.

- 2016. Simon Suster. Empirical studies on word representations.

- 2013. Gideon Kotzé. Complementary Approaches to Tree Alignment: combining statistical and rule-based methods.

- 2013. Daniël de Kok. Reversible Stochastic Attribute-Value Grammar. *Cum laude.*

- 2012. Kostadin Cholakov. Lexical Acquisition for Computational Grammars - A Unified Model.

- 2011. Barbara Plank. Domain Adaptation for Parsing. *Cum laude.*

Co-promotor

- 2010. Tim van de Cruys. Mining for Meaning. The Extraction of Lexico-Semantic Knowledge from Text.

- 2010. Francisco Borges. Parse Selection with Support Vector Learning.

- 2005. Leonoor van der Beek. Topics in Corpus-Based Dutch Syntax.

- 2005. Robbert Prins. Finite-State Pre-Processing for Natural Language Analysis.

- 2005. Begoña Villada. Data-driven Identification of Fixed Expressions and Their Modifiability.

- 2004. Tanja Gaustad. Linguistic Knowledge and Word Sense Disambiguation.

- 2002. Rob Koeling. Dialogue-based Disambiguation: Using Dialogue Status to Improve Speech Understanding.

PhD Thesis Examiner (selection)

- 2021. Michael Schlichtkrull. Incorporating Structure into Neural Models for Language Processing. UvA.

- 2021. De Passos Morgado da Costa Luis Guilherme. Using Rich Models of Language in Grammatical Error Detection. Nanyang Technological University, Singapore.

- 2018. Ke Tran. Predicting and Discovering Linguistic Structure with Neural Networks. UvA Amsterdam.

- 2015. Oscar Strik. Modelling analogical change. Groningen.

- 2015. Valerio Basile. From Logic to Language: Natural Language Generation. Groningen.

- 2014. Jan de Belder. Linear Programming Optimization for Natural Language Processing Problems. KU Leuven. KU Leuven.

- 2013. Fridah Katushemererwe. Computational morphology and Bantu language learning: an implementation for Ruyakitara. Groningen.

- 2012. Vincent Van Asch. Domain Similarity Measures: On the use of distance metrics in natural language processing. Universiteit Antwerpen.

- 2010. Lionel Nicolas. Efficient Production of Linguistic Resources: The Victoria Project. Université Nice Sophia Antipolis.

- 2008. Wouter van Atteveldt. Semantic Network Analysis. Techniques for Extracting, Representing and Querying Media Content. VU Amsterdam

Research Evaluation

- 2018. Habilitation Defense of Benoit Sagot. Member of the jury. Paris.

- 2015. Evaluation of INRIA theme "Language, Speech and Audio". Paris.

- Various memberships of NWO committees for selection of research grants.

Teaching

- Undergraduate courses on Prolog, Formal Language Theory, Problem Solving in Artificial Intelligence, Text Processing, Natural Language Processing, Introduction Alfa-informatica, Constraint Logics, Corpus Linguistics, XML, Search Engines, Unix utilities, Language Technology, Text Processing. Advisor for numerous Bachelor-theses.

- Master courses Natural Language Processing, Corpus Linguistics, Research seminar. Learning from Data. Advisor for numerous MA-theses.

- Invited course at the Winter School of LOT (Netherlands Graduate School of Linguistics), Tilburg 1995 [with Gosse Bouma].

- Invited course at the OzsL Spring School (Dutch Graduate School in Logic), Amsterdam 1998 [with Gosse Bouma].

- Invited course at the ELSNET Summer School, Barcelona 1998.

- Invited course at the Summer School of LOT (Netherlands Graduate School of Linguistics), Tilburg 2000.

- Invited tutorial at the Annual Meeting of the ACL (Association of Computational Linguistics), Sapporo Japan, 2003.

- Guest lecturer at the University of Malta, april/may 2009.

- Invited course at the Winter School of LOT (Netherlands Graduate School of Linguistics), Amsterdam 2015.

Professional Activities

- Association for Computational Linguistics (ACL). Elected as member of the Executive Board. Vice-President Elect in 2012. Vice-President in 2013. President of the ACL in 2014. Member of various ACL committees in 2012-2017.

- Member of the NWO CLARIAH advisory board.

- Co-founder and member of the CLIN working group (Computational Linguistics in the Netherlands), 1990-present. Maintainer of the CLIN website, 1994-present.

- Theme group leader NWO-programme Language and Speech Technology, 1995-2001.

- Programme Committee NWO-programme Interactive Multi-modal Information Extraction, 2003-2008.

- Chair EACL (European Chapter of the Association of Computational Linguistics). 2005-2006.

- Member nominating committee EACL. 2007-2010.

- Editor of the EACL Newsletter. 2003-2005.

Editorial Board, etc.:

- Editor-in-chief of the Computational Linguistics in the Netherlands Journal. 2011-present.

- Editorial Board Computational Linguistics, 1997-2000;

- Editorial Board WEB-SLS, 1996-present.

- Editorial Board Computer Speech and Language 2003-2016.

- Editorial Board Linguistic Issues in Language Technology, 2008-present.

- Editor Special Issue *Natural Language Engineering* on Finite State Methods in NLP (with Karttunen and Koskenniemi). 2001.

Selected Software Packages

All software packages are available free of charge, and available from the webpage `http://www.let.rug.nl/~vannoord/software.html`

Alpino Alpino is a collection of tools and programs for parsing Dutch sentences into dependency structures. It is the de-facto standard robust wide-coverage high-accuracy parser for Dutch.

Fsa Utilities The FSA Utilities is a collection of utilities to construct finite automata from regular expressions; manipulate finite automata; visualise finite automata; apply finite automata, and compile finite automata (to C, C++, Java, Prolog).

TextCat Language Gueser TextCat is a language guesser: given a few lines of text it attempts to decide in which natural language the text is written. TextCat knows about seventy different languages. TextCat implements the text categorization algorithm presented in a paper by Cavnar and Trenkle. TextCat is part of the SpamAssassin spam filter programme.

Hdrug Hdrug is a graphical user environment for the development of logic grammars and related tools.

Publications

Full list of publications, including some 150 refereed publications, is available as `http://www.let.rug.nl/vannoord/papers/`. The Google Scholar profile `http://https://scholar.google.com/citations?user=mBfMmZ4AAAAJ&hl=en&oi=ao` lists over 4000 citations, with an h-index of 35.