

CURRICULUM VITAE Prof. Veronique Hoste

➤ PERSONAL INFORMATION

Family name, First name: Hoste, Veronique

Researcher unique identifier (ORCID): <https://orcid.org/0000-0002-0539-4630>

Date of birth: 12.10.1972

Nationality: Belgian

URL for website: <https://www.lt3.ugent.be/people/veronique-hoste/>

➤ EDUCATION

- 2005 *Title of the dissertation*: Optimization Issues in Machine Learning of Coreference Resolution. Antwerp University, Belgium. PhD supervisor: Prof. Dr. Walter Daelemans
- 1998 Postgraduate degree GGS Taalwetenschap, interuniversity degree (RUG, UIA, KUL, VUB). Grade obtained: cum laude.
- 1997 Licentiate in translation (German, Portuguese, Danish). Grade obtained: magna cum laude. Higher Institute for Translators and Interpreters, Antwerp, Belgium

➤ CURRENT AND PREVIOUS POSITION(S)

- 2018 - : Research director of the Faculty of Arts and Philosophy, Ghent University.
- 2013 - : Head of the Department of Translation, Interpreting and Communication, Ghent University.
- 2013 - : Senior Full Professor in computational linguistics, Ghent University.
- 2006 - : Founder and Research Director of the LT3 (Language and Translation Technology Team) research group. See also: <https://www.lt3.ugent.be> for more information on the team, our publications, demos, etc.
- 2011-2013: Full Professor, University College Ghent.
- 2006-2011: Assistant Professor, University College Ghent.
- 2005-2006: Postdoctoral research assistant, Antwerp University, COREA, Coreference Resolution for Extracting Answers (STEVIN)
- 2005: postdoctoral research assistant, Antwerp University, BioMinT, Biological Text Mining (EU)
- 2001-2004: Research assistant, Antwerp University, PROSIT, Prosody from Information in Text (FWO-NWO).
- 1998-2000: Research assistant, Antwerp University, LINGUADUCT, An Inductive Machine Learning Environment for Corpus Annotation (FWO)

➤ FELLOWSHIPS AND AWARDS

- 2024 - ...: Member of the Royal Flemish Academy of Belgium for Science and the Arts.
- 2023-2024: Francqui chair at the Faculté de Lettres, Traduction et Communication, Université Libre de Bruxelles. Title of inaugural lecture: "Putting the L back in the limelight: a reappraisal of the linguistic perspective in NLP".
- 2010-2013: Visiting professor, Faculty of Arts and Philosophy, Ghent University.
- 2007-2013: Associate researcher, Department of Applied Mathematics and Computer Science, Ghent University. Co-founder of the association research group Language Technology and Computational Intelligence.
- 2006-2015: Visiting professor, Department of Linguistics, Antwerp University.

➤ SUPERVISION OF GRADUATE STUDENTS

Current PhD supervision as promotor or co-promotor: Marzieh Maleki (2024-2028), Nataliia Evgrafova (2023-2027), Serafina Van Geertruyen (2022-2026), Quang Du (2022-2026), Aaron Maladry (2021-2025), Sofie Labat (2021-2025), Sadegh Jafari (2025-2029)

Completed PhDs as supervisor or co-supervisor:

- Loic De Langhe (2024-09-13), "Demystifying discourse in Dutch: a study on event coreference resolution". Main supervisor. With Orphée de Clercq.
- Johnatan Bonilla (2024-05-27): "Universal dependencies for spoken Spanish". Main supervisor at Ugent. With Miriam Bouzouita (Humboldt University).
- Lore De Greve (2024-04-30): "The wisdom of the cloud : an aspect-based sentiment analysis of professional and online literary criticism through the prism of four literary prizes". Co-supervisor. With Gunther Martens, Daan Vandenhoute, Lars Bernaerts and Henk Roose.
- Rosa Lilia Segundo Díaz (2024-04-19): "Juegos del español : iterative design, evaluation and implementation of games with a purpose to enhance parts-of-speech tagging in a corpus of European Spanish dialects". Joint PhD with Hasselt University. Main supervisor at Ugent. With Karin Coninx, Gustavo Ruelo Ruiz and Miriam Bouzouita.
- Olha Kaminska (2023-09-01): "Fuzzy Rough Set Methods for Emotion Detection and Sentiment Analysis". Co-supervisor. With Chris Cornelis.
- Amir Hadifar (2023-06-23): "Towards Efficient NLP Models: Strategies for Improving Data Selection, Training, and Inference". Co-supervisor. With Chris Devellder and Thomas Demeester.
- Luna De Bruyne (2022-12-01): "Feeling EmotioNL: Automatically detecting emotions in Dutch texts". Main supervisor. Co-supervised by Orphée De Clercq.
- Gilles Jacobs (2021-12-09): "Extracting fine-grained events and sentiment from economic news". Main supervisor. Co-supervised by Els Lefever and Diane Breesch.
- Ayla Rigouts Terryn (2021-05-28): "D-TERMINE: data-driven term extraction methodologies investigated". Main Supervisor. Co-supervised by Els Lefever.
- Melissa Farasyn (2018-05-18): "Fitting in or standing out? Subject agreement phenomena in Middle Low German". Co-supervisor. With Anne Breitbarth.
- Arda Tezcan (2018-03-14): "Informative quality estimation of machine translation output". Co-supervisor. With Lieve Macken.
- Nils Smeuninx (2018-09-28): "Dear stakeholder: exploring the language of sustainability reporting : a closer look at readability, sentiment and perception." Co-supervisor. With Bernard De Clerck and Walter Aerts.
- Cynthia Van Hee (2017-09-22): "Can machines sense irony?: exploring automatic irony detection on social media." Main supervisor. Co-supervised by Els Lefever.
- Sabien Hanoulle (2017-07-05): "Translating documentaries : does the integration of a bilingual glossary of domain-specific terminology into the translation process reduce the translators' workload?" Co-supervisor. With Aline Remael.
- Orphée De Clercq (2015-07-09): "Tipping the scales: exploring the added value of deep semantic processing on readability prediction and sentiment analysis." Main supervisor. Co-supervised by Timothy Colleman.
- Bart Desmet (2014-10-17): "Finding the online cry for help : automatic text classification for suicide prevention". Main supervisor. Co-supervised by Johan De Caluwe.
- Els Lefever (2012-09-24): "ParaSense: parallel corpora for word sense disambiguation." Main supervisor. Co-supervised by Martine De Cock.
- Timur Fayruzov (2010-10-29): "Mining and modelling interaction networks for systems biology." Co-supervisor. With Martine De Cock and Chris Cornelis.

Since 2013: promotor of 3-5 master/manama theses per year.

➤ **TEACHING ACTIVITIES (last 5 years)**

When I started my appointment in the department of Translation, Interpreting and Communication, there were no courses integrating language or translation technology. Throughout the years, I took the initiative for different courses integrating technologies in the bachelor and master programmes, leading today to a new bachelor programme in language technology which is unique in Flanders.

- 2024-: New bachelor programme in language technology in the framework of the bachelor in applied linguistics (<https://www.ugent.be/lw/vtc/nl/actueel/nieuws/nieuw-traject-taaltechnologie>).
- 2022-: "Digitale tekstanalyse" (Digital text analytics). (Bachelor 1 in applied linguistics). In collaboration with Els Lefever.
- 2016- : "Digital Humanities (Master in translation, interpreting and communication). In collaboration with Els Lefever.
- 2016- : "Digital Communication" (Bachelor 3 in applied linguistics). In collaboration with Orphée De Clercq and Cynthia Van Hee/Sofie Labat.
- 2019- : "Natural language processing" (postgraduate CALM, Master of Advanced Studies in Linguistics, EMTTI). In collaboration with Cynthia Van Hee and Loic De Langhe.
- 2016-2024: "Project management" (postgraduate Computer-Assisted Language Mediation (CALM)¹, European Master in Technology for Translation and Interpreting (EMTTI)²). In collaboration with Cynthia Van Hee.
- 2020-2024: Local coordinator of the Erasmus Mundus EMTTI programme on Technologies for Translation and Interpreting with different NLP courses.
- 2016-2021: "Technical Writing" (postgraduate CALM). In collaboration with Joke Daems.

➤ **ORGANISATION OF SCIENTIFIC MEETINGS (last 5 years)**

- 2025 Organisation of the European Language Data Space Workshop for Belgium and the Netherlands (together with Carole Tiberius)
https://language-data-space.ec.europa.eu/events/lds-country-workshop-belgium-and-netherlands-2025-01-23_en
- 2024 Program chair for the 2024 LREC-COLING conference (Turin): <https://lrec-coling-2024.org/>
- 2024 Organization of the ECONLP – 4th Workshop on Economics and Natural Language Processing. Co-organization with the Joint Workshop of the 7th Financial Technology and Natural Language Processing (FinNLP) and the 5th Knowledge Discovery from Unstructured Data in Financial Services (KDF) at LREC-COLING 2024.
<https://lt3.ugent.be/econlp/>

¹ The CALM postgraduate programme prepares students for a career in the language, translation and communication industry. <https://studiekiezer.ugent.be/postgraduate-certificate-computer-assisted-language-mediation-en/2022>

² The EMTTI Erasmus Mundus programme addresses the need for a new generation of translators and interpreters who are up-to-date with the latest applications which support their daily work, as well as future developers of such applications. De two-year programme is offered by a consortium of the [University of Wolverhampton, UK](#) (the Coordinator); [University of Malaga, Spain](#); [New Bulgarian University, Bulgaria](#) and [Ghent University, Belgium](#). Erasmus Mundus is a European Union cooperation and mobility programme in the field of higher education. It aims to promote the European Union worldwide as a centre of excellence for learning and education. See: <https://em-tti.eu/>

- 2024 Co-organisation of the shared task on "Cross-lingual Classification of Corporate Social Responsibility (CSR) Themes and Topics". With Ecovadis.
<https://sites.google.com/view/csr-theme-recognition?pli=1>
- 2021 Organization of the ECONLP – 3rd Workshop on Economics and Natural Language Processing in collocation with the EMNLP conference (Hybrid online/Punta Cana). Co-organization with Udo Hahn and Amanda Stent.
- 2021 Organization of the WASSA workshop in collocation with the EACL conference (online). Co-organization with Orphee De Clercq, Alexandra Balahur, Joao Sedoc, Valentin Barriere, Shabnam Tafreshi, Sven Buechel.
<https://aclanthology.org/2021.wassa-1.0.pdf>
- 2019 Organization of the ECONLP - 2nd Workshop on Economics and Natural Language Processing in collocation with the EMNLP conference (Hong Kong). Co-organization with Udo Hahn and Zhu (Drew) Zhang.
<https://www.aclweb.org/anthology/D19-5100.pdf>
- 2019 Organization of the WASSA workshop in collocation with the NAACL conference (Minneapolis). Co-organization with Roman Klinger, Alexandra Balahur and Carlo Strapparava and Orphee De Clercq.
<https://www.aclweb.org/anthology/W19-1300.pdf>
- 2011/2013/
2016/2018 Organization of SemEval international evaluation exercises in 2011, 2013, 2016 and 2018 to build benchmark datasets on which research teams world-wide evaluate their systems. In 2011 and 2013, the focus was on word sense disambiguation, in 2016 on multilingual ABSA and in 2018, the task consisted in automatic irony detection (70 participating teams).
- 2018 Organization of the WASSA workshop in collocation with the EMNLP conference (Brussels). Co-organization with Alexandra Balahur, Saif Mohammad, and Roman Klinger.
<https://www.aclweb.org/anthology/volumes/W18-62/>
- 2018 Organization of the ECONLP - 1st Workshop on Economics and Natural Language Processing in collocation with the EMNLP conference (Melbourne). Co-organization with Udo Hahn and Ming-Feng Tsai.
<https://www.aclweb.org/anthology/W18-3100/>
- 2015/2018/2021 Organization of the first, second and third European Language Resource Coordination (ELRC) workshop in Belgium. In collaboration with Stijn Desmeyere and Ayla Rigouts Terryn.

➤ COMMISSIONS OF TRUST

- Evaluator of the PhD thesis:

Anders Noklestad ("A Machine Learning Approach to Anaphora Resolution Including Named Entity Recognition, PP Attachment Disambiguation, and Animacy Detection"). University of Oslo, 2009. // *Oier Lopez de Lacalle* ("Domain-specific Word Sense Disambiguation"), Universidad del País Vasco, 2009. // *Stefan Raaijmakers* ("Multinomial Language Learning. Investigations into the geometry of language"), University of Tilburg, 2009. // *Lieve Macken* ("Sub-sentential Alignment in Translational Correspondences"), Antwerp University, 2010. // *Kim Luyckx* ("Scalability Issues in Authorship Attribution"), Antwerp University, 2010. // *Alexandra Balahur Dobrescu* ("Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types"), Universidad de Alicante, 2011. // *Sofie Van Landeghem* ("Playing hide and seek on the genomic playground: Unveiling biological function from literature"), Ghent University, 2012. // *Emili Sapena Masip* ("A constraint-based hypergraph partitioning approach to coreference resolution"), University of Barcelona, 2012. // *Klaar Vanopstal* ("Impact of language skills and system experience on medical information retrieval"), Ghent

University, 2013. // *German Hurtado Martin* ("Intelligent Methods for Information Filtering of Research Resources"), Ghent University, 2013. // *Maral Dadvar* ("Experts and Machines United Against Cyberbullying"), University of Twente, 2014. // *Frederik Vaassen* ("Measuring emotion. Exploring the feasibility of automatically classifying emotional text"), Antwerp University, 2014. // *Pawel Mander* ("Psycholinguistics on a large scale: combining text corpora, megastudies and distributional semantics to investigate human language processing"). Doctor in Psychology. Faculty of Psychology and Educational Sciences, Ghent University, 2016. Supervised by Prof. Marc Brysbaert and dr. Emmanuel Keuleers. // *Joke Daems* ("A translation robot for each translator? A comparative study of manual translation and post-editing of machine translations: process, quality and translator attitude"), Doctor in Translation Studies, Faculty of Arts and Philosophy, Ghent University, 2016. Supervised by Prof. Lieve Macken and Prof. Sonia Vandepitte. // *Janneke van de Loo* ("Machine learning approaches to modeling language use of individuals and demographic groups. Applications in societally relevant natural language processing"), Doctor in Linguistics, Faculty of Arts and Humanities, Antwerp University. Supervised by dr. Guy De Pauw and Prof. Walter Daelemans. // *Golnoosh Farnadi* ("User Modeling in Social Media"), Doctor in Computer Science, Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 2017. Joint PhD with KULeuven. Supervised by Prof. Martine De Cock (UGent) and Prof. Marie-Francine Moens (KULeuven). // *Lea Canales Zaragoza* ("Tackling the Challenge of Emotion Annotation in Text"), Doctor in Informatics, Department of Languages and Information Systems, University of Alicante, 2018. Supervised by Prof. Patricio Martinez Barco. // *Friedel Wolff* ("Skoner en kleiner vertaalgeheues (Cleaner and smaller translation memories)"), Doctor in Computer Science. School of Computing, University of South Africa, 2018. Supervised by Prof. Laurette Pretorius and Prof. Paul Buitelaar. // *Lucas Sterckx* ("Methods for Efficient Supervision in Natural Language Processing"), Doctor of Computer Science Engineering, Faculty of Engineering and Architecture, Ghent University, 2018. Supervised by Prof. Chris Develder and Prof. Thomas Demeester. // *Lionel Pigou* ("Gesture and Sign Language Recognition with Deep Learning"), Doctor of Computer Science Engineering, Faculty of Engineering and Architecture, Ghent University, 2018. Supervised by Prof. Joni Dambre and Prof. Mieke Van Herreweghe. // *Suzanne Kleijn* ("Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing"), Faculty of Humanities, Utrecht University, 2019. Supervised by Ted Sanders and Henk Pander Maat. // *Julie Mennes* ("Facilitating crossdisciplinarity Joining forces of philosophy of science and computational linguistics"), Doctor in Philosophy and Linguistics. Faculty of Arts and Philosophy, Ghent University, 2019. Supervised by Prof. Erik Weber and Prof. Els Lefever. // *Ioannis Bekoulis* ("Neural Approaches to Sequence Labeling for Information Extraction"), Doctor of Computer Science Engineering, Faculty of Engineering and Architecture, Ghent University, 2019. Supervised by Prof. Chris Develder and Prof. Thomas Demeester. // *Marten Postma* ("The Meaning of Word Sense Disambiguation Research"), Faculty of Humanities, Free University of Amsterdam, 2019. Supervised by Prof. Piek Vossen and Prof. Jacopo Urbani. // *Frederic Godin* ("Improving and Interpreting Neural Networks for Word-Level Prediction Tasks in Natural Language Processing"), Doctor of Computer Science Engineering, Faculty of Engineering and Architecture, Ghent University, 2019. Supervised by Prof. Wesley De Neve and Prof. Joni Dambre. // *Alessandra Teresa Cignarella* ("Dependency Syntax in the Automatic Detection of Irony and Stance"), Doctor in Computer Science, Joint PhD between the Universities of Turin and Valencia, 2021. Supervised by Prof. Cristina Bosco and Prof. Paolo Rosso. // *Akbar Karimi* ("Text Analysis with Deep Learning and Data Augmentation"), Doctor in Information Technology, University of Parma, 2022. Supervised by Prof. Andrea Prati. // *Klim Zaporozets* ("Neural approaches to entity-centric information extraction"), Doctor of Computer Science Engineering, Faculty of Engineering and Architecture, Ghent University, 2022. Supervised by Prof. Chris Develder and Prof. Thomas Demeester. // *Tulika Bose* ("Transfer Learning for Abusive Language Detection"), Docteur Mention Informatique, Université de Lorraine, 2023. Supervised by Prof. Irina Illina and dr. Dominique Fohr. // *Robbert De Troij* ("Natiolectal variation in Dutch grammar: A data-driven approach"), Joint Doctorate Degree between Radboud University and the Catholic University of Leuven. Supervised by Prof. Antal van den Bosch, Prof. Dirk Speelman, Prof. Stefan Grondelaers. // *Maxime De Bruyn* ("Studies in Conversational AI: Multilingual Capabilities, World Knowledge, and Evaluation Strategies."), Doctor in de Taalkunde, Universiteit Antwerpen,

2024. Supervised by Walter Daelemans // Andries Pieter Wolfert ("Co-speech gestures: Evaluation and Generation"), Doctor in computer science engineering. Supervised by Tony Belpaeme and Francis Wyffels.
- Reviewer of full papers for the major NLP conferences (ACL, EACL, NAACL, COLING, LREC, EMNLP, etc.).
Reviewer of full papers or member editorial board: Expert Systems with Applications, NLE, LRE, ...).
Senior Area chair for EMNLP 2021. Senior Area chair for EACL 2023. Senior Area chair for EMNLP 2024.
 - Frequent evaluation of project proposals (VLAIO, Swedish Research Foundation, Swiss Research Foundation, NWO, Narodowe Centrum Nauki Poland, Academia Sinica Taiwan, Slovak Academy of Sciences, FNRS, Research Council of Finland, ...)
 - Evaluation of teaching programmes: ICT en Vertalen (VU); opleidingscluster Taal- en letterkunde (Uantwerpen).
 - Advisory board member of the Institute for the Dutch Language (2016-2024).
 - Member of the advisory board of EACL (<https://eacL.org/general/>) (since 2023)
 - Ghent University:
 - Head of the Department of Translation, Interpreting and Communication (since 2013).
 - Member of the education commission of all teaching programmes in the Department of Translation, Interpreting and Communication (since 2013).
 - Faculty representative in the university ICT user commission (since 2013)
 - Chair of the Commission of Scientific Research of the Faculty of Arts and Philosophy (as Research Director of the Faculty of Arts and Philosophy (since 2018)).
 - Member of the Ethics committee of the Faculty of Arts and Philosophy, Ghent University (since 2019).
 - Member of the evaluation committee of the Industrial Research Fund, Ghent University (since 2022).
 - Member of the HR commissions of all professors in the departments of Translation, Interpreting and Communication, Linguistics and Archaeology.³

➤ **OVERVIEW OF ONGOING/CONCLUDED RESEARCH GRANTS**

ONGOING RESEARCH GRANTS

- "Parsed Corpus of Spoken Dutch Dialects+" (GCND+). Period: 01/04/2024 – 30/04/2027. Budget: 710.114,62 euro. Co-supervisor. (FWO medium scale research infrastructure).
- Grand Challenge 2, WP on interaction and cognitive systems. Period: 2024 – 2029. Budget: 48 person months. PhD student: Sadegh Jafari. (FAIR 2, Flemish AI research programme from the Department of Economy, Science and Innovation).
- "Modeling emotion trajectories in customer service dialogues" (METRICS). Period: 01/11/2021 – 30/10/2025. Budget: 48 person months. PhD student: Sofie Labat. Main supervisor. (FWO-SB - strategic research funding channel of the Research Foundation Flanders).
- "Modeling common-sense knowledge in automatic irony detection". Period: 01/10/2021 – 30/09/2025. Budget: € 209,000. PhD student: Aaron Maladry. Main supervisor. (BOF - Ghent University research fund).
- "CLARIAH-VL, Advancing the Open Humanities Service Infrastructure (CLARIAH-VL is the Flemish contribution to the European research infrastructures DARIAH and CLARIN". Period: 2021-2024. Budget: € 3,621,431 (211.000 for NLP research at UGent). Co-supervisor. (FWO-ESFRI Research Foundation Flanders - European Strategy Forum on Research Infrastructures).
- "A multilingual, learning-based platform for aspect-based sentiment and emotion analysis." Period: 01/01/2021 – 31/12/2022. Budget: € 370,191.12. Researcher: Ellen De Geyndt. Main supervisor. (VLAIO TETRA

³ Since 2020, Ghent University has a new career model for its professors, focusing on a more sustainable guidance of the colleague professor during his/her academic career. Each professor is assigned a HR commission which is composed of faculty board members, peers and a HR consultant.

- funding channel for application-oriented projects from the Flanders agency on Innovation and Entrepreneurship).
- "ENCORE: Investigating Cross-Document Event Coreference Resolution for Dutch". Period: 01/09/2020 – 31/08/2024. Budget: € 243,606. PhD student: Loic De Langhe. Main supervisor. (FWO Research Foundation Flanders).
- "Grand challenge 4, WP on Conversational Agents". Period: 2019 – 2023. Budget: 48 person months. PhD students: Sofie Labat (main supervisor- until 31/10/2021; Sofie started as a PhD candidate with a FWO-SB scholarship starting 01/11/2021), Klim Zaporozets (co-supervisor), Quanqi Du (main supervisor - as from 01/05/2022). (Flemish AI research programme from the Department of Economy, Science and Innovation).
- "A respeaking and collaborative game-based approach to building a parsed corpus of European Spanish dialects". Period: 01/01/2019 – 31/12/2024. Budget: Ca. 300,000 euro. PhD students: Johnatan Bonilla (co-financing Colombian government), Rosa Lilia Segundo Diaz (co-financing Hasselt University). Main supervisor. (FWO medium scale research infrastructure).

COMPLETED RESEARCH GRANTS

- "NewsDNA. Diversity in the news through algorithmization". Period: 01/02/2018 - 31/10/2022. Budget: 1 PhD student 48 months + 1 postdoc (24 months). Researchers: Camiel Colruyt, Cynthia Van Hee, Thierry Desot. Copromotor. (BOF-GOA Ghent University research fund – concerted research actions)
- "EmotionNL: Emotion detection for Dutch". Period: 01/01/2019 – 31/12/2022. Budget: 48 person months. PhD student: Luna De Bruyne. Main supervisor. (FWO SB strategic research funding channel of the Research Foundation Flanders).
- "Comprehensive Feature Integration in a Human Inspired Speech architecture". Period: June 2017 – 2022. Budget: two times 48 person months. PhD students: Claudia Matos Veliz (main supervisor), Geoffroy Vanderreydt (co-supervisor). (BOF-IOP University research fund – Interdisciplinary research projects)
- "EXTRACT. Extraction of terminology from comparable texts". Period: 2017-2021. FWO Fellowship Ayla Rigouts Terryn (FWO - Research Foundation Flanders). Co-supervisor.
- "SENTiVENT. Event extraction and sentiment analysis for financial applications". Period: 2017-2021. SB fellowship Gilles Jacobs. (FWO - Research Foundation Flanders)
- "A parsed corpus of Historical Low German", Period: 2014-2018. Budget: 2 PhD students (FWO medium scale research infrastructure)
- "SCATE, Smart Computer-Aided Translation Environment", in cooperation with the Universities of Leuven and Hasselt. Period: 01/01/2014 - 31/12/2017. Budget LT3: 519,250 euro (FWO-SB0, Research Foundation Flanders strategic basic research funding channel)
- "AMiCA, Automatic Monitoring for Cyberspace Applications", in cooperation with the Universities of Antwerp (CLiPS), Ghent (IBCN research group) and Leuven (VISICS research group). Period: 01/01/2013 - 31/12/2016. Budget LT3: 793,823 euro (FWO-SB0, Research Foundation Flanders strategic basic research funding channel)
- "SentiFM, Sentiment Mining for the Financial Markets". Period: 01/01/2012 - 31/12/2015. Budget LT3: PhD project of Marjan Van de Kauter, supervised by Veronique Hoste and Diane Breesch (IWT – Agency for Innovation through Science and Technology)
- "PARIS, Personalised Advertisements built from web Sources", in cooperation with K.U.Leuven (LIIR, HDMB and VISICS research groups), Ghent University (LTCl research group), Antwerp University (IELP research group) and VLERICK. Business school. Period: 01/05/2012 - 30/04/2016. Budget LT3: 221,205 euro. (FWO-SB0, Research Foundation Flanders strategic basic research funding channel)

- "SUBTLE, Subjectivity Tagging and Learning", in cooperation with Ghent University (Johan De Caluwe) and the Suicide Prevention Centre. Period: 01/10/2010 - 30/09/2014. Budget: 1 PhD student (Bart Desmet) (Research Fund, University College Ghent)
- "MuST: Multilingual Corpora for the automatic Structuring of Terms". Period: 01/01/2013- 31/12/2014. Budget: postdoc project of Els Lefever.
- "TExSIS, Terminology Extraction for Semantic Interoperability and Standardization". Coördinator. In cooperation with Ghent University (IBCN research group) and with the financial support of a user group. Period: 01/01/2011 - 31/12/2012. Budget: 392,149 euro. (IWT - Agency for Innovation through Science and Technology)
- "HENDI, een Automatische Helderheidstool voor Engels en Nederlands Discours in Parallelle en Vergelijkbare Teksten" ("HENDI, an Automatic Clarity Tool for English and Dutch Discourse in Parallel and Comparable Texts"). Period: 01/10/2008 - 31/10/2012. Budget LT3: 335,785 euro. (Research Fund, University College Ghent)
- "ParaSense or how to use parallel corpora for Word Sense Disambiguation", in cooperation with Ghent University (Martine De Cock). Period: 01/01/2007-31/12/2012. Budget: one PhD student (Els Lefever).
- "Koppeling van schrijfprocesdata aan lexica" (Merging writing process data with lexica), in cooperation with Antwerp University (Luuk Van Waes). Period: 01/01/2010 - 31/12/2012. Budget LT3: 11,700 euro (FWO - Research Foundation Flanders).
- "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 Antwerp University (CLIPS). Period: 01/01/2010 - 30/09/2012. Budget: 120,000 euro. (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: 01/01/2010 - 30/09/2012. Budget LT3: 50,000 euro. (Department of Economy, Science and Innovation, Flemish government)
- "OptiPIL, a readability Optimizer for Patient Information Leaflets, in cooperation with Artesis University College Antwerp. Period: 01/01/2010-30/12/2010. Budget LT3: 49,980 euro. (IOF - Industrial Research Fund, Ghent University)
- "ABOP, Automatische Bijsluiter Optimalisator. Een auteursomgeving voor de correctie en productie van bijsluiterteksten" (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. (IWT - Agency for Innovation through Science and Technology)
- "SoNaR, Stevin Nederlandstalig Referentiecorpus." (SoNaR⁴, 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 + 26,400 (Phase 1). (Stimulation programme for the development of resources for Dutch language and speech technology)
- "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. (Stimulation programme for the development of resources for Dutch language and speech technology)

⁴ The SoNaR project was a project financed by the Dutch-Flemish STEVIN programme. In the project, at least 500 million words of written Dutch were collected; 1 million words were manually annotated with syntactic and semantic information (named entities, semantic roles and coreference).

- "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. (Special Research Fund, Ghent University)
- "Etude doublons et synonymes thésaurus APV". Period: 01/02/2008-30/06/2009. Budget: 54,828 euro (Private funding: Peugeot Citroen SA (LT3 as subcontractor of Telelingua international))
- Server for data storage and computationally expensive processing. Budget: 21,267 euro. (Research Fund, University College Ghent)
- "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. (Special Research Fund, Antwerp University)

➤ **IMPACT – resources (data and software):**

Natural language processing in the last decades has drastically shifted from a discipline in which linguistic rules were handcrafted by experts to a discipline heavily dominated by machine learning and more recently, end-to-end neural approaches. I truly believe it is our duty as computational linguists to exploit our interdisciplinary, and thus our linguistic expertise to contribute to the domain by developing novel methodologies starting from the state-of-the-art in NLP, but also by developing resources, i.e. real-life and high-quality curated data which can serve as the basis for the investigation and development of NLP models. The 500-million word corpus of contemporary written Dutch, for which we collected SMS data from people all over Flanders and the Netherlands (through a radio and television media campaign) and on the basis of which we developed tools for Dutch named-entity recognition, semantic role labelling and coreference resolution (all available through CLARIN⁵) is a key resource in Dutch NLP research and applications. Furthermore, I also co-organized different shared tasks in the framework of the the SemEval international evaluation exercises, viz. on coreference resolution (2010), cross-lingual word-sense disambiguation (2010, 2013), L2 writing (2014), multilingual aspect-based sentiment analysis (2016), irony detection (2018) and corporate social responsibility classification (2024). The data and papers for these tasks have been downloaded and widely referenced exhibiting the great interest in methodologically sound dataset construction. The annotation guidelines for the data construction were written down in LT3 technical reports, e.g.

- Labat, Sofie, Thomas Demeester, and Veronique Hoste. 2020. "Guidelines for Annotating Fine-Grained Emotion Trajectories in Customer Service Dialogues (Version 1.0)." *LT3 Technical Report – LT3 20-01*. Ghent: LT3, Faculty of Arts, Humanities and Law, Ghent University (Ghent, Belgium).
- Colruyt, Camiel, Orphée De Clercq, and Veronique Hoste. 2019. "EventDNA : Guidelines for Entities and Events in Dutch News Texts (v1.0)." *LT3 Technical Report - LT3 19-01*.
- Van Hee, Cynthia, Els Lefever, and Veronique Hoste. 2016. "Guidelines for Annotating Irony in Social Media Text, Version 2.0." *LT3 Technical Report Series*.
- Van Hee, Cynthia, Ben Verhoeven, Els Lefever, Guy De Pauw, Walter Daelemans, and Veronique Hoste. 2015. "Guidelines for the Fine-Grained Analysis of Cyberbullying, Version 1.0." *LT3 Technical Report Series*.

Apart from data construction, we also invested in the release of software with a wide applicability. Some demo versions of our software are available on the LT3 website: <https://lt3.ugent.be/demos/>. All recently published academic papers are accompanied with the underlying code.

⁵ The corpus is available through the Institute for the Dutch Language (INT), which is part of the European CLARIN (Common Language Resources and Technology Infrastructure). The goal of this European programme is to offer researchers in the humanities a stable, permanent, accessible and extendable research infrastructure.

➤ **IMPACT – beyond academia**

Apart from research projects with a wide variety of scientific partners in and beyond Flanders, I'm also very keen to set up research projects in collaboration with **companies or non-profit organizations**:

- Projects with company cofinancing:
 - A research valorization project on improving the readability of patient information leaflets in collaboration with pharmaceutical and other partners (funded by IWT - Agency for Innovation through Science and Technology)
 - A research valorization project on both monolingual and multilingual terminology extraction cooperation with Flemish companies from a variety of sectors i.e. Jan De Nul (dredging), Telenet (telecom), Mediargus (publisher). (funded by IWT - Agency for Innovation through Science and Technology)
 - A research valorization project on aspect-based sentiment and emotion analysis in collaboration with Flemish companies such as Telenet (telecom), Infrabel (public transportation), Cronos (software), Suazio (consultancy). See sentemo.org. (funded by VLAIO - Flanders agency on Innovation and Entrepreneurship)
- LT³ is also frequently a partner in SME innovation projects: LT³ collaborated with the company TheMatchBox in the framework of a SME feasibility study (VLAIO) on terminology and concept extraction for vacancy and cv parsing. LT³ also developed an aspect-based sentiment analysis (ABSA) system for the company Hello Customer in an SME innovation project funded by VLAIO. This has led to the delivery of both a lexicon- and learning-based system. A commercial license was also sold (to be renewed on a yearly basis) for the Dutch part of the LeTS preprocessing pipeline.
- Consultancy: e.g. multilingual terminology extraction for Peugeot Citroen (result: term banks), medical terminology extraction for EBMPracticeNet (result: medical term banks), hate speech detection for MetaChat to monitor social media reactions to Flemish news content (result: software).
- Societally relevant collaboration:
 - Cooperation with the Belgian Suicide Prevention Centre (CPZ) in the framework of the SubTLe and AMiCA projects (see infra) on suicidality detection. LT³ has closely worked together with the CPZ for data collection, annotation and subsequent system development. This has led to the development of a prototype suicidality monitoring system and an API co-funded by ING and Rotary.
 - Project with the city library of Bruges for sentiment and emotion analysis on the letters and poems of Guido Gezelle⁶. As our systems are mainly trained on social media language, this will also involve a "domain adaptation". For doing so, we developed a platform in which the larger crowd will be asked to correct the system results in order to optimize the system.
 - A similar "citizen science" approach is also taken in the "Burgerpraat!" project (2019-2022; <http://burgerpraat.be>) in collaboration with Kris Demuyne. The goal of this project is to engage Flemish citizens in collecting and annotating personal video data along with personality data and to inform the broader public on the capabilities of natural language processing.

➤ **Communication** of research results to the broader public:

⁶ Guido Gezelle (1830-1899) is considered one of the most important poets in the history of Dutch literature.

Press campaigns for data donation: e.g. in the AMiCA project a press campaign was set up by to motivate people to donate their cyberbullying data with large coverage in the written press, on radio and in the television news. A similar effort was done in the framework of the 500-million word corpus of Dutch written text (SoNaR) for the collection of SMS messages, which yielded a massive donation of messages showcasing the large interest of the general public in research in natural language processing and AI.

Apart from that, I also frequently give presentations on NLP for all types of organizations (e.g. Rotary, Alforgrowth, etc.), contributed to the “Z into AI” programme⁷ on the Flemish television channel Kanaal Z, ...

- **Efforts to also intergrate language technology in secondary school programmes:** I participated in the “AI at school” project (see: <https://www.aiopschool.be/chatbot/>), took part in panels on AI in education (organisation by het Talemlatform, Vlaams Parlement), collaborated on the development of new learning outcomes for language technology in secondary school programmes, give frequent tutorials and workshops to teachers, ...

Recent book publication: “Taaltechnologie ontrafeld. Hoe taal en technologie hand in hand gaan” (editor together with Cynthia Van Hee). This is the first book publication on language technology in Dutch, targeted at a non-expert audience. See: <https://www.pelckmansuitgevers.be/taaltechnologie-ontrafeld.html>. The book will be used in different introductory courses on NLP.

- **Invited presentations⁸**

Frequent presentations at the major NLP conferences and collocated workshops (EMNLP, NAACL, EACL, ACL, LREC). Invited talks on coreference resolution (ILPS seminar 2006, DAARC 2009, CBA 2008), terminology extraction and lexical semantics (RIILP 2008, CENTAL 2011), cyberbullying and suicidality detection on social media (Twente seminar on Cyberbullying 2014, RANLP 2017), sentiment analysis (AI4growth2022).

Keynotes at conferences: CLIC-it 2020, EcoFIN 2022 (“Detecting Explicit and Implicit sentiment evoked by fine-grained economic news events”), JEP-TALN-RECITAL 2024 (“Towards a Reappraisal of the Linguistic Perspective in the LLM Era”), AitLA 2025 (“The different faces of emotion: Navigating complex emotions in natural language and across modalities”).

Lecture series in the framework of the Francqui chair: <https://ltc.ulb.be/la-faculte-de-lettres-traduction-et-communication-accueil-veronique-hoste-de-lugent-pour-la-chaire-francqui-2023-2024>.

Invited organization of tutorials: Digital Humanities Summer school (2014, 2020, 2023), RANLP (readability prediction).

- **PEER-REVIEWED PUBLICATIONS (full papers)**

Du, Q., & Hoste, V. (2024). A Multi-task Framework with Enhanced Hierarchical Attention for Sentiment Analysis on Classical Chinese Poetry: Utilizing Information from Short Lines. In M. Härmäläinen, E. Öhman, S. Miyagawa, K. Alnajjar, & Y. Bizzoni (Eds.), *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities* (pp. 113–122). Miami, USA: Association for Computational Linguistics.

⁷ https://kanaalz.knack.be/business-communities/z-into-ai/groupement-normal-1664553.html?cookie_check=1650173507

⁸ Abbreviations: ILPS = Information and Language Processing Systems; CBA = Broadband Communications Systems and Architectures Research Group; RIILP = Research Institute of Information and Language Processing; CENTAL = Center for Natural Language Processing, UCLouvain; RANLP = Recent Advances in Natural Language Processing conference; CLIC-it = Italian Conference on Computational Linguistics.

Van Hee, C., & Hoste, V. (Eds.). (2024). *Taaltechnologie ontrafeld*. Kalmthout: Pelckmans.

Nayekoo, Y., Katrenko, S., Hoste, V., Maladry, A., & Lefever, E. (2024). Shared Task for Cross-lingual Classification of Corporate Social Responsibility ({CSR}) Themes and Topics. In C.-C. Chen, X. Liu, U. Hahn, A. Nourbakhsh, Z. Ma, C. Smiley, ... H.-H. Chen (Eds.), *Proceedings of the Joint Workshop of the 7th Financial Technology and Natural Language Processing, the 5th Knowledge Discovery from Unstructured Data in Financial Services, and the 4th Workshop on Economics and Natural Language Processing @ LREC-COLING 2024* (pp. 283–291). Torino, Italia: ELRA and ICCL.

Evgrafova, N., Hoste, V., & Lefever, E. (2024). Analysing pathos in user-generated argumentative text. In H. Afli, H. Bouamor, C. Blasi Casagran, & S. Ghannay (Eds.), *Proceedings of the Second Workshop on Natural Language Processing for Political Sciences @ LREC-COLING 2024* (pp. 39–44). ELRA and ICCL.

Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., & Xue, N. (Eds.). (2024). *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia: ELRA and ICCL.

Debaene, F., van der Haven, K., & Hoste, V. (2024). Early modern Dutch comedies and farces in the spotlight : introducing EmDComF and its emotion framework. In R. Sprugnoli & M. Passarotti (Eds.), *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024* (pp. 144–155). Torino, Italia: ELRA and ICCL.

Maladry, A., Cignarella, A. T., Lefever, E., Van Hee, C., & Hoste, V. (2024). Human and system perspectives on the expression of irony : an analysis of likelihood labels and rationales. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 8372–8382). Torino, Italia: ELRA and ICCL.

De Bruyne, L., van der Meer, T., De Clercq, O., & Hoste, V. (2024). Using State-of-the-art Emotion Detection Models in a Crisis Communication Context. *Computational Communication Research*, 6(1), 1. <https://doi.org/10.5117/ccr2024.1.1.debr>

Kruijsbergen, J., Van Geertruyen, S., Hoste, V., & De Clercq, O. (2024). Exploring LLMs' Capabilities for Error Detection in Dutch L1 and L2 Writing Products. *Computational Linguistics in the Netherlands Journal*, 13, 173–191.

Segundo Díaz, R. L., Rovelo, G., Bouzouita, M., Hoste, V., & Coninx, K. (2024). The influence of personality traits and game design elements on player enjoyment : an empirical study on gwaps for linguistics. *International Conference on Games and Learning Alliance*, 14475, 204–213. https://doi.org/10.1007/978-3-031-49065-1_20

Labat, S., Demeester, T., & Hoste, V. (2024). EmoTwICS : a corpus for modelling emotion trajectories in Dutch customer service dialogues on Twitter. *LANGUAGE RESOURCES AND EVALUATION*. <https://doi.org/10.1007/s10579-023-09700-0>

Maladry, A., Lefever, E., Van Hee, C., & Hoste, V. (2024). The limitations of irony detection in Dutch social media. *LANGUAGE RESOURCES AND EVALUATION*. <https://doi.org/10.1007/s10579-023-09656-1>

De Langhe, L., De Clercq, O., & Hoste, V. (2023). What Does {BERT} actually Learn about Event Coreference? Probing Structural Information in a Fine-Tuned {D}utch Language Model. In S. Tafreshi, A. Akula, J. Sedoc, A. Drozd, A. Rogers, & A. Rumshisky (Eds.), *Proceedings of the Fourth Workshop on Insights from Negative Results in NLP* (pp. 103–108). Dubrovnik, Croatia: Association for Computational Linguistics.

Segundo Díaz, R. L., Rovelo, G., Bouzouita, M., Hoste, V., & Coninx, K. (2023). The influence of personality traits and game design elements on player enjoyment : a demo on GWAPs for part-of-speech tagging. In M. Haahr, A. Rojas-Salazar, & S. Göbel (Eds.), *Serious games* (Vol. 14309, pp. 353–361). https://doi.org/10.1007/978-3-031-44751-8_28

- Du, Q., Labat, S., Demeester, T., & Hoste, V. (2023). Unimodalities count as perspectives in multimodal emotion annotation. In G. Abercrombie, V. Basile, D. Bernardi, S. Dudy, S. Frenda, L. Havens, ... S. Tonelli (Eds.), *Proceedings of the 2nd Workshop on Perspectivist Approaches to NLP (NLPerspectives 2023), co-located with the 26th European Conference on Artificial Intelligence (ECAI 2023)* (Vol. 3494). CEUR-WS.org.
- Maladry, A., Lefever, E., Van Hee, C., & Hoste, V. (2023). A fine line between irony and sincerity : identifying bias in transformer models for irony detection. In J. Barnes, O. De Clercq, & R. Klinger (Eds.), *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis* (pp. 315–324). <https://doi.org/10.18653/v1/2023.wassa-1.28>
- De Langhe, L., Desot, T., De Clercq, O., & Hoste, V. (2023). A benchmark for dutch end-to-end cross-document event coreference resolution. *ELECTRONICS*, 12(4). <https://doi.org/10.3390/electronics12040850>
- De Langhe, L., De Clercq, O., & Hoste, V. (2023). Leveraging structural discourse information for event coreference resolution in Dutch. *Proceedings of the 4th Workshop on Computational Approaches to Discourse (CODI 2023)*, 48–53. <https://doi.org/10.18653/v1/2023.codi-1.5>
- Hadifar, A., Bitew, S. K., Deleu, J., Hoste, V., Develder, C., & Demeester, T. (2023). Diverse content selection for educational question generation. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics : Student Research Workshop*, 123–133. Association for Computational Linguistics (ACL).
- Kaminska, O., Cornelis, C., & Hoste, V. (2023). Fuzzy rough nearest neighbour methods for aspect-based sentiment analysis. *ELECTRONICS*, 12(5). <https://doi.org/10.3390/electronics12051088>
- Kaminska, O., Cornelis, C., & Hoste, V. (2023). Fuzzy rough nearest neighbour methods for detecting emotions, hate speech and irony. *INFORMATION SCIENCES*, 625, 521–535. <https://doi.org/10.1016/j.ins.2023.01.054>
- Colruyt, C., De Clercq, O., Desot, T., & Hoste, V. (2023). EventDNA : a dataset for Dutch news event extraction as a basis for news diversification. *LANGUAGE RESOURCES AND EVALUATION*, 57(1), 189–221. <https://doi.org/10.1007/s10579-022-09623-2>
- De Langhe, L., De Clercq, O., & Hoste, V. (2023). Constructing a cross-document event coreference corpus for Dutch. *LANGUAGE RESOURCES AND EVALUATION*, 57(2), 819–848. <https://doi.org/10.1007/s10579-022-09597-1>
- De Langhe, L., De Clercq, O., & Hoste, V. (2022). Investigating cross-document event coreference for Dutch. *Proceedings of the Fifth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2022)*, 88–98. Gyeongju, Republic of Korea: Association for Computational Linguistics.
- De Langhe, L., De Clercq, O., & Hoste, V. (2022). Towards fine(r)-grained identification of event coreference resolution types. *COMPUTATIONAL LINGUISTICS IN THE NETHERLANDS JOURNAL*, 12, 183–205.
- De Bruyne, L., Singh, P., De Clercq, O., Lefever, E., & Hoste, V. (2022). How language-dependent is emotion detection? Evidence from multilingual BERT. *Proceedings of the 2nd Workshop on Multi-Lingual Representation Learning (MRL)*, 76–85. Association for Computational Linguistics.
- Labat, S., Hadifar, A., Demeester, T., & Hoste, V. (2022). An emotional journey : detecting emotion trajectories in Dutch customer service dialogues. *Proceedings of the Eighth Workshop on Noisy User-Generated Text (W-NUT 2022)*, 106–112. Gyeongju, Republic of Korea: Association for Computational Linguistics (ACL).
- Vandenbussche, N., Van Hee, C., Hoste, V., & Paemeleire, K. (2022). Using natural language processing to automatically classify written self-reported narratives by patients with migraine or cluster headache. *JOURNAL OF HEADACHE AND PAIN*, 23(1). <https://doi.org/10.1186/s10194-022-01490-0>
Impact factor: 7.4, category: CLINICAL NEUROLOGY, rank: 19/212.

Kaminska, O., Cornelis, C., & Hoste, V. (2022). LT3 at SemEval-2022 task 6 : fuzzy-rough nearest neighbor classification for sarcasm detection. *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Presented at the 16th International Workshop on Semantic Evaluation (SemEval-2022), Seattle, United States. <https://doi.org/10.18653/v1/2022.semeval-1.138>

De Bruyne, L., Karimi, A., De Clercq, O., Prati, A., & Hoste, V. (2022). Aspect-based emotion analysis and multimodal coreference : a case study of customer comments on Adidas instagram posts. In N. Calzolari, B. Frédéric, P. Blache, K. Choukri, C. Cieri, T. Declerck, ... S. Piperidis (Eds.), *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)* (pp. 574–580). ELRA.

Rigouts Terryn, A., Hoste, V., & Lefever, E. (2022). D-terminer : online demo for monolingual and bilingual automatic term extraction. In R. Costa, S. Carvalho, A. Ostroski Anic, & A. F. Khan (Eds.), *Proceedings of the Workshop on Terminology in the 21st century : many faces, many places* (pp. 33–40). Marseille, France: European Language Resources Association (ELRA).

Labat, S., Ackaert, N., Demeester, T., & Hoste, V. (2022). Variation in the expression and annotation of emotions : a Wizard of Oz pilot study. In G. Abercrombie, V. Basile, S. Tonelli, V. Rieser, & A. Uma (Eds.), *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP @LREC2022* (pp. 66–72). Marseille, France: European Language Resources Association (ELRA).

Maladry, A., Lefever, E., Van Hee, C., & Hoste, V. (2022). Irony detection for Dutch : a venture into the implicit. In J. Barnes, O. De Clercq, V. Barriere, S. Tafreshi, S. Alqahtani, J. Sedoc, ... A. Balahur (Eds.), *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis* (pp. 172–181). <https://doi.org/10.18653/v1/2022.wassa-1.16>

De Geyndt, E., De Clercq, O., Van Hee, C., Lefever, E., Singh, P., Parent, O., & Hoste, V. (2022). SentEMO : a multilingual adaptive platform for aspect-based sentiment and emotion analysis. In J. Barnes, O. De Clercq, V. Barriere, S. Tafreshi, S. Alqahtani, J. Sedoc, ... A. Balahur (Eds.), *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis* (pp. 51–61). <https://doi.org/10.18653/v1/2022.wassa-1.5>

Rigouts Terryn, A., Hoste, V., & Lefever, E. (2022). Tagging terms in text : a supervised sequential labelling approach to automatic term extraction. *TERMINOLOGY*, 28(1), 157–189. <https://doi.org/10.1075/term.21010.rig> Impact factor: 0.8, category: LINGUISTICS, rank: 126/194.

Jacobs, G., & Hoste, V. (2022). SENTiVENT : enabling supervised information extraction of company-specific events in economic and financial news. *LANGUAGE RESOURCES AND EVALUATION*, 56(1), 225–257. <https://doi.org/10.1007/s10579-021-09562-4>

Jacobs, G., Van Hee, C., & Hoste, V. (2022). Automatic classification of participant roles in cyberbullying : can we detect victims, bullies, and bystanders in social media text? *NATURAL LANGUAGE ENGINEERING*, 28(2), 141–166. <https://doi.org/10.1017/s135132492000056X>

Kaminska, O., Cornelis, C., & Hoste, V. (2021). Fuzzy-rough nearest neighbour approaches for emotion detection in tweets. *ROUGH SETS (IJCRS 2021)*, 12872, 231–246. https://doi.org/10.1007/978-3-030-87334-9_20

De Bruyne, L., De Clercq, O., & Hoste, V. (2021). Prospects for Dutch emotion detection : insights from the new EmotioNL dataset. *COMPUTATIONAL LINGUISTICS IN THE NETHERLANDS JOURNAL*, 11, 231–255.

De Bruyne, L., De Clercq, O., & Hoste, V. (2021). Mixing and matching emotion frameworks : investigating cross-framework transfer learning for Dutch emotion detection. *ELECTRONICS*, 10(21). <https://doi.org/10.3390/electronics10212643>

Impact factor: 2.69, category: ENGINEERING, ELECTRICAL & ELECTRONIC, rank: 139/276.

- Hahn, U., Hoste, V., & Stent, A. (Eds.). (2021). *Proceedings of the third workshop on economics and natural language processing (ECONLP 2021)*. Association for Computational Linguistics (ACL).
- Desot, T., De Clercq, O., & Hoste, V. (2021). Event prominence extraction combining a knowledge-based syntactic parser and a BERT classifier for Dutch. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, 346–357. https://doi.org/10.26615/978-954-452-072-4_040
- De Bruyne, L., De Clercq, O., & Hoste, V. (2021). Emotional RobBERT and insensitive BERTje : combining transformers and affect lexica for Dutch emotion detection. *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (EACL 2021)*, 257–263. Association for Computational Linguistics.
- Jacobs, G., & Hoste, V. (2021). Fine-grained implicit sentiment in financial news : uncovering hidden bulls and bears. *ELECTRONICS*, 10(20). <https://doi.org/10.3390/electronics10202554>
Impact factor: 2.69, category: ENGINEERING, ELECTRICAL & ELECTRONIC, rank: 139/276.
- Gesquière, N., Hoste, V., Van Hee, C., & wyffels, F. (2021). *Chatbot : leerkrachtenhandleiding*. Dwengo.
- Rigouts Terryn, A., Hoste, V., & Lefever, E. (2021). HAMLET : Hybrid Adaptable Machine Learning approach to Extract Terminology. *TERMINOLOGY*, 27(2), 290–329. <https://doi.org/10.1075/term.20017.rig>
Impact factor: 1.474, category: LINGUISTICS, rank: 88/194.
- Van Hee, C., De Clercq, O., & Hoste, V. (2021). *Guidelines for annotating implicit sentiment evoked by fine-grained news events*.
- Kaminska, O., Cornelis, C., & Hoste, V. (2021). Nearest neighbour approaches for emotion detection in tweets. *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (EACL 2021)*, 203–212. Association for Computational Linguistics.
- Hadifar, A., Labat, S., Hoste, V., Develder, C., & Demeester, T. (2021). A million tweets are worth a few points : tuning transformers for customer service tasks. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, ... Y. Zhou (Eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies* (pp. 220–225). <https://doi.org/10.18653/v1/2021.naacl-main.21>
- Van Hee, C., De Clercq, O., & Hoste, V. (2021). Exploring implicit sentiment evoked by fine-grained news events. *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (EACL 2021)*, 138–148. Association for Computational Linguistics.
- De Bruyne, L., De Clercq, O., & Hoste, V. (2021). Annotating affective dimensions in user-generated content : comparing the reliability of best-worst scaling, pairwise comparison and rating scales for annotating valence, arousal and dominance. *LANGUAGE RESOURCES AND EVALUATION*, 55(4), 1017–1045.
<https://doi.org/10.1007/s10579-020-09524-2>
Impact factor: 1.835, category: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS, rank: 93/113.
- Matos Veliz, C., De Clercq, O., & Hoste, V. (2021). Is neural always better? SMT versus NMT for Dutch text normalization. *EXPERT SYSTEMS WITH APPLICATIONS*, 170. <https://doi.org/10.1016/j.eswa.2020.114500>
Impact factor: 8.665, category: ENGINEERING, ELECTRICAL & ELECTRONIC, rank: 23/276.
- Emmery, C., Verhoeven, B., De Pauw, G., Jacobs, G., Van Hee, C., Lefever, E., ... Daelemans, W. (2021). Current limitations in cyberbullying detection : on evaluation criteria, reproducibility, and data scarcity. *LANGUAGE RESOURCES AND EVALUATION*, 55(3), 597–633. <https://doi.org/10.1007/s10579-020-09509-1>
Impact factor: 1.835, category: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS, rank: 93/113.

- Jacobs, G., & Hoste, V. (2020). Extracting fine-grained economic events from business news. In M. El-Haj, V. Athanasakou, S. Ferradans, C. Salzedo, A. Elhag, H. Bouamor, ... N. Pittaras (Eds.), *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation* (pp. 235–245). COLING.
- Labat, S., Demeester, T., & Hoste, V. (2020). Guidelines for annotating fine-grained emotion trajectories in customer service dialogues (version 1.0). Ghent: LT3, Faculty of Arts, Humanities and Law, Ghent University (Ghent, Belgium).
- De Clercq, O., & Hoste, V. (2020). It's absolutely divine! Can fine-grained sentiment analysis benefit from coreference resolution? *Proceedings of the Third Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2020)*, 11–21. Association for Computational Linguistics (ACL).
- De Clercq, O., De Bruyne, L., & Hoste, V. (2020). News topic classification as a first step towards diverse news recommendation. *COMPUTATIONAL LINGUISTICS IN THE NETHERLANDS JOURNAL*, 10, 37–55.
- Vanroy, B., Labat, S., Kaminska, O., Lefever, E., & Hoste, V. (2020). LT3 at SemEval-2020 Task 7 : comparing feature-based and transformer-based approaches to detect funny headlines. *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval 2020)*, 1033–1040. International Committee for Computational Linguistics.
- De Bruyne, L., De Clercq, O., & Hoste, V. (2020). An emotional mess! Deciding on a framework for building a Dutch emotion-annotated corpus. *PROCEEDINGS OF THE 12TH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC 2020)*, 1636–1644. European Language Resources Association (ELRA).
- Rigouts Terryn, A., Hoste, V., Drouin, P., & Lefever, E. (2020). TermEval 2020 : shared task on automatic term extraction using the Annotated Corpora for term Extraction Research (ACTER) dataset. In B. Daile, K. Kageura, & A. Rigouts Terryn (Eds.), *6th International Workshop on Computational Terminology (COMPUTERM 2020), Proceedings* (pp. 85–94). European Language Resources Association (ELRA).
- Rigouts Terryn, A., Hoste, V., & Lefever, E. (2020). In no uncertain terms : a dataset for monolingual and multilingual automatic term extraction from comparable corpora. *LANGUAGE RESOURCES AND EVALUATION*, 54(2), 385–418. <https://doi.org/10.1007/s10579-019-09453-9>
Impact factor: 1.358, category: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS, rank: 98/112.
- Joris, G., Colruyt, C., Vermeulen, J., Vercoutere, S., De Grove, F., Van Damme, K., ... Martens, L. (2020). News diversity and recommendation systems : setting the interdisciplinary scene. In M. Friedewald, M. Önen, E. Lievens, S. Krenn, & S. Fricker (Eds.), *Privacy and identity management : data for better living : AI and privacy* (Vol. 576, pp. 90–105). https://doi.org/10.1007/978-3-030-42504-3_7
- Tezcan, A., Hoste, V., & Macken, L. (2020). Estimating word-level quality of statistical machine translation output using monolingual information alone. *NATURAL LANGUAGE ENGINEERING*, 26(1), 73–94.
<https://doi.org/10.1017/S1351324919000111>
Impact factor: 1.07, category: LINGUISTICS, rank: 105/193.
- Hahn, U., Hoste, V., & Zhang, Z. (Eds.). (2019). *Proceedings of the Second Workshop on Economics and Natural Language Processing (ECONLP 2019)*. Hong Kong: Association for Computational Linguistics (ACL).
- Colruyt, C., De Clercq, O., & Hoste, V. (2019). Leveraging syntactic parsing to improve event annotation matching. *Aggregating and Analysing Crowdsourced Annotations for NLP : Proceedings of the First Workshop on Aggregating and Analysing Crowdsourced Annotations for NLP*, 15–23. <https://doi.org/10.18653/v1/d19-5903>
- Matos Veliz, C., De Clercq, O., & Hoste, V. (2019). Benefits of data augmentation for NMT-based text normalization of user-generated content. *Proceedings of the 5th Workshop on Noisy User-Generated Text (W-NUT 2019)*, 275–285. <https://doi.org/10.18653/v1/D19-5536>

Matos Veliz, C., De Clercq, O., & Hoste, V. (2019). Comparing MT approaches for text normalization. *Proceedings of Recent Advances in Natural Language Processing (RANLP 2019) : Natural Language Processing in a Deep Learning World*, 740–749.

Rigouts Terryn, A., Drouin, P., Hoste, V., & Lefever, E. (2019). Analysing the impact of supervised machine learning on automatic term extraction : HAMLET vs TernoStat. In G. Angelova, R. Mitkov, I. Nikolova, & I. Temnikova (Eds.), *Proceedings of Recent Advances in Natural Language Processing (RANLP 2019) : natural language processing in a deep learning world* (pp. 1013–1022). https://doi.org/10.26615/978-954-452-049-6_117

Folds, D., Lefever, E., Gera, R., & Hoste, V. (Eds.). (2019). *Proceedings of HUSO 2019, The fifth international conference on human and social analytics*. IARIA, International Academy, Research, and Industry Association.

De Bruyne, L., De Clercq, O., & Hoste, V. (2019). Towards an empirically grounded framework for emotion analysis. *Proceedings of HUSO 2019, the Fifth International Conference on Human and Social Analytics*, 11–16. IARIA, International Academy, Research, and Industry Association.

Balahur, A., Klinger, R., Hoste, V., Carlo Strapparava, C., & De Clercq, O. (Eds.). (2019). *Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Minneapolis, USA: Association for Computational Linguistics.

Vandeghinste, V., Vanallemeersch, T., Augustinus, L., Bulté, B., Van Eynde, F., Pelemans, J., ... Luyten, K. (2019). Improving the translation environment for professional translators. *INFORMATICS-BASEL*, 6(2). <https://doi.org/10.3390/informatics6020024>

Rigouts Terryn, A., Hoste, V., Buyschaert, J., Vander Stichele, R., Van Campen, E., & Lefever, E. (2019). Validating multilingual hybrid automatic term extraction for search engine optimisation : the use case of EBM-GUIDELINES. *ARGENTINIAN JOURNAL OF APPLIED LINGUISTICS*, 7(1), 93–108.

Colruyt, C., De Clercq, O., & Hoste, V. (2019). EventDNA : guidelines for entities and events in dutch news texts (v1.0).

Bauwelinck, N., Jacobs, G., Hoste, V., & Lefever, E. (2019). LT3 at SemEval-2019 Task 5 : multilingual detection of hate speech against immigrants and women in Twitter (hatEval). *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)*, 436–440. <https://doi.org/10.18653/v1/S19-2077>

Tezcan, A., Hoste, V., & Macken, L. (2019). Estimating post-editing time using a gold-standard set of machine translation errors. *COMPUTER SPEECH AND LANGUAGE*, 55, 120–144. <https://doi.org/10.1016/j.csl.2018.10.005> Impact factor: 2.116, category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE, rank: 71/137.

Hoste, V., Vanopstal, K., Rigouts Terryn, A., & Lefever, E. (2019). The trade-off between quantity and quality : comparing a large web corpus and a small focused corpus for medical terminology extraction. *ACROSS LANGUAGES AND CULTURES*, 20(2), 197–211. <https://doi.org/10.1556/084.2019.20.2.3> Impact factor: 0.36, category: LINGUISTICS, rank: 150/187.

Hahn, U., Hoste, V., & Tsai, M.-F. (Eds.). (2018). *Proceedings of the First Workshop on Economics and Natural Language Processing (ECONLP 2018)*. Melbourne, Australia: Association for Computational Linguistics (ACL).

Matos Veliz, C., De Clercq, O., & Hoste, V. (2018). Varying background corpora for SMT-based text normalization. In R. Vandekerckhove, D. Fišer, & L. Hilde (Eds.), *Proceedings of the 6th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora (CMC-corpora 2018)* (pp. 76–76).

Rigouts Terryn, A., Hoste, V., Buyschaert, J., & Lefever, E. (2018). Multilingual hybrid automatic term extraction : the use case of ebpracticenet. In T. Read, S. Montaner, & B. Sedano (Eds.), *Technological innovation for specialized linguistic domains : languages for digital lives and cultures, proceedings of TISLID'18* (pp. 151–162). Beau Bassin, Mauritius: Editions Universitaires Europeennes.

- Van Hee, C., Jacobs, G., Emmery, C., Desmet, B., Lefever, E., Verhoeven, B., ... Hoste, V. (2018). Automatic detection of cyberbullying in social media text. *PLOS ONE*, 13(10). <https://doi.org/10.1371/journal.pone.0203794>
Impact factor: 2.776, category: MULTIDISCIPLINARY SCIENCES, rank: 24/69.
- Van Hee, C., Lefever, E., & Hoste, V. (2018). We usually don't like going to the dentist : using common sense to detect irony on Twitter. *COMPUTATIONAL LINGUISTICS*, 44(4), 793–832. https://doi.org/10.1162/coli_a_00337
Impact factor: 2.13, category: LINGUISTICS, rank: 19/184.
- Jacobs, G., Lefever, E., & Hoste, V. (2018). Economic event detection in company-specific news text. In U. Hahn, V. Hoste, & M.-F. Tsai (Eds.), *ECONOMICS AND NATURAL LANGUAGE PROCESSING (ECONLP 2018)* (pp. 1–10). <https://doi.org/10.18653/v1/W18-3101>
- Van Hee, C., Lefever, E., & Hoste, V. (2018). SemEval-2018 task 3 : irony detection in English tweets. *Proceedings of The 12th International Workshop on Semantic Evaluation*, 39–50. New Orleans, Louisiana: Association for Computational Linguistics.
- De Bruyne, L., De Clercq, O., & Hoste, V. (2018). LT3 at SemEval-2018 Task 1 : a classifier chain to detect emotions in tweets. *Proceedings of The 12th International Workshop on Semantic Evaluation*, 123–127. <https://doi.org/10.18653/v1/S18-1016>
- Joris, G., De Grove, F., Hoste, V., Lievens, E., Martens, L., & De Marez, L. (2018). Promoting news diversity : an interdisciplinary investigation into algorithmic design, personalization and the public interest. *68th Annual ICA Conference, Abstracts*. Presented at the 68th Annual ICA Conference : Voices, Prague, Czech Republic.
- Rigouts Terryn, A., Hoste, V., & Lefever, E. (2018). A gold standard for multilingual automatic term extraction from comparable corpora : term structure and translation equivalents. *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, 1803–1808. Paris: European Language Resources Association (ELRA).
- Van Hee, C., Lefever, E., & Hoste, V. (2018). Exploring the fine-grained analysis and automatic detection of irony on Twitter. *LANGUAGE RESOURCES AND EVALUATION*, 52(3), 707–731. <https://doi.org/10.1007/s10579-018-9414-2>
- Desmet, B., & Hoste, V. (2018). Online suicide prevention through optimised text classification. *INFORMATION SCIENCES*, 439, 61–78. <https://doi.org/10.1016/j.ins.2018.02.014>
Impact factor: 5.524, category: COMPUTER SCIENCE, INFORMATION SYSTEMS, rank: 9/155.
- Van Hee, C., Van de Kauter, M., De Clercq, O., Lefever, E., Desmet, B., & Hoste, V. (2017). Noise or music? Investigating the usefulness of normalisation for robust sentiment analysis on social media data. *TRAITEMENT AUTOMATIQUE DES LANGUES*, 58(1), 63–87.
- Tezcan, A., Hoste, V., & Macken, L. (2017). SCATE taxonomy and corpus of machine translation errors. In G. C. Pastor & I. Durán-Muñoz (Eds.), *Trends in E-tools and resources for translators and interpreters* (Vol. 45, pp. 219–244). https://doi.org/10.1163/9789004351790_012
- Daelemans, W., De Clercq, O., & Hoste, V. (2017). STYLENE : an environment for stylometry and readability research for Dutch. In J. Odiijk & A. Van Hessen (Eds.), *CLARIN in the Low Countries* (pp. 195–210). London: Ubiquity Press.
- De Clercq, O., Lefever, E., Jacobs, G., Carpels, T., & Hoste, V. (2017). Towards an integrated pipeline for aspect-based sentiment analysis in various domains. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 136–142. Association for Computational Linguistics.
- Tezcan, A., Hoste, V., & Macken, L. (2017). A neural network architecture for detecting grammatical errors in statistical machine translation. *THE PRAGUE BULLETIN OF MATHEMATICAL LINGUISTICS*, (108), 133–145. <https://doi.org/10.1515/pralin-2017-0015>

Koleva, M., Farasyn, M., Desmet, B., Breitbarth, A., & Hoste, V. (2017). An automatic part-of-speech tagger for Middle Low German. *INTERNATIONAL JOURNAL OF CORPUS LINGUISTICS*, 22(1), 108–141.
<https://doi.org/10.1075/ijcl.22.1.05kol>

Impact factor: 0.976, category: LINGUISTICS, rank: 75/181.

De Clercq, O., & Hoste, V. (2016). Rude waiter but mouthwatering pastries! An exploratory study into Dutch aspect-based sentiment analysis. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, ... S. Piperidis (Eds.), *LREC 2016 - TENTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION* (pp. 2910–2917). ELRA.

Van Hee, C., Lefever, E., & Hoste, V. (2016). Monday mornings are my fave :) #not exploring the automatic recognition of irony in English tweets. In N. Calzolari, Y. Matsumoto, & R. Prasad (Eds.), *Proceedings of COLING 2016, 26th International Conference on Computational Linguistics : technical papers* (pp. 2730–2739). ACL.

Hoste, V., Van Hee, C., & Poels, K. (2016). Towards a framework for the automatic detection of crisis emotions on social media : a corpus analysis of the tweets posted after the crash of Germanwings flight 9525. *Proceedings of HUSO 2016, The Second International Conference on Human and Social Analytics*, 29–32.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., ... Eryiğit, G. (2016). SemEval-2016 task 5 : aspect based sentiment analysis. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 19–30. Association for Computational Linguistics.

Hoste, V., Lefever, E., van der Waart van Gulik, S., & Desmet, B. (2016). TripleSent: a triple store of events associated with their prototypical sentiment. *Proceedings of The Eighth International Conference on Information, Process, and Knowledge Management*, 91–93. Venice, Italy: IARIA.

Tezcan, A., Hoste, V., & Macken, L. (2016). UGENT-LT3 SCATE Submission for WMT16 Shared Task on Quality Estimation. *Proceedings of the First Conference on Machine Translation, Volume 2: Shared Task Papers*, 843–850. Association for Computational Linguistics.

Hoste, V. (2016). The mention-pair model. In M. Poesio, R. Stuckardt, & Y. Versley (Eds.), *Anaphora resolution : algorithms, resources and applications* (pp. 281–295). Berlin Heidelberg: Springer-Verlag.

Lefever, E., & Hoste, V. (2016). A classification-based approach to economic event detection in Dutch news text. *LREC 2016 - TENTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION*, 330–335. ELRA.

Tezcan, A., Hoste, V., & Macken, L. (2016). Detecting grammatical errors in machine translation output using dependency parsing and treebank querying. *BALTIC JOURNAL OF MODERN COMPUTING*, 4(2), 203–217.

Desmet, B., Jacobs, G., & Hoste, V. (2016). Mental distress detection and triage in forum posts: the LT3 CLPsych 2016 shared task system. *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 148–152. Association for Computational Linguistics.

Van Hee, C., Lefever, E., & Hoste, V. (2016). Exploring the realization of irony in Twitter data. *LREC 2016 - TENTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION*, 1795–1799. ELRA.

Smeuninx, N., De Clerck, B., Hoste, V., & Aerts, W. (2016). The Readability of Sustainability Reporting and its Interaction with Company Performance. *Association For Business Communication 2016 Conference*. Presented at the Association For Business Communication 2016 Conference, Cape Town.

De Clercq, O., & Hoste, V. (2016). All mixed up? Finding the optimal feature set for general readability prediction and its application to English and Dutch. *COMPUTATIONAL LINGUISTICS*, 42(3), 457–490.

https://doi.org/10.1162/COLI_a_00255

Impact factor: 2.528, category: LINGUISTICS, rank: 5/180.

Van Hee, C., Lefever, E., & Hoste, V. (2016). Guidelines for Annotating Irony in Social Media Text, version 2.0.

- Schulz, S., De Pauw, G., De Clercq, O., Desmet, B., Hoste, V., Daelemans, W., & Macken, L. (2016). Multi-modular text normalization of Dutch user-generated content. *ACM TRANSACTIONS ON INTELLIGENT SYSTEMS AND TECHNOLOGY*, 7(4). <https://doi.org/10.1145/2850422>
Impact factor: 3.196, category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE, rank: 26/133.
- Farasyn, M., Koleva, M., Hoste, V., & Breitbarth, A. (2015). A parsed corpus of historical Low German. *25th Meeting of Computational Linguistics in the Netherlands (CLIN25), Abstracts*. Presented at the 25th Meeting of Computational Linguistics in the Netherlands (CLIN25), Antwerp, Belgium.
- Lavid, J., Arús, J., De Clerck, B., & Hoste, V. (2015). Creation of a high-quality, register-diversified parallel (English-Spanish) corpus for linguistic and computational investigations. *PROCEDIA SOCIAL & BEHAVIORAL SCIENCES*, 198, 249–256.
- Van Hee, C., Verhoeven, B., Lefever, E., De Pauw, G., Daelemans, W., & Hoste, V. (2015). Guidelines for the fine-grained analysis of cyberbullying, version 1.0.
- Van Hee, C., Lefever, E., Verhoeven, B., Mennes, J., Desmet, B., De Pauw, G., ... Hoste, V. (2015). Automatic detection and prevention of cyberbullying. In P. Lorenz & C. Bourret (Eds.), *International Conference on Human and Social Analytics, Proceedings* (pp. 13–18). IARIA.
- Hanoulle, S., Hoste, V., & Remael, A. (2015). The translation of documentaries: can domain-specific, bilingual glossaries reduce the translators' workload? An experiment involving professional translators. *NEW VOICES IN TRANSLATION STUDIES*, 23, 25–49.
- Hanoulle, S., Hoste, V., & Remael, A. (2015). The efficacy of terminology-extraction systems for the translation of documentaries. *PERSPECTIVES-STUDIES IN TRANSLATOLOGY*, 23(3), 359–374.
<https://doi.org/10.1080/0907676X.2015.1010549>
- Van Hee, C., Lefever, E., Verhoeven, B., Mennes, J., Desmet, B., De Pauw, G., ... Hoste, V. (2015). Detection and fine-grained classification of cyberbullying events. In G. Angelova, K. Bontcheva, & R. Mitkov (Eds.), *Proceedings of Recent Advances in Natural Language Processing, Proceedings* (pp. 672–680).
- Van Hee, C., Lefever, E., & Hoste, V. (2015). LT3: sentiment analysis of figurative tweets: piece of cake #NotReally. *SemEval: 9th International Workshop on Semantic Evaluations at NaacL 2015, Proceedings*, 684–688. <https://doi.org/10.18653/v1/s15-2115>
- Van de Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *EXPERT SYSTEMS WITH APPLICATIONS*, 42(11), 4999–5010.
<https://doi.org/10.1016/j.eswa.2015.02.007>
Impact factor: 2.981, category: OPERATIONS RESEARCH & MANAGEMENT SCIENCE, rank: 6/82.
- Van de Kauter, M., Desmet, B., & Hoste, V. (2015). The good, the bad and the implicit: a comprehensive approach to annotating explicit and implicit sentiment. *LANGUAGE RESOURCES AND EVALUATION*, 49(3), 685–720.
<https://doi.org/10.1007/s10579-015-9297-4>
Impact factor: 0.975, category: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS, rank: 81/104.
- De Clercq, O., Van de Kauter, M., Lefever, E., & Hoste, V. (2015). Applying hybrid terminology extraction to aspect-based sentiment analysis. *International Workshop on Semantic Evaluation, Proceedings*, 719–724.
<https://doi.org/10.18653/v1/s15-2122>
- Desmet, B., & Hoste, V. (2014). Recognising suicidal messages in Dutch social media. *LREC 2014 - NINTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION*, 830–835.
- Lefever, E., & Hoste, V. (2014). Parallel corpora make sense: bypassing the knowledge acquisition bottleneck for word sense disambiguation. *INTERNATIONAL JOURNAL OF CORPUS LINGUISTICS*, 19(3), 333–367.

<https://doi.org/10.1075/ijcl.19.3.02lef>

Impact factor: 0.65, category: LINGUISTICS, rank: 70/172.

Lefever, E., Van de Kauter, M., & Hoste, V. (2014). HypoTerm detection of hypernym relations between domain-specific terms in Dutch and English. *TERMINOLOGY*, 20(2), 250–278. <https://doi.org/10.1075/term.20.2.06lef>
Impact factor: 0.636, category: LINGUISTICS, rank: 74/172.

De Clercq, O., Schuhmacher, M., Ponzetto, S. P., & Hoste, V. (2014). Exploiting frameNet for content-based book recommendation. *Proceedings of the 1st Workshop on New Trends in Content-Based Recommender Systems Co-Located with the 8th ACM Conference on Recommender Systems : CBRecSys at ACM RecSys, 1245*, 14–21.

Van Hee, C., Van de Kauter, M., De Clercq, O., Lefever, E., & Hoste, V. (2014). LT3: sentiment classification in user-generated content using a rich feature set. *Proceedings of the 8th International Workshop on Semantic Evaluation*, 406–410. <https://doi.org/10.3115/v1/s14-2070>

De Clercq, O., Hertling, S., Hoste, V., Ponzetto, S. P., & Paulheim, H. (2014). Identifying disputed topics in the news. *Proceedings of the LD4KD Workshop at ECML/PKDD2014*, 37–48. CEUR.

Lefever, E., Van de Kauter, M., & Hoste, V. (2014). Evaluation of automatic hypernym extraction from technical corpora in English and Dutch. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.), *LREC 2014 - NINTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION* (pp. 490–497). European Language Resources Association (ELRA).

Oostdijk, N., Reynaert, M., Hoste, V., van den Heuvel, H., De Clercq, O., & Sanders, E. (2014). *SoNaR nieuw media corpus*. Centrale voor Taal- en Spraaktechnologie.

Oostdijk, N., Reynaert, M., Hoste, V., van den Heuvel, H., Monachesi, P., Desmet, B., ... Schuurman, I. (2014). *SoNaR-1*. Centrale voor Taal- en Spraaktechnologie.

Oostdijk, N., Hoste, V., de Jong, F., Reynaert, M., De Clercq, O., Desmet, B., & van den Heuvel, H. (2014). *SoNaR-500*. Centrale voor Taal- en Spraaktechnologie.

De Clercq, O., Schulz, S., Desmet, B., & Hoste, V. (2014). Towards shared datasets for normalization research. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.), *LREC 2014 - NINTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION* (pp. 1218–1223). European Language Resources Association (ELRA).

Macken, L., De Clercq, O., Desmet, B., & Hoste, V. (2014). Dutch parallel corpus en SoNaR. In S. Evenepoel, P. Goethals, & L. Joonen (Eds.), *Beschouwingen uit een talenhuis : opstellen over onderwijs en onderzoek in de vakgroep Vertalen, Tolken en Communicatie aangeboden aan Rita Godyns* (pp. 163–170). Gent: Academia Press.

Desmet, B., De Clercq, O., Van de Kauter, M., Schulz, S., Van Hee, C., & Hoste, V. (2014). Taaltechnologie 2.0: sentimentanalyse en normalisatie. In S. Evenepoel, P. Goethals, & L. Joonen (Eds.), *Beschouwingen uit een talenhuis : opstellen over onderwijs en onderzoek in de vakgroep Vertalen, Tolken en Communicatie aangeboden aan Rita Godyns* (pp. 157–161). Ghent, Belgium: Academia Press.

De Clercq, O., & Hoste, V. (2014). Hoe meetbaar is leesbaarheid? In S. Evenepoel, P. Goethals, & L. Joonen (Eds.), *Beschouwingen uit een talenhuis : opstellen over onderwijs en onderzoek in de vakgroep Vertalen, Tolken en Communicatie aangeboden aan Rita Godyns* (pp. 147–155). Ghent, Belgium: Academia Press.

Vanopstal, K., Macken, L., Lefever, E., Van de Kauter, M., Buysschaert, J., & Hoste, V. (2014). Terminologie: op het snijvlak van ambacht en technologie. In S. Evenepoel, P. Goethals, & L. Joonen (Eds.), *Beschouwingen uit een talenhuis : opstellen over onderwijs en onderzoek in de vakgroep Vertalen, Tolken en Communicatie aangeboden aan Rita Godyns* (pp. 179–189). Gent: Academia Press.

Desmet, B., & Hoste, V. (2014). Fine-grained Dutch named entity recognition. *LANGUAGE RESOURCES AND EVALUATION*, 48(2), 307–343. <https://doi.org/10.1007/s10579-013-9255-y>

Impact factor: 0.619, category: COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS, rank: 89/102.

De Clercq, O., Hoste, V., Desmet, B., van Oosten, P., De Cock, M., & Macken, L. (2014). Using the crowd for readability prediction. *NATURAL LANGUAGE ENGINEERING*, 20(3), 293–325.

<https://doi.org/10.1017/S1351324912000344>

Hendrickx, I., Bouma, G., Daelemans, W., & Hoste, V. (2013). COREA: coreference resolution for extracting answers for Dutch. In *Essential speech and language technology for Dutch : results by the STEVIN-programme* (pp. 97–108). Berlin, Germany: Springer.

Lefever, E., & Hoste, V. (2013). SemEval-2013 task 10: cross-lingual word sense disambiguation. *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2 : Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, 158–166. Association for Computational Linguistics (ACL).

Schropp, G., Lefever, E., & Hoste, V. (2013). A combined pattern-based and distributional approach for automatic hypernym detection in Dutch. *Proceedings of Recent Advances in Natural Language Processing*, 593–600. Shoumen, Bulgaria: INCOMA.

Desmet, B., & Hoste, V. (2013). Emotion detection in suicide notes. *EXPERT SYSTEMS WITH APPLICATIONS*, 40(16), 6351–6358. <https://doi.org/10.1016/j.eswa.2013.05.050>

Impact factor: 1.965, category: OPERATIONS RESEARCH & MANAGEMENT SCIENCE, rank: 11/79.

Van de Kauter, M., Coorman, G., Lefever, E., Desmet, B., Macken, L., & Hoste, V. (2013). LeTs preprocess: the multilingual LT3 linguistic preprocessing toolkit. *COMPUTATIONAL LINGUISTICS IN THE NETHERLANDS JOURNAL*, 3, 103–120.

De Clercq, O., Desmet, B., Schulz, S., Lefever, E., & Hoste, V. (2013). Normalization of Dutch user-generated content. *Proceedings of Recent Advances in Natural Language Processing*, 179–188. Shoumen, Bulgaria: INCOMA.

Lefever, E., Hoste, V., & De Cock, M. (2013). Five languages are better than one: an attempt to bypass the data acquisition bottleneck for WSD. In A. Gelbukh (Ed.), *Lecture Notes in Computer Science* (Vol. 7816, pp. 343–354). https://doi.org/10.1007/978-3-642-37247-6_28

Macken, L., Lefever, E., & Hoste, V. (2013). TExSIS: bilingual terminology extraction from parallel corpora using chunk-based alignment. *TERMINOLOGY*, 19(1), 1–30. <https://doi.org/10.1075/term.19.1.01mac>

Impact factor: 0.375, category: LINGUISTICS, rank: 105/169.

Lefever, E., Hoste, V., & De Cock, M. (2012). Discovering missing Wikipedia inter-language links by means of cross-lingual word sense disambiguation. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.), *LREC 2012 : eight international conference on language resources and evaluation* (pp. 841–846). Paris, France: European Language Resources Association (ELRA).

Reynaert, M., Schuurman, I., Hoste, V., Oostdijk, N., & van Gompel, M. (2012). Beyond SoNaR: towards the facilitation of large corpus building efforts. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.), *LREC 2012 : eight international conference on language resources and evaluation* (pp. 2897–2904). Paris, France: European Language Resources Association (ELRA).

Desmet, B., & Hoste, V. (2012). Combining lexico-semantic features for emotion classification in suicide notes. *BIOMEDICAL INFORMATICS INSIGHTS*, 5(suppl. 1), 125–128. <https://doi.org/10.4137/BII.S8960>

De Clercq, O., Hoste, V., & Monachesi, P. (2012). Evaluating automatic cross-domain Dutch semantic role annotation. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.),

LREC 2012 : eight international conference on language resources and evaluation (pp. 88–93). Paris, France: European Language Resources Association (ELRA).

Macken, L., Hoste, V., Leijten, M., & Van Waes, L. (2012). From keystrokes to annotated process data: enriching the output of Inputlog with linguistic information. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, ... S. Piperidis (Eds.), *LREC 2012 : eight international conference on language resources and evaluation* (pp. 2224–2229). Paris, France: European Language Resources Association (ELRA).

Leijten, M., Macken, L., Hoste, V., Van Horenbeeck, E., & Van Waes, L. (2012). From character to word level: enabling the linguistic analyses of Inputlog process data. *Proceedings of the EACL 2012 Workshop on Computational Linguistics and Writing*. Presented at the EACL 2012 workshop on Computational Linguistics and Writing (CL&W 2012), Avignon, France.

Lefever, E., & Hoste, V. (Eds.). (2011). Computational linguistics in the Netherlands. CLIN.

Hendrickx, I., De Clercq, O., & Hoste, V. (2011). Analysis and reference resolution of Bridge Anaphora across different text genres. In I. Hendrickx, S. Lalitha Devi, A. Branco, & R. Mitkov (Eds.), *Lecture Notes in Artificial Intelligence* (Vol. 7099, pp. 1–11). https://doi.org/10.1007/978-3-642-25917-3_1

De Clercq, O., Hoste, V., & Hendrickx, I. (2011). Cross-domain Dutch coreference resolution. In G. Angelova, K. Bontcheva, R. Mitkov, & N. Nikolov (Eds.), *International conference Recent advances in natural language processing* (pp. 186–193). Shoumen, Bulgaria: INCOMA Ltd.

Lefever, E., Hoste, V., & De Cock, M. (2011). Using parallel corpora for word sense disambiguation. In P. De Causmaecker, J. Maervoet, T. Messelis, K. Verbeeck, & T. Vermeulen (Eds.), *BNAIC : Belgian/Netherlands Artificial Intelligence Conference* (pp. 407–408).

Lefever, E., & Hoste, V. (2011). Examining the validity of cross-lingual word sense disambiguation. *POLIBITS*, (43), 29–35.

Lefever, E., Hoste, V., & De Cock, M. (2011). ParaSense or how to use parallel corpora for word sense disambiguation. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Short Papers*, 317–322. Association for Computational Linguistics (ACL).

Lefever, E., & Hoste, V. (2011). An evaluation and possible improvement path for current SMT behavior on ambiguous nouns. *Proceedings of SSST-5 : Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 52–60. Association for Computational Linguistics (ACL).

van Oosten, P., & Hoste, V. (2011). Readability annotation: replacing the expert by the crowd. In J. Tetreault, J. Burstein, & C. Leacock (Eds.), *Workshop on innovative use of NLP for building educational applications, Proceedings of the workshop* (pp. 120–129). Portland, OR, USA: Association for Computational Linguistics.

Lefever, E., Macken, L., & Hoste, V. (2011). Taal- en spraaktechnologie: een stand van zaken. *OVER TAAL (KORTRIJK-HEULE)*, 50(1), 20–22.

van Oosten, P., Hoste, V., & Tanghe, D. (2011). A posteriori agreement as a quality measure for readability prediction systems. In A. Gelbukh (Ed.), *Lecture Notes in Computer Science* (Vol. 6609, pp. 424–435). https://doi.org/10.1007/978-3-642-19437-5_35

Van Vaerenbergh, L., Peersman, C., Franck, M., Delaere, I., Hoste, V., & Velaerts, P. (2010). Van wetenschappelijke naar populaire term: de ABOP-tool. In C. Wermuth, E. Ruijsendaal, & O. Koornwinder (Eds.), *Terminologie in het Nederlandse Taalgebied* (Vol. 1, pp. 81–93). Gent.

Lefever, E., & Hoste, V. (2010). SemEval-2010 Task 3: cross-lingual word sense disambiguation. In K. Erk & C. Strapparava (Eds.), *SemEval 2010 : 5th International workshop on Semantic Evaluation : proceedings of the workshop* (pp. 15–20). Stroudsburg, PA, USA: Association for Computational Linguistics (ACL).

Recasens, M., Màrquez, L., Sapena, E., Martí, M. A., Taulé, M., Hoste, V., ... Versley, Y. (2010). Semeval-2010 task 1: coreference resolution in multiple languages. In K. Erk & C. Strapparava (Eds.), *SemEval 2010 : 5th International workshop on Semantic Evaluation : proceedings of the workshop*. Stroudsburg, PA, USA: Association for Computational Linguistics (ACL).

Desmet, B., & Hoste, V. (2010). Dutch named entity recognition using ensemble classifiers. In E. Westerhout, T. Markus, & P. Monachesi (Eds.), *Computational Linguistics in the Netherlands 2010 : selected papers from the twentieth CLIN meeting (CLIN 2010)* (pp. 29–41). Utrecht, The Netherlands: Landelijke Onderzoeksschool Taalwetenschap (LOT).

Schuurman, I., Hoste, V., & Monachesi, P. (2010). Interacting semantic layers of annotation in SoNaR, a reference corpus of contemporary written Dutch. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, ... D. Tapias (Eds.), *LREC 2010 : seventh conference on international language resources and evaluation* (pp. 2471–2477). Paris, France: European Language Resources Association (ELRA).

Desmet, B., & Hoste, V. (2010). Towards a balanced named entity corpus for Dutch. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, ... D. Tapias (Eds.), *LREC 2010 : seventh conference on international language resources and evaluation* (pp. 535–541). Paris, France: European Language Resources Association (ELRA).

Vanopstal, K., Desmet, B., & Hoste, V. (2010). Towards a learning approach for abbreviation detection and resolution. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, ... D. Tapias (Eds.), *LREC 2010 : seventh conference on international language resources and evaluation* (pp. 1043–1049). Paris, France: European Language Resources Association (ELRA).

van Oosten, P., Tanghe, D., & Hoste, V. (2010). Towards an improved methodology for automated readability prediction. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, ... D. Tapias (Eds.), *LREC 2010 : seventh conference on international language resources and evaluation* (pp. 775–782). Paris, France: European Language Resources Association (ELRA).

Lefever, E., & Hoste, V. (2010). Construction of a benchmark data set for cross-lingual word sense disambiguation. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, ... D. Tapias (Eds.), *LREC 2010 : seventh conference on international language resources and evaluation* (pp. 1584–1590). Paris, France: European Language Resources Association (ELRA).

Lefever, E., Fayruzov, T., Hoste, V., & De Cock, M. (2010). Clustering web people search results using fuzzy ants. *INFORMATION SCIENCES*, 180(17), 3192–3209. <https://doi.org/10.1016/j.ins.2010.05.018>
Impact factor: 2.833, category: COMPUTER SCIENCE, INFORMATION SYSTEMS, rank: 9/125.

Hoste, V., Vanopstal, K., Lefever, E., & Delaere, I. (2010). Classification-based scientific term detection in patient information. *TERMINOLOGY*, 16(1), 1–29. <https://doi.org/10.1075/term.16.1.01hos>
Impact factor: 0.4, category: LINGUISTICS, rank: 87/141.

Delaere, I., Hoste, V., Peersman, C., Van Vaerenbergh, L., & Velaerts, P. (2009). ABOP, automatic optimization of patient information leaflets. In S. Cardey (Ed.), *Proceedings of the international symposium on data and sense mining, machine translation and controlled languages* (pp. 74–81). Besançon, France: Université de Franche-Comté.

Delaere, I., Hoste, V., & Monachesi, P. (2009). Cultivating trees: adding several semantic layers to the Lassy treebank in SoNaR. In F. Van Eynde, A. Frank, K. De Smedt, & G. van Noord (Eds.), *Proceedings of the seventh international workshop on treebanks and linguistic theories* (pp. 135–146). Utrecht, The Netherlands: LOT (Landelijke Onderzoeksschool Taalwetenschap).

- Daelemans, W., & Hoste, V. (Eds.). (2009). Evaluation of translation technology. Antwerp, Belgium: University Press Antwerp.
- Daelemans, W., & Hoste, V. (2009). Evaluation of translation technology. *LINGUISTICA ANTVERPIENSIA NEW SERIES-THEMES IN TRANSLATION STUDIES*, 8, 9–13.
- Hendrickx, I., & Hoste, V. (2009). Coreference resolution on blogs and commented news. In S. L. Devi, A. Branco, & R. Mitkov (Eds.), *Lecture Notes in Artificial Intelligence* (Vol. 5847, pp. 43–53). https://doi.org/10.1007/978-3-642-04975-0_4
- Lefever, E., & Hoste, V. (2009). SemEval-2010 Task 3: cross-lingual word sense disambiguation. *Proceedings of the NAACL HLT Workshop on Semantic Evaluations : Recent Achievements and Future Directions*, 82–87. Association for Computational Linguistics (ACL).
- Lefever, E., Macken, L., & Hoste, V. (2009). Language-independent bilingual terminology extraction from a multilingual parallel corpus. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 496–504. Association for Computational Linguistics (ACL).
- Fayruzov, T., De Cock, M., Cornelis, C., & Hoste, V. (2009). Linguistic feature analysis for protein interaction extraction. *BMC BIOINFORMATICS*, 10. <https://doi.org/10.1186/1471-2105-10-374>
Impact factor: 3.428, category: MATHEMATICAL & COMPUTATIONAL BIOLOGY, rank: 4/29.
- Lefever, E., Fayruzov, T., Hoste, V., & De Cock, M. (2009). Fuzzy ants clustering for web people search. *Proceedings of WePS2 : 2nd Web People Search Evaluation Workshop, Workshop at WWW2009*. Presented at the 2nd Web People Search Evaluation Workshop (WePS 2009) ; 18th WWW Conference, Madrid, Spain.
- Fayruzov, T., De Cock, M., Cornelis, C., & Hoste, V. (2008). The role of syntactic features in protein interaction extraction. *Data and Text Mining in Bioinformatics, 2nd International Workshop, Proceedings*, 61–68. <https://doi.org/10.1145/1458449.1458463>
- Hoste, V., Lefever, E., Vanopstal, K., & Delaere, I. (2008). Learning-based detection of scientific terms in patient information. *LREC 2008 : Sixth International Conference on Language Resources and Evaluation*, 585–591. Paris, France: European Language Resources Association (ELRA).
- Hendrickx, I., Bouma, G., Coppens, F., Daelemans, W., Hoste, V., Kloosterman, G., ... Verschelde, J.-L. (2008). A coreference corpus and resolution system for Dutch. *LREC 2008 : Sixth International Conference on Language Resources and Evaluation*, 144–149. Paris, France: European Language Resources Association (ELRA).
- Fayruzov, T., De Cock, M., Cornelis, C., & Hoste, V. (2008). DEEPER: a full parsing based approach to protein relation extraction. In E. Marchiori & J. H. Moore (Eds.), *Lecture Notes in Computer Science* (Vol. 4973, pp. 36–47). https://doi.org/10.1007/978-3-540-78757-0_4
- Hendrickx, I., Hoste, V., & Daelemans, W. (2008). Semantic and syntactic features for Dutch coreference resolution. In A. Gelbukh (Ed.), *Lecture Notes in Computer Science* (Vol. 4919, pp. 351–361). https://doi.org/10.1007/978-3-540-78135-6_30
- 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*, 529–536. Stroudsburg, PA, USA: Association for Computational Linguistics (ACL).
- Hoenkamp, E., De Cock, M., & Hoste, V. (Eds.). (2008). *Proceedings of the 8th Dutch-Belgian information retrieval workshop (DIR 2008)*. Oce-Nederland.
- Fayruzov, T., De Cock, M., Cornelis, C., & Hoste, V. (2008). Exploiting grammatical relations for protein relation extraction and role labeling. In E. Hoenkamp, M. De Cock, & V. Hoste (Eds.), *Proceedings of the 8th Dutch-Belgian information retrieval workshop (DIR 2008)* (pp. 37–44). Oce-Nederland.

- Hoste, V., Hendrickx, I., & Macken, L. (2007). The referential versus non-referential use of the neuter pronoun in Dutch and English. *Proceedings from the Corpus Linguistics Conference Series*. Presented at the Corpus Linguistics conference 2007, Birmingham, UK.
- Hoste, V., & van den Bosch, A. (2007). A modular approach to learning Dutch coreference. In C. Johansson (Ed.), *Proceedings from the first Bergen workshop on Anaphora resolution* (pp. 51–75). Cambridge, UK: Cambridge Scholars Publishing.
- Hendrickx, I., Hoste, V., & Daelemans, W. (2007). Evaluating hybrid versus data-driven coreference resolution. In A. Branco (Ed.), *Lecture Notes in Computer Science* (Vol. 4410, pp. 137–150). Berlin, Germany: Springer.
- Hoste, V., Hendrickx, I., & Daelemans, W. (2007). Disambiguation of the neuter pronoun and its effect on pronominal coreference resolution. In V. Matousek & P. Mautner (Eds.), *Lecture Notes in Artificial Intelligence* (Vol. 4629, pp. 48–55). <https://doi.org/10.1007/978-3-540-74628-7>
- Hoste, V., Vanopstal, K., & Lefever, E. (2007). The automatic detection of scientific terms in patient information. In V. Pekar, D. Inkpen, & A. Mulloni (Eds.), *Proceedings of the international workshop on acquisition and management of multilingual lexicons* (pp. 41–48).
- Lefever, E., Fayruzov, T., & Hoste, V. (2007). AUG: a combined classification and clustering approach for web people disambiguation. *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, 105–108. Association for Computational Linguistics (ACL).
- Hoste, V., & De Pauw, G. (2006). KNACK-2002 : a richly annotated corpus of Dutch written text. *LREC 2006 : Fifth International Conference on Language Resources and Evaluation*, 1432–1437. Paris, France: European Language Resources Association (ELRA).
- Hoste, V., & Daelemans, W. (2006). Comparing learning approaches to language learning : there is more to it than “bias.” *Proceedings of BeNeLearn 2006*, 131–138. Ghent, Belgium: Ghent University.
- Hoste, V., & Daelemans, W. (2005). Learning Dutch coreference resolution. *Proceedings of the Fifteenth Computational Linguistics in the Netherlands Meeting (CLIN 2004)*, 133–148.
- Hoste, V., & Daelemans, W. (2005). Comparing learning approaches to coreference resolution : there is more to it than “bias.” *Proceedings of the ICML 2005 Workshop on Meta-Learning*, 20–27.
- Hoste, V. (2005). *Optimization issues in machine learning of coreference resolution*. Universiteit Antwerpen. Faculteit Letteren en Wijsbegeerte., Antwerp, Belgium.
- Decadt, B., Hoste, V., & Depauw, G. (Eds.). (2004). *Proceedings of the fourteenth computational linguistics in the Netherlands meeting*. Antwerp, Belgium: University of Antwerp.
- Decadt, B., Hoste, V., Daelemans, W., & van den Bosch, A. (2004). GAMBL, genetic algorithm optimization of memory-based WSD. *Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, 108–112. Association for Computational Linguistics.
- Hoste, V., Daelemans, W., & Gillis, S. (2004). Using rule-induction techniques to model pronunciation variation in Dutch. *COMPUTER SPEECH AND LANGUAGE*, 18(1), 1–23. [https://doi.org/10.1016/S0885-2308\(03\)00027-5](https://doi.org/10.1016/S0885-2308(03)00027-5)
Impact factor: 0.512, category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE, rank: 58/78.
- Daelemans, W., Hoste, V., De Meulder, F., & Naudts, B. (2003). Combined optimization of feature selection and algorithm parameters in machine learning of language. *Lecture Notes in Artificial Intelligence*, 2837, 84–95. https://doi.org/10.1007/978-3-540-39857-8_10
- Marsi, E., Reynaert, M., van den Bosch, A., Daelemans, W., & Hoste, V. (2003). Learning to predict pitch accents and prosodic boundaries in Dutch. *41st Annual Meeting of the Association for Computational Linguistics : Proceedings of the Conference*, 489–496. Somerset, NJ, USA: Association for Computational Linguistics.

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.
<https://doi.org/10.1017/S1351324902003005>

Daelemans, W., & Hoste, V. (2002). Evaluation of machine learning methods for natural language processing tasks. *LREC 2002 : Third International Conference on Language Resources and Evaluation*. Presented at the 3rd International conference on Language Resources and Evaluation (LREC 2002), Las Palmas, Spain.

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 ICSLP 2002*, 1273–1276.

De Meulder, F., Daelemans, W., & Hoste, V. (2002). A named entity recognition system for Dutch. In M. Theune, A. Nijholt, & H. Hondrop (Eds.), *Language and Computers : Studies in Practical Linguistics* (Vol. 45, pp. 77–88). Amsterdam, The Netherlands: Rodopi.

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. In P. Edmonds, R. Mihalcea, & P. Saint-Dizier (Eds.), *Word sense disambiguation : recent successes and future directions : proceedings of the ACL-02 workshop* (pp. 95–101). New Brunswick, NJ, USA: Association of Computer Linguistics (ACL).

Hendrickx, I., van den Bosch, A., Hoste, V., & Daelemans, W. (2002). Dutch word sense disambiguation: optimizing the localness of context. In P. Edmonds, R. Mihalcea, & P. Saint-Dizier (Eds.), *Word sense disambiguation : recent successes and future directions : proceedings of the ACL-02 workshop* (pp. 61–66). New Brunswick, NJ, USA: Association of Computer Linguistics (ACL).

Hoste, V., Kool, A., & Daelemans, W. (2001). Classifier optimization and combination in the English all words task. In J. Preiss & D. Yarowsky (Eds.), *Proceedings of the second international workshop on evaluating word sense disambiguation systems (SENSEVAL-2)* (pp. 83–86).

Hoste, V., & De Pauw, G. (Eds.). (2001). *Proceedings of the eleventh Belgian-Dutch conference on machine learning*. Antwerp, Belgium: University of Antwerp.

Hoste, V., & Daelemans, W. (2000). Comparing bagging and boosting for natural language processing tasks, a typicality approach. In A. Feelders (Ed.), *Proceedings 10th BeNeLearn meeting* (pp. 101–109).

Hoste, V., Daelemans, W., Tjong Kim Sang, E., & Gillis, S. (2000). Meta-learning for phonemic annotation of corpora. In A. van den Bosch & H. Weigand (Eds.), *Proceedings of the twelfth Belgium-Netherlands artificial intelligence conference* (pp. 331–332).

Hoste, V., Gillis, S., & Daelemans, W. (2000). Machine learning for modeling Dutch pronunciation variation. In P. Monachesi (Ed.), *Selected papers from the tenth CLIN meeting* (pp. 73–84). Utrecht, The Netherlands: Utrecht Institute of Linguistics OTS.

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. In S. Gilis, J. Nuyts, J. Taeldeman, & G. De Schutter (Eds.), *Met taal om de tuin geleid : een bundel opstellen voor Georges De Schutter ter gelegenheid van zijn pre-meritaat* (pp. 81–93). Antwerpen: Universitaire Instelling Antwerpen. Reprografie.

Hoste, V., Gillis, S., & Daelemans, W. (2000). A rule induction approach to modeling regional pronunciation variation. *COLING 2000 : 18th International Conference on Computational Linguistics : Proceedings of the Conference*, 327–333. Morgan Kaufmann.

Hoste, V., Daelemans, W., Tjong Kim Sang, E., & Gillis, S. (2000). Meta-learning for phonemic annotation of corpora. In P. Langley (Ed.), *Proceedings of ICML-2000* (pp. 375–382). San Francisco, CA, USA: Morgan Kaufmann.

