

CURRICULUM VITAE

Véronique Hoste

Name	Véronique Hoste
Address	Faculty of Translation Studies University College Ghent Groot-Brittanniëlaan 45 B-9000, Ghent Belgium Email: veronique.hoste@hogent.be

1 Academic titles and degrees

B.A. Higher Institute for Translators and Interpreters. German, Portugese, Danish. 1995. [magna cum laude]

M.A. Higher Institute for Translators and Interpreters. German, Portugese, Danish. 1997. [magna cum laude]

Masters thesis: “Ein kontrastives niederländisch-deutsches Valenzlexikon. Zum Problem der laut- und sinnverwandten Wörter für niederländischsprachige Germanistikstudenten”.

Postgraduate degree GGS Taalwetenschap, interuniversity degree (RUG, UIA, KUL, VUB), 1998. [cum laude]

Thesis: “Example-Based Machine Translation. The application of memory-based reasoning in translation technology”.

Ph.D. Computational Linguistics (Doctoraat Letteren en Wijsbegeerte), Antwerp University. “Optimization Issues in Machine Learning of Coreference Resolution”, 2005.

2 Scientific experience

May 2006 - ... Assistant Professor (docent), University College Ghent. Founder and Research Director of the LT3 (Language and Translation Technology Team) research group (<http://veto.hogent.be/lt3>). Director of the “language technology” section.

September 2007 - ... Associate researcher, Department of Applied Mathematics and Computer Science, Ghent University. Co-founder of the association research group “Language Technology and

Computational Intelligence”.

May 2006 - ... Guest professor, Antwerp University

July 2005 - April 2006 (postdoctoral research assistant, Antwerp University): ‘COREA, Coreference Resolution for Extracting Answers’. (STEVIN)

April 2005 - June 2005 (postdoctoral research assistant, Antwerp University): ‘BioMinT, Biological Text Mining’. (EU)

2001-2004 (research assistant, Antwerp University): “PROSIT, Prosody from Information in Text”. In this project different techniques are investigated to generate a prosodic structure for a Dutch speech synthesis system. (FWO Flanders - NWO)

1998-2000 (research assistant, Antwerp University): “LINGUADUCT, An Inductive Machine Learning Environment for Corpus Annotation”. This project investigates how learning algorithms can be incorporated in the annotation of existing corpora. (FWO Flanders)

3 Project management

Current project management

- HOF (Research Fund, University College Ghent)
“SUBTLE, SUBjectivity Tagging and Learning”, in cooperation with Ghent University and the Suicide Prevention Centre.
Period: summer 2010 (4-year project)
Budget: 1 PhD student
- FWO (Flemish Research Fund, “Krediet voor navorsers”) “Koppeling van schrijfprocesdata aan lexica” (Merging writing process data with lexica), in cooperation with Luuk Van Waes (Antwerp University). Period: 2010
Budget LT3: 11,700 euro
- EWI (Department of Economy, Science and Innovation, Flemish government)
“Stylene, Een web service voor stylometrie- en leesbaarheidsonderzoek voor het Nederlands” (Stylene: a web service for stylometry and readability research for Dutch), in cooperation with Walter Daelemans (Antwerp University).
Period: 01/01/2010 - 31/12/2011.
Budget: 120,000 euro.
- EWI (Department of Economy, Science and Innovation, Flemish government)
“TTNWW, TST Tools voor het Nederlands als Webservices in een Workflow” (TTNWW: web services for Dutch as web services in a work flow), in cooperation with the Universities of Antwerp, Tilburg, Leuven, Utrecht and Groningen.
Period: 2010 - 2012
Budget LT3: 50,000 euro.

- IOF StarTT (Industrial Research Fund, Ghent University)
 “OptiPIL, a readability Optimizer for Patient Information Leaflets”, in cooperation with Artesis University College Antwerp
 Period: 01/01/2011-30/12/2011.
 Budget LT3: 49,980 euro.
- STEVIN
 “SoNaR, Stevin Nederlandstalig Referentiecorpus.” (SoNaR, Stevin Dutch Reference Corpus), in cooperation with the University of Utrecht and the Catholic University of Leuven.
 Period: 01/01/2009-31/03/2012.
 Budget LT3: 191,121 euro.
- STEVIN
 “DuOMAn, Dutch Language Online Media Analysis” in cooperation with the University of Amsterdam, the University of Groningen, GridLine and TrendLight.
 Period: 01/05/2009-31/10/2011.
 Budget LT3: 44,070 euro.
- BOF (Special Research Fund, Ghent University) “AutoWeb: een Semantisch Web waarin Agenten zich Gedragen als Autodidact” (AutoWeb: a Semantic Web in which the agents behave as Autodidacts”), in cooperation with M. De Cock (Ghent University).
 Period: 01/11/2006-31/10/2009.
- HOF (Research Fund, University College Ghent)
 “HENDI, een Automatische Helderheidstool voor Engels en Nederlands Discours in Parallele en Vergelijkbare Teksten” (“HENDI, an Automatic Clarity Tool for English and Dutch Discourse in Parallel and Comparable Texts”), in cooperation with S. Vandepitte, S. Marzo, B. Defrancq, Hildegard Vermeiren and Roland Maerivoet (University College Ghent), W. Daelemans (Antwerp University) and Gilles-Maurice de Schryver (Ghent University)
 Period: 01/10/2008-30/09/2011.
 Budget: 335.785 euro.
- HOF (Research Fund, University College Ghent)
 Server for data storage and computationally expensive processing.
 Budget LT3: 21.267 euro.

Previous projects

- BOF (Special Research Fund, Antwerp University)
 “APRO: Annotatie van genuscongruentie en van het pleonastisch vs. anaforisch gebruik van pronomina in het Nederlands” (“APRO: Annotation of gender agreement and of the pleonastic versus anaphoric use of pronouns in Dutch”).
 Period: 2006
 Budget: 7,410 euro.
- IWT TETRA (Institute for the Promotion of Innovation by Science and Technology in Flanders)
 “ABOP, Automatische Bijsluiter Optimalisator. Een auteursomgeving voor de correctie en productie van bijsluiter teksten” (“ABOP, Automatic Leaflet Optimizer. An authoring environment for the correction and production of patient leaflets”), in cooperation with University College Antwerp, Antwerp University and the Catholic University of Leuven.
 Period: 01/12/2007-30/12/2009.
 Budget LT3: 157,458 euro.

- Private funding: Peugeot Citroen SA (LT3 as subcontractor of Telelingua international)
“Étude doublons et synonymes thésaurus APV”.
Period: 01/02/2008-31/12/2009.
Budget LT3: 54,828 euro.
- STEVIN (Stimulation programme for the development of resources for Dutch language and speech technology)
“SoNaR, Stevin Nederlandstalig Referentiecorpus. Fase 1” (SoNaR, Stevin Dutch Reference Corpus. Phase 1), in cooperation with the University of Twente, Tilburg University and the University of Nijmegen.
Period: 01/01/2008-31/12/2008.
Budget LT3: 26,400 euro.

4 Teaching experience

- Flanders Language Valley, course on machine translation, Ieper, 2001
- *Language technology (Computerlinguïstiek)*, course on machine translation, Antwerp University, 1999-2001.
- *Language technology (Computerlinguïstiek)*, course on anaphora resolution, Antwerp University, 2004-2006.
- *Text Mining*, guest lecture on anaphora resolution, Tilburg University, 3 May 2005.
- *Introduction to Language technology (BA3)*, School of Translation Studies, University College Ghent, 2006-...
- *Language Technology (MA)*, School of Translation Studies, University College Ghent, 2007-...
- *Introduction to Language Technology (Postgraduate Dutch and Translation)*, School of Translation Studies, University College Ghent, 2007-....
- *Introduction to Language Technology*, Doctoral School of Arts, Humanities and Law, Doctoral School of Natural Sciences, Ghent University, January 2009.
- *Computer aided translation*, March 2009. In cooperation with Telelingua International.
- *Natural Text Understanding (MA)*, University of Antwerp, 2009-2010. In cooperation with Guy De Pauw.
- *Advanced Methods for Corpus Linguistics*, Doctoral School of Arts, Humanities and Law, December 2009. In cooperation with Evie Coussé and Gert De Sutter.

5 Organisation of conferences and workshops

- Benelearn 2001, “Eleventh Belgian-Dutch Conference on Machine Learning”, Hof Van Liere, Antwerp, December 21, 2001.
(in cooperation with Guy De Pauw)

- Atila workshop, Middelheim Castle, Antwerp, September 3, 2003.
- CLIN 2003, “Computational linguistics in the Netherlands meeting”, UA, Antwerp, December 19, 2003.
(in cooperation with Bart Decadt, Guy De Pauw and Walter Daelemans)
- “Workshop on the analysis and generation of prosody”, Tilburg, December 21, 2004.
(in cooperation with Walter Daelemans, Antal van den Bosch, Erwin Marsi and Martin Reynaert)
- “Workshop on Machine Learning of Information Extraction”, UA, Hof van Liere, Antwerp, March 10, 2005.
(in cooperation with Walter Daelemans)
- Atila workshop, Priorij Corsendonk, November 28-29, 2005.
- DIR 2008, “Dutch-Belgian Information Retrieval Workshop”, Maastricht, April 14-15, 2008.
(in cooperation with Martine De Cock and Eduard Hoenkamp)
- ABOP workshop (“Automatische BijsluiterOptimalisator”), Gent, January 20, 2010.
- SemEval-2 task “Cross-Lingual Word Sense Disambiguation”,
<http://semeval2.fbk.eu/semeval2.php?location=tasks>.
(in cooperation with Els Lefever)
- SemEval-2 task “Coreference Resolution in Multiple languages”,
<http://semeval2.fbk.eu/semeval2.php?location=tasks>.
(in cooperation with Lluís Marquez, M. Antonia Martí, Massimo Poesio, Marta Recasens, Emili Sapena, Mariona Taule, Yannick Versley. (Universitat de Barcelona, Universitat Politècnica de Catalunya, Università di Trento, Universität Tübingen)
- CLIN 2011, “Computational linguistics in the Netherlands meeting”.

6 Scientific advice

Member “visitatiecommissie Taalwetenschappen” (2008), (presided by Prof. Dr. J. De Caluwé).
Member of the Ghent University Association Hercules commission (2008-2010).

7 Programme committee/reviewing

- **Member of the programme committee (reviewing of full papers):**
EACL 99 workshop “Linguistically Interpreted Corpora” (Bergen, 1999), ACL-2004 Annual Meeting of the Association of Computational Linguistics (Barcelona, 2004), CONLL-2005 Ninth Conference on Computational Natural Language Learning (Ann Arbor, 2005), EMNLP-2006 Empirical Methods in Natural Language Processing (Sydney, 2006), COLING-ACL International Conference on Computational Linguistics-Annual Meeting of the Association of Computational Linguistics (Sydney, 2006), ESSLLI workshop “Ambiguity in Anaphora” (Malaga,

2006), ICSC-2007 IEEE International Conference on Semantic Computing (Irvine, 2007), SemEval 2007 International Workshop on Semantic Evaluations (Prague, 2007), RALP-2007 Recent Advances in Natural Language Processing (Borovets, 2007), LREC-2008 Language Resources and Evaluation Conference (Marrakech, 2008), ACL-2008 Annual Meeting of the Association of Computational Linguistics (Ohio, 2008), SEW Semantic Evaluations Workshop (Ohio, 2008), DIR-2008 Dutch-Belgian Information Retrieval Workshop (Maastricht, 2008), EMNLP-2008 Conference on Empirical Methods in Natural Language Processing (Waikiki, Honolulu, 2008), EACL-2009 European Chapter of the Annual Meeting of the Association of Computational Linguistics (Athens, 2009), DIR-2009 Dutch-Belgian Information Retrieval Workshop (Twente, 2009), DAARC-2009 Discourse Anaphora and Anaphor Resolution Colloquia (Goa, 2009), SEW-2009 Semantic Evaluations: Recent Achievements and Future Directions (Colorado, 2009), ACL-IJCNLP 2009 Annual Meeting of the Association of Computational Linguistics (Singapore, 2009), NAACL-HLT 2009 North American Chapter of the Association for Computational Linguistics - Human Language Technologies (Colorado, 2009), ECML-PKDD 2009 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (Bled, 2009), RANLP-2009 Recent Advances in Natural Language Processing (Borovets, 2009), ACL-2010 Annual Meeting of the Association of Computational Linguistics (Uppsala, 2010), SemEval 2010 International Workshop on Semantic Evaluations (Uppsala, 2010), LREC-2010 Language Resources and Evaluation Conference (Valletta, 2010), WASSA Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (Lisbon, 2010), NeSp-NLP 2010 Negation and Speculation in Natural Language Processing workshop (Uppsala, 2010).

- **Reviewing for:**

- Journal of Artificial Intelligence Research (2003, 2006, 2007)
- Computational Linguistics in the Netherlands (2004, 2006, 2007, 2010)
- Journal of Machine Learning Research (2006)
- Research on Language and Computation (2007, 2010)
- Festschrift for David Dowty ‘Theory and Evidence in Semantics’ (2007)

- **Project Proposal Refereeing:**

- Hercules 1 and 2 (infrastructure) (2008, 2009)
- Erasmus Hogeschool research fund (Linguistics) (2008)
- ESRC Economic and Social Research Council (UK, 2009)
- NWO (The Netherlands, 2009)
- FWO (2010)

8 Ph.D. students

- Timur Fayruzov (2006-2010) prepares a Ph.D in the framework of the BOF special research fund of the University of Ghent. The project is co-supervised by Martine De Cock (Ghent University), Veronique Hoste (University College Ghent) and Chris Cornelis (Ghent University). The project aims at the development of technologies for the (semi-)automatic maintenance and enrichment of ontologies. This will lead to a system that suggests adaptations and corrections to an existing ontology on the basis of information extracted from evolving document collections. As a test case, we investigate the maintenance and enrichment of the cell cycle ontology on the basis of the electronically available biomedical scientific publications.
- Els Lefever (2007-2012) prepares a Ph.D in the framework of the HOF research fund of the

University College Ghent. The project is co-supervised by Veronique Hoste (University College Ghent) and Martine De Cock (Ghent University). The project aims at the development of the first automatic robust word sense disambiguation system for Dutch. Lacking large annotated text collections, large quantities of parallel texts are used to induce lexical knowledge in an unsupervised way (by means of clustering techniques).

- Philp van Oosten (2009-) prepares a Ph.D in the framework of the HENDI BOF research project which aims at the development of an intelligent readability metric. Philip is co-supervised by Veronique Hoste and Martine De Cock. The goal of the project is to go far beyond existing length-based approaches and to also automatically model syntactic, semantic and pragmatic textual complexity in order to obtain a holistic readability score.
- Member of the exam commission:
 - Anders Noklestad: "A Machine Learning Approach to Anaphora Resolution Including Named Entity Recognition, PP Attachment Disambiguation, and Animacy Detection" (University of Oslo, June 18-19, 2009)
 - Oier Lopez de Lacalle: "Domain-specific Word Sense Disambiguation" (Universidad del País Vasco, December 2009)
 - Stefan Raaijmakers: "Multinomial Language Learning. Investigations into the geometry of language" (University of Tilburg, 1st December 2009)
 - Lieve Macken: "Sub-sentential Alignment of Translational Correspondences" (University of Antwerp, May 28, 2010)

9 Membership

- Co-promotor (in cooperation with Martine De Cock) of the Ghent University Association research group "Language Technology and Computational Intelligence". The AOG comprises the following groups: Computational Web Intelligence (Ghent University), Language and Translation Technology Team LT3 (University College Ghent), Computational Linguistics for African Languages (Ghent University), Department of Industrial Management (University College Ghent), Department of Applied Engineering (University College Ghent).
- Secretary of CLIF, Computational Linguistics in Flanders, FWO sponsored research community on Language and Speech Technology, 2004.
- Member of the CLIF coordination group (2007-...).
- Member of ACL (Association for Computational Linguistics).
- Member of the organisation committee of Semeval-1/Senseval-4 (2007).
- Member of the research unit of the faculty of translation studies (2007-...)

10 Publications

- Articles

- A1
 - * Lefever, E., Fayruzov, T., Hoste, V. & De Cock, M.: 2010, ‘Clustering Web People Search Results Using Fuzzy Ants’. *Information Sciences*, to appear.
 - * Hoste, V., Vanopstal, K., Lefever, E. & Delaere, I.: 2010, ‘Classification-Based Scientific Term Detection in Patient Information’. *Terminology*, 16(1).
 - * Fayruzov, T., De Cock, M., Cornelis, C. & Hoste, V.: 2009, ‘Linguistic Feature Analysis of Protein Interaction Extraction’. *BMC Bioinformatics*, 10:374. (SCI: 3.78)
 - * Hoste, V., Daelemans, W. and Gillis, S.: 2004, ‘Using rule-induction techniques to model pronunciation variation in Dutch’, *Computer Speech and Language*, 18(1), 1-23. (SCI: 1.413)
 - * Daelemans, W., Hoste, V., De Meulder, F. and Naudts, B.: 2003, ‘Combined optimization of feature selection and algorithm parameters in machine learning of language’, *Lecture Notes in Artificial Intelligence*, 2837, 84-95. (SCI2004: 0.251)
- A2
 - * Hoste, V., Daelemans, W., Hendrickx, I. & van den Bosch, A.: 2002, ‘Parameter Optimization for Machine-Learning of Word Sense Disambiguation’, *Natural Language Engineering*, 8(4), 311-325.
- Books
 - B2
 - * Daelemans, W. & Hoste, V.: 2009, ‘Introduction. Evaluation of Translation Technology’. *Linguistica Antverpiensia New Series 8*. University Press Antwerp, 9-13.
 - * Daelemans, W., De Pauw, G., Durieux, G., Gillis, S., Hoste, V. & Tjong Kim Sang, E.: 2000, ‘Zelflerende Systemen als Instrument voor de Taalkunde en de Taaltechnologie, Met taal om de tuin geleid’, 81-93.
 - B3
 - * Daelemans, W. & Hoste, V. (eds.): 2009, *Evaluation of Translation Technology. Linguistica Antverpiensia New Series 8*. University Press Antwerp.
 - * Hoenkamp, E., De Cock, M. & Hoste, V. (eds.): 2008, *Proceedings of the 8th Belgian-Dutch Information Retrieval Workshop (DIR-2008)*, Maastricht, the Netherlands, 100 pages.
 - * Decadt, B., Hoste, V. & Depauw, G. (eds.): 2004, *Computational Linguistics in the Netherlands 2003. Selected Papers from the Fourteenth CLIN meeting*, Antwerp, Belgium, 184 pages.
 - * Hoste, V. & De Pauw, G. (eds.): 2001, *Proceedings of the Eleventh Belgian-Dutch Conference on Machine Learning*, Antwerp, Belgium.
- Conference Proceedings
 - P1 (ISI Web of Science)
 - * Hendrickx, I. & Hoste, V.: 2009, ‘Coreference Resolution on Blogs and Commented News’, *Lecture Notes in Artificial Intelligence*, 5847, Berlin: Springer Verlag, 43-53.
 - * Hendrickx, I., Hoste, V. & Daelemans, W.: 2008, ‘Semantic and syntactic features for Dutch coreference resolution’, *Lecture Notes in Computer Science*, 4919, Berlin: Springer Verlag, 351-361.

- * Fayruzov, T., De Cock, M., Cornelis, C. & Hoste, V.: 2008, ‘DEEPER: A Full Parsing Based Approach to Protein Relation Extraction’, *Lecture Notes in Computer Science*, 4973, Berlin: Springer Verlag, 36-47.
- * Hoste, V., Hendrickx, I. & Daelemans, W.: 2007, ‘Disambiguation of the neuter pronoun and its effect on pronominal coreference resolution’, *Lecture Notes in Artificial Intelligence*, 4629, Berlin: Springer Verlag, 48-55.
- * Hendrickx, I., Hoste, V. & Daelemans, W.: 2007, ‘Evaluating hybrid versus data-driven coreference resolution’, *Lecture Notes in Computer Science*, 4410, Berlin: Springer Verlag, 137-150.
- * Marsi, E., Reynaert, M., van den Bosch, A., Daelemans, W. & Hoste, V.: 2003, ‘Learning to Predict Pitch Accents and Prosodic Boundaries in Dutch’, *Proceedings of the 41st annual meeting of the association for computational linguistics*, 489-496.
- * De Meulder, F., Daelemans, W. & Hoste, V.: 2002, ‘A Named Entity Recognition System for Dutch’, *Language and Computers: Studies in Practical Linguistics*, 45, 77-88.

– C1

- * Vanopstal, K. & Hoste, V.: 2010, ‘Towards a learning approach for abbreviation detection and resolution’, *Proceedings of the International Conference on Language Resources and Evaluation*, Valletta, Malta.
- * Van Oosten, P., Tanghe, D. & Hoste, V.: 2010, ‘Towards an Improved Methodology for Automated Readability Prediction’, *Proceedings of the International Conference on Language Resources and Evaluation*, Valletta, Malta.
- * Desmet, B. & Hoste, V.: 2010, ‘Towards a Balanced Named Entity Corpus for Dutch’, *Proceedings of the International Conference on Language Resources and Evaluation*, Valletta, Malta.
- * Lefever, E. & Hoste, V.: 2010, ‘Construction of a Benchmark Data Set for Cross-lingual Word Sense Disambiguation’, *Proceedings of the International Conference on Language Resources and Evaluation*, Valletta, Malta.
- * Ineke Schuurman, I., Hoste, V. & Monachesi, P.: 2010, ‘Interacting semantic layers of annotation in SoNaR, a reference corpus of contemporary written Dutch’, *Proceedings of the International Conference on Language Resources and Evaluation*, Valletta, Malta.
- * Lefever, E., Macken, L. & Hoste, V.: 2009, ‘Language-independent bilingual terminology extraction from a multilingual parallel corpus’, *Proceedings of the European Chapter of the Annual Meeting of the Association of Computational Linguistics*. Athens, Greece, 496-504.
- * Delaere, I., Hoste, V., Peersman, C., Van Vaerenbergh, L. & Velaerts, P.: 2009, ‘ABOP, Automatic Optimization of Patient Information Leaflets’, *Proceedings of the International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages*, Besancon, France.
- * Schuurman, I., Hoste, V. & Monachesi, P.: 2009, ‘Cultivating Trees: Adding Several Semantic Layers to the Lassy Treebank in SoNaR’, *Proceedings of the 7th International Workshop on Treebanks and Linguistic Theories*, Groningen, The Netherlands, 135-146.
- * Lefever, E., Fayruzov, T., Hoste, V. & De Cock, M.: 2009, ‘Fuzzy Ants Clustering for Web People Search’, *Proceedings of the 2nd Web People Search Evaluation Workshop (WePS 2009)*, 18th WWW Conference, Madrid, Spain.

- * Lefever, E. & Hoste, V.: 2009, ‘SemEval-2010 Task 3: Cross-lingual Word Sense Disambiguation’, *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*, Boulder, Colorado, 82-87.
- * Hoste, V., Lefever, E., Vanopstal, K. & Delaere, I.: 2008, ‘Learning-based Detection of Scientific Terms in Patient Information’, *Proceedings of the sixth international conference on Language Resources and Evaluation*. Marrakech, Morocco.
- * Macken, L., Lefever, E. & Hoste, V.: 2008, ‘Linguistically-based sub-sentential alignment for terminology extraction from a bilingual automotive corpus’, *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, Manchester, United Kingdom, 529-536.
- * Fayruzov, T., De Cock, M., Cornelis, C. & Hoste, V.: 2008, ‘Exploiting Grammatical Relations for Protein Relation Extraction and Role Labeling’, *Proceedings of the Dutch-Belgian Information Retrieval workshop (DIR-2008)*, Maastricht, the Netherlands, 37-44.
- * Hendrickx, I., Bouma, G., Coppens, F., Daelemans, W., Hoste, V., Kloosterman, G., Mineur, A.-M., Van Der Vloet, J. & Vershelde, J.-L.: 2008, ‘Coreference Resolution for Extracting Answers for Dutch’, *Proceedings of the sixth international conference on Language Resources and Evaluation*, Marrakech, Morocco.
- * Lefever, E., Fayruzov, T. & Hoste, V.: 2007, ‘A combined classification and clustering approach for web people disambiguation’, *Proceedings of the 4th International Workshop on Semantic Evaluations*, Prague, Czech Republic, 105-108.
- * Hoste, V., Vanopstal, K. & Lefever, E.: 2007, ‘The Automatic Detection of Scientific Terms in Patient Information’, *Proceedings of the Workshop on Acquisition and Management of Multilingual Lexicons (in conjunction with RANLP-2007)*, Borovetz, Bulgaria, 41-48.
- * Hoste, V. & van den Bosch, A.: 2007, ‘A Modular Approach to Learning Dutch Co-reference Resolution’, *Proceedings from the first Bergen Workshop on Anaphora Resolution (WAR I)*, Mjolfell, Norway, 51- 75.
- * Hoste, V. & Daelemans, W.: 2006, ‘Comparing Learning Approaches to Language Learning. There is More to it Than Bias’, *Proceedings of Benelearn 2006*, Ghent, Belgium, 131-138.
- * Hoste, V. & De Pauw, G.: 2006, ‘KNACK-2002: a Richly Annotated Corpus of Dutch Written Text’, *Proceedings of the fifth international conference on Language Resources and Evaluation*, Genova, Italy.
- * Hoste, V. & Daelemans, W.: 2005, ‘Learning Dutch Coreference Resolution’, *Proceedings of the Fifteenth Computational Linguistics in the Netherlands Meeting (CLIN 2004)*.
- * Hoste, V. & Daelemans, W.: 2005, ‘Comparing Learning Approaches to Coreference Resolution. There is More to it Than ’Bias’’, *Proceedings of the Workshop on Meta-Learning (held in conjunction with ICML-2005)*, Bonn, Germany, 20-27.
- * Decadt, B., Hoste, V., Daelemans, W. & van den Bosch, A.: 2004, ‘GAMBL, Genetic Algorithm Optimization of Memory-Based WSD’, *Proceedings of SENSEVAL-3*, Barcelona, Spain, 108-112.
- * Hoste, V., Daelemans, W., Hendrickx, I. & van den Bosch, A.: 2002, ‘Evaluating the Results of a Memory-Based Word-Expert Approach to Unrestricted Word Sense Disambiguation’, *Proceedings of the Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, PA, USA, 95-101.
- * Daelemans, W. & Hoste, V.: 2002, ‘Evaluation of Machine Learning Methods for Natural Language Processing Tasks’, *Proceedings of the Third International Conference*

on Language Resources and Evaluation (LREC 2002), Las Palmas, Gran Canaria, 755-760.

- * Hendrickx, I., van den Bosch, A., Hoste, V. & Daelemans, W.: 2002, 'Dutch Word Sense Disambiguation: Optimizing the Localness of Context', *Proceedings of the Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, PA, USA, 61-66.
- * Marsi, E., Busser, B., Daelemans, W., Hoste, V., Reynaert, M., van den Bosch, A.: 2002, 'Combining Information Sources for Memory-Based Pitch Accent Placement', *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)*.
- * Hoste, V., Kool, A. & Daelemans, W.: 2001, 'Classifier Optimization and Combination in the English All Words Task', *Proceedings of SENSEVAL-2. Second International Workshop on Evaluating Word Sense Disambiguation Systems*, New Brunswick: ACL, 83-86.
- * Hoste, V., Daelemans, W., Tjong Kim Sang, E. & Gillis, S.: 2000, 'Meta-Learning for Phonemic Annotation of Corpora', *Proceedings of the 17th International Conference on Machine Learning*, San Francisco: Morgan Kaufmann, 375-382.
- * Hoste, V., Gillis, S. & Daelemans, W.: 2000, 'A Rule Induction Approach to Modeling Regional Pronunciation Variation', *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken, Germany, 327-333.
- * Hoste, V., Gillis, S. & Daelemans, W.: 2000, 'Machine learning for modeling Dutch pronunciation variation', *Computational Linguistics in the Netherlands 1999. Selected papers from the Tenth CLIN Meeting*, Utrecht, the Netherlands, 73-84.
- * Hoste, V., Daelemans, W., Tjong Kim Sang, E. & Gillis, S.: 2000, 'Meta-Learning for Phonemic Annotation of Corpora', *Proceedings of the Twelfth Belgium-Netherlands Artificial Intelligence Conference*, Tilburg, the Netherlands, 331-332.
- * Hoste, V. & Daelemans, W.: 2000, 'Comparing Bagging and Boosting for Natural Language Processing Tasks, a Typicality Approach', *Proceedings of the Tenth Belgian-Dutch Conference on Machine Learning*, Tilburg, the Netherlands, 101-109.

- Dissertation D1

Véronique Hoste, *Optimization Issues in Machine Learning of Coreference Resolution*, Ph.D. thesis, 246 pages, Antwerp University, 2005.

11 Lectures/conference contributions

- accepted abstract on 'A Machine Learning Approach to Phonemic Corpus Annotation' (with Walter Daelemans, Steven Gillis), Tenth Meeting of Computational Linguistics in the Netherlands, Utrecht, Netherlands, December 10, 1999.
- accepted paper on 'Meta-Learning for Phonemic Annotation of Corpora' (with Walter Daelemans, Steven Gillis), 17th International Conference on Machine Learning, Stanford, VS, 29 June - 2 July, 2000.
- accepted paper on 'A Rule Induction Approach to Modeling Regional Pronunciation Variation' (with Walter Daelemans, Steven Gillis), 18th International Conference on Computational Linguistics, Saarbrücken, Germany, 31 July - 4 August, 2000.

- accepted paper on ‘Comparing Bagging and Boosting for Natural Language Processing Tasks: A Typicality Approach’ (with Walter Daelemans, Steven Gillis), BNAIC, Tilburg, the Netherlands, November 3, 2000.
- accepted abstract on ‘Comparing Bagging and Boosting for Natural Language Processing Tasks: A Typicality Approach’ (with Walter Daelemans), Eleventh Meeting of Computational Linguistics in the Netherlands, Tilburg, the Netherlands, December 13, 2000.
- accepted abstract on ‘Memory-based all-words word-sense disambiguation for Dutch and English’ (with Walter Daelemans, Antal van den Bosch and Iris Hendrickx), Twelfth Meeting of Computational Linguistics in the Netherlands, Enschede, the Netherlands, November 30, 2001.
- accepted abstract on ‘Interaction between named-entity recognition and base NP chunking for Dutch (with Walter Daelemans and Fien De Meulder), Twelfth Meeting of Computational Linguistics in the Netherlands, Enschede, November 30, 2001.
- accepted abstract on ‘Optimization in Machine Learning of Natural Language’ (with Walter Daelemans), CLIF symposium, Brussels, April 25, 2003.
- ‘Limitations of comparative machine learning methodology. (with Walter Daelemans), ATILA meeting, Tilburg, The Netherlands, June 19, 2002.
- accepted paper on ‘Dutch Word Sense Disambiguation: Optimizing the Localness of Context. (with Iris Hendrickx, Walter Daelemans, Antal van den Bosch), ACL 2002 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions, Philadelphia, USA, July 11, 2002.
- accepted paper on ‘Evaluation of Machine Learning Methods for Natural Language Processing Tasks. (with Walter Daelemans), LREC 2002, Las Palmas de Gran Canaria, May 29, 2002.
- accepted paper on ‘Evaluating the results of a memory-based word-expert approach to unrestricted wsd. (with iris Hendrickx, Walter Daelemans, Antal van den Bosch), ACL 2002 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions, Philadelphia, USA, July 11, 2002.
- accepted abstract on ‘Limitations of comparative machine learning methodology for language technology (with Walter Daelemans), 13th Meeting of Computational Linguistics in the Netherlands, Groningen, The Netherlands, November 29, 2002.
- accepted abstract on ‘Predicting pitch accents and phrase breaks from text. (with Erwin Marsi, Martin Reynaert, Walter Daelemans, Antal van den Bosch), 13th Meeting of Computational Linguistics in the Netherlands, Groningen, The Netherlands, November 29, 2002.
- accepted paper on ‘Combined Optimization of Feature Selection and Algorithm Parameters in Machine Learning of Language’, (with Walter Daelemans, Fien De Meulder and Bart Naudts) 14th European Conference On Machine Learning, Cavtat, Croatia, 22 - 26 September, 2003.
- accepted paper on ‘GAMBL Genetic Algorithm Optimization of Memory-Based WSD’ (with Bart Decadt, Walter Daelemans, Antal van den Bosch), Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (Senseval- 3), Barcelona, Spain, July 26, 2004.
- accepted abstract on ‘Learning Dutch Coreference Resolution’ (with Walter Daelemans), Fifteenth Meeting of Computational Linguistics in the Netherlands, Leiden, the Netherlands, December 17, 2004.

- accepted paper on ‘Comparing Learning Approaches to Coreference Resolution. There is More to it Than Bias’ (with Walter Daelemans), Workshop on Meta-Learning, Bonn, Germany, August 11, 2005.
- accepted paper on ‘Comparing Learning Approaches to Learning Language. There is More to it Than Bias’ (with Walter Daelemans), Benelearn conference, Ghent, 11-12 May, 2006.
- accepted abstract on ‘The automatic resolution of “het” in a machine learning approach to Dutch coreference resolution’ (with Walter Daelemans and Iris Hendrickx), CLIN17, Leuven, January 12, 2007.
- accepted poster presentation on ‘A combined classification and clustering approach for web people disambiguation.’ (with Els Lefever, Timur Fayruzov), Workshop Machine Learning for Natural Language Processing, Amsterdam, The Netherlands, May 16, 2007.
- accepted paper on ‘Evaluating hybrid versus data-driven coreference resolution.’ (with Iris Hendrickx, Walter Daelemans), 6th Discourse Anaphora and Anaphor Resolution Colloquium, Lagos, Portugal, March 30, 2007.
- accepted paper on ‘Disambiguation of the neuter pronoun and its effect on pronominal coreference resolution’ (with Walter Daelemans and Iris Hendrickx), 10th International Conference TSD 2007, Plzen, Czech republic, September 2007.
- accepted paper on ‘The Automatic Detection of Scientific Terms in Patient Information’, Workshop on Acquisition and Management of Multilingual Lexicons (in conjunction with RANLP-2007), Borovetz, Bulgaria, October 1, 2007.
- accepted abstract on ‘COREA Wrap up: the corpus and evaluation’ (with Iris Hendrickx, Gosse Bouma, Frederik Coppens, Walter Daelemans, Anne-Marie Mineur, Joeri Van Der Vloet, Jean-Luc Vershelde), CLIN 2007, Nijmegen, The Netherlands, December 7, 2007.
- accepted abstract on ‘Learning-based Detection of Scientific Terms in Patient Information’ (with Els Lefever, Klaar Vanopstal), CLIN 2007, Nijmegen, The Netherlands, December 7, 2007.
- accepted paper on ‘Exploiting Grammatical Relations for Protein Relation Extraction and Role Labeling’ (with Timur Fayruzov, Martine De Cock and Chris Cornelis), DIR-2008, Maastricht, The Netherlands, April 14, 2008.
- accepted paper on ‘Coreference Resolution for Extracting Answers for Dutch’ (with Iris Hendrickx, Gosse Bouma, Frederik Coppens, Walter Daelemans, Geert Kloosterman, Anne-Marie Mineur, Joeri Van Der Vloet and Jean-Luc Vershelde), LREC-2008, Marrakech, Morocco, May 29, 2008.
- accepted poster on ‘Learning-based Detection of Scientific Terms in Patient Information’ (with Els Lefever, Klaar Vanopstal and Isabelle Delaere), LREC-2008, Marrakech, Morocco, May 28, 2008.
- accepted paper on ‘Linguistically-based sub-sentential alignment for terminology extraction from a bilingual automotive corpus’ (with Lieve Macken and Els Lefever), Coling 2008, Manchester, United Kingdom, August 18, 2008.
- accepted abstract on ‘ABOP: Automatic PIL Optimizer’ (with Isabelle Delaere), LTCI-2008 (Second AUGent Workshop on Language Technology and Computational Intelligence), Ghent, August 25, 2008.

- accepted abstract on ‘Language-independent bilingual terminology extraction from a multilingual parallel corpus’ (with Els Lefever and Lieve Macken), CLIN 2009, Groningen, The Netherlands, January 22, 2009.
- accepted paper on ‘Cultivating Trees: Adding Several Semantic Layers to the Lassy Treebank in SoNaR’ (with Ineke Schuurman and Paola Monachesi), International Workshop on Treebanks and Linguistic Theories, Groningen, The Netherlands, January 23, 2009.
- accepted paper on ‘Language-independent bilingual terminology extraction from a multilingual parallel corpus’ (with Els Lefever and Lieve Macken), Annual Meeting of the Association of Computational Linguistics. Athens, Greece, April 2, 2009.
- accepted paper on ‘Fuzzy Ants Clustering for Web People Search’ (with Els Lefever, Timur Fayruzov and Martine De Cock), Second Web People Search Evaluation Workshop, April 21, Madrid, Spain, 2009.
- accepted paper on ‘Task 3: Cross-lingual Word Sense Disambiguation’ (with Els Lefever), Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009) in collocation with NAACL-2009, June 4, Boulder, USA, 2009.
- accepted paper on ‘ABOP, Automatic Optimization of Patient Information Leaflets’ (with Isabelle Delaere, Claudia Peersman, Leona Van Vaerenbergh and Peter Velaerts), International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages, July 1st, Besancon, France, 2009.
- accepted paper on ‘Coreference Resolution on Blogs and Commented News’ (with Iris Hendrickx), 7th Discourse Anaphora and Anaphor Resolution Colloquium, November 5, Goa, India, 2009.
- accepted abstract on ‘Towards a Gold Standard for Readability Prediction’, ATILA-2009, November 9, Corsendonk, Belgium, 2009.
- abstract on ‘ABOP, an Automatic optimizer for Patient Information Leaflets’ (with Isabelle Delaere and Peter Velaerts), ATILA-2009, November 10, Corsendonk, Belgium, 2009.
- abstract on ‘SemEval-2: Cross-lingual Word Sense Disambiguation’ (with Els Lefever), ATILA-2009, November 10, Corsendonk, Belgium, 2009.
- abstract on ‘Coreference Resolution for Dutch Online Media Analysis’ (with Iris Hendrickx), LTCI-2009, September 14, Gent, Belgium, 2009.
- abstract on ‘Automatic Scoring of Text Readability’ (with Philp Van Oosten and Dries Tanghe), LTCI-2009, September 14, Gent, Belgium, 2009.
- abstract on ‘PRISE, a PRotein Interaction Search Engine’ (with Timur Fayruzov, Martine De Cock, Chris Cornelis, et al.) , LTCI-2009, September 14, Gent, Belgium, 2009.
- abstract on ‘Hybrid Shallow Parsing Based Information Extraction for Glycosylation Curation’ (with Violaine Pillet, Anne-Lise Veuthy and Walter Daelemans), BioTM workshop, May 10-11, Gent, Belgium, 2010.
- accepted paper on ‘Towards a learning approach for abbreviation detection and resolution’ (with Klaar Vanopstal and Bart Desmet), LREC-2010, May 19, Valletta, Malta, 2010.
- accepted paper on ‘Towards an Improved Methodology for Automated Readability Prediction’ (with Philip Van Oosten and Dries Tanghe), LREC-2010, May 19, Valletta, Malta, 2010.

- accepted paper on ‘Towards a Balanced Named Entity Corpus for Dutch’ (with Bart Desmet), LREC-2010, May 19, Valletta, Malta, 2010.
- accepted paper on ‘Construction of a Benchmark Data Set for Cross-lingual Word Sense Disambiguation’ (with Els Lefever), LREC-2010, May 20, Valletta, Malta, 2010.
- accepted paper on ‘Interacting semantic layers of annotation in SoNaR, a reference corpus of contemporary written Dutch’, LREC-2010, May 20, Valletta, Malta, 2010.

Invited

- ‘Learning Noun Phrase Coreference Resolution’, TKI Colloquium, Tilburg, the Netherlands, May 19, 2004. (invitation by Erwin Marsi)
- ‘Optimization issues in Machine Learning of Coreference Resolution’, Workshop on Anaphora Resolution (WAR), Mjolfjell, Norway, September 28, 2005. (invitation by Christer Johansson)
- ‘Optimization in machine learning of language. A focus on the task of coreference resolution’, ILPS seminar, University of Amsterdam, Amsterdam, the Netherlands, June 9, 2006. (invitation by David Ahn)
- ‘State-of-the-art translation technology: introduction and perspectives’, Telelingua Seminar, keynote speaker, Brussels, June 15, 2006. (invitation by Luc Van Haute)
- ‘Framing Discourse as a classification approach’, Machine Learning for NLP 2007, Amsterdam, the Netherlands, May 16, 2007. (invitation by Maarten van Someren and Sophia Katrenko)
- ‘NLP challenges in biomedical text mining’, VIB, Ghent, November 21, 2007. (invitation by Yvan Saeys)
- ‘Introduction into to the Foundations and Challenges of Text Data Mining’, LTCI-2008, August 25, 2008. (invitation by Martine de Cock and Patricia Victor)
- ‘Language independent bilingual terminology extraction from a parallel corpus. The PSA test case’, University of Wolverhampton, November 5, 2008. (invitation by Ruslan Mitkov)
- ‘Dutch coreference resolution. Issues and applications’, Corpus-Based Approaches to Coreference Resolution in Romance Languages (CBA-2008), November 14, 2008. (invitation by Toni Marti)
- ‘ABOP, Automatische Bijsluiter-Optimalisator’, Taal en Bedrijf, November 19, 2008. (invitation by Peter Spyns)
- ‘Taaltechnologie’, Rotary Deinze, February 1, 2010.
- ‘ABOP, Automatische BijsluiterOptimalisator’, workshop rond ‘Valorisatie van onderzoek in de HSW’. Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten, February 12, 2010. (invitation by Godelieve Laureys)