CURRICULUM VITAE

ORPHÉE DE CLERCQ

Contact	Department of Translation, Interpreting and Communication
	Faculty of Arts and Philosophy – Ghent University
	Groot-Brittanniëlaan 45
	B - 9000, Ghent
	Belgium
E–mail	orphee.declercq@ugent.be
Website	http://www.lt3.ugent.be/people/orphee-de-clercq/

About me

I'm a researcher in the field of Natural Language Processing with considerable experience of readability prediction and automatic analysis of subjective user-generated content using machine learning techniques. I'm very excited to live in a time where we can teach computers how to better understand human language and love how the current progress in the field often blows my mind. Besides that, I am intrigued by the many societal challenges arising from this digital (r)evolution.

1   Education

PhD Ghent University. (Computational) Linguistics (doctoraat in de taalkunde). 2015. Dissertation: Tipping the scales. Exploring the added value of deep semantic processing on readability prediction and sentiment analysis.

M.A. University College Ghent. Multilingual Communication (English, French). 2008. Master's Thesis: The use of metaphors in the language of two conservative movements: the American Christian Right and the Flemish nationalist party Vlaams Belang.

B.A. University College Ghent. Applied Linguistics (English, French). 2007.

2   Work Experience

August 2015 - current
Senior researcher (post-doctoral assistant) at Ghent University. I have been given the opportunity to further explore the research domains of:

1. Readability prediction: optimise Dutch and English machine learning systems, investigate domain portability, explore the link between readability and translatability and move towards Automated Writing Evaluation.

2. Fine-grained sentiment and emotion analysis: continue working on different domains and languages and try to incorporate transfer learning to this purpose. In 2017, I have also been able to valorize my research on aspect-based sentiment analysis during a VLAIO innovation project together with the Flemish company Hello Customer.

Currently, I also take an active research role in the following research projects:

- NewsDNA (2018-2022): Promoting news diversity: An interdisciplinary investigation into algorithmic design, personalization and the public interest.

- EmotioNL (2018-2022): Multilingual fine-grained emotion detection for business intelligence.

- PreDicT (2017-2021): Predicting Difficulty in Translation.

- COFI (2017-2021): Comprehensive Integration of Linguistic Features in a Human Inspired Speech Recognition Architecture.

October 2013 - July 2015
Research Assistant, Ghent University. Continuation on the PARIS project (IWT-SBO). Responsible for text mining – deep linguistic processing – of user-generated content in order to build personalized advertisements from web sources.

February 2014 - June 2014
Visiting Researcher, University of Mannheim. Worked at the Research Group Data and Web Science in the framework of a DAAD Research Project. Supervised by prof. dr. Simone Paolo Ponzetto. Worked on content-based recommender systems and 'semantifying' text using Linked Open Data (LOD).

January 2012 - September 2013
Research Assistant, University College Ghent. Worked on two research projects:

1. HENDI HOF project (January – October 2012). Responsible for creating a full-fledged readability prediction system and tool for Dutch.

2. PARIS IWT-SBO project (November 2012 – September 2013). "Personalised advertisements built from web sources" (IWT-SBO), start PhD project.

January 2011 - December 2011
Research Assistant, Utrecht University (0.8 fte). Worked on the STEVIN-funded SoNaR project. Responsible for the manual annotation of 500K text with Dutch semantic roles and the creation of an automatic semantic role labeler. Recruitment and supervision of student assistants.

Research Assistant, University College Ghent (0.2 fte). Worked on the Stylene project (EWI Flanders): "A web service for research in stylometry and readability prediction".

October 2008 - December 2010
Research Assistant, University College Ghent. Involvement in two STEVIN-funded research projects to consolidate the position of Dutch in the modern information and communication society:

1. Dutch Parallel Corpus Project DPC (October 2008 - June 2009). Responsible for data acquisition and semi-automatic control of each processing step (sentence alignment, tokenization, lemmatization, part-of-speech tagging) for the English-Dutch and Dutch-English language pairs. Supervision and recruitment of student assistants.

2. STEVIN Nederlandstalig Referentiecorpus Project SoNaR (October 2008 - December 2010). Responsible for the acquisition of Flemish text material. Responsible for the manual annotation of a one-million-word subset with coreference information and re-training and improving the automatic coreference resolution system COREA. Supervision and recruitment of student assistants.

3   Teaching Experience

- *Computer-Assisted Language Learning*

  (Master in Teaching - elective course, ca. 5 students), Department of Translation, Interpreting and Communication, Ghent University.
  October-December 2019: coordinator, all sessions, creation of all course material, individual feedback and evaluation.

- *Digital Communication Management*

  (Postgraduate course Computer-Assisted Language Mediation, ca. 10 students), Department of Translation, Interpreting and Communication, Ghent University.
  October-December 2016, 2017, 2019: coordinator, 6 out of 9 sessions, school Ambassador Digital Masters Academy or DMA program (extra-curricular training and student coaching).
  → *In 2017 my students won the second prize (out of 4,300 students) in the DMA program.*

- *Digital Communication - Blended learning*

  (Bachelor 3, Applied Language Studies, ca. 140 students), Department of Translation, Interpreting and Communication, Ghent University

  - March-May 2017, 2018, 2019 & 2020: creation of 4 e-learning environments, provide individual feedback to all students, 2 theoretical sessions and 2 follow-up sessions.
    → *Received a student evaluation score of 10 in 2018*
  - March-May 2016: creation of 4 e-learning environments, provide individual feedback to all students, all 4 follow-up sessions.

- *Introduction to Translation Technology - Practical course*

(Postgraduate course Dutch and Translation, ca. 30 students – all 6 sessions), Department of Translation, Interpreting and Communication, Ghent University, October-December 2015.

- *Introduction to Language Technology - Practical course*

  (Postgraduate Dutch and Translation, ca. 30 students – 4 out of 6 sessions), Faculty of Applied Language Studies, University College Ghent, October-December 2010 and 2013

- *Introduction to Language Technology - Practical course*

  (Bachelor 3, Applied Language Studies, ca. 140 students – all 6 sessions), Faculty of Applied Language Studies, University College Ghent, March-May 2013.

  → *Received an average of 8.4/10 out of 124 student evaluations.*

- *Master's Thesis advisor*

  - Jitse Criquielion (2020). The influence of topic interest on teens' reading comprehension.
  - Liesel Maertens (2020). Human rater versus automated writing evaluation: imminent clash or harmonious fusion?
  - Siel Debouver (2020). l'Analyse des sentiments appliquée sur des articles de journaux.
  - Tine Dhaene (2020). Demandez-le à Siri: La variation régionale linguistique influe-t-elle sur le fonctionnement de la version Flamande de Siri?
  - Senne Van Hoecke (2019). Towards automated précis grading. Linguistic features that might be good predictors: an exploratory study.
  - Silke Bekaert (2019). Whether you want it or not, bias does play a role in evaluation". An assessment of the reliability and validity of the writing evaluation method used for the course Nederlands Schrijfvaardigheid at Ghent University.
  - Manon Vermeersch (2018). Hotel agréable et bien situé. L'analyse des sentiments au niveau des aspects pour des avis sur des hotels en néerlandais et français.
  - Marie D'Hondt (2018). Verspätet – En retard. L'analyse de sentiments appliquée aux tweets au sujet de la SNCB et de la Deutsche Bahn.
  - Veera Voordeckers (2018). Un dîner interculturel. Une étude sur l'analyse de sentiments au niveau des aspects en français et russe.
  - Guillaume Goemanne (2017). Précise. Research into the possibility of a consistent error typology for the précis text genre as a first step towards AWE.
  - Enzo Neve (2017). Lekker, yummy, délicieux. Towards aspect-based sentiment analysis in Dutch, English and French.
  - Lenie Van Hecke (2016). Verdict de la Cour: Illisible. Rechercher l'influence des connaissances préalables pour juger la lisibilité.
  - Sofie Vandenhoven (2016). What does the bird say? A corpus study on what Twitter can teach us about personality.

- Margot Dedeyne (2015). LeTs get personal. Comparing manually created user profiles with automatic text processing techniques to derive personal interests from tweets.
- Valerie Dumortier (2014). How can a thorough linguistic analysis contribute to automatic sentiment analysis in Twitter?
- Rob Braet (2014). Careful what you tweet for! Can we determine personal preferences based on tweets using NLP tools and sentiment analysis.
- Lana Eeckhout (2013). From conversation to conversion. An explorative study on the adoption of social media and the application of sentiment analysis in Flemish media and web shops.
- Xenia Van Cotthem (2011). Est-il possible d'entraîner un étiqueteur morphosyntaxique aux textes non-normalisés? Une étude expérimentale du processus de pré-processing appliqué aux textos."

- Prepared sessions/workshops for prospecting high school students on "The thrills of NLP" , "Linguistics in the digital age", ... (2013-2019).

4 Grants and Awards

- 2020: granted an FWO award for a senior research project fundamental research (co-promotor) - G013820N. "ENCORE: Investigating cross-document EveNt Coreference REsolution for Dutch"

- 2020: granted an FWO Travel Fellowship to attend the LREC2020 conference at Marseille, France (conference cancelled because of COVID-19).

- 2016: granted an FWO Travel Fellowship to attend the NAACL2016 and SemEval2016 conference at San Diego, USA.

- 2014: granted a DAAD Contact Fellowship (Kontaktstipendien) for a 5-month research stay at the University of Mannheim to establish international collaborations between universities.

5 Organisation of Conferences, Workshops and Shared Tasks

- Co-organizer of WASSA 2019: Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis

- Co-organizer of the WASSA 2018 Implicit Emotions Shared Task.

- Co-organizer of SemEval 2016 - Task 5 Aspect-Based Sentiment Analysis.

- PARIS workshop, Heverlee, May 10, 2016.

- Atila 2014, Ghent, November 20-21, 2014.

- Atila 2010, Ostend, September 23-24, 2010.

- DPC workshop ("Dutch Parallel Corpus"), Antwerp, September 18, 2009.

6 Reviewing and Program Committees

- Full paper reviewing

  - 2021: area chair for EACL2021.
  - 2020: Reviewer for LREC2020, IJCAI 2020, EMNLP 2020, Language@Internet journal, Future Generation Computer Systems journal, PC member of the BEA 2020 workshop.
  - 2019: Reviewer for Applied Computing and Informatics journal. PC member of IJCAI 2019, ACL 2019, WASSA 2019, ESSLLI 2019 Student Session, RANLP 2019, EMNLP-IJCNLP 2019.
  - 2018: Reviewer for LREC2018, the ISI-NLP2 workshop, PLOS ONE journal, SemEval 2018. PC member of IJCAI-ECAI 2018, ESSLLI 2018 Student Session, COLING 2018, CMC-corpora 2018, WASSA 2018. Guest Editor for PLOS ONE journal.
  - 2017: PC member of The 3rd International Workshop on Natural Language Processing for Informal Text, PC member of the EACL 2017 Student Research Workshop, Reviewer for HUSO2017, RANLP2017, Journal of Information Processing and Management.
  - 2016: Reviewer for EMOSEDE Session at HUSO2016 conference, PC member of the ISWC Poster & Demos Session, Reviewer for LREC2016, SemEval2016, PC member of The 2nd International Workshop on Natural Language Processing for Informal Text, Dutch data provider for SemEval 2016 Task 5 on Aspect-based Sentiment Analysis.
  - 2015: Reviewer for SemEval2015, RANLP2015, Journal of Natural Language Engineering.
  - 2014: Reviewer for LREC2014, Journal of Natural Language Engineering.
  - 2013: Reviewer for RANLP2013, secondary reviewer for CICLING2013, ACL2013, SocInfo2013.
  - 2012: PC member of Lingvisticae Investigationes (Special Issue on SMS Communication: A linguistic approach), secondary reviewer for PITR2012.

- Project reviewing

  - 2016: Scientific reviewer for the Flanders Innovation & Entrepeneurship government agency (VLAIO).

7 PhD Guidance and Commissions

- Guidance

  - Co-promotor of Luna De Bruyne who prepares an FWO-SB PhD on transfer learning for emotion detection.
  - Co-promotor of Bram Vanroy who prepares a PhD in the framework of the PreDicT project (Predicting difficulty in translation).

6

- Examination Commission

  - Jury member, examination commission of Nils Smeuninckx (UGent): Dear Stakeholder. Exploring the language of sustainability reporting, a closer look at readability, sentiment and perception. (2018)
  - Jury member, examination commission of Suzanne Kleijn (Utrecht University): Clozing in on readability. How linguistic features affect and predict text comprehension and on-line processing. (2018)
  - Rapporteur, examination commission of Arda Tezcan (UGent): Informative Quality Estimation of Machine Translation Output. (2018)
  - Jury member, examination commission of Cynthia Van Hee (UGent): Can machines sense irony? Exploring automatic irony detection on social media. (2017)
  - Rapporteur, examination commission of Sabien Hanoulle (UGent): Translating documentaries: Does the integration of a bilingual glossary of domain-specific terminology into the translation process reduce the translator's workload? (2017)
  - Rapporteur, examination commission of Joke Daems (UGent): A translation robot for each translator? A comparative study of manual translation and post-editing of machine translations: process, quality and translator attitude. (2016)

8 Additional Training

- 2018:*Research School 'Facets of Writing and Writing Research: Substantive and Methodological Issues'* (Sig Writing 2018, EARLI special interest group on writing)

- 2017: *Online Marketing Fundamentals Certificate, Google Analytics Certificate* (Google DMA), Basisdocententraining (UGent), Academic Leadership Programme for Postdocs (True Colours cvba, UGent postdoc training)

- 2016: *Creativity, Pitching & Resilience* (Streetwize, UGent postdoc training)

- 2015: *Technical Writing* (Teka Info Pilots)

- 2013: *Project Management, Effective Slide Design* (UGent Doctoral Schools)

- 2012: *Reference and Coherence in Discourse* and *Using Bayesian methods to model cognition* (LOT Summer School); *Introduction to R* and *Introductory Statistics* (IPVW – UGent Doctoral Schools)

- 2009: *Introduction to Language Technology, Advanced Methods for Corpus Linguistics* (UGent Doctoral Schools).

9 Membership and Service

- 2018-: Member of the Study Programme Committee, Department of Translation, Interpreting and Communication, Ghent University

- 2018-: Member of the Departmental Board, Department of Translation, Interpreting and Communication, Ghent University

- 2018-: Member of DH Research Lab, Ghent Centre for Digital Humanities, Ghent University

10   Publications

- Journal articles

  - A1 (ISI Web of Science)

    * Bram Vanroy, **Orphée De Clercq** and Lieve Macken (2019). Correlating Process and Product Data to Get an Insight into Translation Difficulty. *Perspectives: Studies in Translation Theory and Practice*, 27:6, 924-941.

    * Catherine Bouko, July De Wilde, Sofie Decock, **Orphée De Clercq**, Valentina Manchia and David Garcia (to appear). Reactions to Brexit in images: a multimodal content analysis of shared visual content on Flickr. *Visual Communication*, available online: `http://journals.sagepub.com/doi/10.1177/1470357218780530`

    * Gert De Sutter, Bert Cappelle, **Orphée De Clercq**, Rudy Loock and Koen Plevoets (2017). Towards a corpus-based, statistical approach of translation quality : measuring and visualizing linguistic deviance in student translations. *Linguistica Antverpiensia New Series – Themes in Translation Studies*, 16, 25-39.

    * Joke Daems, **Orphée De Clercq** and Lieve Macken (2017). Translationese and Post-editese: How Comparable is Comparable Quality? *Linguistica Antverpiensia New Series – Themes in Translation Studies*, 16, 89-103.

    * Catherine Bouko, Laura Calabrese and **Orphée De Clercq** (2017). Cartoons as interdiscourse: a quali-quantitative analysis of social representations based on collective imagination in cartoons produced after the Charlie Hebdo attack. *Discourse, Context & Media*, 15, 24-33.

    * **Orphée De Clercq** and Véronique (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. IF 2016: 2.528

    * Sarah Schulz, Guy De Pauw, **Orphée De Clercq**, Bart Desmet, Véronique Hoste, Walter Daelemans, Lieve Macken (2016). Multi-Modular Text Normalization of User-Generated Content. *Transactions on Intelligent Systems and Technology*, 7(4), 1-22 . IF 2016: 3.196

    * **Orphée De Clercq**, Véronique Hoste, Bart Desmet, Philip van Oosten, Martine De Cock and Lieve Macken (2014) Using the Crowd for Readability Prediction. *Natural Language Engineering*, 20(3), 293-325. Cambridge University Press. IF 2014: 0.639

    * Lieve Macken, **Orphée De Clercq** and Hans Paulussen (2011) Dutch Parallel Corpus: a Balanced Copyright-Cleared Parallel Corpus. *Meta*, 56(2), 374-390. Les Presses de l'Université de Montréal.

  - A2

    * Cynthia Van Hee, Marjan Van de Kauter, **Orphée De Clercq**, Els Lefever, Bart Desmet, Véronique Hoste (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.

- A3

  * Maribel Montero Perez, **Orphée De Clercq**, Piet Desmet, Geert Peeters and Serge Verlinde (2009) Dutch parallel corpus: un nouveau corpus parallèle multilingue disponible en ligne. Romaneske, 34 (4), 2-8.

- A4

  * **Orphée De Clercq** (2016). Van tekstherkenning naar tekstbegrip. Computers en plagiaatdetectie. Karakter, 53.

- Books

  - B2

    * Walter Daelemans, **Orphée De Clercq** and Véronique Hoste (2018). STYLENE: an Environment for Stylometry and Readability Research for Dutch. In J. Odijk (ed.), CLARIN in the Low Countries. Ubiquity Press.

    * **Orphée De Clercq** and Véronique Hoste (2014). Hoe meetbaar is leesbaarheid? In S. Evenepoel, P. Goethals and L. Jooken (eds.), Beschouwingen uit een talenhuis, 147-155. Academia Press, Ghent, Belgium.

    * Bart Desmet, **Orphée De Clercq**, Marjan Van de Kauter, Sarah Schulz, Cynthia Van Hee and Véronique Hoste (2014). Taaltechnologie 2.0: sentimentanalyse en normalisatie. In S. Evenepoel, P. Goethals and L. Jooken (eds.), Beschouwingen uit een talenhuis, 157-161. Academia Press, Ghent, Belgium.

    * Lieve Macken, **Orphée De Clercq**, Bart Desmet and Véronique Hoste (2014). Dutch Parallel Corpus en SoNaR. In S. Evenepoel, P. Goethals and L. Jooken (eds.), Beschouwingen uit een talenhuis, 163-170. Academia Press, Ghent, Belgium.

  - B3

    * Alexandra Balahur, Roman Klinger, Véronique Hoste, Carlo Strapparava and **Orphée De Clercq** (2019). Proceedings of the 10th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis - NAACL, Minneapolis, USA.

- Conference Proceedings

  - P1 (ISI Web of Science)

    * **Orphée De Clercq** and Véronique Hoste (2016), Rude waiter but mouthwatering pastries! An exploratory study into Dutch Aspect-Based Sentiment Analysis. Proceedings of the 10th Language Resources and Evaluation Conference (LREC2016).

    * **Orphée De Clercq**, Sarah Schulz, Bart Desmet and Véronique Hoste (2014), Towards Shared Datasets for Normalization Research. Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014), Reykjavik, Iceland.

* **Orphée De Clercq**, Véronique Hoste and Paola Monachesi (2012), Evaluating Automatic Cross-Domain Semantic Role Annotation. Proceedings of the 8th Language Resources and Evaluation Conference (LREC2012), Istanbul, Turkey.

* Maaske Treurniet, **Orphée De Clercq**, Henk van den Heuvel and Nelleke Oostdijk (2012), Collecting a Corpus of Dutch SMS. Proceedings of the 8th Language Resources and Evaluation Conference (LREC2012), Istanbul, Turkey.

* **Orphée De Clercq** and Maribel Montero Perez (2010), Data Collection and IPR in Multilingual Parallel Corpora: Dutch. Parallel Corpus. Proceedings of the 7th Language Resources and Evaluation Conference (LREC2010), Valletta, Malta.

* Martin Reynaert, Nelleke Oostdijk, **Orphée De Clercq**, Henk van den Heuvel and Franciska de Jong (2010), Balancing SoNaR: IPR versus Processing Issues in a 500-Million-Word Written Dutch Reference Corpus. Proceedings of the 7th Language Resources and Evaluation Conference (LREC2010), Valletta, Malta.

– C1

* **Orphée De Clercq** and Senne Van Hoecke (2020). An Exploratory Study into Automated Précis Grading. Proceedings of the 12th Language Resources and Evaluation Conference (LREC2020).

* Luna De Bruyne, **Orphée De Clercq** and Véronique Hoste (2020). An Emotional Mess! Deciding on a Framework for Building a Dutch Emotion-Annotated Corpus. Proceedings of the 12th Language Resources and Evaluation Conference (LREC2020).

* Joris, Glen, Camiel Colruyt, Judith Vermeulen, Stefaan Vercoutere, Frederik De Grove, Kristin Van Damme, **Orphe De Clercq**, et al. (2020). News Diversity and Recommendation Systems?: Setting the Interdisciplinary Scene. In Privacy and Identity Management. Data for Better Living?: AI and Privacy, 576:90105. Springer.

* Claudia Matoz Veliz, **Orphée De Clercq** and Véronique Hoste (2019). Benefits of Data Augmentation for NMT-Based Text Normalization of User-Generated Content. Proceedings of the 2019 EMNLP-IJCNLP Workshop W-NUT: The 5th Workshop on Noisy User-Generated Text.

* Camiel Colruyt, **Orphée De Clercq** and Véronique Hoste (2019). Leveraging syntactic parsing to improve event annotation matching. Proceedings of the 2019 EMNLP-IJCNLP Workshop AnnoNLP: The First Workshop on Aggregating and Analysing Crowdsourced Annotations for NLP.

* Claudia Matos Veliz, **Orphée De Clercq** and Véronique Hoste (2019). Comparing MT Approaches for Text Normalization. Proceedings of the 12th International Conference on Recent Advances in Natural Language Processing (RANLP2019).

* Luna De Bruyne, **Orphée De Clercq** and Véronique Hoste (2019). Towards an Empirically-Grounded Framework for Emotion Analysis. Proceedings of the Second International Conference on Human and Social Analytics (HUSO).

* Roman Klinger, **Orphée De Clercq**, Saif M. Mohammad and Alexandra Balahur (2018). IEST: WASSA-2018 Implicit Emotions Shared Task. Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis at EMNLP 2018.

* Bram Vanroy, **Orphée De Clercq** and Lieve Macken (2018). Predicting difficulty in translation?: a pilot study. In T. Read, S. Montaner, B. Sedano (Eds.), Technological innovation for specialized linguistic domains: languages for digital lives and cultures, proceedings of TISLID18 (pp. 163-174). Presented at the Third conference on Technological Innovation for Specialized Linguistic Domains (TISLID18), Beau Bassin, Mauritius: Éditions Universitaires Européennes.

* Luna De Bruyne, **Orphée De Clercq** and Véronique Hoste (2018). LT3 at SemEval-2018 Task 1: A classifier chain to detect emotions in tweets. Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2018) at NAACL 2018.

* **Orphée De Clercq**, Els Lefever, Gilles Jacobs, Tijl Carpels and Véronique Hoste (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 at EMNLP 2017.

* **Orphée De Clercq** (2016). The many aspects of fine-grained sentiment analysis. An overview of the task and its main challenges. Proceedings of the Second International Conference on Human and Social Analytics (HUSO). ISSN 2519-8351.

* Sofie Vandenhoven and **Orphée De Clercq** (2016). What does the bird say? Exploring the Link Between Personality and Language Use in Dutch Tweets. Proceedings of HUSO 2016, The Second International Conference on Human and Social Analytics. ISSN 2519-8351.

* Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, **Orphée De Clercq**, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukakevitch, Evgeny Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra and Gülşen Eryiğit (2016). SemEval-2016 Task 5: Aspect Based Sentiment Analysis. Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) at NAACL-HTL.

* **Orphée De Clercq**, Marjan Van de Kauter, Els Lefever and Véronique Hoste (2015), LT3: Applying Hybrid Terminology Extraction to Aspect-Based Sentiment Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval-2015) at NAACL.

* **Orphée De Clercq**, Michael Schuhmacher, Simone Paolo Ponzetto and Véronique Hoste (2014), Exploiting FrameNet for Content-Based Book Recommendation. Proceedings of the CBRecsys workshop at ACM RecSys2014. CEUR-1245.

* **Orphée De Clercq**, Sven Hertling, Véronique Hoste, Simone Paolo Ponzetto and Heiko Paulheim (2014), Identifying Disputed Topics in the News. Proceedings of the LD4KD workshop at ECML/PKDD2014. CEUR-1232.

* Cynthia Van Hee, Marjan Van de Kauter, **Orphée De Clercq**, Els Lefever and Véronique Hoste (2014), LT3: Sentiment Classification in user-generated content using a rich feature set. Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval-2014), Dublin, Ireland.
* **Orphée De Clercq**, Sarah Schulz, Bart Desmet, Els Lefever and Véronique Hoste (2013), Normalisation of Dutch User-Generated Content. Proceedings of the 9th International Conference on Recent Advances in Natural Language Processing (RANLP2013).
* **Orphée De Clercq**, Véronique Hoste and Iris Hendrickx (2011), Cross-Domain Dutch Coreference Resolution. Proceedings of the 8th International Conference on Recent Advances in Natural Language Processing (RANLP 2011).
* Iris Hendrickx, **Orphée De Clercq** and Véronique Hoste (2011), Analysis and Reference Resolution of Bridge Anaphora across Different Genres. Lecture Notes in Artificial Intelligence, 7099. Springer – Verlag.

- Dissertation

  - **Orphée De Clercq** (2015), Tipping the scales. Exploring the added value of deep semantic processing on readability prediction and sentiment analysis. Ghent University, Faculty of Arts and Philosophy.

- Other

  - **Orphée De Clercq**, Bart Desmet and Véronique Hoste (2014). Guidelines for Normalizing Dutch and English User Generated Content. LT3 Technical Report LT3 14.01. Version 1.0

  - Nelleke Oostdijk, Véronique Hoste, Franciska de Jong, Martin Reynaert, **Orphée De Clercq**, Bart Desmet, Henk van den Heuvel (2014). SoNaR-500. Centrale voor Taal- en Spraaktechnologie, ISBN 9789491325014.

  - Nelleke Oostdijk, Martin Reynaert, Véronique Hoste, Henk van den Heuvel, Paola Monachesi, Bart Desmet, **Orphée De Clercq**, Ineke Schuurman (2014). SoNaR-1. Centrale voor Taal- en Spraaktechnologie, ISBN 9789491325021.

  - Nelleke Oostdijk, Martin Reynaert, Véronique Hoste, Henk van den Heuvel, **Orphée De Clercq**, Eric Sanders (2014). SoNaR Nieuw Media Corpus. Centrale voor Taal- en Spraaktechnologie, ISBN 9789491325038.

  - Piet Desmet, Willy Vandeweghe, Hans Paulussen, Lieve Macken, Maribel Montero Perez, **Orphée De Clercq**, Lidia Rura and Julia Trushkina (2011). Dutch Parallel Corpus. Instituut voor Nederlandse Lexicologie, ISBN 9789491325007.

  - **Orphée De Clercq** and Martin Reynaert (2010), SoNaR Acquisition Manual. LT3 Technical Report LT3 10-02. Version 1.0

11 Presentations, Conference Contributions

- **Invited:**

- poster presentation on "Intricate Sentiment Analysis of Customer Feedback" at AI4Growth Session 4 - Retail, B2C & Customer exp., Ghent, Belgium, January 28, 2019.
- presentation on "Aspect-based sentiment analysis for customer reviews" at Hands-on Digital Day 2019 (Department of Communication Sciences, Ghent University), Ghent, Belgium, May 6, 2019.
- presentation on "Generic Readability Prediction" at the workshop on Predicting and Improving Readability, Utrecht, The Netherlands, April 5, 2018.
- presentation on "Staan linguïsten en juristen lijnrecht tegenover elkaar wanneer het aankomt op leesbaarheid?" at the symposium Begrijpelijke taal in het privaatrecht, Leiden, December 1, 2017.

  tutorial on "From Easy-to-calculate Formulas to Holistic Readability Prediction" (together with Véronique Hoste) at the RANLP conference, Varna, Bulgaria, September 2, 2017.

  presentation on "Lekker, yummy, délicieux. Fine-grained sentiment analysis of customer reviews" at the 3rd Belgium NLP Meetup, Ghent, Belgium, May 18, 2017.

  keynote on "The many aspects of fine-grained sentiment analysis. An overview of the task and its main challenges" at HUSO conference, Barcelona, Spain, November 15, 2016.

  presentation on "*UCL: great beer at the 'cercles' but very dirrrty!* Aspect-based sentiment analysis of customer reviews." at Séminaires du CENTAL, UCL, Louvain-la-Neuve, Belgium, October 28, 2016.

  presentation on "Overview of my readability research", DWS Colloquium, Mannheim, Germany, February 14, 2014.

- **Main presenter:**

  - presentation on "Uncovering Machine Translationese: an experiment on 4 MT systems for English-French translations" (co-presentation with Rudy Loock) at TQ2020, Roubaix, France, January 31, 2020.
  - accepted abstract, presentation on "Towards Dutch Automated Writing Evaluation" at CLIN30, Utrecht, the Netherlands, January 30, 2020.
  - accepted abstract, presentation on "EventDNA: Identifying event mention spans in Dutch-language news text" at CLIN30, Utrecht, the Netherlands, January 30, 2020.
  - accepted paper, poster presentation on "Benefits of Data Augmentation for NMT-Based Text Normalization of User-Generated Content" at EMNLP-IJCNLP 2019 - W-NUT workshop, Hong Kong, November 4, 2019.
  - accepted abstract (on invitation), presentation on "Machine learning for readability prediction" at DICOEN 2019 conference, Panel 5: Readability: methodological advances and practical applications, Leuven, Belgium, June 4, 2019.
  - accepted abstract, presentation on "OMG het was zooo lekker! Can fine-grained sentiment analysis benefit from coreference resolution?" at CLIN 28, Nijmegen, The Netherlands, January 26, 2018.

– poster presentation on "Towards an integrated pipeline for aspect-based sentiment analysis" at the LW Research Day, Ghent, Belgium, November 29, 2017.

– accepted abstract, presentation on "Hello ABSA! Collaboration between LT3 and Hello Customer to develop an aspect-based sentiment analysis pipeline" at the Atila Meeting, Tilburg, The Netherlands, November 17, 2017.

– presentation of demo "Readability prediction" at the Atila Meeting, Tilburg, The Netherlands, November 16, 2017.

– accepted paper, poster presentation on "Towards an integrated pipeline for aspect-based sentiment analysis in various domains" at the 8th WASSA workshop at EMNLP2017, Kopenhagen, Denmark, September 8, 2017.

– accepted abstract, poster presentation on "Court verdict: unreadable" and presentation on "Translationese and Posteditese: Can users and computers detect the difference?" (co-presentation with Joke Daems) at CLIN 27, Leuven, Belgium, February 9, 2017.

– two accepted papers, presentation on "What does the bird say? Exploring the Link Between Personality and Language Use in Dutch Tweets" at HUSO conference, EMOSEDE session, Barcelona, Spain, November 15, 2016.

– accepter paper, poster presentation, shared task co-organizer of SemEval 2016 ABSA: aspect-based sentiment analysis at NAACL2016, San Diego, USA, June 16, 2016.

– accepted paper, poster presentation on "Rude waiter but mouthwatering pastries! An exploratory study into Dutch Aspect-Based Sentiment Analysis" at LREC2016, Portorez, Slovenia, May 26, 2016.

– two presentations on "Wot r U talkin bout?! Transform noisy user-generated text to something readable" and "Lovely dress but such a mess: Discover how customers talk about your company/store/product at PARIS workshop, Heverlee, Belgium, May 10, 2016.

– accepted abstract, poster presentation on "Sustainability report readability prediction: an NLP expansion on the tried-and-true formulae" (co-presentation with Nils Smeuninx) at CLIN 26, Amsterdam, The Netherlands, December 18, 2015.

– lightning talk on "Participate in this year's SemEval 2016 Task on multilingual aspect-based sentiment analysis" at Atila meeting, Antwerp, Belgium, October 15, 2015.

– presentation on "Terminology Extraction to Aspect-Based Sentiment Analysis" at CLIN 2015, Antwerp, Belgium, February 6, 2015.

– presentation on, "Content-Based Book Recommendation", Atila meeting, Ghent, Belgium, November 20, 2014.

– accepted paper, presentation on, "Exploiting FrameNet for Content-Based Book Recommendation" at the CBRecsys workshop, Recsys 2014, Foster City, USA, October 5, 2014.

– accepted paper, presentation on "Identifying Disputed Topics in the News at the LD4KD workshop, ECML 2014, Nancy, France, September 19, 2014.

- accepted paper, presentation on "Towards Shared Datasets for Normalization Research" at (co-presentation with Sarah Schulz) at LREC 2014, Reykjavik, Iceland, May 28, 2014.

- presentation on "Crash-testing text processing on user-generated content" at CLIN 2014, Leiden, The Netherlands, January 17, 2014.

- poster presentation on "Normalization of Dutch User-Generated Content" (co-presentation with Sarah Schulz and Bart Desmet) at CLIN 2014, Leiden, The Netherlands, January 17, 2014.

- presentation on "Automatic Normalization of Dutch User-Generated Content" at Onderzoeksforum Faculty of Applied Language Studies, Ghent, Belgium, December 10, 2013.

- presentation on "Mining through the noise" at LTCI seminar, Ghent, Belgium, December 6, 2013.

- presentation on "Plain Language: Investigating the effect of different text characteristics on text readability" at 3rd Translation Studies Day@European Commission, Brussels, Belgium, October 25, 2013.

- presentation on "First steps in normalizing Dutch UGC" at Atila Meeting, Corsendonk, Belgium, September 27, 2013.

- accepted paper, presentation on "Normalisation of Dutch User-Generated Content" at RANLP2013, Hissar, Bulgaria, September 10, 2013.

- presentation on "HENDI: holistic readability prediction" at Onderzoeksforum Faculty of Applied Language Studies, Ghent, Belgium, April 22, 2013.

- accepted abstract, presentation on "Features for automatic readability prediction" at CLIN 2013, Enschede, The Netherlands, January 18, 2013.

- presentation on "Dutch readability prediction" at LTCI Meeting, Ghent, Belgium, December 14, 2012.

- presentation on "Comparing features for automatic readability prediction" at Atila Meeting, Groesbeek, The Netherlands, November 23, 2012.

- poster presentation on "SoNaR1: a one-million-word Dutch semantic resource", LOT Summer School, Utrecht, The Netherlands, July 2, 2012.

- accepted paper, poster presentation on "Collecting a corpus of Dutch SMS" at LREC 2012, Istanbul, Turkey, May 24, 2012.

- accepted paper, presentation on "Evaluating automatic cross-domain Dutch semantic role annotation" at LREC 2012, Istanbul, Turkey, May 23, 2012.

- accepted abstract, presentation on "Evaluating automatic cross-domain Dutch semantic role annotation" at CLIN 2012, Tilburg, The Netherlands, January 20, 2012.

- presentation on "Coreference resolution across different genres and the special case of bridge relations" at Atila Meeting, Antwerp, Belgium, December 1, 2011.

- accepted paper, presentation on "Cross-Domain Dutch Coreference Resolution" at RANLP 2011, Hissar, Bulgaria, September 13, 2011.

- poster presentation on "Semantic annotation in SoNaR" at Doctorandidag Hogeschool Gent, Ghent, Belgium, March 10, 2011.

- accepted abstract, presentation on "Cross-Domain Dutch Coreference Resolution" at CLIN 2011, Belgium, Ghent, February 11, 2011.

- presentation on "SoNaR Stevin Nederlandstalig Referentiecorpus" at STEVIN-programmadag, Antwerp, Belgium, October 14, 2010.

- accepted paper, poster presentation on "Data Collection and IPR in Multilingual Parallel Corpora" at LREC 2010, Valletta, Malta, May 21, 2010.

- accepted paper, poster presentation on "Balancing SoNaR: IPR versus Processing Issues in a 500-Million-Word Written Dutch Reference Corpus" at LREC 2010, Valletta, Malta, May 20, 2010.

- accepted abstract, presentation on "Acquiring a Reference Corpus of Written Dutch" at CLIN 2010, Utrecht, The Netherlands, February 5, 2010.

- accepted abstract, presentation on "Dutch Parallel Corpus: a new parallel corpus with Dutch as central language" at CLIN 2010, Utrecht, The Netherlands, February 5, 2010.

- presentation on "Acquisition and IPR" (co-presentation with Maribel Montero Perez), DPC workshop, Antwerp, Belgium, September 18, 2009.

- **Not main presenter**

  - accepter paper, presentation on "Leveraging syntactic parsing to improve event annotation matching" (with Camiel Colruyt and Véronique Hoste) at EMNLP-IJCNLP 2019 - AnnoNLP workshop, Hong Kong, November 3, 2019.

  - accepted paper, poster presentation on "Comparing MT Approaches for Text Normalization" (with Claudia Matoz Veliz and Véronique Hoste) at RANLP 2019, Varna, Bulgaria, September 2, 2019.

  - accepted abstract, presentation on "Modelling Translation Entropy and Syntactic Equivalence" (with Bram Vanroy and Lieve Macken) at the MEMENTO workshop - Machine Translation Summit, Dublin, Ireland, August 20, 2019.

  - accepted paper, presentation on "Towards an Empirically Grounded Framework for Emotion Analysis" (with Luna De Bruyne and Véronique Hoste) at EMOSEDE - HUSO, Rome, Italy, July 1, 2019.

  - accepted abstract, poster presentation on "Transfer Learning for Spanish Emotion Detection" (with Luna De Bruyne and Véronique Hoste) at CLIN 29, Groningen,The Netherlands, January 31, 2019.

  - accepted abstract, presentation on "Varying Background Corpora for SMT-based Text Normalization" (with Claudia Matos Veliz and Véronique Hoste) at CLIN 29, Groningen,The Netherlands, January 31, 2019.

  - accepted abstract, poster presentation on "Varying Background Corpora for SMT-based Text Normalization" (with Claudia Matos Veliz and Véronique Hoste) at CMC and Social Media Corpora 2018, Antwerp, Belgium, September 18, 2018.

- accepted paper, poster presentation on "LT3 at SemEval-2018 Task 1: A classifier chain to detect emotions in tweets." (with Luna De Bruyne, Véronique Hoste) at SemEval 2018, New Orleans, USA, June 5, 2018.

- accepted abstract, presentation on "Predicting Difficulty in Translation: A Pilot Study" (with Bram Vanroy and Lieve Macken) at the TISLID'18 conference, Ghent, Belgium, May 24, 2018.

- poster presentation on "Predicting difficulty in translation" (with Bram Vanroy and Lieve Macken) at the LW Research Day, Ghent, Belgium, November 29, 2017.

- accepted abstract, presentation on "A corpus-based, statistical approach to translation quality" (with Bert Cappelle, Gert De Sutter, Rudy Loock and Koen Plevoets) at Traduction Qualité 2016: corpus qualité, Lille, France, February 5, 2016.

- accepted paper, poster presentation on "LT3: Applying Hybrid Terminology Extraction to Aspect-Based Sentiment Analysis" (with Els Lefever and Véronique Hoste), SemEval2015, Denver, USA, June 4, 2015.

- accepted paper, poster presentation on "LT3: Sentiment Classification in user-generated content using a rich feature set" (with Cynthia Van Hee, Marjan Van de Kauter, Els Lefever and Véronique Hoste), Dublin, Ireland, August 24, 2014.

- presentation on "Crowdsourcing for Dutch Readability Prediction" (with Véronique Hoste), Workshop on Writing and Readability Tools, Utrecht, The Netherlands, January 31, 2014.

- accepter paper, presentation on "Analysis and Reference Resolution of Bridge Anaphora across Different Genres" (with Iris Hendrickx and Véronique Hoste), DAARC 2011, Faro, Portugal, October 7, 2011.

- **Chair**

  - session chair *Named Entity Recognition and Term Extraction* at CLIN 28, Nijmegen, The Netherlands, January 26, 2018.

  - session chair *Language acquisition & psycholinguistics* at CLIN 27, Leuven, Belgium, February 9, 2017.

  - session chair *SRL/Semantic Parsing* at CLIN 26, Amsterdam, The Netherlands, December 18, 2015.

  - session chair *Neural Networks* at Atila Meeting, Antwerp, Belgium, October 15, 2015.

  - session chair at Atila Meeting, Corsendonk, Belgium, September 27, 2013.