

Curriculum Vitae, Gertjan van Noord

November 2016

Born 8 May 1961, Culemborg, Netherlands

Married, 3 children. Current work address:

Computational Linguistics CLCG & Informatiekunde

Faculteit der Letteren

Rijksuniversiteit Groningen

Postbus 716, NL 9700 AS Groningen

+31 (0)50-3637811

Current home address:

Burgemeester Seinenstraat 44, NL 9831 PX Aduard

+31 (0)50-4032209

G.J.M.van.Noord@rug.nl

<http://www.let.rug.nl/~vannoord>

Education

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

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

Employment

2017–present Rijksuniversiteit Groningen. Director of the CLCG.

1/2011–present Rijksuniversiteit Groningen. Professor in Language Technology.

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

1/1992–3/1999 Rijksuniversiteit Groningen. Assistant Professor (Universitair Docent) 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 Rijksuniversiteit Utrecht. Researcher at the General Linguistics dept. for the EC-funded Eurotra project. Responsible for the MiMo2 sub-project.

Project management and Ph.D.-student supervision

2019 PaQuUd. Project funded by Clariah.

2019- Co-supervision of PhD project of 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 coordinator. 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-assistants (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.

Involvement Ph.D. projects (selection)

- 2018. Ke Tran. Predicting and Discovering Linguistic Structure with Neural Networks. UvA Amsterdam. [member reading committee]
- 2015. Oscar Strik. Modelling analogical change. Groningen. [member reading committee]
- 2015. Valerio Basile. From Logic to Language: Natural Language Generation. Groningen. [member reading committee]
- 2014. Jan de Belder. Linear Programming Optimization for Natural Language Processing Problems. KU Leuven. KU Leuven. [member PhD jury]
- 2013. Fridah Katushemerewe. Computational morphology and Bantu language learning: an implementation for Ruyakitara. Groningen. [reading committee]
- 2012. Vincent Van Asch. Domain Similarity Measures: On the use of distance metrics in natural language processing. Universiteit Antwerpen. [member PhD jury]
- 2010. Lionel Nicolas. Efficient Production of Linguistic Resources: The Victoria Project. Université Nice Sophia Antipolis. [member PhD jury]
- 2008. Wouter van Atteveldt. Semantic Network Analysis. Techniques for Extracting, Representing and Querying Media Content. VU Amsterdam [member PhD jury]

Involvement in Research Evaluation

- 2018. Habilitation Defense of Benoit Sagot. Member of the jury. Paris.
- 2015. Evaluation of INRIA theme “Language, Speech and Audio”. Paris.

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

- Elected as member of the Executive Board of the ACL. Vice-President Elect in 2012. Vice-President in 2013. President in 2014.
- Member of the 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.

Conference Chair/Organizer:

CLIN 1, Utrecht 1990 (co-chair). CLIN 4, Groningen 1993 (co-chair). NWO TST Workshop ‘Disambiguation in Spoken Dialogue Systems’, Nijmegen 1997 (organizer). EACL, Bergen Norway 1999 (workshops chair). COLING, Saarbrücken Germany 2000 (area chair), ESSLLI workshop Finite State Methods in NLP, Helsinki Finland 2001 (co-chair). ACL 2001 Toulouse France (tutorials chair), ACL, Sapporo Japan 2003 (area chair). Eight International Workshop on Parsing Technologies, Nancy France 2003 (chair). ACL workshop Deep Linguistic Processing 2007 (co-chair). Treebanks and Linguistic Theories 2009 (local chair). Distributional Semantics Workshop 2010 Groningen (organiser). EACL 2012 (mentoring co-chair). LREC 2020 session REPROLANG.

Conference Programme Committee Member:

ACL workshop Reversible Grammar in Natural Language Processing 1991. Third International Workshop on Natural Language Generation 1991. COLING 1992. EACL 1993. COLING 1996. ACL/EACL 1997. Formal Grammar Conference 1997. ACL/EACL 1997. Workshop Finite State Methods in Natural Language Processing 1998. Formal Grammar Conference/HPSG Conference, 1998. EACL 1999. COLING 2000 workshop ‘Using Toolsets and Architectures to build NLP systems’. COLING 2000 workshop ‘Finite State Phonology’. COLING 2000. ACL 2000. TAG+ 2000. ESSLLI workshop Finite State Methods in NLP 2001. HPSG 2001. TAG+ 2001. IWPT 2001. ACL 2002. NLULP 2002. TAG+ 2002. Mathematics of Language 2003. EACL 2003. IWPT 2003. ACL 2003. IJCNLP 2004. LREC 2004. FSMNLP 2005. IJCNLP 2005. ACL 2005. SIGSEM 2005. IWPT 2005. EACL 2006. HLT/NAACL 2006. IWPT 2007. EACL 2009. EMNLP 2009. IWPT 2009. KI 2009. GEAF 2009. COLING 2010. ACL 2010. ECAI 2010. EMNLP 2010. DepLing 2011. TLT10. ACL-HLT 2011. ACL-MWE 2011. IJCNLP 2011. NAACL 2012. ACL 2012. NLDB 2012. TLT 2012. TLT 10. ACL 2013. EMNLP 2014. SLSP 2014. MWE 2014. LREC 2014. TLT 13. DEPLING 2015. GEAF 2015. DMTW 2015. NAACL 2016. COLING 2016. DMTW2016. EACL 2017. MWE 2017.

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

Reviewer for Computational Linguistics, WEB-SLS, Computer Speech and Language, Journal of Logic Programming, Language and Computation, Natural Language Engineering, New Generation Computing, Computational Intelligence, Traitement Automatique des Langues. Transactions of Speech and Language Processing. Language Resources and Evaluation. ACM Transactions on Speech and Language Processing.

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

Local duties

- Director of CLCG, 2017-present
- chair instruction committee Research Master Linguistics 2013-2017
- member Expert-group ICT 2012-present
- member department board CIW 2012-2017
- member “Raad van Toezicht” Informatiekunde 2012-2016
- member department board Informatiekunde 2010-2012
- department chair Informatiekunde 2006-2007
- member department board Informatiekunde 2001-2007
- coordinator research group Computational Linguistics of the CLCG 2003-2016
- Member Advisory Board Rekencentrum / Donald Smits Center for Information Technology 2005-2016
- department chair Informatiekunde 2001-2002

Selected Invited Lectures

- October 2016. Invited lecture at Rich Parsing Workshop. “Disambiguation and selectional preferences in Alpino”. University of Amsterdam.
- September 2016. Invited lecture OTS Colloquium, “Large-scale Automated Syntactic Analysis of Dutch”. University of Utrecht.
- January 2016. Invited lecture, “Improving Automatically Parsed Dutch Treebanks”. University of Antwerp.
- December 2014. Invited lecture at TLT 13, Tübingen.
- December 2013. Invited lecture, “Large Scale Syntactic Annotation of Written Dutch: Lassy”. University of Twente.
- June 14, 2012, “Towards Understanding Dutch Automatically”. Guest Lecture at INCAS3, Assen.
- May 14, 2009, “Parsing to improve Parsing”. Artificial Intelligence colloquium, University of Malta.

- march 30, 2009, “Parsed Corpora for Linguists”. At the EACL Workshop “The Interaction between Linguistics and Computational Linguistics: Virtuous, Vicious or Vacuous?”. Athens.
- 10 juli 2008. “Self-trained Bilexical Preferences for Improved Syntactic Disambiguation”. CoLi Colloquium, University of the Saarland, Saarbrücken.
- 25 april 2008. “Large Scale Syntactic Annotation for Dutch”. Q-go linguïstendag. Q-go Diemen.
- 24 juni 2007. with Gosse Bouma. “Mining Ontological Knowledge from Syntactically Annotated Corpora”. Workshop: What can Natural Language Processing and Semantic Web technologies do for eLearning? ACL 2007. Prague.
- 23 juni 2007. with Timothy Baldwin, Mark Dras, Julia Hockenmaier and Tracy Holloway King. “The Impact Of Deep Linguistic Processing”. International Workshop on Parsing Technologies (IWPT). ACL 2007, Prague.
- 15 february 2007. “Large Scale Syntactic Annotation for Dutch: How and Why”. KU Leuven, Leuven.
- 11 september 2006. “LASSY: Large Scale Syntactic Annotation of Written Dutch”. STEVIN-programmadag. Antwerpen.
- 29 juni 2006. “Improving Knowledge-based Parsing with Corpus-based Methods”. Israeli Seminar in Computational Linguistics (ISCOL). University of Haifa, Haifa.
- 25 juni 2006. “Robust Parsing, Error Mining and Automated Lexical Acquisition in Alpino”. Research Workshop Large Scale Grammar Development and Grammar Engineering. University of Haifa, Haifa.
- 23 juni 2006. “Disambiguation in the Alpino parser for Dutch”. Symposium Ambiguity in Language: Theoretical, Behavioral and Neuroimaging Perspectives. University of Groningen, Groningen.
- 10 april 2006. “At Last Parsing Is Now Operational”. Traitement Automatique des Langues (TALN) 2006. KU Leuven, Leuven.
- 2 april 2006. “Robust Parsing, Error Mining, Automated Lexical Acquisition, and Evaluation”. ROMAND workshop. EACL 2006, Trento.
- 23 maart 2006. “Syntactische Annotatie in D-Coi en LASSY”. TST-dag. Rotterdam.
- 11 november 2005. “ALP Is Not Over”. Algorithms for Linguistic Processing Workshop. University of Groningen.
- 13 juli 2005. “On Passives”. 2nd International Workshop on Constraint-based Grammar. University of Bremen.

- 24 juni 2004. “Error Mining for Wide-Coverage Grammar Engineering”. CoLi Colloquium. University of the Saarland, Saarbrücken.
- 7 maart 2003. “Grammar Engineering Using Very Large Corpora”. Symposium, University of Edinburgh.
- 9 april 2002. “Alpino: Wide Coverage Computational Analysis of Dutch”. Seminar Johns Hopkins University, Baltimore.
- 21 februari 2002. “Wide Coverage Computational Analysis of Dutch”. Seminar University of Sussex. Brighton.
- 15 juni 2001. “Alpino: Wide Coverage Computational Analysis of Dutch”. Computing with LLL Seminar. UvA, Amsterdam.
- 6 augustus 2000. with Dale Gerdemann. “Approximation and Exactness in Finite State Optimality Theory”. Invited lecture at ACL Workshop on Computational Phonology, Luxembourg.
- 12 december 1997. Colloquium, vakgroep Taal & Spraak, Nijmegen. “A hybrid and robust parser for OVIS2”.
- 13 juni 1997. T&I Colloquium Tilburg. “Grammatical Analysis in the OVIS2 spoken dialogue system”.
- 21 maart 1997. CLIF meeting Leuven. “Grammar-based NLP in the NWO Priority Programme on Language and Speech Technology”.
- 29 maart 1996. Nederlandse Vereniging voor Fonetische Wetenschappen, workshop Determinisme en Statistiek in Spraakonderzoek. Nijmegen. Title: “Desambiguatie in OVIS”.
- 4 maart 1993. “Head-driven Parsing for Lexicalist Grammars”. Linguistics Colloquium, Universite de Geneve, Geneva.
- 20 januari 1993. “The Parsing and Generation Problem for Unification Grammars”. Workshop Grammar, Proof Theory and Complexity, Amsterdam 20-22 januari 1993.

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

Journal publications

- [1] Stuart M. Shieber, Gertjan van Noord, Robert C. Moore, and Fernando C. N. Pereira. Semantic-head-driven generation. *Computational Linguistics*, 16(1):30–42, 1990.
- [2] Gertjan van Noord, Joke Dorrepaal, Pim van der Eijk, Maria Florenza, Herbert Ruessink, and Louis des Tombe. An overview of MiMo2. *Machine Translation*, 6:201–214, 1991.
- [3] Gertjan van Noord. Head corner parsing for TAG. *Computational Intelligence*, 10(4):525–534, 1994.
- [4] Gertjan van Noord. An efficient implementation of the head corner parser. *Computational Linguistics*, 23(3):425–456, 1997.
- [5] Gertjan van Noord and Günter Neumann. Syntactic generation. *Linguistica Computazionale*, 13:147–150, 1997. Survey of the State of the Art in Human Language Technology.
- [6] Gertjan van Noord, Gosse Bouma, Rob Koeling, and Mark-Jan Nederhof. Robust grammatical analysis for spoken dialogue systems. *Journal of Natural Language Engineering*, 5(1):45–93, 1999.
- [7] Gertjan van Noord. The treatment of epsilon moves in subset construction. *Computational Linguistics*, 26(1):61–76, 2000.
- [8] Gertjan van Noord and Dale Gerdemann. Finite state transducers with predicates and identities. *Grammars*, 4:263–286, 2001.
- [9] Leonoor van der Beek, Gosse Bouma, and Gertjan van Noord. Een brede computationele grammatica voor het nederlands. *Nederlandse Taalkunde*, 7(4):353–374, 2002.
- [10] Robbert Prins and Gertjan van Noord. Reinforcing parser preferences through tagging. *Traitement Automatique des Langues*, 44(3):121–139, 2003.
- [11] Jan Daciuk and Gertjan van Noord. Finite automata for compact representation of tuple dictionaries. *Theoretical Computer Science*, 313(1):45–56, 2004.

- [12] Gosse Bouma, Ismail Fahmi, Jori Mur, Gertjan van Noord, Lonneke van der Plas, and Jörg Tiedeman. Linguistic knowledge and question answering. *Traitement Automatique des Langues*, 2(46):15–39, 2005.
- [13] Valia Kordoni, Gertjan van Noord. Passives in Germanic Languages: the case of Dutch and German. *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)*, 49:77–96, 2009.

Books and edited collections

- [1] Gertjan van Noord. *Reversibility in Natural Language Processing*. PhD thesis, Utrecht University, 1993.
- [2] Gosse Bouma and Gertjan van Noord, editors. *CLIN IV, Papers from the Fourth Clin Meeting*. Vakgroep Alfa-informatica, RUG, Groningen, 1994.
- [3] Jean-Claude Junqua and Gertjan van Noord, editors. *Robustness in Language and Speech Technology*. Kluwer Academic Publishers, Dordrecht, 2001.
- [4] Lauri Karttunen, Kimmo Koskenniemi, and Gertjan van Noord. Special issue: Finite state methods in language language processing. *Natural Language Engineering*, 9(1), 2003.
- [5] Frank van Eynde, Anette Frank, Koenraad de Smedt, and Gertjan van Noord, editors. *Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories (TLT 7), January 23-24, 2009, Groningen, The Netherlands*. LOT Occasional Series. LOT, Utrecht, 2009.

ACL/COLING conference publications

- [1] Stuart M. Shieber, Gertjan van Noord, Robert C. Moore, and Fernando C. N. Pereira. A semantic-head-driven generation algorithm for unification based formalisms. In *27th Annual Meeting of the Association for Computational Linguistics*, pages 7–17, Vancouver, 1989.
- [2] Gertjan van Noord. Reversible unification-based machine translation. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING)*, pages 299–304, Helsinki, 1990.
- [3] Gertjan van Noord. Head corner parsing for discontinuous constituency. In *29th Annual Meeting of the Association for Computational Linguistics*, pages 114–121, Berkeley, 1991.
- [4] Günter Neumann and Gertjan van Noord. Self-monitoring with reversible grammars. In *Proceedings of the 15th [sic] International Conference on Computational Linguistics (COLING)*, pages 700–706, Nantes, 1992.

- [5] Gosse Bouma and Gertjan van Noord. Head-driven parsing for lexicalist grammars: Experimental results. In *Sixth Conference of the European Chapter of the Association for Computational Linguistics*, pages 71–80, Utrecht, 1993.
- [6] Gosse Bouma and Gertjan van Noord. Constraint-based categorial grammar. In *32th Annual Meeting of the Association for Computational Linguistics*, pages 147–154, New Mexico, 1994.
- [7] Gertjan van Noord and Gosse Bouma. Adjuncts and the processing of lexical rules. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING)*, pages 250–256, Kyoto, 1994.
- [8] Gertjan van Noord. The intersection of finite state automata and definite clause grammars. In *33th Annual Meeting of the Association for Computational Linguistics*, pages 159–165, MIT Cambridge Mass., 1995.
- [9] Dale Gerdemann and Gertjan van Noord. Transducers from rewrite rules with backreferences. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 126–133, Bergen Norway, 1999.
- [10] Gertjan van Noord. Error mining for wide-coverage grammar engineering. In *ACL2004*, Barcelona, 2004. ACL.
- [11] Gertjan van Noord. Learning efficient parsing. In *EACL 2009, The 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 817–825, Athens, Greece, 2009.
- [12] Kostadin Cholakov, Gertjan van Noord. Acquisition of Unknown Word Paradigms for Large Scale Grammars. In *COLING2010*, Beijing, 2010.
- [13] Kostadin Cholakov, Gertjan van Noord, Valia Kordoni, Yi Zhang. An empirical comparison of Unknown Word Prediction Methods. In *IJCNLP2011*, Thailand, 2011.
- [14] Barbara Plank and Gertjan van Noord. Effective Measures of Domain Similarity for Parsing. In *ACL2011*, Portland, 2011.
- [15] Daniel de Kok and Barbara Plank and Gertjan van Noord. Reversible Stochastic Attribute-value Grammars. In *ACL2011*, Portland, 2011.
- [16] Simon Suster and Gertjan van Noord. From neighborhood to parenthood: the advantages of dependency representation over bigrams in Brown clustering. In *COLING 2014*, Dublin, 2014.
- [17] Simon uster, Ivan Titov, Gertjan van Noord. Bilingual Learning of Multi-sense Embeddings with Discrete Autoencoders. In *NAACL*, San Diego, 2016.

Book chapters

- [1] Gertjan van Noord. An overview of head-driven bottom-up generation. In Robert Dale, Chris Mellish, and Michael Zock, editors, *Current Research in Natural Language Generation*, pages 141–165. Academic Press, 1990.
- [2] Gertjan van Noord. Head corner parsing. In C. J. Rupp, Mike Rosner, and Rod Johnson, editors, *Constraints, Language and Computation*, pages 315–338. Academic Press, London, 1993.
- [3] Gosse Bouma and Gertjan van Noord. A lexicalist account of the Dutch verb cluster. In Gosse Bouma and Gertjan van Noord, editors, *CLIN IV, Papers from the Fourth Clin Meeting*, Groningen, 1994.
- [4] Gertjan van Noord. FSA Utilities: A toolbox to manipulate finite-state automata. In Darrell Raymond, Derick Wood, and Sheng Yu, editors, *Automata Implementation*, pages 87–108. Springer Verlag, 1997. Lecture Notes in Computer Science 1260.
- [5] Gertjan van Noord and Gosse Bouma. Dutch verb clustering without verb clusters. In Patrick Blackburn and Maarten de Rijke, editors, *Specifying Syntactic Structures*, pages 213–243. CSLI Publications / Folli, Stanford, 1997.
- [6] Gosse Bouma and Gertjan van Noord. Word order constraints on verb clusters in German and Dutch. In Erhard Hinrichs, Tsuneko Nakazawa, and Andreas Kathol, editors, *Complex Predicates in Nonderivational Syntax*, pages 43–72. Academic Press, New York, 1998.
- [7] Gert Veldhuijzen van Zanten, Gosse Bouma, Khalil Sima'an, Gertjan van Noord, and Remko Bonnema. Evaluation of the NLP components of the OVIS2 spoken dialogue system. In Frank van Eynde, Ineke Schuurman, and Ness Schelkens, editors, *Computational Linguistics in the Netherlands 1998*, pages 213–229. Rodopi Amsterdam, 1999.
- [8] Gertjan van Noord and Dale Gerdemann. An extendible regular expression compiler for finite-state approaches in natural language processing. In O. Boldt and H. Juergensen, editors, *Automata Implementation. 4th International Workshop on Implementing Automata, WIA '99*. Springer, 2000. Springer Lecture Notes in Computer Science 2214.
- [9] Gosse Bouma, Gertjan van Noord, and Robert Malouf. Wide coverage computational analysis of Dutch. In W. Daelemans, K. Sima'an, J. Veenstra, and J. Zavrel, editors, *Computational Linguistics in the Netherlands 2000*, 2001.
- [10] Gertjan van Noord. Robust parsing of word graphs. In Jean-Claude Junqua and Gertjan van Noord, editors, *Robustness in Language and Speech Technology*. Kluwer Academic Publishers, Dordrecht, 2001.
- [11] Gertjan van Noord. Finite state processing. In Lynn Nadel (editor-in chief), editor, *Encyclopedia of Cognitive Science*, pages 130–134. Nature Publishing Group; Wiley, 2003.

- [12] Martijn Wieling, Mark-Jan Nederhof, and Gertjan van Noord. Parsing partially bracketed input. In Khalil Sima'an, Maarten de Rijke, Remko Scha, and Rob van Son, editors, *CLIN 2005 Proceedings of the 16th Meeting of Computational Linguistics in the Netherlands*, pages 1–16, Universiteit van Amsterdam, Amsterdam, 2006.
- [13] Gosse Bouma, Jori Mur, Gertjan van Noord, Lonneke van der Plas, and Jörg Tiedemann. Question answering for dutch using dependency relations. In Carol Peters et al., editor, *Accessing Multilingual Information Repositories (Lecture Notes in Computer Science 4022)*, pages 370–379. Springer, Berlin, 2006.
- [14] Gosse Bouma, Ismail Fahmi, Jori Mur, Gertjan van Noord, Lonneke van der Plas, and Jörg Tiedemann. Using syntactic knowledge for QA. In Carol Peters et al., editor, *Evaluation of Multilingual and Multi-modal Information Retrieval (Lecture Notes in Computer Science 4730)*, pages 318–327. Springer, Berlin, 2007.
- [15] Gosse Bouma, Geert Kloosterman, Jori Mur, Gertjan van Noord, Lonneke van der Plas, and Jörg Tiedemann. Question answering with Joost at QA@CLEF 2007. In Carol Peters et al., editor, *Advances in Multilingual and Multimodal Information Retrieval (Lecture Notes in Computer Science 5152)*, pages 257-260. Springer, Berlin, 2008.
- [16] Gertjan van Noord. Self-trained Bilexical Preferences to Improve Disambiguation Accuracy. In Harry Bunt, Paola Merlo, Joakim Nivre, editor, *Trends in Parsing Technology. Dependency Parsing, Domain Adaptation, and Deep Parsing*. Springer Verlag. 2010.
- [17] Gertjan van Noord, Gosse Bouma, Frank van Eynde, Daniel de Kok, Jelmer van der Linde, Ineke Schuurman, Erik Tjong Kim Sang, Vincent Vandeghinste. Large Scale Syntactic Annotation of Written Dutch: Lassy. In *Essential Speech and Language Technology for Dutch: the STEVIN Programme*. Springer, in press.
- [18] Vincent Vandeghinste, Scott Martens, Gideon Kotze, Jorg Tiedemann, Joachim Van den Bogaert, Koen De Smet, Frank Van Eynde, and Gertjan van Noord. Parse and Corpus-based Machine Translation. In *Essential Speech and Language Technology for Dutch: the STEVIN Programme*. Springer, in press.
- [19] Jan De Belder, Daniel de Kok, Gertjan van Noord, Fabrice Nauze, Leonoor van der Beek, and Marie-Francine Moens. Question Answering of Informative Web Pages: How Summarisation Technology Helps. In *Essential Speech and Language Technology for Dutch: the STEVIN Programme*. Springer, in press.

Other reviewed publications

- [1] Gertjan van Noord. BUG: A directed bottom-up generator for unification based formalisms. *Working Papers in Natural Language Processing, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht*, 4, 1989.

- [2] Gertjan van Noord. Towards uniform processing of constraint-based categorial grammars. In *Proceedings of ACL SIG workshop Reversible Grammar in Natural Language Processing*, pages 12–19, Berkeley, 1991.
- [3] Günter Neumann and Gertjan van Noord. Reversible grammars for self-monitoring and generation of paraphrases. In Tomek Strzalkowski, editor, *Reversible Grammar in Natural Language Processing*, pages 59–95. Kluwer, 1993.
- [4] Gosse Bouma and Gertjan van Noord. Word order constraints on German verb clusters. In Geert-Jan Kruijff, Glynn Morrill, and Dick Oehrle, editors, *Proceedings Formal Grammar*, pages 15–28, Prague, 1996.
- [5] Gertjan van Noord. FSA Utilities: Manipulation of finite-state automata implemented in Prolog. In *WIA '96 First International Workshop on Implementing Automata*, pages 47–66, 1996. Technical Report #495, Department of Computer Science, University of Western Ontario, London Ontario.
- [6] Gertjan van Noord. Robust parsing with the head-corner parser. In John Carroll, editor, *Workshop on Robust Parsing*, pages 83–92, Prague, 1996. These proceedings are also available as Cognitive Science Research Paper #435; School of Cognitive and Computing Sciences, University of Sussex.
- [7] Gosse Bouma, Rob Koeling, Mark-Jan Nederhof, and Gertjan van Noord. Grammatical analysis in a spoken dialog system. In Roel Jonkers, Edith Kaan, and Anko Wiegel, editors, *Language and Cognition 5. Yearbook 1995*, pages 27–44. University of Groningen, Groningen, 1996.
- [8] Gertjan van Noord and Gosse Bouma. Hdrug, A flexible and extendible development environment for natural language processing. In *Proceedings of the EACL/ACL workshop on Environments for Grammar Development, Madrid, 1997*.
- [9] Mark-Jan Nederhof, Gosse Bouma, Rob Koeling, and Gertjan van Noord. Grammatical analysis in the OVIS spoken-dialogue system. In *Proceedings of the ACL/EACL Workshop on Spoken Dialog Systems*, pages 66–73, Madrid, Spain, 1997.
- [10] Gertjan van Noord. The treatment of epsilon moves in subset construction. In *Finite-state Methods in Natural Language Processing*, Ankara, 1998.
- [11] Dale Gerdemann and Gertjan van Noord. Approximation and exactness in finite state optimality theory. In Jason Eisner, Lauri Karttunen, and Alain Thériault, editors, *Finite-State Phonology. Proceedings of the Fifth Workshop of the ACL SPecial Interest Group in Computational Phonology*, pages 34–45, Luxembourg, 2000.
- [12] Jan Daciuk and Gertjan van Noord. Finite automata for compact representation of language models in nlp. In Burce Watson and Derick Wood, editors, *Proceedings of the 6th Conference on Implementations and Applications on Automata (CIAA)*, pages 45–55, Pretoria, South Africa, 2001.

- [13] Robbert Prins and Gertjan van Noord. Unsupervised pos-tagging improves parsing accuracy and parsing efficiency. In *Proceedings of the Seventh International Workshop on Parsing Technologies (IWPT)*, pages 154–165, Beijing, China, 2001.
- [14] Tony Mullen, Robert Malouf, and Gertjan van Noord. Statistical parsing of Dutch using maximum entropy models with feature merging. In J. Tsujii, editor, *NLPRS2001, Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium*, pages 481–486, Tokyo, 2001. University of Tokyo Press.
- [15] Leonoor van der Beek, Gosse Bouma, Robert Malouf, and Gertjan van Noord. The Alpino dependency treebank. In *Computational Linguistics in the Netherlands*, 2002.
- [16] Gertjan van Noord and Robert Malouf. Wide coverage parsing with stochastic attribute value grammars. Draft available from the authors. A preliminary version of this paper was published in the Proceedings of the IJCNLP workshop Beyond Shallow Analyses, Hainan China, 2004., 2005.
- [17] Gertjan van Noord and Valia Kordoni. A raising analysis of the dutch passive. In Stefan Müller, editor, *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar; HPSG05, University of Lisbon, Lisbon*, pages 410–426. CSLI Publications, 2005.
- [18] Gosse Bouma, Jori Mur, and Gertjan van Noord. Reasoning over dependency relations for QA. In Farah Benamarah, Marie-Francine Moens, and Patrick Saint-Dizier, editors, *Knowledge and Reasoning for Answering Questions*, pages 15–21, 2005. Workshop associated with IJCAI 05.
- [19] Gertjan van Noord, Ineke Schuurman, and Vincent Vandeghinste. Syntactic annotation of large corpora in STEVIN. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, Genoa, Italy, 2006.
- [20] Gertjan van Noord. **At Last Parsing Is Now Operational**. In *TALN 2006 Verbum Ex Machina, Actes De La 13e Conference sur Le Traitement Automatique des Langues naturelles*, pages 20–42, Leuven, 2006.
- [21] Gertjan van Noord. Using self-trained bilexical preferences to improve disambiguation accuracy. In *Proceedings of the International Workshop on Parsing Technology (IWPT)*, ACL 2007 Workshop, pages 1–10, Prague, 2007. Association for Computational Linguistics, ACL.
- [22] Timothy Baldwin, Mark Dras, Julia Hockenmaier, Tracy Holloway King, and Gertjan van Noord. The impact of deep linguistic processing on parsing technology. In *Proceedings of the International Workshop on Parsing Technology (IWPT)*, ACL 2007 Workshop, pages 36–38, Prague, 2007. Association for Computational Linguistics, ACL.

- [23] Nelleke Oostdijk, Martin Reynaert, Paola Monachesi, Gertjan van Noord, Roland Ordeman, Ineke Schuurman, and Vincent Vandeghinste. From D-Coi to SoNaR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, 2008.
- [24] Barbara Plank and Gertjan van Noord. Exploring an auxiliary distribution based approach to domain adaptation of a syntactic disambiguation model. In *Coling Workshop Cross Framework and Cross Domain Parser Evaluation*, Manchester, 2008.
- [25] Gertjan van Noord. Huge parsed corpora in Lassy. In Frank van Eynde, Anette Frank, Koenraad De Smedt, and Gertjan van Noord, editors, *Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories (TLT 7)*, number 12 in LOT Occasional Series, pages 115–126, Utrecht, The Netherlands, 2009. Netherlands Graduate School of Linguistics.
- [26] Gertjan van Noord and Gosse Bouma. Parsed corpora for linguistics. In *EACL 2009 Workshop The Interaction between Linguistics and Computational Linguistics: Virtuous, Vicious or Vacuous?*, pages 33–39, Athens, Greece, 2009.
- [27] Kostadin Cholakov and Gertjan van Noord. Combining finite state and corpus-based techniques for unknown word prediction. In *Recent Advances in Natural Language Processing (RANLP)*, pages 1–5, Borovets, Bulgaria, September 2009.
- [28] Daniël de Kok, Jianqiang Ma, and Gertjan van Noord. A generalized method for iterative error mining in parsing results. In *Proceedings of the 2009 Workshop on Grammar Engineering Across Frameworks (GEAF 2009)*, pages 71–79, Suntec, Singapore, August 2009. Association for Computational Linguistics.
- [29] Yan Zhao, Gertjan van Noord. POS Multi-tagging Based on Combined Models. In *LREC 2010*, pages 1249–1252.
- [30] Barbara Plank, Gertjan van Noord. Grammar-driven versus data-driven: Which Parsing System is More Affected by Domain Shifts? In *ACL workshop NLP and Linguistics: Finding the Common Ground*. Uppsala, Sweden, 2010. Association for Computational Linguistics.
- [31] Kostadin Cholakov, Gertjan van Noord. Using Unknown Word Techniques To Learn Known Words. In *EMNLP 2010*. 2010.
- [32] Danil de Kok, Gertjan van Noord. A Sentence Generator for Dutch. In *Proceedings of the 20th Meeting of Computational Linguistics in the Netherlands*.
- [33] Barbara Plank, Gertjan van Noord. Dutch Dependency Parser Performance Across Domains. In *Proceedings of the 20th Meeting of Computational Linguistics in the Netherlands*.

- [34] Kostadin Cholakov, Gertjan van Noord, Valia Kordoni, Yi Zhang. Adaptability of Lexical Acquisition for Large-scale Grammars. In RANLP 2011.
- [35] Angelina Ivanova and Gertjan van Noord. Treelet Probabilities for HPSG Parsing and Error Correction. In LREC 2014, Reykjavik, 2014
- [36] Michal Novk, Dieke Oele and Gertjan van Noord. Comparison of Coreference Resolvers for Deep Syntax Translation. In EMNLP 2015 Workshop on Discourse in Machine Translation, September 2015, Lisbon.
- [37] Dieke Oele and Gertjan van Noord. Lexical Choice in Abstract Dependency Trees In Deep Machine Translation Workshop 2015. September 2015, Prague.
- [38] Rob van der Goot and Gertjan van Noord. ROB: Using Semantic Meaning to Recognize Paraphrases. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)
- [39] Luis Gomes, Gertjan van Noord, Antonio Branco, Steven Neale. Seeking to Reproduce "Easy Domain Adaptation". In 4REAL workshop at LREC 2016.
- [40] Antnio Branco, Hans Uszkoreit, Aljoscha Burchardt, Jan Hajic, Martin Popel, Kiril Simov, Petya Osenova, Markus Egg, Eneko Agirre, Gertjan van Noord, Filipe Barrancos and Rosa Del Gaudio. QTLep: A European scientific research project on machine translation by deep language engineering approaches. In: META-FORUM 2016.
- [41] Rosa Gaudio, Gorka Labaka, Eneko Agirre, Petya Osenova, Kiril Simov, Martin Popel, Dieke Oele, Gertjan van Noord, Joo Silva, Joo Antnio Rodrigues, Steven Neale, Lus Gomes, Nuno Rendeiro, Andreia Querido and Antnio Branco. SMT and Hybrid systems of the QTLep project in the WMT16 IT-task. In WMT16, 2016.
- [42] Dieke Oele, Gertjan van Noord. Choosing lemmas from Wordnet synsets in Abstract Dependency Trees. Second Workshop on Deep Language Processing for Quality Machine Translation. Varna, Bulgaria, 2016.