

# CURRICULUM VITAE

April 2009

name	Antonius Petrus Johannes (Antal) van den Bosch
date of birth	5 May 1969
place of birth	Made, the Netherlands
sex	male
marital status	married, two children
nationality	Dutch
e-mail	<a href="mailto:Antal.vdnBosch@uvt.nl">Antal.vdnBosch@uvt.nl</a>
URL	<a href="http://ilk.uvt.nl/~antalb">http://ilk.uvt.nl/~antalb</a>

## Education

### Ph.D.

Universiteit Maastricht (The Netherlands), Faculty of General Sciences, Department of Computer Science. Thesis *Learning to pronounce written words. A study in inductive language learning*. Defended 11 December 1997; cum laude.

### M.A.

Taal- en literatuurwetenschap, i.h.b. taal en informatica (Arts, with emphasis on computational linguistics), Faculty of Arts, Tilburg University, The Netherlands. Diploma received 26 May 1992.

### Secondary school

Gymnasium  $\beta$ , Gertrudislyceum, Roosendaal, The Netherlands. Graduation 1987.

## Positions

**Full professor** – Faculty of Humanities, Tilburg centre for Creative Computing, Department of Communication and Information Sciences, Tilburg University, the Netherlands. January 2008 – present.

**Associate professor** – Faculty of Humanities, Department of Communication and Information Sciences (until January 2007: Faculty of Arts, Dept. of Language and Information Sciences), Tilburg University, the Netherlands. February 2006 – December 2007.

**Guest professor** – Faculty of Arts, Department of Linguistics, University of Antwerp, Antwerp, Belgium. January 2005 – present.

**Visiting faculty** – WhizBang! Labs – Research, Pittsburgh, PA, USA. September 2001 – December 2001.

**Assistant professor** – Faculty of Arts (Computational Linguistics and Artificial Intelligence research unit), Tilburg University, The Netherlands. February 2001 – January 2006.

**Academie–Onderzoeker** – Fellow of the Royal Dutch Academy of Arts and Sciences (KNAW, Koninklijke Nederlandse Academie van Wetenschappen), Faculty of Arts, Computational Linguistics and Artificial Intelligence research unit, Tilburg University, the Netherlands. July 1999 – January 2001.

**Postdoc researcher** – financed by NWO, the Dutch Organisation for Scientific Research, as part of the “Induction of Linguistic Knowledge” (ILK) project at the Faculty of Arts, Computational Linguistics and Artificial Intelligence research unit, Tilburg University, the Netherlands. September 1997 – June 1999.

**Assistant in opleiding** – Ph.D. student, Department of Computer Science, Faculty of General Sciences, Universiteit Maastricht, The Netherlands. April 1994 – August 1997.

**Research assistant** – Laboratoire de Psychologie Expérimentale, Université Libre de Bruxelles, Brussels, Belgium. January 1994 – March 1994.

**Research assistant** – 0.5 fte, Institute for Language Technology and AI, Tilburg University, the Netherlands. June 1992 – December 1993.

**Research assistant** – 0.5 fte, Department of Experimental and Theoretical Psychology, Faculty of Social Sciences, Tilburg University, the Netherlands. June 1992 – December 1993.

**Student assistant** – 0.2 fte, Institute for Language Technology and AI, Tilburg University, the Netherlands. September 1991 – May 1992.

## Publications

Publications are listed in the categories **thesis, books, journal articles, book contributions and papers in proceedings, edited volumes, reports and long abstracts, non-reviewed popularizing publications, media productions, and forthcoming**. Except in the final category, publications are listed in reversed chronological order.

### Thesis

Van den Bosch, A. (1997) *Learning to pronounce written words. A study in inductive language learning*. Ph.D. Thesis, Universiteit Maastricht, the Netherlands. Cadier en Keer: Phidippides. ISBN 9080157724.

### Books

1. Daelemans, W., and Van den Bosch, A. (2005). *Memory-based language processing*. Cambridge, UK: Cambridge University Press. ISBN 0521808901.
2. Razenberg, K., and Van den Bosch, A. (1997). *Het principe van de quaterniteit als algemeen denkraam*. Best, The Netherlands: Damon. ISBN 9055730238. In Dutch.

### Journal articles

1. Van den Bosch, A., Van Erp, M., and Sporleder, C. (2009). Making a clean sweep of cultural heritage. *IEEE Intelligent Systems*, **34:2**, pp. 54–63.

2. Van den Bosch, A., and Berck, P. (2009). Memory-based machine translation and language modeling. *The Prague Bulletin of Mathematical Linguistics*, **91**, pp. 17–26.
3. Van den Bosch, A. (2006). Spelling space: A computational testbed for phonological and morphological changes in Dutch spelling. *Written Language and Literacy*, **9:1**, pp. 25–44.
4. Maruster, L., Weijters, A., Van der Aalst, W., and Van den Bosch, A. (2006). A rule-based approach for process discovery: Dealing with noise and imbalance in process logs. *Data Mining and Knowledge Discovery*, **13**, pp. 67–87.
5. Van den Bosch, A. (2005). Scalable classification-based word prediction and confusable correction. *Traitement Automatique des Langues*, **46:2**, pp. 39–63.
6. Hoste, V., Hendrickx, I., Daelemans, W., and Van den Bosch, A. (2002). Parameter optimization for machine learning of word sense disambiguation. *Natural Language Engineering*, **8:4**, pp. 311–325.
7. Maruster, L., Weijters, A., De Vries, G., Van den Bosch, A., and Daelemans, W. (2002). Logistic-based patient grouping for multi-disciplinary treatment. *Artificial Intelligence in Medicine*, **26:1–2**, pp 87–107.
8. Veenstra, J., Van den Bosch, A., Buchholz, S., Daelemans, W., and Zavrel, J. (2000) Memory-based word sense disambiguation. *Computers and the Humanities*, special issue on Senseval, Word Sense Disambiguations, Ed. Adam Kilgarriff and Martha Palmer, **34:1-2**.
9. Van den Bosch, A. (1999). Careful abstraction from instance families in memory-based language learning. *Journal of Experimental and Theoretical Artificial Intelligence*, **11:3**, special issue on Memory-Based Language Processing, W. Daelemans, guest ed., pp. 339–368.
10. Daelemans, W., Van den Bosch, A., and Zavrel, J. (1999). Forgetting exceptions is harmful in language learning. *Machine Learning*, **34**, pp. 11–43.
11. Weijters, A., and Van den Bosch, A. (1999). Interpreting knowledge representations in BP-SOM. *Behaviormetrika*, **26:1**, pp. 107–128.
12. Vroomen, J., Van den Bosch, A., and De Gelder, B. (1998). A connectionist model for bootstrap learning of syllabic structure. *Language and Cognitive Processes*, special issue on Language Acquisition and Connectionism, **13:2/3**, pp. 193–220.
13. Weijters, A., Van den Bosch, A., and Van den Herik, H.J. (1997). Behavioural aspects of BP-SOM. *Connection Science*, **9:3**, pp. 253–252.
14. Daelemans, W., Van den Bosch, A., and Weijters, A. (1997). IGTrees: using trees for compression and classification in lazy learning algorithms. *Artificial Intelligence Review*, special issue on Lazy Learning, **11:1–5**, pp. 407–423.
15. Van den Bosch, A., Content, A., Daelemans, W., and De Gelder, B. (1995). Measuring the complexity of writing systems. *Journal of Quantitative Linguistics*, **1:3**, pp. 178–188.

## Book contributions and papers in peer-reviewed proceedings

2009

1. Stehouwer, H., and Van den Bosch, A. (2009). Putting the t where it belongs. In S. Verberne, H. van Halteren, and P.-A. Coppen (Eds.), *Computational Linguistics in the Netherlands 2007: Selected papers from the 18th CLIN meeting*, Nijmegen, The Netherlands, pp. 21–36.
2. Bogers, T., and Van den Bosch, A. (2009). Using language modeling for spam detection in social reference manager websites. In R. Aly, C. Hauff, I. den Hamer, D. Hiemstra, T. Huibers, and F. de Jong (Eds.), *Proceedings of the 9th Belgian-Dutch Information Retrieval Workshop (DIR-2009)*, pp 87–94.

3. Van Erp, M., Lendvai, P., and Van den Bosch, A. (2009). Comparing alternative data-driven ontological vistas of natural history. In Bunt, H.C., Petukhova, O., and Wubben, S. (Eds.), *Proceedings of the Eighth International Conference on Computational Semantics (IWCS-2009)*, Tilburg, The Netherlands, pp. 282–285.
4. Wubben, S., and Van den Bosch, A. (2009). A semantic similarity metric based on free link structure. In Bunt, H.C., Petukhova, O., and Wubben, S. (Eds.), *Proceedings of the Eighth International Conference on Computational Semantics (IWCS-2009)*, Tilburg, The Netherlands, pp. 355–358.
5. Wubben, S., Van den Bosch, A., Krahmer, E., and Marsi, E. (2009). Clustering and matching headlines for automatic paraphrase acquisition. In E. Krahmer and M. Theune (Eds.), *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, Athens, Greece, pp. 122–125.
6. Van Erp, M., Van den Bosch, A., Wubben, S., and Hunt, S. (2009). Instance-driven discovery of ontological relation labels. In P. Lendvai and L. Borin (Eds.), *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education (LaTeCH-SHELT&R 2009)*, Athens, Greece, pp. 60–68.
7. Van den Bosch, A. (2009). Machine Learning. In A. Lüdeling and M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*, Vol. 2, pp. 855–872. Berlin: Walter de Gruyter. 2008
8. Hoste, V., and Van den Bosch, A. (2008). A modular approach to learning Dutch co-reference. In C. Johansson (Ed.), *Proceedings from the First Bergen Workshop on Anaphora Resolution (WAR I)*, Bergen, Norway, pp. 51–75.
9. Bogers, T., and Van den Bosch, A. (2008). Using language models for spam detection in social bookmarking. In *Proceedings of the 2008 ECML/PKDD Discovery Challenge Workshop*, pp. 1–12. Antwerp, Belgium, September 2008.
10. Bogers, T., Kox, K., and Van den Bosch, A. (2008). Using citation analysis for expert retrieval in workgroups. In E. Hoenkamp, M. de Cock, and V. Hoste (Eds.), *Proceedings of the 8th Belgian-Dutch Information Retrieval Workshop (DIR 2008)*, pp 21–28. Maastricht, April 2008.
11. Puerta Melguizo, M.C., Muoz Ramos, O., Boves, L., Bogers, T., and Van den Bosch, A. (2008). A personalized recommender system for writing in the internet age. In *Proceedings of the LREC 2008 workshop on Natural Language Processing Resources, Algorithms, and Tools for Authoring Aids*, Marrakech, Morocco, 2008
12. Van den Bosch, A. and Bogers, T. (2008). Efficient context-sensitive word completion for mobile devices. In *MobileHCI 2008: Proceedings of the 10th International Conference on Human-Computer Interaction with Mobile Devices and Services, IOP-MMI special track*, pp 465-470. Amsterdam, September 2008.
13. Bogers, T., and Van den Bosch, A. (2008). Recommending scientific articles using CiteULike. In *RecSys '08: Proceedings of the 2008 ACM Conference on Recommender Systems*, pp 287–290, ACM Press, October 2008. 2007
14. Van den Bosch, A., Stroppa, N., and Way, A. (2007). A memory-based classification approach to marker-based EBMT. In F. Van Eynde, V. Vandeghinste, and I. Schuurman (Eds.), *Proceedings of the METIS-II Workshop on New Approaches to Machine Translation*, pp. 63–72. January 11, 2007, Leuven, Belgium.
15. Stroppa, N., Van den Bosch, A., and Way, A. (2007). Exploiting source similarity for SMT using context-informed features. In A. Way and B. Gawronska (Eds.), *Proceedings of the 11th International Conference on Theoretical Issues in Machine Translation (TMI 2007)*, Skövde University Studies in Informatics 2007:1, Skövde, Sweden, ISSN 1653-2325, pp. 231–240.

16. Balog, K., Bogers, T., Azzopardi, L., De Rijke, M., and Van den Bosch, A. (2007). Broad expertise retrieval in sparse data environments. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 551–558. ACM Press.
17. Van den Bosch, A., Busser, G.J., Canisius, S., and Daelemans, W. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. In P. Dirix, I. Schuurman, V. Vandeghinste, and F. Van Eynde (Eds.), *Computational Linguistics in the Netherlands 2006: Selected Papers of the Seventeenth CLIN Meeting*. Utrecht: LOT, pp. 191–206.
18. Stevens, G., Monachesi, P., and Van den Bosch, A. (2007). A pilot study for automatic semantic role labeling in Dutch. In P. Dirix, I. Schuurman, V. Vandeghinste, and F. Van Eynde (Eds.), *Computational Linguistics in the Netherlands 2006: Selected Papers of the Seventeenth CLIN Meeting*. Utrecht: LOT, pp. 99–114.
19. Puerta Melguizo, M.C., Bogers, T., Deshpande, A., Boves, L., and Van den Bosch, A. (2007). What a proactive recommendation system needs: Relevance, non-intrusiveness, and a new long-term memory. In *Proceedings of the 9th International Conference on Enterprise Information, ICEIS-2007*.
20. Bogers, T., and Van den Bosch, A. (2007). Comparing and evaluating information retrieval algorithms for news recommendation. In *the Proceedings of the 2007 ACM Conference on Recommender Systems*, pp. 141–144. ACM Press.
21. Morante, R. and Van den Bosch, A. (2007). Memory-based semantic role labeling of Catalan and Spanish. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP-2007*, pp. 388–394. Borovets, Bulgaria.
22. Canisius, S., and Van den Bosch, A. (2007). Recompiling a knowledge-based dependency parser into memory. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP-2007*, pp. 104–108. Borovets, Bulgaria.
23. Van den Bosch, A., Marsi, E, and Soudi, A. (2007). Memory-based morphological analysis and part-of-speech tagging of Arabic. In A. Soudi, G. Neumann, and A. van den Bosch (Eds.), *Arabic Computational Morphology: Knowledge-based and Empirical Methods*, Chapter 11. Berlin: Springer, pp. 203–219.
24. Van den Bosch, A., and Van der Sloot, K. (2007). Superlinear parallelisation of the  $k$ -nearest neighbor classifier. In P. Adriaans, M. van Someren, and S. Katrenko (Eds.), *Proceedings of the 18th BENELEARN Conference*, Amsterdam, The Netherlands, pp. 7–14.
25. Hendrickx, I., Morante, R., Sporleder, S, and Van den Bosch, A. (2007). ILK: Machine learning of semantic relations with shallow features and almost no data. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic, pp. 187–190.
26. Van den Bosch, A., and Van der Sloot, K. (2007). Superlinear parallelization of  $k$ -nearest neighbor retrieval. In M. Dastani and E. de Jong (Eds.), *Proceedings of the 19th Belgian-Dutch Artificial Intelligence Conference (BNAIC-2007)*, Utrecht, The Netherlands, pp. 65–72.
27. Bogers, T, Thoonen, W., and Van den Bosch, A. (2006). Expertise classification: Collaborative classification vs. automatic extraction. In *Proceedings of the 17th Annual ASIS&T SIG/CR Workshop on Classification Research*, Austin, TX.
28. Canisius, S., Van den Bosch, A., and Daelemans, W. (2006). Discrete versus probabilistic sequence classifiers for domain-specific entity chunking. In *Proceedings of the Eighteenth Belgian-Dutch Conference on Artificial Intelligence (BNAIC-2006)*, Namur, Belgium, pp. 75–82.

2006

29. Canisius, S., Bogers, T., Van den Bosch, A., Geertzen, J., and Tjong Kim Sang, E. (2006). Dependency parsing by inference over high-recall dependency predictions. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, New York City, NY, USA.
30. Van den Bosch, A., and Canisius, S. (2006). Improved morpho-phonological sequence processing with constraint satisfaction inference. In *Proceedings of the Eighth Meeting of the ACL Special Interest Group in Computational Phonology (SIGPHON-06)*, June 2006, New York City, NY, USA, pp. 61–69.
31. Van den Bosch, A. (2006). All-word prediction as the ultimate confusable disambiguation. In *Proceedings of the HLT-NAACL Workshop on Computationally hard problems and joint inference in speech and language processing*, June 2006, New York City, NY, USA.
32. Sporleder, C., Van Erp, M., Porcelijn, T., Van den Bosch, A, and Arntzen, P. (2006). Identifying named entities in text databases from the natural history domain. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-2006)*, Genoa, Italy.
33. Van den Bosch, A., Schuurman, I, and Vandeghinste, V. (2006). Transferring PoS-tagging and lemmatization tools from spoken to written Dutch corpus development. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-2006)*, Genoa, Italy.
34. Sporleder, C., Van Erp, M., and Van den Bosch, A. (2006). Correcting ‘wrong-column’ errors in text databases. In *Proceedings of the Annual Machine Learning Conference of Belgium and The Netherlands (BENELEARN-06)*, Ghent, Belgium.
35. Sporleder, C., Van Erp, M., Porcelijn, T., and Van den Bosch, A. (2006). Spotting the ‘odd-one-out’: Data-driven error detection and correction in textual databases. In *Proceedings of the EACL-2006 Workshop on Adaptive Text Mining (ATEM-06)*, Trento, Italy.
36. Bogers, T., and Van den Bosch, A. (2006). Authoritative re-ranking of search results. In *Proceedings of the 28th European Conference on Information Retrieval (ECIR 2006)*, vol. 3936 of Lecture Notes on Computer Science, pp. 519–522. Berlin: Springer Verlag.
37. Canisius, S., Van den Bosch, A., and Daelemans, W. (2006). Constraint satisfaction inference: Non-probabilistic global inference for sequence labelling. In *Proceedings of the EACL 2006 Workshop on Learning Structured Information in Natural Language Applications*, Trento, April 2006.
38. Bogers, T., and Van den Bosch, A. (2006). Authoritative re-ranking in fusing authorship-based subcollection search results. In F. de Jong and W. Kraaij (Eds.), *Proceedings of the Sixth Belgian-Dutch Information Retrieval Workshop, DIR-2006*, pp 49–55. Enschede: Neslia Paniculata.
39. Hendrickx, I, and Van den Bosch, A. (2005). Hybrid algorithms with instance-based classification. In *Proceedings of the 19th European Conference on Machine Learning, ECML-2005*, Porto, Portugal.
40. Van den Bosch, A., and Daelemans, W. (2005). Improving sequence segmentation learning by predicting trigrams. In *Proceedings of the Ninth Conference on Natural Language Learning, CoNLL-2005*, Ann Arbor, MI, pp. 80–87.
41. Tjong Kim Sang, E., Canisius, S, Van den Bosch, A., and Bogers, T. (2005). Applying spelling error correction techniques for improving semantic role labelling. In *Proceedings of the Ninth Conference on Natural Language Learning, CoNLL-2005*, Ann Arbor, MI.

2005

42. Marsi, E., Van den Bosch, A., and Souidi, A. (2005). Memory-based morphological analysis generation and part-of-speech tagging of Arabic. In *Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages*, Ann Arbor, MI.
43. Lendvai, P., and Van den Bosch, A. (2005). Robust ASR lattice representation types in pragma-semantic processing of spoken input. In *Proceedings of the AAAI Workshop on Spoken Language Understanding*, Pittsburgh, PA, USA, pp. 15–22.
44. Canisius, S., Van den Bosch, A., and Daelemans, W. (2005). Rule meta-learning for trigram-based sequence processing. In J. Cussens and C. Nedellec (Eds.), *Proceedings of the Fourth Learning Language in Logic Workshop*, pp. 3–10, Bonn, August 2005.
45. Van den Bosch, A. (2005). Memory-based understanding of user utterances in a spoken dialogue system: Effects of feature selection and co-learning. In *Workshop Proceedings of the 6th International Conference on Case-Based Reasoning*, Chicago, August 2005, pp. 85–94. 2004
46. Van den Bosch, A. (2004). Feature transformation through rule induction: A case study with the  $k$ -NN classifier. In J. Fürnkranz (Ed.), *Proceedings of the ECML/PKDD-2004 Workshop on Advances in Inductive Rule Learning*, Pisa, Italy, pp. 1–16.
47. Lendvai, P., Van den Bosch, A., Krahmer, E., and Canisius, S. (2004). Memory-based robust interpretation of recognised speech. In: *Proceedings of SPECOM '04, 9th International Conference "Speech and Computer"*, St. Petersburg, Russia, pp. 415–422.
48. Van den Bosch, A., Canisius, S., Daelemans, W., Hendrickx, I., and Tjong Kim Sang, E. (2004). Memory-based semantic role labeling: Optimizing features, algorithm, and output. In H.T. Ng and E. Riloff (Eds.), *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004)*, Boston, MA, USA.
49. Decadt, B., Hoste, V., Daelemans, W., and Van den Bosch, A. (2004). GAMBL, genetic algorithm optimization of memory-based WSD. In: R. Mihalcea and P. Edmonds (eds.), *Proceedings of the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (Senseval-3)*, Barcelona, Spain, July 2004, pages 108-112.
50. Hendrickx, I., and Van den Bosch, A. (2004). Maximum-entropy parameter estimation for the  $k$ -NN modified value-difference kernel. In R. Verbrugge, N. Taatgen, and L. Schomaker (Eds.), *Proceedings of the 16th Belgian-Dutch Conference on Artificial Intelligence*, Groningen, The Netherlands, pp. 19–26.
51. Van den Bosch, A. (2004). Wrapped progressive sampling search for optimizing learning algorithm parameters. In R. Verbrugge, N. Taatgen, and L. Schomaker (Eds.), *Proceedings of the 16th Belgian-Dutch Conference on Artificial Intelligence*, Groningen, The Netherlands, pp. 219–226.
52. Canisius, S., and Van den Bosch, A. (2004). A memory-based shallow parser for spoken Dutch. In Decadt, B., De Pauw, G. and Hoste, V. (Eds.), *Selected papers from the Thirteenth Computational Linguistics in the Netherlands Meeting*, Antwerp, Belgium, pp. 31–45. 2003
53. Van Herwijnen, O., Van den Bosch, A., Terken, J., and Marsi, E. (2003). Learning PP attachment for filtering prosodic phrasing. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*. New Brunswick, NJ: ACL, pp. 139–146.
54. Lendvai, P., Van den Bosch, A., and Krahmer, E. (2003). Machine learning for shallow interpretation of user utterances in spoken dialogue systems. In *Proceedings of the EACL-03 Workshop on Dialogue Systems: Interaction, Adaptation and Styles of Management*. Budapest, 2003. pp. 69–78.

55. Lendvai, P., Van den Bosch, A., and Krahmer, E. (2003). Memory-based disfluency chunking. In R. Eklund (Ed.), *Proceedings of DISS'03, Disfluency in Spontaneous Speech Workshop*, Göteborg University, Sweden, 2003. pp. 63–66.
56. Hendrickx, I., and Van den Bosch, A. (2003). Memory-based one-step named-entity recognition: Effects of seed list features, classifier stacking, and unannotated data. In: *Proceedings of the Seventh Conference on Natural Language Learning*, CoNLL-2003, Edmonton, Canada, 2003, pp. 176–179.
57. Marsi, E., Reynaert, M., Van den Bosch, A., Daelemans, W., and Hoste, V. (2003). Learning to predict pitch accents and prosodic boundaries in Dutch. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, pp. 489–496. 2002
58. Van den Bosch, A., and Buchholz, S. (2002). Shallow parsing on the basis of words only: A case study. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics*. New Brunswick, NJ: ACL, pp. 433–440.
59. Hendrickx, I., Van den Bosch, A., Hoste, V., and Daelemans, W. (2002). Optimizing the localness of context. In *Proceedings of the Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, PA, USA. New Brunswick, NJ: ACL, pp. 61–65.
60. Maruster, L., Weijters, A., Van der Aalst, W., and Van den Bosch, A. (2002). Process mining: discovering direct successors in process logs. In S. Lange, K. Satoh, C.H. Smith (Eds.): *Discovery Science, 5th International Conference*. Lecture Notes in Computer Science 2534. Berlin: Springer, pp. 364–373.
61. Hoste, V., Daelemans, W., Hendrickx, I, and Van den Bosch, A. (2002). Evaluating the results of a memory-based word expert approach in unrestricted word sense disambiguation. In *Proceedings of the Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, PA, USA. New Brunswick, NJ: ACL, pp. 95–101.
62. Lendvai, P., Van den Bosch, A., Krahmer, E, and Swerts, M. (2002). Improving machine-learned detection of miscommunications in human-machine dialogues through informed data splitting. In *Proceedings of the ESSLLI-2002 Workshop on Machine Learning Approaches in Computational Linguistics*, Trento, Italy.
63. Lendvai, P., Van den Bosch, A., Krahmer, E., and Swerts, M. (2002). Multi-feature error detection in spoken dialogue systems. In *Proceedings of the 12th Computational Linguistics in the Netherlands Meeting*, Twente, the Netherlands.
64. Marsi, E., Busser, G.J., Daelemans, W., Hoste, V., Reynaert, M., and Van den Bosch, A. (2002). Combining information sources for memory-based pitch accent placement. In *Proceedings of the International Conference on Spoken Language Processing ICSLP-2002*, pp. 1273–1276.
65. Van den Bosch, A. (2002). Expanding  $k$ -NN analogy with instance families. In R. Skousen, D. Lonsdale, and D. Parkinson (Eds.), *Analogical Modeling: An exemplar-based approach to language*. Amsterdam: John Benjamins, pp. 209–224. 2001
66. Van den Bosch, A., Krahmer, E., and Swerts, M. (2001). Detecting problematic turns in human-machine interactions: Rule-induction versus memory-based learning approaches. In *Proceedings of the 39th Meeting of the Association for Computational Linguistics*. New Brunswick, NJ: ACL, pp. 499–506.
67. Maruster, L., Van der Aalst, W., Weijters, A., Van den Bosch, A., and Daelemans, W. (2001). Automatic Discovery of Workflow Models from Hospital Data. In *Proceedings of the Eleventh Belgium-Netherlands Conference on Artificial Intelligence*, pp. 183–194.

68. Busser, G.J., Daelemans, W., and Van den Bosch, A. (2001). Predicting phrase breaks with memory-based learning. In *Proceedings of the Fourth ISCA Tutorial and Research Workshop on Speech Synthesis*.
69. Daelemans, W. and Van den Bosch, A. (2001). TreeTalk: Memory-Based Word Phonemisation. In R. I. Dampier (Ed.), *Data-Driven Techniques in Speech Synthesis*. Dordrecht: Kluwer Academic Publishers, pp. 149–172.
70. Hendrickx, I., and Van den Bosch, A. (2001). Dutch word sense disambiguation: Data and preliminary results. In *Proceedings of the Second Workshop on Word Sense Disambiguation, SENSEVAL-2*, Toulouse, France. 2000
71. Van den Bosch, A. and Zavrel, J. (2000). Unpacking multi-valued symbolic features and classes in memory-based language learning. In P. Langley (Ed.), *Proceedings of the Seventeenth International Conference on Machine Learning*. San Francisco, CA: Morgan Kaufmann, pp. 1055–1062.
72. Van den Bosch, A., and Daelemans, W. (2000). A distributed, yet symbolic model of text-to-speech processing. In P. Broeder and J.M.J. Murre (Eds.), *Models of Language Acquisition: inductive and deductive approaches*. Oxford University Press, pp. 76–99.
73. Van den Bosch, A. (2000). Using induced rules as complex features in memory-based language learning. In *Proceedings of the Fourth Conference on Computational Natural Language Learning and of the Second Learning Language in Logic Workshop (CoNLL/LLL)*. New Brunswick, NJ: ACL, pp.73–78.
74. Buchholz, S. and Van den Bosch, A. (2000). Integrating seed names and n-grams for a named entity list and classifier. In: *Proceedings of the Second International Conference on Language Resources and Evaluation, LREC-2000*, Athens, Greece, June 2000, pp. 1215–1221.
75. Weijters, A., Van den Bosch, A., and Postma, E. (2000). Learning statistically neutral tasks without expert guidance. In: S.A. Solla, T.K. Leen and K.R. Müller (Eds.), *Advances in Neural Information Processing*, Vol. 12. Cambridge, MA: The MIT Press, pp. 73–79.
76. Maruster, L., Weijters, A., and Van den Bosch, A. (2000). Local learning methods for makespan estimating in batch process industries. In A. Feelders (Ed.), *Proceedings of the Tenth Dutch-Belgian Conference on Machine Learning (BENELEARN 2000)*. Tilburg: Dept. of Information Management, pp. 69-76.
77. Van den Bosch, A. (2000). Induced rules as complex features in memory-based learning. In A. Feelders (Ed.), *Proceedings of the Tenth Dutch-Belgian Conference on Machine Learning (BENELEARN 2000)*. Tilburg: Dept. of Information Management, pp. 69-76. 1999
78. Van den Bosch, A., and Daelemans, W. (1999). Memory-based morphological analysis. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, ACL'99*, University of Maryland, USA, 20-26 July 1999, pp. 285–292.
79. Van den Bosch, A. (1999). Instance-family abstraction in memory-based language learning. In I. Bratko and S. Dzeroski (Eds.), *Machine Learning: Proceedings of the Sixteenth International Conference, ICML'99*, Bled, Slovenia, June 27-30, 1999, pp. 39–48.
80. Busser, G.J., Daelemans, W., and Van den Bosch, A. (1999). Machine learning of word pronunciation: the case against abstraction. In *Proceedings of the Sixth European Conference on Speech Communication and Technology, Eurospeech99*, Budapest, Hungary, Sept. 5-10, 1999, pp. 2123–2126. 1998
81. Weijters, A., Van den Bosch, A., and Van den Herik, H.J. (1998). Interpretable neural networks with BP-SOM. In C. Nedellec and C. Rouveirol (Eds.), *Machine Learning: ECML-98, Lecture Notes in Artificial Intelligence*, Vol. 1398. Berlin: Springer Verlag, pp. 406–411.

82. Van den Bosch, A., and Daelemans, W. (1998). Do not forget: Full memory in memory-based learning of word pronunciation. In *Proceedings of the Third Conference on New Methods in Language Processing, NeMLaP3, and the Second Conference on Natural Language Learning, CoNLL98*, Sydney, Australia, pp. 195–204.
83. Van den Bosch, A., Weijters, A., and Daelemans, W. (1998). Modularity in inductively-learned word pronunciation systems. In *Proceedings of the Third Conference on New Methods in Language Processing, NeMLaP3, and the Second Conference on Natural Language Learning, CoNLL98*, Sydney, Australia, pp. 185–194.
84. Weijters, A., and Van den Bosch, A. (1998). Interpretable Neural Networks with BP-SOM. In A. P. del Pobil, J. Mira and M. Ali (Eds.), *Tasks and Methods in Applied Artificial Intelligence: Proceedings of the 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Vol II, Lecture Notes in Artificial Intelligence 1416. Berlin: Springer Verlag, pp 564–573.
85. Daelemans, W., Durieux, G, and Van den Bosch, A. (1998). Toward inductive lexicons: a case study. In P. Velardi (Ed.), *Proceedings Workshop on Adapting Lexical and Corpus Resources to Sublanguages and Applications*, Granada, Spain, May 1998, pp. 29–35
86. Daelemans, W., and Van den Bosch, A. (1997). Language-independent data-oriented grapheme-to-phoneme conversion. In J.P.H. van Santen, R.W. Sproat, J.P. Olive, and J. Hirschberg (Eds.), *Progress in Speech Synthesis*, pp. 77–89. New York, NY: Springer Verlag.
87. Daelemans, W., Van den Bosch, A., and Weijters, A. (1997). Empirical learning of natural language processing tasks. In M. van Someren and G. Widmer (Eds.), *Machine Learning: ECML-97*. Lecture Notes in Artificial Intelligence 1224. Berlin: Springer Verlag, pp. 337–344.
88. Wolters, M. and Van den Bosch, A. (1997). Automatic phonetic transcription of words based on sparse data. In Daelemans, W., Van den Bosch, A., and Weijters, A. (Eds.), *Workshop notes of ECML/MLnet Familiarization Workshop on Empirical Learning of Natural Language Processing Tasks*, April 1997, Prague, Czech Republic, pp. 61–70.
89. Van den Bosch, A., Weijters, A., Van den Herik, H.J., and Daelemans, W. (1997). When small disjuncts abound, try lazy learning: A case study. In P. Flach, W. Daelemans, and A. van den Bosch (Eds.), *Proceedings of the Seventh Belgian-Dutch Conference on Machine Learning, BENELEARN-97*, pp. 109–118.
90. Daelemans, W., van den Bosch, A., and Zavrel, J. (1997). A feature-relevance heuristic for indexing and compressing large case bases. In M. van Someren and G. Widmer (Eds.), *Poster Papers of the Ninth European Conference on Machine Learning*, pp. 29–38.
91. Weijters, A., Van den Bosch, A., and Van den Herik, H.J. (1997). Intelligible neural networks with BP-SOM. In *Proceedings of the Ninth Dutch Conference on Artificial Intelligence, NAIC-97*, pp. 27–36.
92. Weijters, A., Van den Herik, H.J., Van den Bosch, A., and Postma, E. (1997). Avoiding overfitting with BP-SOM. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, IJCAI-97*, pp. 1140–1145.
93. Van den Bosch, A., and Weijters, A. (1996). Stretching the limits of learning without modules. In M. van der Heyden, J. Mrcic-Flögel and K. Weigl (Eds.), *Proceedings of the HELNET International Workshop on Neural Networks*, Vol. I/II, Amsterdam: VU University Press, pp. 177–185.
94. Van den Bosch, A., Daelemans, W., and Weijters, A. (1996). Morphological analysis as classification: An inductive-learning approach. In *Proceedings of the Second Conference on New Methods in Language Processing, NeMLaP-2*, Bilkent University, Turkey.

1997

1996

95. Van den Bosch, A., Daelemans, W., and Weijters, A. (1996). An inductive-learning approach to morphological analysis. In Durieux, G., Daelemans, W., and Gillis, S. (Eds.), *Papers from the Sixth Computational Linguistics in the Netherlands Meeting*, December 1996, University of Antwerp, Belgium, pp. 213–230.
96. Weijters, A., Van den Bosch, A., Postma, E., and Van den Herik, H.J. (1996). Avoiding overfitting in BP-SOM. In H.J. van den Herik and A. Weijters (Eds.), *Proceedings of BENELEARN-96*, Maastricht, pp. 157–166. 1995
97. Van den Bosch, A., Weijters, A., and van den Herik, H.J. (1995). Scaling effects with greedy and lazy machine-learning algorithms. In *Proceedings of the Seventh Dutch Conference on AI, NAIC-95*, Erasmus University, Rotterdam, pp. 211–218.
98. Van den Bosch, A., Weijters, A., Van den Herik, H.J., and Daelemans, W. (1995). The profit of learning exceptions. In *Proceedings of the Fifth Belgian-Dutch Conference on Machine Learning, BENELEARN-96*, Brussels, pp. 118–126.
99. Weijters, A., and Van den Bosch, A. (1995). Connectionism. In Verschueren, J., Östman, J.O., and Blommaert, J. (Eds.), *Handbook of Pragmatics: Learning Manual*, pp. 165–171. Amsterdam: Benjamins. 1994
100. Van den Bosch, A., Content, A., Daelemans, W., and De Gelder, B. (1994). Analysing orthographic depth of different languages using data-oriented algorithms. In A. A. Polikarpov (Ed.), *Proceedings of the Second International Conference on Quantitative Linguistics, QUALICO-94*, Moscow, pp. 26–31.
101. Daelemans, W. and van den Bosch, A. (1994). A language-independent, data-oriented architecture for grapheme-to-phoneme conversion. In *Proceedings of ESCA-IEEE Speech Synthesis Conference '94*, New York, pp. 199–203. 1993
102. Daelemans, W. and van den Bosch, A. (1993). Tabtalk: Reusability in data-oriented grapheme-to-phoneme conversion. In *Proceedings of Eurospeech'93*, pp. 1459–1466. Berlin: T.U. Berlin.
103. Van den Bosch, A., and Daelemans, W. (1993). Data-oriented methods for grapheme-to-phoneme conversion. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics, EACL-93*, pp. 45–53. Also available as ITK Research Report 42, 1993, and in W. Sijtsma and O. Zweekhorst (Eds.), *Computational Linguistics in the Netherlands: Papers from the third CLIN meeting 1992*, Tilburg: ITK, pp. 13–26.
104. Gillis, S., Durieux, G., Daelemans, W., and van den Bosch, A. (1993). Learnability and markedness: Dutch stress assignment. In *Proceedings of the 15th Conference of the Cognitive Science Society 1993*, Boulder, CO, pp. 452–457.
105. Daelemans, W., Gillis, S., Durieux, G., and Van den Bosch, A. (1993). Learnability and markedness in data-driven acquisition of stress. In T.M. Ellison and J.M. Scobbie (Eds.), *Computational Phonology*. Edinburgh Working Papers in Cognitive Science, 8, pp. 157–178. Also available as ITK Research Report 43, Tilburg: ITK.
106. Daelemans, W., Van den Bosch, A., Gillis, S., and Durieux, G. (1993). A data-driven approach to stress acquisition. In P. Adriaans (Ed.), *ECML-93 Workshop Notes on Machine Learning Techniques and Text Analysis*, Vienna, pp. 15–24. 1992
107. Daelemans, W. and van den Bosch, A. (1992). A neural network for hyphenation. In: I. Aleksander and J. Taylor (Eds.), *Artificial Neural Networks 2*, Vol. 2. Amsterdam: North-Holland, pp. 1647–1650.

108. Van den Bosch, A., and Daelemans, W. (1992). Linguistic pattern matching capabilities of connectionist networks. In: J. van Eijk and W. Meyer Viol (Eds.), *Proceedings of the Computational Linguistics in the Netherlands meeting 1991*. Utrecht: OTS, pp. 40–53 (also appeared in W. Daelemans and D. Powers (Eds.), *Proceedings First SHOE Workshop*, Tilburg: ITK, pp. 183–196.)
109. Daelemans, W. and van den Bosch, A. (1992). Generalization performance of backpropagation learning on a syllabification task. In M.F.J. Drossaers and A Nijholt (Eds.), *Proceedings of TWLT3: Connectionism and Natural Language Processing*, pp. 27–37. Enschede: University of Twente.
110. Gillis, S., Durieux, G., Daelemans, W. and van den Bosch, A. (1992). Exploring artificial learning algorithms: Learning to stress Dutch simplex words. *Antwerp Papers in Linguistics* 71.

### Edited volumes

1. Souidi, A., Van den Bosch, A., and Neumann, G. editors (2007). *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Berlin: Springer. ISBN 978-1-4020-6045-8.
2. Zaenen, A., and Van den Bosch, A., editors (2007). *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, ACL-2007*. Prague, Czech Republic.
3. Sporleder, C., Van den Bosch, A., and Grover, C., editors (2007). *Proceedings of the Workshop on Language Technology for Cultural Heritage, LaTeCH-2007*. Prague, Czech Republic.
4. Roth, D., and Van den Bosch, A., editors (2002). *Proceedings of the Sixth Conference on Language Learning, CoNLL-2002*. New Brunswick, NJ: ACL.
5. Van den Bosch, A., and Weigand, H., editors (2000). *Proceedings of the Twelfth Belgium-Netherlands Artificial Intelligence Conference, BNAIC'00*. Tilburg: ILK/Infolab.
6. Flach, P., Daelemans, W., and Van den Bosch, A., editors (1997). *Proceedings of the Seventh Belgian-Dutch Conference on Machine Learning, BENELEARN-97*. Tilburg: ILK/Infolab.
7. Daelemans, W., Weijters, A., and Van den Bosch, A., editors (1997). *Workshop notes of ECML/MLnet Familiarization Workshop on Empirical Learning of Natural Language Processing Tasks*, April 1997, Prague, Czech Republic.

### Reports, extended abstracts

1. Daelemans, W., Zavrel, J., Van der Sloot, K., and Van den Bosch, A. (2007). *TiMBL: Tilburg Memory-Based Learner, version 6.1, Reference Guide*. ILK Technical Report no. ILK-0707. Earlier versions 1998–2007.
2. Daelemans, W., Zavrel, J., Van der Sloot, K., and Van den Bosch, A. (2007). *MBT: Memory-Based Tagger, version 3.1, Reference Guide*. ILK Technical Report no. ILK-0708. Earlier versions 2002–2007.
3. Van den Bosch, A., Sporleder, C, Van Erp, M., and Hunt, S. (2007). Automatic techniques for generating and correcting cultural heritage collection metadata. In *Proceedings of the 19th Joint International Conference of the Association for Computers and the Humanities, and the Association for Literary and Linguistic Computing, DH-2007*, University of Illinois, Urbana-Champaign, pp. 223–224.
4. Van den Bosch, A. (2004). *Fambl 2.3 Reference Guide*. ILK Technical Report ILK-0403.

5. De Smedt, K., Black, B., Van den Bosch, A., Lavid Lopez, J., McKeivitt, P., Way, A. (2000). European studies on computational linguistics. In K. De Smedt, H. Gardiner, E. Ore, T. Orlandi, H. Short, J. Souillot, and W. Vaughan (Eds.), *Computing in the Humanities Education: A European Perspective*. Bergen, Norway: University of Bergen, pp. 89–154.
6. Weijters, A., Van den Bosch, A., and Van den Herik, H.J. (1996). *Applying Ockham's razor to back-propagation*. Technical Reports in Computer Science CS 96-03, Department of Computer Science, University of Maastricht, the Netherlands.
7. Van den Bosch, A. (1994). A distributed, yet symbolic model of text-to-speech processing. Extended abstract in the *Proceedings of the Cognitive Models of Language Acquisition Workshop*, Tilburg University, pp. 91–93.
8. Daelemans, W., Gillis, S., Van den Bosch, A., and Durieux, G. (1993). Learning linguistic mappings: an instance-based learning approach. In *Conference Abstracts of the ACH-ALLC 1993 International Joint Conference*, Georgetown University, Washington, DC, pp. 38–40.

### Non-reviewed and popularizing publications

1. Van den Bosch, A. (2008). *Het volgende woord*. Inaugural address (in Dutch). Tilburg University, October 10, 2008. Published online at <http://ilk.uvt.nl/hetvolgendewoord>
2. Van den Bosch, A. (2008). Digitale verhalen. In H. van Lierop (Ed.), *Jeugdliteratuur en andere media: Cultuureducatie in woord en beeld*. Leidschendam: Biblion Uitgeverij, pp. 16–26.
3. Daelemans, W., and Van den Bosch, A. (2007). Dat gebeurt mei niet: Computationale modellen voor verwarbare homofonen. In D. Sandra, R. Rymenans, P. Cuvelier, and P. van Petegem (Eds), *Tussen taal, spelling en onderwijs: Essays bij het emeritaat van Frans Daems*. Gent: Academia Press, pp. 199–210.
4. Van den Bosch, A. (2007). Memory-based re-engineering of a knowledge-based dependency parser. In Donkers, J., Mommers, L., Postma, E., and Schmidt, A. (Eds), *Liber amicorum ter gelegenheid van de 60e verjaardag van Prof.dr. H. Jaap van den Herik*. Maastricht: MICC, pp. 168–175.
5. Van den Bosch, A. (2006). Machine learning and natural language processing in Tilburg. *BNVKI Newsletter*, **23:3**, pp. 52–57.
6. Van den Bosch, A. (2005). Informatie zoeken. In H. van Driel (Ed.), *Digitaal Communiceren*. Meppel, the Netherlands: Boom. (in Dutch).
7. Van den Bosch, A. (2002). Text mining for science. In J. Meij (Ed.), *Dealing with the data flood: Mining data, text, and multimedia*. Den Haag, the Netherlands: Stichting Toekomstbeeld der Techniek.
8. Meij, J., and Van den Bosch, A. (2002). Text mining techniques. In J. Meij (Ed.), *Dealing with the data flood: Mining data, text, and multimedia*. Den Haag, the Netherlands: Stichting Toekomstbeeld der Techniek.
9. Van den Bosch, A. (2002). Implementations of memory and analogy for language processing. *NVTI Newsletter, Dutch Foundation for Theoretical Computer Science*, January 2002.
10. Daelemans, W., and Van den Bosch, A. (1992). De voorleesmachine. *Informatica Nieuwsbrief*, **2:9**, pp. 10–15. Alphen aan den Rijn: Samsom. (in Dutch)

## Media productions

1. Van den Bosch, A. (2007). *k-Nearest Neighbor Classification*. PAL Video, with sound. 3m:21s. Tilburg University.
2. Brienen, S., Graaf, F. de, Lupko, E., Maarsseveen, I. van, Veling, K., Vilsteren, P. van, en Van den Bosch, A. (2000). *Kwintessens*, issue 3. PAL video, with sound. 30m. First broadcast March 13, 2000, Teleac-NOT, Dutch public television.

## Forthcoming

1. Canisius, S. and Van den Bosch, A. (to appear). A constraint satisfaction approach to machine translation. To appear in L. Marquez and H. Somers (Eds.), *Proceedings of the Thirteenth Annual Conference of the European Association for Machine Translation (EAMT-2009)*, Barcelona, Spain.

## Other scientific activities

### Invited lectures

1. *Memory-based machine translation*.  
CNTS Colloquium, Centre for Dutch Language and Speech, University of Antwerp, Belgium, April 9, 2009.
2. *Constraint satisfaction inference for dependency parsing and MT*.  
NLTC Seminar Series, Dublin City University, Dublin, Ireland, February 17, 2009.
3. *Implicit Linguistics*.  
CLCG Linguistics Colloquium, Groningen, The Netherlands, September 19, 2008.
4. *Mining natural history*.  
SIMIN Spring Symposium, Municipal Museum, The Hague, The Netherlands, June 6, 2008.
5. *Automatic Part-of-Speech tagging*.  
Diachronic Linguistics Meeting, Meertens Institute, Amsterdam, The Netherlands, May 30, 2008.
6. *Facts are little theories*.  
Semantic Web Seminar, University of Utrecht, Utrecht, The Netherlands, March 9, 2008. Dr. Virginia Dignum.
7. *Explicit versus implicit computational linguistics*.  
XXIII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN-2007), Sevilla, Spain, September 10, 2007.
8. *Natural language processing and machine learning*.  
Free University of Amsterdam AI Group "Natural Language Processing Days", Leusden, The Netherlands, December 12, 2006.
9. *A memory-based learning-plus-inference approach to morphological analysis*.  
Workshop on Flexible Architectures for Large Vocabulary Recognition (FLaVoR), Leuven, Belgium, November 17, 2006.
10. *Constraint Satisfaction Inference for Discrete Sequence Processing in NLP*.  
Dublin City University, National Centre for Language Technology Seminar Series, Dublin, Ireland, April 19, 2006.
11. *Implicit computational linguistics*.  
University of Amsterdam, Institute for Logic, Language, and Computation, Computational Linguistics Seminar, March 22, 2006.

12. *Automatic correction of coreference resolution and chaining.*  
Workshop on Anaphora Resolution, Mjøelvfjell, Norway, September 29, 2005.
13. *Memory-based understanding of user utterances in a spoken dialogue system: Effects of feature selection and co-learning.*  
Textual Case Based Reasoning Workshop at ICCBR-2005, Chicago, IL, August 24, 2005.
14. *Understanding domain-specific questions and answers.*  
Language and Speech Colloquium Series, Nijmegen University, June 1, 2005.
15. *Learning linguistic sequences.*  
Computational Linguistics Seminar Series, University of Bergen, Bergen, Norway, May 2, 2005.
16. *Learning syntactically valid sequences.*  
Workshop on Machine Learning of Information Extraction, Antwerp, Belgium, March 10, 2005.
17. *Understanding the "More Data" Effect: A Closer Look at Learning Curves.*  
University of Edinburgh Division of Informatics Institute for Communicating and Collaborative Systems / Human Communication Research Centre Seminar Series, January 16, 2004.
18. *Unravelling the intermediate representation myth in natural language processing*  
*Computing with LLI Seminar series*, ILLC, University of Amsterdam, May 31, 2002, on invitation by dr. M. de Rijke.
19. *Text mining*  
Symposium *Dealing with the data flood*, Stichting Toekomstbeeld der Techniek, Rotterdam, April 23, 2002.
20. *Machine learning: Tools for natural language processing*  
*Maastricht Machine Learning Day*, January 10, 2002.
21. *Issues in memory-based natural language processing: Memory, representation, and data*  
University of Pittsburgh, Learning Research and Development Center (LRDC), November 30, 2001.
22. *Conjunctive and disjunctive feature construction*  
WhizBang! Labs–Research colloquium, September 27, 2001.
23. *On the usage of word graphs (and more) to detect dialogue problems.*  
Nijmegen colloquium on Language and Speech, Katholieke Universiteit Nijmegen, Nijmegen, The Netherlands, April 4, 2001. With Emiel Krahmer and Marc Swerts.
24. *Situating AI, language and logic in a national Dutch research programme on cognition.*  
NWO Consultation Day "Fruits of Enlightenment: A special programme for the cognitive sciences", Utrecht, February 9, 2001.
25. *Memory-based language processing.*  
IPO Center for User-System Interaction, Eindhoven Technical University, November 10, 2000.
26. *Memory-Based learning and TiMBL.*  
Analogical Modeling of Language (AML) Conference, Brigham Young University, Provo, Utah, March 22, 2000. With Walter Daelemans.
27. *Memory-based language processing: some findings and applications.*  
Text Learning group meeting, School of Computer Science, Carnegie Mellon University, Pittsburgh PA, March 20, 2000.
28. *Robust language technology for speech technology.*  
IPO Center for User-System Interaction Colloquium series, IPO, Eindhoven Technical University, the Netherlands, March 10, 2000.
29. *Instance-family abstraction in memory-based learning.*  
Ninth Dutch-Belgian Conference on Machine Learning (BENELEARN'99), Maastricht, the Netherlands, November 5, 1999.
30. *Learning word pronunciation.*  
Parlevink Colloquium Series, Department of Theoretical Informatics, University of Twente, The Netherlands. February 9, 1998.
31. *Experiments in machine learning of morpho-phonology.*  
ITK Colloquium Series, Institute for Language Technology and AI, Tilburg University, The Netherlands. December 7, 1995.

32. *A language-independent, data-oriented architecture for grapheme-phoneme conversion*.  
Department of Computer Science Colloquium, Universiteit Maastricht, The Netherlands. May 16, 1994.

## Training experience

### Graduate training

#### *Tilburg University*

*Text mining*, block course, curriculum Master *Human Aspects of Information Technology*, Faculty of Humanities, Tilburg University. First semester 2007/2008 (earlier version of course 2004/2005). With Walter Daelemans.

*Digitaal erfgoed* (Digital heritage), semester course, curriculum Bachelor minor *Computing for the Humanities*, Faculty of Arts, Tilburg University. First semester 2005/2006 – 2008/2009.

*ICT voor studie en werk* (ICT for study and work), semester course, curriculum Bachelor ACW and CIW, Faculty of Humanities, Tilburg University. First semester 2006/2007, 2007/2008. With Jan de Vuijst.

*Van tekst naar informatie* (From text to information), semester course, curriculum Bachelor minor “Computing for the Humanities”, Faculty of Arts, Tilburg University. Second semester 2005/2006. With Sander Canisius.

*Language and speech technologies: Advanced*, semester course, Research Master “Language and Communication”, Faculties of Arts, Tilburg University and Radboud University Nijmegen. Second semester 2005/2006.

*Informatietechnologie* (Information technology, before 2005/2006 named *Taal- en informatietechnologie*, Language and information technology), semester block course, curriculum Computational Linguistics and AI / Communication and Digital Media, Faculty of Arts, Tilburg University. Second semester 2001/2002 – 2005/2006. With Emiel Kraemer.

*Lerende systemen* (Machine learning), semester course, curriculum Computational Linguistics and AI, Faculty of Arts, Tilburg University. Second semester 1999/2000, 2000/2001, 2002/2003.

#### *University of Antwerp*

*Language Technology*, guest lecturer, semester course, University of Antwerp, Belgium. Second semester 2001/2002 – 2005/2006.

*Capita selecta Computational Linguistics: Machine Learning of Language*, guest lecturer, semester course, University of Antwerp, Belgium. Second semester 2000/2001.

#### *Universiteit Maastricht*

*Automatisch leren en redeneren* (Machine learning and automated reasoning), trimester project, curriculum Knowledge Technology, Faculty of General Sciences, Universiteit Maastricht. With dr. E. Postma, dr. A. Weijters, and dr. P. Braspenning. Second trimester 1995/1996 and 1996/1997.

*Kunstmatige intelligentie* (Artificial intelligence), trimester course, curriculum Psychology, Faculty of Psychology, Universiteit Maastricht. With dr. E. Postma. Third trimester 1995/1996.

## Post-graduate training

*Computational Linguistics: A Machine-Learning Approach*, four-day Ph.D. course, Universitat de Barcelona, Barcelona, Spain, October 9–13, 2006.

*Machine learning*, half-semester Ph.D. course shared with Eindhoven Technical University, Tilburg and Eindhoven, April–July 2006.

*Issues in empirical machine learning research*, course module, part of School of Information and Knowledge Systems (SIKS) Ph.D. course on Research methods and methodology for information and knowledge systems, Zeist, November 2005 and Lunteren, November 2006.

*Memory-based language processing*, one-week advanced course as part of ESSLLI-2003 (15th European Summer School in Logic, Language, and Information), Vienna, Austria, August 18–22, 2003. With Walter Daelemans.

*Memory-based analogical language processing*, one-week course as part of LOT Summer School (Dutch National Research School on Linguistics), Tilburg University, June 23–27, 2003.

*Machine Learning and Memory Based Learning for Natural Language Processing*, three-day Ph.D. course as part of a Forskerutdanningskurs Statistik Språkbehandling (Linguistics Workshop on Statistical Language Processing), University of Bergen, Norway, September 30 – October 4, 2002.

*Inductive language learning*, one-week introductory course as part of ESSLLI-X (the Tenth European Summer School in Logic, Language and Information), Saarbrücken, Germany, August 23–28, 1998.

## Tutorials

*Machine learning for text and speech processing*, tutorial at the Interspeech 2007 Conference, Antwerp, Belgium, August 27, 2007. With Walter Daelemans.

*Machine learning of natural language*, tutorial at the 17th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD06), Berlin, Germany, September 22, 2006. With Walter Daelemans.

*Models of memory and analogy for natural language learning and processing*, tutorial at the 27th Annual Meeting of the Association for Computational Linguistics (ACL-99), University of Maryland, College Park, Maryland, USA, 26 June 1999.

## Project management and organisation

Research project management:

(abbreviations: NWO = Netherlands Organisation for Scientific Research; SoBU = Cooperation Association of the Brabant Universities; FWO = Flanders Organisation for Scientific Research; EZ = Dutch Ministry for Economic Affairs; OCW = Dutch Ministry for Education, Culture, and Science; Totalling budgets (including matching) are given in Euros (EUR).)

SoBU project 98-D, *Automatic inductive learning for planning*. 2000-2003. With prof. W. Daelemans, dr. A. Weijters, and prof. H. Wortmann (Eindhoven Technical University). Researcher: L. Maruster (Eindhoven Technical University). 100,000 EUR.

- SoBU project 00-C, *Learning to communicate: Machine learning of dialogue strategies*. 2001-2004. With E. Krahmer, J. Terken (Eindhoven Technical University). Researcher: P. Lendvai. 105,000 EUR.
- SoBU project 02-10, *Weighted error minimisation in assigning prosodic structure to synthetic speech*. 2002-2004. With J. Terken (Eindhoven Technical University). Researcher: O. van Herwijnen (Eindhoven Technical University). 90,000 EUR.
- PROSIT, Prosody from Information in Text*. Funded by the joint NWO and FWO (Flanders Organisation for Scientific Research) Vlaams-Nederlands Comité voor Taal en Cultuur. 2001-2004. With W. Daelemans (UIA, Antwerp, Belgium). Researchers: E. Marsi, M. Reynaert, V. Hoste (UIA). 620,000 EUR.
- Memory models of language*. Funded by NWO *Vernieuwingsimpuls* programme, 2001-2005. Researchers: A. van den Bosch, I. Hendrickx, G.J. Busser. 680,000 EUR.
- VindIT: Combining visual and textual information for information retrieval*. Funded by NWO ToKeN-2000 Programme, 2002-2005. With E.O. Postma (Universiteit Maastricht), L. Vuurpijl and E. Hoenkamp (Nijmegen University). Researcher: M. van Zaanen. 686,000 EUR. (Tilburg part: 210,000 EUR).
- NEXTENS, Nederlandse Extensie voor Tekst naar Spraak*. Funded by SST, Stichting Spraaktechnologie (Foundation for Speech Technology), 2002-2003. With T. Rietveld (KUN). Development and research: E. Marsi, A. Russell. 40,000 EUR.
- Prosodic annotation for the *Spoken Dutch Corpus*, student assistant team, data generation, and evaluation. Funded by NWO Spoken Dutch Corpus Programme. 2002-2003. With E. Marsi. 25,000 EUR.
- ROLAQUAD: Robust Language Understanding in Question-Answer Dialogues*. Part of *IMIX: Interactive Multimodal Information Extraction*, an NWO Programme. Funded by NWO, 2004-2008. With W. Daelemans, J. Zavrel (Textkernel B.V.). Researchers: S. Canisius and P. Lendvai. 600,000 EUR.
- A-Propos: Pro-active personalization for professional document writing*. Funded by the EZ IOP-MMI (Innovative research programmes: Human-machine interaction) programme, 2004-2008. With L. Boves (Radboud University) and a consortium of Dutch language technology companies. Researcher: T. Bogers. 830,000 EUR (Tilburg part: 200,000 EUR).
- MITCH: Mining Information from Texts in the Cultural Heritage*. Part of *CATCH: Continuous Access to the Cultural Heritage*, an NWO Programme. Funded by NWO, 2005-2009. With W. Daelemans. Researchers: C. Sporleder (post-doc), P. Lendvai (post-doc), M. van Erp (Ph.D. student). Scientific programmer: T. Porcelijn, S. Hunt. 560,000 EUR.
- D-Coi: Dutch Corpus Initiative*. Funded by *STEVIN: Spraak-en Taaltechnologische Essentiële voorzieningen in het Nederlands*, an NWO/FWO/EZ/OCW/AWI/IWT Dutch Language Union Programme. With partners from other Dutch and Flemish universities, 2005–2006. Post-doc researcher: M. Reynaert. 580,000 EUR (Tilburg part: 84,000 EUR)
- Implicit Linguistics*. Funded by NWO *VICI* program, 2006-2011. Researchers: A. van den Bosch, M. van Zaanen (post-doc researcher), H. Stehouwer (Ph.D. student). Scientific programmer: P. Berck. 1,250,000 EUR.
- Linguistic Data Quality Management*. Funded by NWO *HEFBOOM* program, 2006-2008. Post-doc researcher: M. Reynaert. 81,000 EUR.
- SoNaR: STEVIN Nederlandstalig Referentiecorpus, Phase 1*. Funded by *STEVIN: Spraak-en Taaltechnologische Essentiële voorzieningen in het Nederlands*, an NWO/FWO/EZ/OCW/AWI/IWT Dutch Language Union Programme. With partners from other Dutch and Flemish universities, 2008. Post-doc researcher: T. Gaustad van Zaanen. 100,000 EUR (Tilburg part: 45,000 EUR)

*MEMPHIX: Memory-based Paraphrasing with Implicit and Explicit Semantics*, a Tilburg University Faculty of Humanities Graduate School project. Ph.D. student: S. Wubben. With profs. H. Bunt and E. Krahmer.

*HITIME: Historical Timeline Mining and Extraction*. Part of *CATCH: Continuous Access to the Cultural Heritage*, an NWO Programme. Funded by NWO, 2009-2013. Researchers: one post-doc researcher, one Ph.D. student, one scientific programmer. 614,000 EUR.

*DutchSemCor: A one-million-word text corpus for Dutch with sense-tags*, an NWO GW Investeringsubsidie Middelgroot. With prof. P. Vossen (VU Amsterdam), prof. M. de Rijke (University of Amsterdam). 751,000 EUR (Tilburg part: 140.250 EUR).

President of SIGNLL, *Special Interest Group on Natural Language Learning* of the Association for Computational Linguistics (ACL) (2005–2007; Secretary 2003–2005; Information officer 1995–2003).

Treasurer of STIL, *Stichting Toepassing Inductieve Leertechieken* (Foundation for Inductive Learning Applications) (2008–present; Chair 2002–2008; Founding board member since 1997).

Chair of BNVKI-AIABN: Benelux Association for Artificial Intelligence (Belgisch–Nederlandse Vereniging voor Kunstmatige Intelligentie / Association de l’Intelligence Artificiel Belgique–Néerlandais) (2006–2011; Secretary 2001–2006).

## Awards

*Best Explanation Award*, AI Video Competition, AAI-2007, Vancouver, CA, August 2007.

*Research Award 2002*, Faculty of Arts, Tilburg University. With Hans Broekhuis and Ad Backus.

*SNS Limburg Award 1998*, best Ph.D. thesis Universiteit Maastricht 1997–1998.

## Advice

Advisor (promotor) or co-advisor (co-promotor)

- Sander Canisius, *Structured prediction for natural language processing: A constraint satisfaction approach*, Tilburg University, February 13, 2009. Promotor, with second promotor Walter Daelemans.
- Martin Reynaert, *Text induced spelling correction*, Tilburg University, December 2, 2005. Co-promotor.
- Iris Hendrickx, *Local classification and global estimation: Explorations of the k-nearest neighbor algorithm*, Tilburg University, November 28, 2005. Co-promotor.
- Piroska Lendvai, *Extracting information from spoken input*, Tilburg University, December 20, 2004. Co-promotor.
- Laura Maruster, *A machine learning approach to understand business processes*, Eindhoven Technical University, August 27, 2003. Co-promotor.
- Sabine Buchholz, *Memory-based grammatical relations*, Tilburg University, The Netherlands, December 13, 2002. Co-promotor.

Member of Ph.D. committee

- Jeroen Geertzen, *Dialogue act recognition and prediction: Explorations in computational dialogue modelling*, Tilburg University, February 11, 2009. Committee member.

- Benjamin Torben Nielsen, *Dendritic morphology: Function shapes structure*, Tilburg University, December 3, 2008. Committee member.
- Jesús Giménez Linares, *Empirical machine translation and its evaluation*, Universitat Politècnica de Catalunya, Barcelona, Spain, July 2, 2008. Jury member.
- Grzegorz Chrupala, *Towards a machine-learning architecture for lexical functional grammar parsing*, Dublin City University, Ireland, June 10, 2008. External examiner.
- Emmanuel Keuleers, *Memory-based learning of inflectional morphology*, Universiteit Antwerpen, Belgium, March 6, 2008. Jury member.
- Christian Biemann, *Unsupervised and knowledge-free natural language processing in the structure discovery paradigm*, Universität Leipzig, Germany, 2007. Referee.
- Xavier Carreras, *Learning and inference in phrase recognition: A filtering-ranking architecture using perceptron*, Universitat Politècnica de Catalunya, Barcelona, Spain, October 28, 2005. Committee member.
- Véronique Hoste, *Optimization issues in machine learning of coreference resolution*, University of Antwerp, Belgium, March 11, 2005. Committee member.
- Olga van Herwijnen, *Weighted error minimisation in assigning prosodic structure for synthetic speech*, Eindhoven Technical University, May 17, 2004. Committee member.
- Stefan Frank, *Computational modeling of discourse comprehension*, Tilburg University, February 27, 2004. Committee member.
- Brian Sheppard, *Toward perfect play of scrabble*, Universiteit Maastricht, July 5, 2002. External reviewer.

#### Member of editorial board

- Computational Linguistics (2003–2005)
- Research on Language and Computation (2006–2011)

#### Reviewer for journal (alphabetic)

- ACM Computing Surveys (2008)
- AI Communications (2004, 2005, 2007)
- Artificial Intelligence (2007)
- Annals of Biomedical Engineering (2007)
- Computers in Industry (2003, 2008)
- Computational Intelligence (2003, 2006)
- Computational Linguistics (2003–2007)
- Data and Knowledge Engineering (2002, 2005)
- IEEE Transactions on Knowledge and Data Engineering (2002, 2004)
- IEEE Transactions on Speech and Audio Processing (1998)
- Information Fusion (2008)
- International Computer Games Association Journal (2003)
- Iranian Journal of Electrical and Computer Engineering (2006)
- Journal for Artificial Intelligence Research (2003–2005)
- Journal of Machine Learning Research (2004–2006)
- Language Resources and Evaluation (2005–2008)
- Literator: Tydskrif vir besondere en vergelykende taal- en literatuurstudie (2008)
- Machine Learning Journal (1998, 2004)

- Research on Language and Computation (2003, 2007, 2008)
- Speech Communication (2004, 2005, 2007)
- Written Language and Literacy (2004, 2007)

Scientific programme chairmanship (reversed chronological)

- ACL-2007, 45th Annual Meeting of the Association for Computational Linguistics. June 25–30, 2007, Prague, Czech Republic. Programme co-chair with Annie Zaenen (PARC).
- LaTeCH, ACL-2007 Workshop on Language Technology for Cultural Heritage. June 28, 2007, Prague, Czech Republic. Programme co-chair with Caroline Sporleder (Tilburg University, Saarland University) and Claire Grover (University of Edinburgh).
- ESSLLI-2004, European Summer School on Logic, Language, and Information. August 9–20 2004, Nancy, France. Programme chair.
- CoNLL-2002, Sixth Conference on Computational Natural Language Learning. August 31 – September 1, 2002, Taipei, Taiwan. Programme co-chair with D. Roth (UIUC).
- BNAIC-2000, Twelfth Belgian-Dutch Artificial Intelligence Conference. November 1–2 2000, Kaatsheuvel, The Netherlands. Programme co-chair with H. Weigand (Tilburg University).

Member of programme committee (alphabetic)

- AAAI, Annual Meeting of the Association for the Advancement of Artificial Intelligence (2008). Satellite events:
  - AAAI-2005 Workshop on Spoken Language Understanding for Conversational Systems, Pittsburgh, PA, USA (2005)
- ACL, Annual Meeting of the Association for Computational Linguistics (2003, 2005, 2006, 2008, 2009, area chair for ACL-2004, programme chair for ACL-2007). Satellite events:
  - ACL-2002 Morphological and Phonological Learning Workshop, Philadelphia, PA, USA (2002)
  - ACL-2005 Workshop on Psychocomputational Models of Language Acquisition, Ann Arbor, MI, USA (2005)
  - ACL-2005 Workshop on Software
  - ACL-2007 Workshop on Language Technology for Cultural Heritage, Prague, Czech Republic (2007)
  - ACL-2007 Workshop on Cognitive Aspects of Computational Language Acquisition, Prague, Czech Republic (2007)
- BENELEARN, Belgian-Dutch Conference on Machine Learning (1996 – 2001, 2005 – 2009)
- BNAIC, Belgium-Netherlands Artificial Intelligence Conference (1998 – 2008)
- CLIN, Computational Linguistics in the Netherlands (1999–2008)
- COGSCI, The Annual Meeting of the Cognitive Science Society (2009)
- COLING, International Conference on Computational Linguistics (2002, 2008)
- CoNLL, Conference on Computational Natural Language Learning (1998 – 2009)
- DIR, Dutch-Belgian Information Retrieval Workshop (2008, 2009)
- EACL, Annual Meeting of the European Chapter of the Association for Computational Linguistics (2001, 2009, area chair for EACL-2006). Satellite events:

- EACL-2009 Workshop on Computational Linguistic Aspects of Grammatical Inference, Athens, Greece (2009)
- EACL 2009 Workshop on Cognitive Aspects of Computational Language Acquisition, Athens, Greece (2009)
- ECAI, European Conference on Artificial Intelligence (2008)
- ECML, European Conference on Machine Learning (1998, 2003; area chair for ECML-PKDD-2008). Satellite event:
  - ECML-97 MLNet Familiarisation Workshop on Empirical Learning of Natural Language, Prague, Czech Republic (1997)
- EMNLP, Empirical Methods in Natural Language Processing Conference (2004 – 2009)
- ESSLLI, European Summer School on Logic, Language, and Information (programme chair in 2004). Satellite events:
  - ESSLLI-2002 Workshop on Machine Learning Approaches in Computational Linguistics, Trento, Italy (2002)
  - ESSLLI-2007 Workshop on Example-based Models of Language Acquisition and Use, Dublin, Ireland (2007)
  - ESSLLI Student Session (2005 – 2008)
- ICML, International Conference on Machine Learning (2003, 2007, 2008, area chair for 2005 and 2006)
- IJCAI, International Joint Conference on Artificial Intelligence (2003)
- INFWET, Interdisciplinaire Conferentie Informatiewetenschap (2004)
- Interspeech/ICSLP, International Conference on Spoken Language Processing (2006, 2007)
- LaTeCH, Workshop on Language Technology for Cultural Heritage (2007 – 2009)
- LREC, International Conference on Language Resources and Evaluation (2006, 2008)
- NIPS, Conference on Neural Information Processing Systems (2008)

#### Scientific advice

- Reviewer, NWO *Cognition* program (2001)
- Reviewer, NWO *Vernieuwingsimpuls* program (2002–2004, 2007)
- Reviewer, NWO *Mozaiek* program (2004)
- Reviewer, NWO *Talent* program (2004)
- Reviewer, EPSRC *Advanced Fellowship* program, UK (2005)
- Reviewer, BSF (United States – Israel Bi-national Science Foundation) program (2005)
- Reviewer, ICREA (Institució Catalana de Recerca I Estudis Avançats) *Senior Researcher* program, Spain/Catalunya (2005–2007)
- Reviewer, IWT-Vlaanderen, Belgium (2007)
- Reviewer, NRF South Africa (2007)
- Evaluator, Nederlandse Taalunie (Dutch Language Union) (2007)
- Member of Programme Committee, NWO Programme *IMIX* (Interactive Multimodal Information Extraction) (2002–2008)
- Member of Programme Committee, NWO Programme *CATCH* (Continuous Access to the Cultural Heritage) (2004–present)

## **Memberships**

Senior research fellow of SIKS, the Dutch Research School for Information and Knowledge Systems (from 2009).

Member of the European Association for Machine Translation (EAMT) (from 2009).

Member of SIGNLL, Special Interest Group on Natural Language Learning of the Association for Computational Linguistics (ACL) (from 1995).

Member of BNVKI-AIABN, Benelux Association for Artificial Intelligence (from 1994).

Member of the Association for Computational Linguistics (ACL) (from 1993).