

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

- 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: 01/12/2007-30/11/2009.
Budget: 157,458 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: 26,400 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: 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: 01/05/2009-31/10/2009.

Budget: 35,097 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.
Period: 01/11/2006-31/10/2009.
- Private funding: Peugeot Citroen SA (LT3 as subcontractor of Telelingua international) “Étude doublons et synonymes thésaurus APV”.
Period: 01/02/2008-30/06/2009.
Budget: 54,828 euro.
- 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 and Hildegard Vermeiren.
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: 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.

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.

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)
- SemEval-2 task “Cross-Lingual Word Sense Disambiguation”,
<http://semeval2.fbk.eu/semeval2.php?location=tasks>.
(in cooperation with Els Lefever)

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:**
 - 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, 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)
 - ESSLI 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)
 - RANLP-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)
 - co-editor of *Linguistica Antverpiensia*, New Series (8/2009)
 - 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)
- **Reviewing for:**
 - Journal of Artificial Intelligence Research (2003, 2006, 2007)
 - Computational Linguistics in the Netherlands (2004, 2006, 2007)
 - Journal of Machine Learning Research (2006)
 - Research on Language and Computation (2007)
 - 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)

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).
- Philip 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, 1 December 2009)

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

- **2000**

- Véronique Hoste, Walter Daelemans, Erik Tjong Kim Sang en Steven Gillis. ‘Meta-Learning for Phonemic Annotation of Corpora.’ In: P. Langley (ed.) *Proceedings of the 17th International Conference on Machine Learning*, San Francisco: Morgan Kaufmann, 375-382, 2000.
- Véronique Hoste, Steven Gillis and Walter Daelemans. ‘A Rule Induction Approach to Modeling Regional Pronunciation Variation.’ *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken, Germany, 327-333, 2000.
- Walter Daelemans, Guy De Pauw, Gert Durieux, Steven Gillis, Véronique Hoste and Erik Tjong Kim Sang. ‘Zelflerende Systemen als Instrument voor de Taalkunde en de Taaltechnologie.’ In: Gillis, S., J. Nuyts, J. Taeldeman (eds.) *Met Taal om de Tuin geleid*, Antwerp, UIA, 81-94, 2000.
- Véronique Hoste, Steven Gillis and Walter Daelemans. ‘Machine learning for modeling Dutch pronunciation variation.’ In: P. Monachesi (ed.) *Computational Linguistics in the Netherlands 1999. Selected papers from the Tenth CLIN Meeting*, Utrecht, the Netherlands, 73-84, 2000.
- Véronique Hoste, Walter Daelemans, Erik Tjong Kim Sang en Steven Gillis. ‘Meta-Learning for Phonemic Annotation of Corpora.’ In: van den Bosch & Weigand (eds.) *Proceedings of the Twelfth Belgium-Netherlands Artificial Intelligence Conference*, Tilburg, the Netherlands, 331-332, 2000.
- Véronique Hoste and Walter Daelemans. ‘Comparing Bagging and Boosting for Natural Language Processing Tasks, a Typicality Approach.’ In: Feelders, A. (ed.) *BENELEARN 2000. Proceedings of the Tenth Belgian-Dutch Conference on Machine Learning*, Tilburg, the Netherlands, 101-109, 2000.

- **2001**

- Véronique Hoste and Guy De Pauw. *Proceedings of the Eleventh Belgian-Dutch Conference on Machine Learning (Benelearn 2001)*, Antwerp, 2001.
- Véronique Hoste, Anne Kool and Walter Daelemans. ‘Classifier Optimization and Combination in the English All Words Task.’ In: Judita Preiss and David Yarowsky (eds.) *Proceedings of SENSEVAL-2. Second International Workshop on Evaluating Word Sense Disambiguation Systems*. New Brunswick: ACL, 83-86, 2001.

- **2002**

- Iris Hendrickx, Antal van den Bosch, Véronique Hoste en Walter Daelemans. ‘Dutch Word Sense Disambiguation: Optimizing the Locality of Context.’ In: Phil Edmonds, Rada Mihalcea, Patrick Saint-Dizier (eds.) *Proceedings of the Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*. New Brunswick: ACL, 61-66, 2002.

- Véronique Hoste, Walter Daelemans, Iris Hendrickx en Antal van den Bosch. ‘Evaluating the Results of a Memory-Based Word-Expert Approach to Unrestricted Word Sense Disambiguation.’ In: Phil Edmonds, Rada Mihalcea, Patrick Saint-Dizier (eds.) *Proceedings of the Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*. New Brunswick: ACL, 95-101, 2002.
 - Fien De Meulder, Walter Daelemans and Véronique Hoste, ‘A Named Entity Recognition System for Dutch.’ In: M. Theune, A. Nijholt, and H. Hondrop (eds.) *Computational Linguistics in the Netherlands 2001. Selected Papers from the Twelfth CLIN Meeting*, Amsterdam - New York: Rodopi, 47-61, 2002.
 - Erwin Marsi, Bertjan Busser, Walter Daelemans, Véronique Hoste, Martin Reynaert and Antal van den Bosch. ‘Combining Information Sources for Memory-Based Pitch Accent Placement.’ *Proceedings of ICSLP-2002, International Conference on Spoken Language Processing*, Denver, USA, 1273-1276, 2002.
 - Walter Daelemans and Véronique Hoste. ‘Evaluation of Machine Learning Methods for Natural Language Processing Tasks’, *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, 755-760, Las Palmas, Spain, 2002.
 - Véronique Hoste, Iris Hendrickx, Walter Daelemans en Antal van den Bosch. ‘Parameter Optimization for Machine-Learning of Word Sense Disambiguation’, *Natural Language Engineering, Special Issue on Word Sense Disambiguation Systems*, 311-325, 2002.
- **2003**
 - Erwin Marsi, Martin Reynaert, Antal van den Bosch, Walter Daelemans and Véronique. ‘Learning to Predict Pitch Accents and Prosodic Boundaries in Dutch.’ *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 489-496, 2003.
 - Walter Daelemans, Véronique Hoste, Fien De Meulder en Bart Naudts. ‘Combined Optimization of Feature Selection and Algorithm Parameter Interaction in Machine Learning of Language’, *Lecture Notes in Artificial Intelligence 2837*, 84-95, 2003.
- **2004**
 - Véronique Hoste, Walter Daelemans and Steven Gillis. ‘Using Rule Induction Techniques to Model Pronunciation Variation in Dutch’, *Computer Speech and Language*, 18(1), 1-23, 2004.
 - Bart Decadt, Véronique Hoste, Walter Daelemans and Antal van den Bosch. ‘GAMBL, Genetic Algorithm Optimization of Memory-Based WSD’, *Proceedings of SENSEVAL-3*, Barcelona, Spain, 108-112, 2004.
 - Bart Decadt, Véronique Hoste and Guy De Pauw (eds.), *Proceedings of the Fourteenth Computational Linguistics in the Netherlands Meeting (CLIN 2003)*, 2004.
- **2005**
 - Véronique Hoste, *Optimization Issues in Machine Learning of Coreference Resolution*, Ph.D. thesis, 246 pages, Antwerp University, 2005.
 - Véronique Hoste and Walter Daelemans, ‘Comparing Learning Approaches to Coreference Resolution. There is More to it Than Bias’, *Proceedings of the Workshop on Meta-Learning (in conjunction with ICML-2005)*, Bonn, Germany, 20-27, 2005.
 - Véronique Hoste and Walter Daelemans, ‘Learning Dutch Coreference Resolution’, *Proceedings of the Fifteenth Computational Linguistics in the Netherlands Meeting (CLIN 2004)*, Leiden, the Netherlands, 2005.

- **2006**

- Véronique Hoste and Walter Daelemans, ‘Comparing Learning Approaches to Language Learning. There is More to it Than Bias’, *Proceedings of the Annual Machine Learning Conference of Belgium and the Netherlands (Benelearn 2006)*, Ghent, 131-138, 2006.
- Véronique Hoste and Guy De Pauw, ‘KNACK-2002: a Richly Annotated Corpus of Dutch Written Text’, *Proceedings of LREC 2006*, Genua, Italy, 2006.

- **2007**

- Iris Hendrickx, Véronique Hoste and Walter Daelemans, ‘Evaluating Hybrid versus Data-Driven Coreference Resolution’, *In: Anaphora: Analysis, Algorithms and Applications. Lecture Notes in Computer Science Volume 4410*, Berlin, Heidelberg: Springer Verlag, 137150, 2007.
- Els Lefever, Timur Fayruzov and Véronique Hoste, ‘A combined classification and clustering approach for web people disambiguation’, *Proceedings of the 4th International Workshop on Semantic Evaluations*, Prague, Czech Republic, 105-108, 2007.
- Véronique Hoste, Iris Hendrickx and Walter Daelemans, ‘Disambiguation of the neuter pronoun and its effect on pronominal coreference resolution’, *In: Text, Speech and Dialogue. Proceedings of the 10th International Conference TSD 2007*, Plzen, Czech Republic, Springer Lecture Notes in Computer Science Volume 4629, Berlin, Heidelberg: Springer, 4855, 2007.
- Els Lefever, Timur Fayruzov, and Véronique Hoste. A combined classification and clustering approach for web people disambiguation, *In: Maarten van Someren et al. (eds), Workshop Notes on Machine Learning for Natural Language Processing, Amsterdam, The Netherlands, 1415*, 2007.
- Véronique Hoste and Antal van den Bosch, *A Modular Approach to Learning Dutch Coreference*, *In: Christer Johansson (ed.), Proceedings from the first Bergen Workshop on Anaphora Resolution (WAR I)*, Cambridge Scholars Publishing, 51- 75, 2007.
- Véronique Hoste, Klaar Vanopstal and Els Lefever, ‘The Automatic Detection of Scientific Terms in Patient Information’, *Proceedings of the Workshop on Acquisition and Management of Multilingual Lexicons (in conjunction with RANLP)*, Borovetz, Bulgarije, 41-48, 2007.
- Véronique Hoste, Iris Hendrickx and Lieve Macken, ‘The referential versus non-referential use of the neuter pronoun in Dutch and English’, *Corpus Linguistics 2007*, Birmingham, UK, 2007.
- Iris Hendrickx, Véronique Hoste and Walter Daelemans, ‘The COREA-project: Evaluation Report of the Coreference Resolution Module, Deliverable STEVIN project COREA. 13 pages, 2007.
- Iris Hendrickx, Véronique Hoste and Walter Daelemans. ‘COREA, documentation of the coreference resolution application. Deliverable STEVIN project COREA. 11pages, 2007.
- Gosse Bouma, Walter Daelemans, Iris Hendrickx, Véronique Hoste and Anne-Marie Mineur. ‘The COREA-project. Manual for the annotation of coreference in Dutch texts’. Deliverable STEVIN project COREA. 41 pages, 2007.

- **2008**

- Iris Hendrickx, Véronique Hoste and Walter Daelemans, ‘Semantic and Syntactic features for Anaphora Resolution for Dutch’, *Springer Lecture Notes in Computer Science*, Volume 4919, Berlin: Springer Verlag, pp. 351-361, Proceedings of the CICLing-2008 conference, Haifa, Israel, 2008.

- Timur Fayruzov, Martine De Cock, Chris Cornelis and Véronique Hoste, ‘DEEPER: a Full Parsing based Approach to Protein Relation Extraction’, *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, Springer Lecture Notes in Computer Science, Volume 4973, Springer Berlin / Heidelberg, Naples, Italy, 2008.
- Timur Fayruzov, Martine De Cock, Chris Cornelis and Véronique Hoste, ‘Exploiting Grammatical Relations for Protein Relation Extraction and Role Labeling’, In: *Proceedings of the Dutch-Belgian Information Retrieval workshop (DIR-2008)*, Maastricht, the Netherlands, pp. 37-44, 2008.
- Eduard Hoenkamp, Martine De Cock and Véronique Hoste (eds.), *Proceedings of the Dutch-Belgian Information Retrieval workshop (DIR-2008)*, Maastricht, the Netherlands, 100 pages, 2008.
- Iris Hendrickx, Gosse Bouma, Frederik Coppens, Walter Daelemans, Véronique Hoste, Geert Kloosterman, Anne-Marie Mineur, Joeri Van Der Vloet and Jean-Luc Verschelde, ‘Coreference Resolution for Extracting Answers for Dutch’. In: *Proceedings of the sixth international conference on Language Resources and Evaluation*. Marrakech, Morocco, 2008.
- Véronique Hoste, Els Lefever, Klaar Vanopstal and Isabelle Delaere, ‘Learning-based Detection of Scientific Terms in Patient Information’. In: *Proceedings of the sixth international conference on Language Resources and Evaluation*. Marrakech, Morocco, 2008.
- Lieve Macken, Els Lefever and Véronique Hoste, ‘Linguistically-based sub-sentential alignment for terminology extraction from a bilingual automotive corpus’. In: *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, Manchester, United Kingdom, pp. 529-536, 2008.

• 2009

- Els Lefever, Lieve Macken and Véronique Hoste, ‘Language-independent bilingual terminology extraction from a multilingual parallel corpus’. In: *Proceedings of the European Chapter of the Annual Meeting of the Association of Computational Linguistics*. Athens, Greece, pp. 496-504, 2009.
- Ineke Schuurman, Véronique Hoste and Paola Monachesi. ‘Cultivating Trees: Adding Several Semantic Layers to the Lassy Treebank in SoNaR’. In: *Proceedings of the 7th International Workshop on Treebanks and Linguistic Theories*, Groningen, The Netherlands, 2009
- Els Lefever, Timur Fayruzov, Véronique Hoste and Martine De Cock. ‘Fuzzy Ants Clustering for Web People Search’. In: *Proceedings of the second Web People Search Evaluation Workshop (WePS 2009, in collocation with the 18th WWW Conference*, Madrid, Spain, 2009.
- Isabelle Delaere, Véronique Hoste, Claudia Peersman, Leona Van Vaerenbergh and Peter Velaerts, ‘ABOP, Automatic Optimization of Patient Information Leaflets’. In: *Proceedings of the International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages*, Besancon, France, to appear.
- Els Lefever and Véronique Hoste, ‘SemEval-2010 Task 3: Cross-lingual Word Sense Disambiguation’. In: *Proceedings of the Semantic Evaluations Workshop (SEW-2009, in collocation with NAACL-HLT 2009)*, Boulder, USA, 2009.
- Véronique Hoste, Klaar Vanopstal, Els Lefever and Isabelle Delaere, ‘Classification-Based Scientific Term Detection in Patient Information. *Terminology*, 16(1).
- Timur Fayruzov, Martine De Cock, Chris Cornelis and Véronique Hoste, ‘Linguistic Feature Analysis of Protein Interaction Extraction. *BMC Bioinformatics*, 10:374, 2009.

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 Verschelde), 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 Verschelde), 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.

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)