

## CURRICULUM VITAE

September 2016

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:a.vandenbosch@let.ru.nl">a.vandenbosch@let.ru.nl</a>
URL	<a href="http://antalvandenbosch.ruhosting.nl">http://antalvandenbosch.ruhosting.nl</a>
<i>h</i> -index (Publish or Perish)	35

## 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

**Research director** , Centre for Language Studies, Faculty of Arts, Radboud University, Nijmegen, The Netherlands. From September 2015.

**Full professor** Language and Speech Technology – Faculty of Arts, Department of Communication and Information Sciences, Radboud University, Nijmegen, The Netherlands. From September 2011.

**Full professor** Memory, Language, and Meaning – School of Humanities, Tilburg centre for Cognition and Communication, and Department of Communication and Information Sciences, Tilburg University, the Netherlands. January 2008–August 2011.

**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.

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

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

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

## Other affiliations and fellowships

**Member** – Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands. Elected 2012.

**Integrator** – Netherlands eScience Center, Amsterdam, The Netherlands. January 2012–present.

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

**Fellow** of ECCAI, European Coordinating Committee for Artificial Intelligence. Elected 2015.

**Research fellow** of Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands. 2011–present.

**Senior research fellow** of SIKS, Dutch Research School for Information and Knowledge Systems. 2009–present.

## Publications

Publications are listed in the categories **thesis**, **books**, **journal articles**, **edited volumes**, **work to appear**, **book contributions and papers in proceedings**, **reviews**, **encyclopaedic articles**, **reports and long abstracts**, **non-reviewed and popularizing publications**, and **media productions**.

## 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 9780521808903
2. Razenberg, K., and Van den Bosch, A. (1997). *Het principe van de quaterniteit als algemeen denkraam*. Best, The Netherlands: Damon. ISBN 9789055730230

## Journal articles

1. Karsdorp, F., and Van den Bosch, A. (2016). The structure and evolution of story networks. *Royal Society Open Science*, **3:6**. doi: 10.1098/rsos.160071
2. Hallmann, K., Kunneman, F., Liebrecht, C., Van den Bosch, A., and Van Mulken, M. (2016). Sarcastic soulmates: Intimacy and irony markers in social media messaging. *Linguistic Issues in Language Technology*, **14:7**.
3. Greefhorst, A., and Van den Bosch, A. (2016). Predicting liaison: An example-based approach. *Traitement Automatique des Langues*, **57:1**, pp 13–32.
4. Van Gompel, M., and Van den Bosch, A. (2016) Efficient n-gram, skipgram and flexgram modelling with Colibri Core. *Journal of Open Research Software*, **4:1**, p.e30. doi: 10.5334/jors.105
5. Van den Bosch, A., Bogers, T, and De Kunder, M. (2016). Estimating search engine index size variability: A nine-year longitudinal study. *Scientometrics*, **107:2**, pp. 839–856. doi: 10.1007/s11192-016-1863-z
6. Ittoo, A., Nguyen, L. M., and Van den Bosch, A. (2016). Text analytics in industry: Challenges, desiderata and trends. *Computers in Industry*, **78**, pp. 96–107. doi: 10.1016/j.compind.2015.12.001
7. Kunneman, F., and Van den Bosch, A. (2016). Open-domain extraction of future events from Twitter. *Natural Language Engineering*. Published online. doi: 10.1017/S1351324916000036
8. Willems, R., Frank, S., Nijhof, A., Hagoort, P., and Van den Bosch, A. (2015). Prediction during natural language comprehension. *Cerebral Cortex*, **26:6**, 2506–2516. doi:10.1093/cercor/bhv075
9. Karsdorp, F., Van der Meulen, M., Meder, T., and Van den Bosch, A. (2015). MOMFER: A search engine of Thompson’s Motif-Index of folk literature. *Folklore*, **126:1**, pp. 37–52. doi: 10.1080/0015587X.2015.1006954
10. Kunneman, F., Liebrecht, C., Van den Bosch, A., and Van Mulken, M. (2015). Signaling sarcasm: From hyperbole to hashtag. *Information Processing & Management*, **51:4**, pp. 500–509. doi:10.1016/j.ipm.2014.07.006

2016

2015

- 2014
11. Stoop, W., and Van den Bosch, A. (2014). Improving word prediction for augmentative communication by using idiolects and sociolects. *Dutch Journal of Applied Linguistics*, **3:2**, pp. 136–153.  
doi: 10.1075/dujal.3.2.03sto
  12. Verberne, S., D'hondt, E., Van den Bosch, A., and Marx, M. (2014). Automatic thematic classification of election manifestos. *Information Processing & Management*, **50:4**, pp. 554–567.  
doi:10.1016/j.ipm.2014.02.006
  13. Pander Maat, H., Kraf, R., Van den Bosch, A., Dekker, N., Van Gompel, M., Kleijn, S., Sanders, T., and Van der Sloot, K. (2014). T-Scan: a new tool for analyzing Dutch text. *Computational Linguistics in the Netherlands Journal*, **4**, pp. 53–74.
  14. Kunneman, F., Hürriyetoğlu, A., Oostdijk, N., and Van den Bosch, A. (2014). Timely identification of event start dates from Twitter. *Computational Linguistics in the Netherlands Journal*, **4**, pp. 39–52.
  15. Tellings, A., Vermeer, A., Hulsbosch, M., and Van den Bosch, A. (2014). BasiLex: an 11.5 million words corpus of Dutch texts written for children. *Computational Linguistics in the Netherlands Journal*, **4**, pp. 191–208.

2013

  16. Tjong Kim Sang, E., and Van den Bosch, A. (2013). Dealing with big data: The case of Twitter. *Computational Linguistics in the Netherlands Journal*, **3**, pp. 121–134.
  17. Berendsen, R., De Rijke, M., Balog, K., Bogers, T., and Van den Bosch, A. (2013). On the assessment of expertise profiles. *Journal of the American Society for Information Science and Technology*, **64:10**, pp. 2024–2044.  
doi:10.1002/asi.22908
  18. Van den Bosch, A. and Daelemans, W. (2013). Implicit schemata and categories in memory-based language processing. *Language and Speech*, **56:3**, pp. 308–326.  
doi:10.1177/0023830913484902

2012

  19. Van den Bosch, A., Morante, R., and Canisius, S. (2012). Joint learning of dependency parsing and semantic role labeling. *Computational Linguistics in the Netherlands Journal*, **2**, pp. 97–117.
  20. Van de Camp, M., and Van den Bosch, A. (2012). The socialist network. *Decision Support Systems*, **53:4**, pp. 761–769.  
doi:10.1016/j.dss.2012.05.031

2011

  21. Van den Bosch, A. (2011). Effects of context and recency in scaled word completion. *Computational Linguistics in the Netherlands Journal*, **1**, pp. 79–94.
  22. Haque, R., Naskar, S. K., Van den Bosch, A., and Way, A. (2011). Integrating source-language context into phrase-based statistical machine translation. *Machine Translation*, **25:3**, pp. 239–285.  
doi:10.1007/s10590-011-9100-2
  23. Bogers, T., and Van den Bosch, A. (2011). Fusing recommendations for social bookmarking web sites. *International Journal of Electronic Commerce*, **15:3**, pp. 33–75.

2009

  24. Van den Bosch, A., Lendvai, P., Van Erp, M., Hunt, S., Van der Meij, M., and Dekker, R. (2009). Weaving a new fabric of natural history. *Interdisciplinary Science Reviews*, **34:2–3**, pp. 206–223.

25. Van den Bosch, A., Van den Herik, H.J., and Doorenbosch, P. (2009). Digital discoveries in museums, libraries, and archives: Computer science meets cultural heritage. *Interdisciplinary Science Reviews*, **34:2–3**, pp. 129–138.
26. Van den Bosch, A., Van Erp, M., and Sporleder, C. (2009). Making a clean sweep of cultural heritage. *IEEE Intelligent Systems*, **24:2**, pp. 54–63.
27. 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. 2006
28. 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.
29. 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. 2005
30. Van den Bosch, A. (2005). Scalable classification-based word prediction and confusable correction. *Traitement Automatique des Langues*, **46:2**, pp. 39–63. 2002
31. 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.
32. 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. 2000
33. 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**. 1990s
34. 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.
35. Daelemans, W., Van den Bosch, A., and Zavrel, J. (1999). Forgetting exceptions is harmful in language learning. *Machine Learning*, **34**, pp. 11–43.
36. Weijters, A., and Van den Bosch, A. (1999). Interpreting knowledge representations in BP-SOM. *Behaviormetrika*, **26:1**, pp. 107–128.
37. 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.
38. Weijters, A., Van den Bosch, A., and Van den Herik, H.J. (1997). Behavioural aspects of combining back-propagation and self-organizing maps. *Connection Science*, **9:3**, pp. 253–252.
39. 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.
40. 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.

## Edited volumes

1. Zervanou, K., and Van den Bosch, A., editors (2012). *Proceedings of the EACL 2012 workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, LaTeCH-2012*. Avignon, France.
2. Van den Bosch, A., and Shatkay, H., editors (2012) *Proceedings of the ACL 2012 workshop on Detecting Structure in Scholarly Discourse, DSSD-2012*. Jeju, Korea.
3. Sporleder, C., Van den Bosch, A., and Zervanou, K., editors (2011). *Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series*. Berlin: Springer. ISBN 9783642202261
4. Van den Bosch, A., and Bouma, G., editors (2011). *Interactive multi-modal question answering*. Berlin: Springer. ISBN 9783642175244
5. Soudi, A., Van den Bosch, A., and Neumann, G., editors (2007). *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Berlin: Springer. ISBN 9781402060458
6. 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. ISBN 9781932432886
7. 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. ISBN 9781932432886
8. Roth, D., and Van den Bosch, A., editors (2002). *Proceedings of the Sixth Conference on Language Learning, CoNLL-2002*. Stroudsburg, PA: ACL.
9. Van den Bosch, A., and Weigand, H., editors (2000). *Proceedings of the Twelfth Belgium-Netherlands Artificial Intelligence Conference, BNAIC'00*. Tilburg: ILK/Infolab.
10. 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.
11. 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*.

## Work to appear

1. Hürriyetoğlu, A., Gudehus, C., Oostdijk, N., and Van den Bosch, A. (to appear). Relevancer: Finding and labeling relevant information in tweet collections. To appear in *Proceedings of the 8th International Conference on Social Informatics (SocInfo 2016)*, Bellevue, WA, USA.
2. Sanders, E., De Gier, M., and Van den Bosch, A. (to appear). Using demographics in predicting election results with Twitter. To appear in *Proceedings of the 8th International Conference on Social Informatics (SocInfo 2016)*, Bellevue, WA, USA.
3. Wubben, S., Krahmer, E., Van den Bosch, A., and Verberne, S. (to appear). Abstractive summarization of captions with Attentive Recurrent Neural Networks. To appear in *Proceedings of the 9th International Natural Language Generation conference*.

## Book contributions and papers in peer-reviewed proceedings

2016

1. Hendrickx, I., Lefever, E., Croijmans, I., Majid, A., and Van den Bosch, A. (2016). Very quaffable and great fun: Applying NLP to wine reviews. In *Proceedings of the 54th Conference of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 306–312. Berlin: Association for Computational Linguistics.
2. Onrust, L., Van den Bosch, A., and Van hamme, H. (2016). Improving cross-domain n-gram language modelling with skipgrams. In *Proceedings of the 54th Conference of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 137–142. Berlin: Association for Computational Linguistics.
3. Oostdijk, N., Hürriyetoğlu, A., Puts, M., Daas, P., and Van den Bosch, A. (2016). Information extraction from social media: A linguistically motivated approach. In *Proceedings of the Workshop on Risk and NLP* (pp. 1-11). Paris, France.
4. Sommerdijk, B., Sanders, E., and Van den Bosch, A. (2016). Can tweets predict tv ratings? In Calzolari, N., Choukri, K., Declerck, T., Grobelnik, M., Maegaard, B, Mariani, J., Moreno, A., Odijk, J., and Piperidis, S. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC-2016)*. Portoro, Slovenia: ELRA.
5. Kordoni, V., Van den Bosch, A., Kermanidis, K., Sosoni, V., Cholakov, K., Hendrickx, I., Huck, M., and Way, A. (2016). Enhancing access to online education: Quality machine translation of MOOC content. In Calzolari, N., Choukri, K., Declerck, T., Grobelnik, M., Maegaard, B, Mariani, J., Moreno, A., Odijk, J., and Piperidis, S. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC-2016)*. Portoro, Slovenia: ELRA.
6. Brugman, H., Reynaert, M., Van der Sijs, N., Van Stipriaan, R., Tjong Kim Sang, E., Van den Bosch, A. (2016). Nederlab: Towards a single portal and research environment for diachronic Dutch text corpora. In Calzolari, N., Choukri, K., Declerck, T., Grobelnik, M., Maegaard, B, Mariani, J., Moreno, A., Odijk, J., and Piperidis, S. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC-2016)*. Portoro, Slovenia: ELRA.
7. Koolen, M., Bogers, T., Van den Bosch, A., and Kamps, J. (2015). Looking for books in social media: An analysis of complex search requests. In Hanbury, A., Kazai, G., Rauber, A., Fuhr, N. (Eds.), *Advances in Information Retrieval: 37th European Conference on IR Research*. Lecture Notes in Computer Science, Vol. 9022. Berlin: Springer, pp. 184–196.
8. Van den Bosch, A. and Bresnan, J. (2015). Modeling dative alternations of individual children. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning (CogACLL-2015)*, Lisbon, Portugal, pp. 103–112.
9. Kunneman, F., and Van den Bosch, A. (2015). Automatically identifying periodic social events from Twitter. In *Proceedings of Recent Advances in Natural Language Processing Conference 2015 (RANLP-2015)*, pp. 320–328.
10. Karsdorp, F., Van der Meulen, M., Meder, T., and Van den Bosch, A. (2015). Animacy detection in stories. In M. A. Finlayson, B. Miller, A. Lieto, and R. Ronfard (Eds.), *Proceedings of the 6th Workshop on Computational Models of Narrative (CMN-2015)*, pp. 82–97.
11. Karsdorp, F., Kestemont, M., Schöch, C., and Van den Bosch, A. (2015). The love equation: Computational modeling of romantic relationships in French classical drama. In M. A. Finlayson, B. Miller, A. Lieto, and R. Ronfard (Eds.), *Proceedings of the 6th Workshop on Computational Models of Narrative (CMN-2015)*, pp. 89–107.

2015

12. Van den Bosch, A., Bogers, T., and De Kunder, M. (2015). A longitudinal analysis of estimating search engine index size. In A. A. Salah, Y. Tonta, A. A. A. Salah, C. Sugimoto, and U. Al (Eds.), *Proceedings of the 15th International Society of Scientometrics and Informetrics Conference (ISSI-2015)*, pp. 71–82. Istanbul, Turkey: Boazii University.
13. Wubben, S., Verberne, S., Krahmer, E., and Van den Bosch, A. (2015). Facilitating online discussions by automatic summarization. In *Proceedings of the 27th Benelux Conference on Artificial Intelligence (BNAIC 2015)*, Hasselt, Belgium.
14. Van Gompel, M., and Van den Bosch, A. (2014). Translation assistance by translating L1 fragments in an L2 context. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, USA, pp. 871–880. ACL.
15. Stoop, W., and Van den Bosch, A. (2014). Using idiolects and sociolects to improve word prediction. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg, Sweden, pp. 318–327. ACL.
16. Hürriyetoglu, A., Oostdijk, N., and Van den Bosch, A. (2014). Estimating time to event from tweets using temporal expressions. In *Proceedings of the Workshop on Language Analysis in Social Media, LASM-2014*, Gothenburg, Sweden, pp. 8–16. ACL.
17. Kunneman, F., Liebrecht, C., and Van den Bosch, A. (2014). The (un)predictability of emotional hashtags in Twitter. In *Proceedings of the Workshop on Language Analysis in Social Media, LASM-2014*, Gothenburg, Sweden, pp. 26–34. ACL.
18. Zervanou, K., Hendrickx, I, Düring, M., and Van den Bosch, A. (2014). Documenting social unrest: Detecting strikes in historical daily newspapers. In A. Nadamoto, A. Jatowt, A. Wierzbicki, and J. L. Leidner (Eds.), *Social Informatics. Lecture Notes in Computer Science Vol. 8359*. Berlin: Springer, pp. 120–133.
19. Wubben, S., Krahmer, E., and Van den Bosch, A. (2014). Creating and using large monolingual parallel corpora for sentential paraphrase generation. In *Proceedings of the 9th Language Resources and Evaluation Conference*, Reykjavik, Iceland. ELRA.
20. Van Gompel, M., Van den Bosch, A., and Dijkstra, A. (2014). Oersetter: Frisian-Dutch statistical machine translation. In Boersma, P., Brand, H., and Spoelstra, J. (Eds.), *Philologia Frisica anno 2012*. Ljouwert: Fryske Akademy, pp. 287–296.
21. Van Gompel, M., Hendrickx, I, Van den Bosch, A., Lefever, E., and Hoste, V. (2014). SemEval-2014 Task 5: L2 writing assistant. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval-2014*, Dublin, Ireland, pp. 36–44.
22. Kunneman, F., and Van den Bosch, A. (2014). Event detection in Twitter: A machine-learning approach based on term pivoting. In F. Grootjen, M. Otworowska, and J. Kwisthout (Eds.), *Proceedings of the 26th Benelux Conference on Artificial Intelligence*, pp. 65–72.
23. Düring, M., and Van den Bosch, A. (2014). Multi-perspective event detection in texts documenting the 1944 Battle of Arnhem. In Biemann, C., and Mehler, A. (Eds.), *Text Mining: From Ontology Learning to Automated Text Processing Applications*, Chapter 10. Heidelberg: Springer, pp. 201–219.
24. Hürriyetoglu, A., Kunneman, F, and Van den Bosch, A. (2013). Estimating the time between Twitter messages and future events. In *Proceedings of the 13th Dutch-Belgian Information Retrieval Workshop*, pp. 20–23.
25. Hendrickx, I., Zervanou, K., Düring, M., and Van den Bosch, A. (2013). Searching and finding strikes in the New York Times. In F. Mambrini, M. Passarotti, and C. Sporleder (Eds.), *Proceedings of the Third Workshop on Annotation of Corpora for Research in the Humanities*, Sofia, Bulgaria, pp. 25–36.

2014

2013



26. Sanders, E. and Van den Bosch, A. (2013). Relating political party mentions on Twitter with polls and election results. In *Proceedings of the 13th Dutch-Belgian Information Retrieval Workshop*, pp. 68–71.
27. Karsdorp, F. and Van den Bosch, A. (2013). Identifying motifs in folktales using topic models. In *Proceedings of the 22 Annual Belgian-Dutch Conference on Machine Learning*, pp. 41–49.
28. Liebrecht, C., Kunneman, F., and Van den Bosch, A. (2013). The perfect solution for detecting sarcasm in tweets #not. In A. Balahur, E. van der Goot, and A. Montoyo (Eds.), *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA-2013*, pp. 29–37. New Brunswick, NJ: ACL.
29. Van Gompel, M. and Van den Bosch, A. (2013). WSD2: Parameter optimisation for memory-based cross-lingual word-sense disambiguation. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation, SemEval-2013*. Atlanta, GA: ACL. pp. 183–187.
30. Van den Bosch, A., and Berck, P. (2013). Memory-based grammatical error correction. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, pp. 102–108. ACL.
31. Wubben, S., Krahmer, E., and Van den Bosch, A. (2013). Using character overlap to improve language transformation. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, LaTeCH-2013*, pp. 11–19. ACL.
32. Van den Bosch, A., and Bogers, T. (2013). Memory-based named entity recognition in tweets. In A. Cano, M. Rowe, M. Stankovic, and A.-S. Dadzie (Eds.), *#MSM2013 Workshop Concept Extraction Challenge Proceedings*, pp. 40–43.
33. Tops, H., Kunneman, F., and Van den Bosch, A. (2013). Predicting time-to-event from Twitter messages. In K. Hindriks, M. de Weerd, B. van Riemsdijk, and M. Warnier (Eds.), *Proceedings of the 25th Benelux Artificial Intelligence Conference, BNAIC-2013*. Delft, The Netherlands.
34. Verberne, S., Van den Bosch, A., Strik, H., and Boves, L. (2012). The effect of domain and text type on text prediction quality. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France*, pp. 561–569. ACL.
35. Mos, M., Van den Bosch, A., and Berck, P. (2012). The predictive value of word-level perplexity in human sentence processing: A case study on fixed adjective–preposition constructions in Dutch. In S. Gries and D. Divjak (Eds.), *Frequency Effects in Language Learning and Processing, Vol 1*. Berlin: Mouton De Gruyter, pp. 207–240.
36. Karsdorp, F., Van Kranenburg, P., Meder, T., Trieschnigg, D., and Van den Bosch, A. (2012). In search of an appropriate abstraction level for motif annotations. In M. Finlayson (Ed.), *Proceedings of the 2012 Computational Models of Narrative Workshop, Istanbul, Turkey*, pp. 22–26.
37. Vossen, P., Görög, A., Izquierdo, R., and Van den Bosch, A. (2012). DutchSemCor: Targeting the ideal sense-tagged corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC-2012, Istanbul, Turkey*, pp. 584–589.
38. Van den Bosch, A., and Berck, P. (2012). Memory-based text correction for preposition and determiner errors. In *Proceedings of the 7th Workshop on the Innovative Use of NLP for Building Educational Applications, Montreal, Canada*, pp. 289–294.
39. Wubben, S., Krahmer, E., and Van den Bosch, A. (2012). Sentence simplification by monolingual machine translation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, ACL-2012, Jeju, Korea*, pp. 1015–1024. ACL.

40. Kunneman, F., and Van den Bosch, A. (2012). Leveraging unscheduled event prediction through mining scheduled event tweets. In N. Roos, M. Winands, and J. Uiterwijk (Eds.), *Proceedings of the 24th Benelux Conference on Artificial Intelligence*, Maastricht, The Netherlands, pp. 147–154.
41. Karsdorp, F., Van Kranenburg, P., Meder, T., and Van den Bosch, A. (2012). Casting a spell: Identification and ranking of actors in folktales. In F. Mambrini, M. Passarotti, and C. Sporleder (Eds.), *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities*, pp. 39–50. Lisbon, Portugal: Edi ções Colibri.
42. Van Gompel, M., Van den Bosch, A., and Berck, P. (2011). Extending memory-based machine translation to phrases. In T. Markus, P. Monachesi, and E. Westerhout (Eds.), *Computational Linguistics in the Netherlands 2010: Selected Papers from the Twentieth CLIN Meeting*. Utrecht: LOT, pp. 45–58.
43. Wubben, S., Van den Bosch, A., and Krahrmer, E. (2011). Paraphrasing headlines by machine translation. In T. Markus, P. Monachesi, and E. Westerhout (Eds.), *Computational Linguistics in the Netherlands 2010: Selected Papers from the Twentieth CLIN Meeting*. Utrecht: LOT, pp. 169–183.
44. Canisius, S., Van den Bosch, A., and Daelemans, W. (2011). Constraint-satisfaction inference for entity recognition. In Van den Bosch, A., and Bouma, G. (Eds.), *Interactive multi-modal question answering*. Berlin: Springer, pp. 199–221.
45. Wubben, S., Marsi, E., Van den Bosch, A., and Krahrmer, E. (2011). Comparing phrase-based and syntax-based paraphrase generation. In K. Filippova and S. Wan (Eds.), *Proceedings of the ACL-HLT Workshop on Text-to-Text Generation*, Portland, OR, pp. 27–33.
46. Van de Camp, M., and Van den Bosch, A. (2011). A link to the past: Constructing historical social networks. In A. Balahur, E. Boldrini, A. Montoyo, and P. Martinez-Barco (Eds.), *Proceedings of the ACL-HLT Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA-2011)*, Portland, OR, pp. 61–69.
47. Zervanou, K., Korkontzelos, I., Van den Bosch, A., and Ananiadou, S. (2011). Enrichment and structuring of archival description metadata. In K. Zervanou and P. Lendvai (Eds.), *Proceedings of the Fifth ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH-2011)*, Portland, OR, pp. 44–53.
48. Van Erp, M., Van den Bosch, A., Hunt, S., Van der Meij, M., Dekker, R., and Lendvai, P. (2011). Natural selection: Finding specimens in a natural history collection. In O. Grillo and G. Venora (Eds.), *Changing Diversity in Changing Environment*. Rijeka, Croatia: InTech. ISBN 979-953-307-255-4
49. Vossen, P., Görög, A., Laan, F., Van Gompel, M., Izquierdo, R., and Van den Bosch, A. (2011). DutchSemCor: Building a semantically annotated corpus for Dutch. In I. Kosem and K. Kosem (Eds.), *Electronic lexicography in the 21st century: New applications for new users; Proceedings of eLex 2011*, pp. 286–296. ISBN 978D961D92983D3D6.
50. Daelemans, W., and Van den Bosch, A. (2010). *Memory-based learning*. In A. Clark, C. Fox, and S. Lappin (Eds.), *Handbook of Computational Linguistics and Natural Language Processing*. Oxford, UK: Wiley-Blackwell Publishers, pp. 154–179.
51. Wubben, S., Van den Bosch, A., and Krahrmer, E. (2010). Paraphrase generation as monolingual translation: Data and evaluation. In J. Kelleher, B. Mac Namee, and I. van der Sluis (Eds.), *Proceedings of the 10th International Workshop on Natural Language Generation (INLG 2010)*, pp. 203–207, Dublin, July 2010.

2011

2010

52. De Rijke, M., Balog, K., Bogers, T., and Van den Bosch, A. (2010). On the evaluation of entity profiles. In M. Agosti, N. Ferro, C. Peters, M. de Rijke, and A. Smeaton (Eds.), *CLEF2010: Conference on Multilingual and Multimodal Information Access Evaluation*, pp. 94–99.
53. Van den Hoven, M., Van den Bosch, A., and Zervanou, K. (2010). Beyond reported history: Strikes that never happened. In S. Darányi and P. Lendvai (Eds.), *Proceedings of the First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, Vienna, Austria, pp. 20–28.
54. Haque, R., Kumar Naskar, S., Van den Bosch, A., and Way, A. (2010). Supertags as source language context in hierarchical phrase-based SMT. In *Proceedings of the Ninth Conference of the Association for Machine Translation in the Americas*, Denver, CO, USA, pp. 210–219.
55. Van den Bosch, A., Nauts, P., and Eckhardt, N. (2010). A kid’s Open Mind Common Sense. In C. Havasi, D. Lenat, and B. Van Durme (Eds.), *Commonsense Knowledge: Papers from the AAAI Fall Symposium*, Arlington, VA. Menlo Park, CA: AAAI Press, pp. 114–119.
56. 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.
57. 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)*, Workshop Proceedings Series 09-01, Centre for Telematics and Information Technology, Enschede. ISSN 0929-0672, pp 87–94.
58. 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.
59. 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.
60. 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.
61. 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.
62. 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.
63. Canisius, S., and Van den Bosch, A. (2009). A constraint satisfaction approach to machine translation. In H. Somers and L. Màrquez (Eds.), *Proceedings of the 13th Annual Meeting of the European Association for Machine Translation (EAMT-2009)*, pp. 182–189. ISBN 9788469239438
64. Morante, R., Van Asch, V., and Van den Bosch, A. (2009). Joint memory-based learning of syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL): Shared Task*, pp. 25–30. Stroudsburg, PA: Association for Computational Linguistics.

65. Morante, R., Van Asch, V., and Van den Bosch, A. (2009). Dependency parsing and semantic role labeling as a single task. In *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing, RANLP-2009*, pp. 275–280.
66. Morante, R., and Van den Bosch, A. (2009). Feature construction for memory-based semantic role labeling of Catalan and Spanish. In N. Nicolov, G. Angelova and R. Mitkov (Eds.), *Recent Advances in Natural Language Processing V: Selected papers from RANLP 2007. Current Issues in Linguistic Theory 309*. Amsterdam: John Benjamins, pp. 131–142. ISBN 9789027248251
67. Van Gompel, M., Van den Bosch, A., and Berck, P. (2009). Extending memory-based machine translation to phrases. In M. Forcada and A. Way (Ed.), *Proceedings of the Third Workshop on Example-Based Machine Translation*, Dublin, Ireland, pp. 79–86.
68. Haque, R., Naskar, S., Van den Bosch, A., and Way, A. (2009). *Dependency relations as source context in phrase-based SMT*. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation (PACLIC 2009)*, pp. 170–179.
69. Bogers, T., and Van den Bosch, A. (2009). Collaborative and content-based filtering for item recommendation on social bookmarking websites. In D. Jannach, W. Geyer, J. Freyne, S. S. Anand, C. Dugan, B. Mobasher, and A. Kobsa (Eds.), *Proceedings of the ACM RecSys '09 workshop on Recommender Systems and the Social Web*, pp. 9–16, October 2009.
70. 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.
71. 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.
72. 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.
73. Puerta Melguizo, M.C., Muñoz 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
74. 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.
75. 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.
76. 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. ISBN 9789081486101
77. 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.

2008

2007

78. 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.
79. 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.
80. 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.
81. 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*.
82. 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.
83. 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.
84. 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.
85. 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.
86. 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.
87. 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. ISBN 9781932432886
88. 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.
89. Bogers, T, Thoonen, W., and Van den Bosch, A. (2006). Expertise classification: Collaborative classification vs. automatic extraction. In J. Tennis and J. Furner (Eds.), *Proceedings of the 17th Annual ASIS&T SIG/CR Classification Research Workshop*, Austin, TX, pp. 1–20.
90. 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

91. 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.
92. 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.
93. 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.
94. 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.
95. 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.
96. 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.
97. 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.
98. 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)*, Lecture Notes on Computer Science, Vol. 3936. Berlin: Springer, pp. 519–522.
99. 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.
100. 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.
101. 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.
102. 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.
103. 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

104. Marsi, E., Van den Bosch, A., and Soudi, 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.
105. 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.
106. 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.
107. 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
108. 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.
109. Lendvai, P., Van den Bosch, A., Kraemer, 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.
110. 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.
111. 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.
112. 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.
113. 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.
114. 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
115. 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*. Stroudsburg, PA: ACL, pp. 139–146.
116. Lendvai, P., Van den Bosch, A., and Kraemer, 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.

117. 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.
118. 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.
119. 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
120. 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*. Stroudsburg, PA: ACL, pp. 433–440.
121. 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. Stroudsburg, PA: ACL, pp. 61–65.
122. 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, Vol. 2534. Berlin: Springer, pp. 364–373.
123. 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.
124. 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.
125. 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.
126. 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.
127. 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
128. 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*. Stroudsburg, PA: ACL, pp. 499–506.
129. 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.



130. 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*.
131. 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.
132. 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
133. 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.
134. 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.
135. 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)*. Stroudsburg, PA: ACL, pp.73–78.
136. 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.
137. 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.
138. 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.
139. 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
140. 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.
141. 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.
142. 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
143. Weijters, A., Van den Bosch, A., and Van den Herik, H.J. (1998). Interpretable neural networks with BP-SOM. In C. Nedellec and C. Rouveïrol (Eds.), *Machine Learning: ECML-98, Lecture Notes in Artificial Intelligence*, Vol. 1398. Berlin: Springer, pp. 406–411.

144. 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.
145. 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.
146. 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, Vol. 1416. Berlin: Springer, pp 564–573.
147. 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
148. 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.
149. 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, Vol. 1224. Berlin: Springer Verlag, pp. 337–344.
150. 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.
151. 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.
152. 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.
153. 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.
154. 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.
155. 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.
156. 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.

157. 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.
158. 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
159. 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.
160. 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.
161. 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
162. 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.
163. 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
164. 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.
165. 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.
166. 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.
167. 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.
168. 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
169. 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.

170. 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.)
171. 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.
172. 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.

## Reviews

1. Van den Bosch, A. (2014). Peter Spyns and Jan Odijk (Editors), *Essential Speech and Language Technology for Dutch: Results by the STEVIN programme* Springer, 2013, ISBN: 978-3-642-30909-0, xvii + 413 pp. *Machine Translation*, **28:1**, 57–60.  
doi: 10.1007/s10590-014-9147-y.

## Encyclopaedic articles

1. Van der Beek, L., and Van den Bosch, A. (2015). Translation technology in the Netherlands and Belgium. In S.-W. Chan (Ed.), *The Routledge Encyclopedia of Translation Technology*, Chapter 21, pp. 352–363. New York, NY: Routledge. ISBN 978-0-415-52484-1
2. Van den Bosch, A. (2010). *Hidden markov models*. In C. Sammut and G. I. Webb (Eds.), *Encyclopedia of Machine Learning*. Berlin: Springer Verlag, pp. 493–494. Also contributed cross-referencing entries on *Markov process*, *Viterbi algorithm*, *Baum-Welch algorithm*, and *Expectation-Maximization algorithm*.

## Reports, extended abstracts

1. Koolen, M., Bogers, T., Van den Bosch, A., and Kamps, J. (2015). What makes book search in social media complex? In *Proceedings of the 14th Dutch-Belgian Information Retrieval Workshop, DIR-2015*, Amsterdam, The Netherlands, pp. 21.
2. Reynaert, M., Van Gompel, M., Van der Sloot, K., and Van den Bosch, A. (2015). PICCL: Philosophical Integrator of Computational and Corpus Libraries. In *Proceedings of CLARIN Annual Conference 2015 – Book of Abstracts*, pp. 75–79.
3. Kordoni, V., Cholakov, K., Egg, M., Way, A., Birch, L., Kermanidis, K., Sosoni, V., Tsoumakos, D., Van den Bosch, A., Hendrickx, I., Papadopoulos, M., Georgakopoulou, P., Gialama, M., Van Zaanen, M., Buliga, I., Jermol, M., and Orlic, D. (2015). TraMOOC: Translation for Massive Open Online Courses. In *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*.
4. Hürriyetoglu, A., Tsiagkas, M., and Van den Bosch, A. (2013). Real-time storage and processing of all geo-located tweets in MongoDB. In *Proceedings of the Dutch-Belgian Database Day*, Rotterdam, November 13, 2013.

5. Van Gompel, M., Van der Sloot, K., and Van den Bosch, A. (2012). *Ucto: Unicode Tokenizer, Version 0.5.3, Reference Guide*. ILK Technical Report no. ILK-1205.
6. Van den Bosch, A., Sporleder, C., Lendvai, P., Van Erp, M., and Hunt, S. (2009). New perspectives on natural history collections through automated discovery. In *Abstracts of the 24th Annual Meeting of the Society for the preservation of natural history collections (SPNHC-2009)*, Leiden, The Netherlands, p. 41.
7. 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.
8. 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.
9. 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.
10. Van den Bosch, A. (2004). *Fambl 2.3 Reference Guide*. ILK Technical Report ILK-0403.
11. 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.
12. 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.
13. 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.
14. 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. (2015). What's hot in . . . Computational Linguistics? *Benelux A.I. Newsletter*, **1:2**, pp. 6–8.
2. Van den Bosch, A. (2012). *Taal in uitvoering*. Inaugural address (In Dutch). Radboud University, November 9, 2012. ISBN 978-90-9027-167-5. Published online at <http://antalvandenbosch.ruhosting.nl/oratie.pdf>
3. Van den Bosch, A. (2012). Valkuil.net: Spellingcorrectie met gevoel voor context. *Tekstblad*, **18:2**, pp. 16–18.
4. Bogers, T., and Van den Bosch, A. (2012). Onderzoekers die dit artikel lezen, lezen ook... *Dixit*, **8**, p. 42.
5. Van den Bosch, A. (2012). Contextgebaseerde spellingcorrectie met Valkuil.net. *Dixit*, **8**, p. 17.

6. Sporleder, C., Van den Bosch, A., and Zervanou, K. (2011). Language technology for cultural heritage, social sciences and humanities: Chances and challenges. In C. Sporleder, A. van den Bosch, and K. Zervanou (Eds.), *Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series*. Berlin: Springer, pp. xxi–xxxi.
7. Meijers, J., and Van den Bosch, A. (2011). Voor-onze-neus-weggekaapt.nl: Nederlandse woorden gegijzeld als domeinnaam. *Onze Taal*, **80:5**, pp. 126–128.
8. Van den Bosch, A. (2011). Ook de burger kan snel schatgraven in WikiLeaks. *Brabants Dagblad*, January 26, 2011.
9. Van den Bosch, A. (2010). Zo had het ook kunnen gaan: Graven in de geschiedenis van de sociale beweging met het HiTiME project. *Dixit*, **7:1**, pp 14–15.
10. Van den Bosch, A. (2010). Van tekst naar tekst. *De Connectie*, **4:4**, pp. 6–9.
11. Van den Bosch, A. (2009). Informatie achter slot en grendel. *.ego Magazine*, **8:1**, pp. 16–18.
12. 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>
13. Van den Bosch, A. (2008). Digitale verhalen. In H. van Lierop (Ed.), *Jeugdletteratuur en andere media: Cultuureducatie in woord en beeld*. Leidschendam: Biblion Uitgeverij, pp. 16–26.
14. 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.
15. 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.
16. Van den Bosch, A. (2006). Rolaquad. *De Connectie*, **3:3**, pp. 20–23.
17. Van den Bosch, A. (2006). Machine learning and natural language processing in Tilburg. *BNVKI Newsletter*, **23:3**, pp. 52–57.
18. Van den Bosch, A. (2005). Informatie vinden. In H. van Driel (Ed.), *Digitaal Communiceren*. Meppel, the Netherlands: Boom. (in Dutch).
19. 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.
20. 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.
21. Van den Bosch, A. (2002). Implementations of memory and analogy for language processing. *NVTI Newsletter, Dutch Foundation for Theoretical Computer Science*, January 2002.
22. Daelemans, W., and Van den Bosch, A. (1992). De voorleesmachine. *Informatica Nieuwsbrief*, **2:9**, pp. 10–15. Alphen aan den Rijn: Samsom. (in Dutch)
23. Van den Bosch, A. (1992). *A hybrid model for text-to-speech conversion*. Unpublished MA thesis, Katholieke Universiteit Brabant.

## 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. 24m. First broadcast March 13, 2000, Teleac-NOT, Dutch public television.

## Other scientific activities

### Invited lectures, keynotes

1. *Listening to millions of voices: Text analytics of historical and new media*  
Hans-Kilian-Vorlesungen, Ruhr Univesität Bochum, German, December 14, 2015.
2. *Artificial intelligence and natural language processing*  
Studium Generale, Erasmus University, Rotterdam, the Netherlands, November 30, 2015.
3. *Prediction and surprise in the brain during narrative comprehension*  
Cognitive Aspects of Computational Language Learning Workshop, Lisbon, September 18, 2015.
4. *Implicit grammar in memory-based models of syntactic variation*  
Gradience in Grammar Workshop, Stanford University, CA, USA, January 18, 2014.
5. *Semantics or making sense: Meaning and entailment in text-to-text processing*  
Intelligent Systems Lab Amsterdam Colloquium, University of Amsterdam, December 17, 2013.
6. *Did it really happen? Discovering events and motifs in historical narratives*  
Digital Humanities Workshop in memory of Emanuele Pianta, Trento, Italy, December 10, 2013.
7. *Text analytics for detecting events and motifs in historical texts*  
The 1st International Workshop on Histoinformatics, Kyoto, Japan, November 25, 2013.
8. *Computational modeling: Dividing or equalizing?*  
Nijmegen Lectures 2013, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, January 30, 2013.
9. *Example-based modeling of syntactic alternations.*  
Symposium "New Ways of Analysing Syntactic Variation", Molenhoek, The Netherlands, November 16, 2012.
10. *Employing source-language context in statistical machine translation.*  
Symposium "Exploiting parallel corpora for lexical disambiguation", Ghent University, September 24, 2012.
11. *Big linguistic data.*  
TLA Nijmegen Lecture Series on E-Humanities, Max Planck Institute for Psycholinguistics, Nijmegen, June 6, 2012.
12. *Back to basics: Skousen's natural statistics for language modeling.*  
Symposium on occasion of Lou Boves' farewell lecture, Nijmegen, The Netherlands, April 7, 2011.
13. *Generating texts from memory.*  
Annual School for Information and Knowledge Systems Day, Veldhoven, The Netherlands, November 2, 2010.
14. *Constraint satisfaction inference for dependency parsing and MT.*  
Intelligent Systems Lab Amsterdam Colloquium, University of Amsterdam, Amsterdam, the Netherlands, May 26, 2009.
15. *Memory-based Language Models for Spelling Correction.*  
Twelfth Annual Research Colloquium of the Special Interest Group for Computational Linguistics in the UK and Ireland (CLUKI), Dublin City University, Dublin, Ireland, April 24, 2009.

16. *Memory-based machine translation.*  
CNTS Colloquium, Centre for Dutch Language and Speech, University of Antwerp, Belgium, April 9, 2009.
17. *Constraint satisfaction inference for dependency parsing and MT.*  
NLTC Seminar Series, Dublin City University, Dublin, Ireland, February 17, 2009.
18. *Implicit Linguistics.*  
CLCG Linguistics Colloquium, Groningen, The Netherlands, September 19, 2008.
19. *Mining natural history.*  
SIMIN Spring Symposium, Municipal Museum, The Hague, The Netherlands, June 6, 2008.
20. *Automatic Part-of-Speech tagging .*  
Diachronic Linguistics Meeting, Meertens Institute, Amsterdam, The Netherlands, May 30, 2008.
21. *Facts are little theories.*  
Semantic Web Seminar, University of Utrecht, Utrecht, The Netherlands, March 9, 2008.
22. *Explicit versus implicit computational linguistics.*  
XXIII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN-2007), Sevilla, Spain, September 10, 2007.
23. *Natural language processing and machine learning.*  
Free University of Amsterdam AI Group "Natural Language Processing Days", Leusden, The Netherlands, December 12, 2006.
24. *A memory-based learning-plus-inference approach to morphological analysis.*  
Workshop on Flexible Architectures for Large Vocabulary Recognition (FLaVoR), Leuven, Belgium, November 17, 2006.
25. *Constraint satisfaction inference for discrete sequence processing in NLP.*  
Dublin City University, National Centre for Language Technology Seminar Series, Dublin, Ireland, April 19, 2006.
26. *Implicit computational linguistics.*  
University of Amsterdam, Institute for Logic, Language, and Computation, Computational Linguistics Seminar, March 22, 2006.
27. *Automatic correction of coreference resolution and chaining.*  
Workshop on Anaphora Resolution, Mjøelvfjell, Norway, September 29, 2005.
28. *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.
29. *Understanding domain-specific questions and answers.*  
Language and Speech Colloquium Series, Nijmegen University, June 1, 2005.
30. *Learning linguistic sequences.*  
Computational Linguistics Seminar Series, University of Bergen, Bergen, Norway, May 2, 2005.
31. *Learning syntactically valid sequences.*  
Workshop on Machine Learning of Information Extraction, Antwerp, Belgium, March 10, 2005.
32. *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.
33. *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.*
34. *Text mining*  
Symposium *Dealing with the data flood*, Stichting Toekomstbeeld der Techniek, Rotterdam, April 23, 2002.
35. *Machine learning: Tools for natural language processing*  
*Maastricht Machine Learning Day, January 10, 2002.*
36. *Issues in memory-based natural language processing: Memory, representation, and data*  
University of Pittsburgh, Learning Research and Development Center (LRDC), November 30, 2001.



37. *Conjunctive and disjunctive feature construction*  
WhizBang! Labs–Research colloquium, September 27, 2001.
38. *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 Kraemer and Marc Swerts.
39. *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.
40. *Memory-based language processing.*  
IPO Center for User-System Interaction, Eindhoven Technical University, November 10, 2000.
41. *Memory-Based learning and TiMBL.*  
Analogical Modeling of Language (AML) Conference, Brigham Young University, Provo, Utah, March 22, 2000. With Walter Daelemans.
42. *Memory-based language processing: some findings and applications.*  
Text Learning group meeting, School of Computer Science, Carnegie Mellon University, Pittsburgh PA, March 20, 2000.
43. *Robust language technology for speech technology.*  
IPO Center for User-System Interaction Colloquium series, IPO, Eindhoven Technical University, the Netherlands, March 10, 2000.
44. *Instance-family abstraction in memory-based learning.*  
Ninth Dutch-Belgian Conference on Machine Learning (BENELEARN’99), Maastricht, the Netherlands, November 5, 1999.
45. *Learning word pronunciation.*  
Parlevink Colloquium Series, Department of Theoretical Informatics, University of Twente, The Netherlands. February 9, 1998.
46. *Experiments in machine learning of morpho-phonology.*  
ITK Colloquium Series, Institute for Language Technology and AI, Tilburg University, The Netherlands. December 7, 1995.
47. *A language-independent, data-oriented architecture for grapheme-phoneme conversion.*  
Department of Computer Science Colloquium, Universiteit Maastricht, The Netherlands. May 16, 1994.

## Teaching experience

### Undergraduate level

#### *Radboud University*

*Intelligent information tools*, block course, curriculum Master *Communicatie- en Informatiewetenschappen*, track *Communicatie en beïnvloeding*, Radboud University. Academic years 2011/2012 – 2014/2015.

*Nieuwe media, nieuwe genres* (New media, new genres), block course, curriculum Master *Communicatie- en Informatiewetenschappen*, track *Nieuwe media, taal en communicatie*, Radboud University. Academic years 2013/2014, 2014/2015. With Wilbert Spooren.

*Text Mining*, semester course, Research Master *Language and Communication*, Radboud University and Tilburg University. Academic years 2013/2014, 2014/2015.

*Example-based Language Modeling*, semester course, Research Master *Language and Communication*, Radboud University and Tilburg University. Second semester 2013/2014, 2014/2015, 2015/2016.

### **Tilburg University**

*Introduction Human Aspects of Information Technology*, quarter year course, curriculum Bachelor *Communication and Information Technology*, School of Humanities, Tilburg University. First semester 2009/2010. With Eric Postma.

*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 2005/2006–2009/2010.

*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. 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 Krahmer.

*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**

*Cognitive Artificial Intelligence*, semester course, University of Antwerp, Belgium. First semester 2009/2010.

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

*Capita selecta Computational Linguistics: Machine Learning of Language*, 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 Eric Postma, Ton Weijters, and Peter Braspenning. Second trimester 1995/1996 and 1996/1997.

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

## Graduate and post-graduate levels

Course modules for the Dutch School of Information and Knowledge Systems (SIKS):

*Machine learning for language modeling*, part of SIKS Advanced Ph.D. course on Computational Intelligence, Utrecht, March 12, 2010.

*Issues in empirical machine learning research*, part of SIKS Basic Ph.D. course on Research methods and methodology for information and knowledge systems, 2005, 2006, 2009–2015.

*Entities in chains*, part of SIKS Advanced Ph.D. course on Probabilistic Methods for Entity Resolution and Entity Ranking, Zeist, April 21, 2009.

*Implicit linguistics in memory-based language modeling*, one-week course as part of the First GLOW Spring School, Brussels, Belgium, April 7–11, 2014.

*Exemplar-based versus abstractionist models of word processing*, one-week course as part of the ESF NetWordS Summer School on Interdisciplinary Approaches to Exploring the Mental Lexicon, with Colin Davis, Dubrovnik, Croatia, July 2–6, 2012.

*Inducing linguistic knowledge from statistical machine translation systems*, one-week course as part of LOT Summer School (Dutch Graduate School in Linguistics), Leuven, Belgium, June 20–24, 2011.

*Memory-based language modeling*, one-week advanced course as part of ESSLLI-2010 (22th European Summer School in Logic, Language, and Information), Copenhagen, Denmark, August 16–20, 2010.

*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.

*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 in 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

*Memory-based models of language learning and processing: A computational account*, tutorial at the 13th International Congress for the Study of Child Language, Amsterdam, July 14, 2014. With Walter Daelemans.

*Text analytics for data journalism*. tutorial at the Annual Meeting of the Association of Investigative Journalists (VVOJ), Ghent, Belgium, November 20, 2010.

*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, June 26, 1999.

## Grant acquisition

### Grants awarded:

(abbreviations: NWO = Netherlands Organisation for Scientific Research; KNAW = Royal Dutch Academy for Arts and Sciences; 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). 100,000 EUR. Ph.D. student: L. Maruster (Eindhoven Technical University).

SoBU project 00-C, *Learning to communicate: Machine learning of dialogue strategies*. 2001-2004. With E. Kraemer, J. Terken (Eindhoven Technical University). 105,000 EUR. Ph.D. student: P. Lendvai.

SoBU project 02-10, *Weighted error minimisation in assigning prosodic structure to synthetic speech*. 2002-2004. With J. Terken (Eindhoven Technical University). 90,000 EUR. Ph.D. student: O. van Herwijnen (Eindhoven Technical University).

*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). 620,000 EUR. Researchers: E. Marsi, M. Reynaert, V. Hoste (UIA).

*Memory models of language*. Funded by NWO *Vernieuwingsimpuls* programme, 2001-2005. 680,000 EUR. Researchers: A. van den Bosch, I. Hendrickx, G.J. Busser.

*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). 686,000 EUR. (Tilburg part: 210,000 EUR). Post-doc researcher: M. van Zaanen.

*NEXTENS, Nederlandse Extensie voor Tekst naar Spraak*. Funded by SST, Stichting Spraaktechnologie (Foundation for Speech Technology), 2002-2003. 40,000 EUR. With T. Rietveld (KUN). Development and research: E. Marsi, A. Russell.

Prosodic annotation for the *Spoken Dutch Corpus*, student assistant team, data generation, and evaluation. Funded by NWO Spoken Dutch Corpus Programme. 2002-2003. 25,000 EUR. With E. Marsi.

*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. 830,000 EUR (Tilburg part: 200,000 EUR). Researcher: T. Bogers.
- 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. 560,000 EUR. Researchers: C. Sporleder (post-doc), P. Lendvai (post-doc), M. van Erp (Ph.D. student). Scientific programmer: T. Porcelijn, S. Hunt.
- 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. 580,000 EUR (Tilburg part: 84,000 EUR). Post-doc researcher: M. Reynaert.
- Implicit Linguistics.* Funded by NWO VICI program, 2006-2011. 1,250,000 EUR. Researchers: A. van den Bosch (PI), M. van Zaanen (post-doc researcher, until September 2009), T. Gaustad van Zaanen (post-doc researcher, from October 2009), H. Stehouwer (Ph.D. student). Scientific programmer: P. Berck.
- Linguistic Data Quality Management.* Funded by NWO HEFBOOM program, 2006-2008. 81,000 EUR. Post-doctoral researcher: Martin Reynaert.
- 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. 100,000 EUR (Tilburg part: 45,000 EUR). Post-doctoral researcher: Tanja Gaustad van Zaanen.
- MEMPHIX: Memory-based Paraphrasing with Implicit and Explicit Semantics*, a Tilburg University Faculty of Humanities Graduate School project. Ph.D. student: Sander Wubben. With profs. Harry Bunt and Emiel Krahmer.
- HITIME: Historical Timeline Mining and Extraction.* Part of *CATCH: Continuous Access to the Cultural Heritage*, an NWO Programme. Funded by NWO, 2009–2013. 614,000 EUR. Researchers: Kalliopi Zervanou (post-doctoral researcher), Matje van de Camp (Ph.D. student), Steve Hunt (scientific programmer).
- DutchSemCor: A one-million-word text corpus for Dutch with sense-tags*, an NWO Humanities Investment Subsidy 'Medium' project, 2009–2012. With prof. P. Vossen (VU Amsterdam), prof. M. de Rijke (University of Amsterdam). 751,000 EUR (Tilburg part: 140,250 EUR). Post-doctoral researcher: Rubén Izquierdo. Scientific programmer: Maarten van Gompel.
- AMICUS: Automated motif discovery in cultural heritage and scientific communication texts*, an NWO Humanities Internationalization networking grant, 2009–2012. With the universities of Borås (Sweden), Pecs (Hungary), the Hungarian Academy of Sciences, and other international partners. 38,000 EUR.
- PoliticalMashup*, an NWO Humanities Investment Subsidy 'Medium' project, 2010–2013. With prof. M. de Rijke, dr. M. Marx (University of Amsterdam), and prof. G. Voerman (University of Groningen). 552,000 EUR (Tilburg part: 98,000 EUR). Post-doctoral researcher: Martin Reynaert.
- TTNWW: TST Tools voor het Nederlands als Webservices in een Workflow*, a Dutch-Flemish CLARIN cooperation (Dutch part funded by NWO). With partners from other Dutch and Flemish universities, 2010–2012. 1,356,000 EUR (Tilburg part: 75,000 EUR). Post-doctoral researcher: Martin Reynaert. Scientific programmer: Maarten van Gompel.
- ADNEXT: Adaptive Information Extraction over Time*, work package 1 of *P1 Infiniti: Information Retrieval for Information Services*, a FES COMMIT project, 2011–2015. 2,845,000 EUR (Nijmegen part: 284,500 EUR). Ph.D. students: Florian Kunneman and Ali Hürriyetoğlu.

*Nederlab*, an NWO Humanities Investment Subsidy ‘Big’ project. With the Meertens Institute and other partners, 2013–2017. 4,000,000 EUR (Nijmegen part: 150,000 EUR). Post-doctoral researcher: Martin Reynaert. Also with Erwin Komen.

*CLARIAH: Common Lab Research Infrastructure for the Arts and Humanities*, funded by the NWO National roadmap for large-scale research facilities program, 2015–2019. 12,000,000 EUR (Nijmegen part: 500,000 EUR). Scientific programmers: Ko van der Sloot. Also with Henk van den Heuvel, Nelleke Oostdijk, Maarten van Gompel.

*DISCOSUMO: Discussion Thread Summarization for Mobile Devices*, funded by NWO Humanities ‘Creative Industry’ Strategic Research programme, 2015–2018. 423,920 EUR (Nijmegen part: 215.500 EUR). Post-doctoral researcher: Suzan Verberne.

*TraMOOC: Translation for Massive Open Online Courses*, a Horizon 2020 ICT-17 project. With Humboldt University and other partners, 2015–2018. 3.270.710 EUR (Nijmegen part: 319.550 EUR). Post-doctoral researcher: Iris Hendrickx.

*FutureTDM*, a Horizon 2020 GARRI3-2014 project. With the National Library of the Netherlands and other partners, 2015–2018. 1,388,003 EUR (Nijmegen part: 170,870 EUR). Post-doctoral researcher: Maria Eskevich.

## Professional organizations

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

Chair of BNVKI-AIABN: Benelux Association for Artificial Intelligence (2006–2011; Secretary 2001–2006).

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).

President of SIGHUM, *Special Interest Group on Language Technologies for the Socio-economic Sciences and the Humanities* of the Association for Computational Linguistics (ACL) (2013–2015).

## Societal organizations

*Stichting Sint Aegten*, Board member.

## Awards

*Best Paper Award*, Sixth International Workshop on Computational Models of Narrative, CMN-2015, Atlanta, Georgia, USA. With Folgert Karsdorp, Marten Van der Meulen, Theo Meder.

*Best Paper Award*, Benelux AI Conference, BNAIC-2014, Nijmegen, The Netherlands, November 2014. With Florian Kunneman.

*Best Explanation Award*, AI Video Competition, AAAI-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) of ongoing Ph.D. projects

1. Martijn Bentum, Radboud University. With Mirjam Ernestus.
2. Peter Berck, Radboud University. With Eric Postma.
3. Maarten van Gompel, Radboud University.
4. Ali Hürriyetoğlu, Radboud University. With Nelleke Oostdijk.
5. Folgert Karsdorp, Radboud University and Meertens Institute, Amsterdam, The Netherlands. With Theo Meder, Mariët Theune, and Franciska de Jong.
6. Florian Kunneman, Radboud University. With Margot van Mulken.
7. Alessandro Lopopolo, Radboud University. With Karl-Magnus Petersson.
8. Louis Onrust, Radboud University and University of Leuven, Leuven, Belgium. With Hugo Van hamme.
9. Véronique Verhagen, Tilburg University. With Ad Backus and Joost Schilperoord.

Advisor (promotor) or co-advisor (co-promotor) of completed Ph.D.s

1. Matje van de Camp, *A link to the past: Constructing historical social networks from unstructured data*, Tilburg University, March 2, 2016. Promotor, with Eric Postma.
2. Sander Wubben, *Text-to-text generation by monolingual machine translation*, Tilburg University, June 5, 2013. Promotor, with Emiel Krahmer.
3. Herman Stehouwer, *Statistical language models for alternative sequence selection*, Tilburg University, December 7, 2011. Promotor, with Jaap van den Herik and Menno van Zaanen (co-promotor).
4. Marieke van Erp, *Accessing natural history: Discoveries in data cleaning, structuring, and retrieval*, Tilburg University, June 30, 2010. Promotor, with co-promotor Piroska Lendvai.
5. Maria Mos, *Complex lexical items*, Tilburg University, May 12, 2010. Promotor, with co-promoters Anne Vermeer and Ad Backus.
6. Toine Bogers, *Recommender systems for social bookmarking*, Tilburg University, December 8, 2009. Promotor.
7. Stephan Raaijmakers, *Multinomial language learning: Investigations into the geometry of language*, Tilburg University, December 1, 2009. Promotor, with Walter Daelemans.
8. Sander Canisius, *Structured prediction for natural language processing: A constraint satisfaction approach*, Tilburg University, February 13, 2009. Promotor, with Walter Daelemans.
9. Martin Reynaert, *Text induced spelling correction*, Tilburg University, December 2, 2005. Co-promotor. Promotor: Walter Daelemans.
10. Iris Hendrickx, *Local classification and global estimation: Explorations of the k-nearest neighbor algorithm*, Tilburg University, November 28, 2005. Co-promotor. Promotor: Walter Daelemans.
11. Piroska Lendvai, *Extracting information from spoken input*, Tilburg University, December 20, 2004. Co-promotor.
12. Laura Maruster, *A machine learning approach to understand business processes*, Eindhoven Technical University, August 27, 2003. Co-promotor, with Emiel Krahmer (co-promotor), Walter Daelemans (first promotor), and Harry Bunt (second promotor).

13. Sabine Buchholz, *Memory-based grammatical relations*, Tilburg University, The Netherlands, December 13, 2002. Co-promotor. Promotores: Walter Daelemans and Harry Bunt.

Member of Ph.D. committee



1. Wen Li, *Understanding geo-spatial information on social media*. Delft University, May 23, 2016. Committee member.
2. Maya Sappelli, *Knowledge work in context: User centered knowledge worker support*. Radboud University, February 19, 2016. Manuscript committee member.
3. Christine Liebrecht, *Intens krachtig: Stilistische intensieveerders in evaluatieve teksten*. Radboud University, July 10, 2015. Committee member.
4. Orphée de Clerck, *Tipping the scales: exploring the added value of deep semantic processing on readability prediction and sentiment analysis*, Ghent University, July 6, 2015. Committee member.
5. Niels Bloom, *Grouping by association: using associative networks for document categorization*, University of Twente, June 10, 2015. Committee member.
6. Binyam Gebrekidan Gebre, *Machine learning for gesture recognition from videos*, Radboud University, February 9, 2015. Manuscript committee member.
7. Jasmina Maric, *Web communities, immigration, and social capital*, Tilburg University, November 18, 2014. Committee member.
8. Eva D'hondt, *Cracking the patent*, Radboud University, June 17, 2014. Committee member.
9. Joost van Doremalen, *Developing automatic speech recognition-enabled language learning applications: From theory to practice*, Radboud University, May 16, 2014. Manuscript committee member.
10. Mena Habib, *Named entity extraction and disambiguation for informal text: The missing link*, Twente University, May 9, 2014. Committee member.
11. Daniel Kuhlwein, *Machine learning for automated reasoning*, Radboud University, April 14, 2014. Manuscript committee member.
12. Thomas Markus, *Where social noise and structure converge*, Utrecht University, January 7, 2014. Committee member.
13. Iris Hanique, *Mental representation and processing of reduced words in casual speech*, Radboud University, November 13, 2013. Committee member.
14. Erwin Komen, *Finding focus: A study of the historical development of focus in English*, Radboud University, June 25, 2013. Committee member.
15. Joel Ribero, *Multidimensional process discovery*, Eindhoven Technical University, March 13, 2013. Committee member.
16. Els Lefever, *Exploiting parallel corpora for lexical disambiguation*, Ghent University, Belgium, September 24, 2012. Jury member.
17. Daphne Theijssen, *Making choices: Modelling the English dative alternation*, Radboud University, June 18, 2012. Committee member.
18. Mike Kestemont, *Het gewicht van de auteur: Een onderzoek naar stylometrische auteursherkenning in de Middelnederlandse epiek*, University of Antwerp, Belgium, May 23, 2012. Jury member.
19. Çağrı Cöltekin, *Catching words in a stream of speech: Computational simulations of segmenting transcribed child-directed speech*, University of Groningen, December 8, 2011. Committee member.
20. Barbara Plank, *Domain adaptation for parsing*, University of Groningen, December 8, 2011. Committee member.
21. Owen Lamont, *Applications of memory based learning to spoken dialogue system development*, Murdoch University, Australia, 2011. External examiner.
22. Raoul Frijters, *Text mining the biomedical literature for application to drug discovery*, Radboud University, September 16, 2011. Committee member.
23. Xiaoyu Mao, *Airport under control: Multiagent scheduling for airport ground handling*, Tilburg University, May 25, 2011. Committee member.
24. Bart Bogaert, *Cloud content contention*, Tilburg University, March 30, 2011. Committee member.
25. Maarten Clements, *Personalised access to social media*, Delft University of Technology, December 6, 2010. Committee member.
26. Eline Westerhout, *Definition extraction for glossary creation: A study on extracting*

- definition for semi-automatic glossary creation in Dutch*, Utrecht University, July 2, 2010. Committee member.
27. Tim van de Cruys, *Mining for meaning: The extraction of lexico-semantic knowledge from text*, University of Groningen, June 24, 2010. Committee member.
  28. Lineke Sneller, *Does ERP add company value? A study for The Netherlands and the United Kingdom*, Nyenrode School of Accountancy & Controlling, June 18, 2010. Committee member.
  29. Suzan Verberne, *In search of the why: Developing a system for answering why-questions*, Radboud University, April 27, 2010. Committee member.
  30. Sander Bakkes, *Rapid adaptation of video game AI*, Tilburg University, March 3, 2010. Committee member.
  31. Laurens van der Maaten, *Feature extraction from visual data*, Tilburg University, June 23, 2009. Committee member.
  32. Fritz Reul, *New architectures in computer chess*, Tilburg University, June 17, 2009. Committee member.
  33. Jeroen Geertzen, *Dialogue act recognition and prediction: Explorations in computational dialogue modelling*, Tilburg University, February 11, 2009. Committee member.
  34. Benjamin Torben Nielsen, *Dendritic morphology: Function shapes structure*, Tilburg University, December 3, 2008. Committee member.
  35. Jesús Giménez Linares, *Empirical machine translation and its evaluation*, Universitat Politècnica de Catalunya, Barcelona, Spain, July 2, 2008. Jury member.
  36. Grzegorz Chrupala, *Towards a machine-learning architecture for lexical functional grammar parsing*, Dublin City University, Ireland, June 10, 2008. External examiner.
  37. Emmanuel Keuleers, *Memory-based learning of inflectional morphology*, Universiteit Antwerpen, Belgium, March 6, 2008. Jury member.
  38. Christian Biemann, *Unsupervised and knowledge-free natural language processing in the structure discovery paradigm*, Universität Leipzig, Germany, 2007. Referee.
  39. 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.
  40. Véronique Hoste, *Optimization issues in machine learning of coreference resolution*, University of Antwerp, Belgium, March 11, 2005. Jury member.
  41. Olga van Herwijnen, *Weighted error minimisation in assigning prosodic structure for synthetic speech*, Eindhoven Technical University, May 17, 2004. Committee member.
  42. Stefan Frank, *Computational modeling of discourse comprehension*, Tilburg University, February 27, 2004. Committee member.
  43. Brian Sheppard, *Toward perfect play of scrabble*, Universiteit Maastricht, July 5, 2002. External reviewer.

Member of editorial board, journal advisory board, or standing reviewer committee

- Computational Linguistics (2003–2005)
- Computational Linguistics in the Netherlands Journal (from 2011)
- FoLLI Publications on Language, Logic, and Information (2005–2011)
- Genocide Studies and Prevention (from 2015)
- Language and Speech Technology Technical Report Series (from 2012)
- Research on Language and Computation (2006–2011)
- Transactions of the Association for Computational Linguistics (from 2014)

Reviewer for journal (alphabetic)

- ACM Computing Surveys
- AI Communications
- Artificial Intelligence
- Annals of Biomedical Engineering
- Computers in Industry
- Computer Speech and Language
- Computational Intelligence
- Computational Linguistics
- Data and Knowledge Engineering
- Frontiers in Psychology
- IEEE Transactions on Knowledge and Data Engineering
- IEEE Transactions on Speech and Audio Processing
- Information Fusion
- Information Systems
- International Computer Games Association Journal
- International Journal of Document Analysis and Recognition
- International Journal of Lexicography
- Iranian Journal of Electrical and Computer Engineering
- Journal for Artificial Intelligence Research
- Journal of Germanic Linguistics
- Journal of Machine Learning Research
- Journal of Management Information Systems
- Journal of Natural Language Engineering
- Koers
- Language Resources and Evaluation
- Literary and Linguistic Computing
- Literator: Tydskrif vir besondere en vergelykende taal- en literatuurstudie
- Machine Learning Journal
- Machine Translation
- PLOS ONE
- Proceedings of the National Academy of Sciences of the United States of America (PNAS)
- Proceedings of the Royal Society A
- Research on Language and Computation
- Science
- Speech Communication
- Tijdschrift voor Taalbeheersing
- Transactions of the Association for Computational Linguistics
- Written Language and Literacy

Programme chair / general chair (reversed chronological)

- ACL-2016, 54th Annual Meeting of the Association for Computational Linguistics. Humboldt University, Berlin, Germany. General chair.
- HistoInformatics-2016, 3rd International Workshop on Computational History. 12–16 July, Krakow, Poland. Programme co-chair with Adam Jatowt, Marten Düring, and Johannes Preiser-Kappeller.
- HistoInformatics-2014, 2nd International Workshop on Computational History. November 10, 2014, Barcelona, Spain. Programme co-chair with Adam Jatowt, Marten Düring, and Gaël Dias.
- BNAIC-2014, 26th Benelux Artificial Intelligence Conference. November 6–7, 2014, Radboud University, Nijmegen, The Netherlands. Programme co-chair with Johan Kwisthout, Tom Heskes, Peter Desain, and Bert Kappen.
- BENELEARN-2013, 22nd Belgian-Dutch Conference on Machine Learning. June 3, 2013, Radboud University, Nijmegen, The Netherlands. Programme co-chair with Tom Heskes and David van Leeuwen.
- LaTeCH-2012, EACL-2012 Sixth Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. April 24, 2012, Avignon, France. Programme co-chair with Kalliopi Zervanou.

- 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 First Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. 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 Dan Roth (UIUC).
- BNAIC-2000, Twelfth Belgian-Dutch Artificial Intelligence Conference. November 1–2 2000, Kaatsheuvel, The Netherlands. Programme co-chair with Hans Weigand (Tilburg University).
- BENELEARN-1997, Seventh Belgian-Dutch Conference on Machine Learning. October 21, 1997, Tilburg University, Tilburg, The Netherlands. Programme co-chair with Peter Flach and Walter Daelemans.

Member of programme committee / reviewer (alphabetic)

- AAI, Annual Meeting of the Association for the Advancement of Artificial Intelligence (2008). Satellite events:
  - AAI-2005 Workshop on Spoken Language Understanding for Conversational Systems, Pittsburgh, PA, USA (2005)
- ABC, European Conference of the Association of Business Communication (2011)
- ACL, Annual Meeting of the Association for Computational Linguistics (2003, 2005, 2006, 2008–2012, area chair for ACL-2004, programme chair for ACL-2007). ACL mentor programme 2009, 2010. 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 (2007, 2015, 2016)
  - ACL-2010 Workshop on Domain Adaptation for Natural Language Processing, Uppsala, Sweden (2010)
- AND, Workshop on Analytics for Noisy Unstructured Text Data (2009–2011)
- BENELEARN, Belgian-Dutch Conference on Machine Learning (1996–2001, 2005–2011)
- BNAIC, Belgium-Netherlands Artificial Intelligence Conference (1998–2010)
- CLIN, Computational Linguistics in the Netherlands (1999–2009)
- 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 (1998–2011)
- DIR, Dutch-Belgian Information Retrieval Workshop (2008–2011)
- EACL, Annual Meeting of the European Chapter of the Association for Computational Linguistics (2001, 2009, 2012, 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)
- EAMT, Annual Meeting of the European Association for Machine Translation (2010–2012, 2015)
- ECAI, European Conference on Artificial Intelligence (2008, area chair for 2010)
- 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 (2004–2011, area chair 2010 and 2013)
- 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)
- IJCNLP, International Joint Conference on Natural Language Processing (2011)
- INFWET, Interdisciplinaire Conferentie Informatiewetenschap (2004)
- Interspeech/ICSLP, International Conference on Spoken Language Processing (2006, 2007)
- LaTeCH, Workshop on Language Technology for Cultural Heritage Data (2007–2014)
- LREC, International Conference on Language Resources and Evaluation (2006, 2008, 2010)
- NIPS, Conference on Neural Information Processing Systems (2008, 2009)
- NLDB, International conference on Application of Natural Language Processing to Information Systems (2012)
- \*SEM, Joint Conference on Lexical and Computational Semantics (2013, 2015)

#### Scientific advice

- Netherlands
  - Humanities integrator, Netherlands eScience Centre (from 2012).
  - Reviewer, NWO programs *Vrije Competitie (Open Competitie)*, *Vernieuwingsimpuls*, *Mozaïek*, *Talent*, *Cognition*
  - Advisor, NWO *Taalportaal* program (2011)
  - Advisor, NWO *CLARIN-NL* program (2010–2011)
  - Advisor, Lorentz Center (2011–2013)
  - Evaluator, Nederlandse Taalunie (Dutch Language Union) (2007)
- International
  - Reviewer, ESF

- Reviewer, EPSRC *Advanced Fellowship* program, UK
- Reviewer, BSF (United States–Israel Bi-national Science Foundation) program
- Reviewer, ICREA (Institució Catalana de Recerca I Estudis Avançats) *Senior Researcher* program, Spain/Catalunya
- Reviewer, IWT-Vlaanderen, Belgium
- Reviewer, NRF South Africa

#### Scientific board memberships

- Chair, board of the Centre for Language and Speech Technology, Faculty of Arts, Radboud University (from 2012, chair from 2014)
- Member, scientific board of NWO Gravity *Language in Interaction* programme (from 2012)
- Member, board of National Roadmap for Large-Scale Research Facilities Programme CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities) (from 2014)

#### Advisory board and steering group memberships

- Member, eHumanities.nl network steering committee (from 2016)
- Member, CLARIAH board (Common Lab Research Infrastructure for the Arts and Humanities) (from 2015)
- Member, KNAW DANS scientific advisory board (Data Archiving and Networked Services) (from 2013)
- Member, KNAW Huygens ING scientific advisory board (Huygens Institute for the History of the Netherlands) (from 2012)
- Chair, Nederlandse Taalunie advisory board for INL (Dutch Language Union) (2011–2015)

#### Scientific programme committee memberships

- Netherlands eScience Center *ASDI* (Accelerating Scientific Discovery), chair of the board (2016)
- KNAW Programme *Computational Humanities*, board member (2009–2016)
- NWO Programme *CATCH* (Continuous Access to Cultural Heritage), board member (2004–2010)
- NWO Programme *IMIX* (Interactive Multimodal Information Extraction), board member (2002–2008)