

## Publications

- 2018 **Egon W. Stemle, Adriane Boyd, Maarten Janssen, Therese Lindström Tiedemann, Nives Mikelić Preradović, Alexandr Rosen, Dan Rosén, and Elena Volodina**, *Working together towards an ideal infrastructure for language learner corpora*.
- 2018 **Jennifer-Carmen Frey, Egon W. Stemle, and A. Seza Doğruöz**, *Comparison of Automatic vs. Manual Language Identification in Multilingual Social Media Texts*.
- 2018 **Willem Elbers, Egon W. Stemle, André Moreira, Alexander König, Luca Cattani, and Martin Palma**, *The CLARIN ERIC deployment infrastructure and its applicability to reproducible research*.
- 2018 **Andrea Abel and Egon W. Stemle**, *On the Detection of Neologism Candidates as Basis for Language Observation and Lexicographic Endeavours: The STyrLogism Project*.
- 06.2018 **Egon Stemle and Alexander Onysko**, *Using Language Learner Data for Metaphor Detection*, In *Proceedings of the Workshop on Figurative Language Processing*, pages 133–138.
- 05.2018 **Lukas Egarter Vigl and Egon Stemle**, *Was darf Forschung mit Social Media Daten?*, magazine, Academia-Interview Titelthema.
- 10.2017 **Michael Beißwenger, Ciara R. Wigham, Carole Etienne, Darja Fišer, Holger Grunt Suárez, Laura Herzberg, Erhard Hinrichs, Tobias Horsmann, Natali Karlova-Bourbonous, Lothar Lemnitzer, Julien Longhi, Harald Lungen, Lydia-Mai Ho-Dac, Christophe Parisse, Céline Poudat, Thomas Schmidt, Egon Stemle, Angelika Storrer, and Torsten Zesch**, *Connecting Resources: Which Issues have to be Solved to Integrate CMC Corpora from Heterogeneous Sources and for Different Languages?*, In Stemle and Wigham, editors, *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities*.
- 10.2017 **Egon W. Stemle and Ciara R. Wigham, editors**, *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities*.
- 05.2017 **Michael Beißwenger, Thierry Chanier, Tomaž Erjavec, Darja Fišer, Axel Herold, Nikola Lubešić, Harald Lungen, Céline Poudat, Egon Stemle, Angelika Storrer, and Ciara Wigham**, *Closing a Gap in the Language Resources Landscape: Groundwork and Best Practices from Projects on Computer-mediated Communication in four European Countries*, In *Selected Papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016, CLARIN Common Language Resources and Technology Infrastructure*, pages 1–18.
- 12.2016 **Egon W Stemle**, *bot.zen @ EVALITA 2016 - A minimally-deep learning PoS-tagger (trained for Italian Tweets)*, In Basile et al., editors, *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016)*.
- 12.2016 **Jennifer-Carmen Frey, Aivars Glaznieks, and Egon W. Stemle**, *The DiDi Corpus of South Tyrolean CMC Data: A multilingual corpus of Facebook texts*, In Basile et al., editors, *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016)*.
- 12.2016 **Andrea Abel, Aivars Glaznieks, Lionel Nicolas, and Egon Stemle**, *An extended version of the KoKo German L1 Learner corpus*, In Basile et al., editors, *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016)*.

- 10.2016 **Michael Beißwenger, Thierry Chanier, Isabella Chiari, Tomaž Erjavec, Darja Fišer, Axel Herold, Nikola Lubešić, Harald Lungen, Céline Poudat, Egon Stemle, Angelika Storrer, and Ciara Wigham**, *Integrating corpora of computer-mediated communication into the language resources landscape: Initiatives and best practices from French, German, Italian and Slovenian projects*, In *Proceedings of the CLARIN Annual Conference 2016*.
- 08.2016 **Egon W. Stemle**, *bot.zen @ EmpiriST 2015 - A minimally-deep learning PoS-tagger (trained for German CMC and Web data)*, In *Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task*, pages 115–119.
- 08.2016 **Paul Cook, Stefan Evert, Roland Schäfer, and Egon Stemle, editors**, *Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task*, Workshop at ACL2016, Berlin, Germany.
- 09.2015 **Jennifer-Carmen Frey, Aivars Glaznieks, and Egon W. Stemle**, *The DiDi Corpus of South Tyrolean CMC Data*, In *Proceedings of the 2nd Workshop on Natural Language Processing for Computer-Mediated Communication / Social Media at GSCL2015 (NLP4CMC2015)*.
- 06.2015 **Lionel Nicolas, Egon Stemle, Aivars Glaznieks, and Andrea Abel**, *A Generic Data Workflow for Building Annotated Text Corpora*, In Castello et al., editors, *Studies in Learner Corpus Linguistics: Research and Applications for Foreign Language Teaching and Assessment*, volume 190 of *Linguistic Insights*, chapter 18, pages 337–351.
- 04.2015 **Egon Stemle and Alexander Onysko**, *Automated L1 identification in English learner essays and its implications for language transfer*, volume 4 of *Hamburg Studies on Linguistic Diversity*.
- 12.2014 **Aivars Glaznieks and Egon Stemle**, *Challenges of building a CMC corpus for analyzing writer's style by age: The DiDi project*, volume 29.
- 12.2014 **Aivars Glaznieks, Andrea Abel, Verena Lyding, Lionel Nicolas, and Egon Stemle**, *Establishing a Standardised Procedure for Building Learner Corpora*, volume 8.
- 12.2014 **Michel Génèreux, Egon W. Stemle, Lionel Nicolas, and Verena Lyding**, *Correcting OCR errors for German in Fraktur font*, In Basili et al., editors, *Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it 2014)*.
- 10.2014 **Jennifer-Carmen Frey, Egon W. Stemle, and Aivars Glaznieks**, *Collecting language data of non-public social media profiles*, In Faaß and Ruppenhofer, editors, *Workshop Proceedings of the 12th Edition of the KONVENS Conference*, pages 11–15.
- 05.2014 **Verena Lyding, Lionel Nicolas, and Egon Stemle**, *'interHist' - an interactive visual interface for corpus exploration*, In Calzolari et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 635–641.
- 05.2014 **Andrea Abel, Aivars Glaznieks, Lionel Nicolas, and Egon Stemle**, *KoKo: An L1 Learner Corpus for German*, In Calzolari et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2414–2421.
- 04.2014 **Verena Lyding, Egon Stemle, Claudia Borghetti, Marco Brunello, Sara Castagnoli, Felice Dell Orletta, Henrik Dittmann, Alessandro Lenci, and Vito Pirrelli**, *The PAISÀ Corpus of Italian Web Texts*, In *Proceedings of the 9th Web as Corpus Workshop (WAC-9)*, pages 36–43.
- 12.2013 **Egon W. Stemle and Alexander Onysko**, *Language as a Detective Story*, magazine, Academia.
- 11.2013 **Verena Lyding, Claudia Borghetti, Henrik Dittmann, Lionel Nicolas, and Egon Stemle**, *Open Corpus Interface for Italian Language Learning*, In *Proceedings of the International Conference ICT for Language Learning*, 6th edition.
- 09.2013 **Lionel Nicolas, Egon W. Stemle, Klara Kranebitter, and Verena Lyding**, *High-Accuracy Phrase Translation Acquisition Through Battle-Royale Selection*, In Angelova et al., editors, *Proceedings of Recent Advances in Natural Language Processing, RANLP 2013*, pages 516–524.
- 07.2013 **Stefan Evert, Egon Stemle, and Paul Rayson, editors**, *Proceedings of the 8th Web as Corpus Workshop (WAC-8)*, Workshop at the seventh international Corpus Linguistics conference (CL2013), Lancaster, UK.

- 06.2013 **Klara Kranebitter and Egon W. Stemle**, *Constructing concept relation maps to support building concept systems in comparative legal terminology*, In Roche et al., editors, *Terminologie & Ontologie: Théories et Applications. Actes de la septième conférence TOTh 2013*, pages 97–116.
- 09.2012 **Lionel Nicolas, Egon W. Stemle, and Klara Kranebitter**, *Towards high-accuracy bilingual phrase acquisition from parallel corpora*, In Jancsary, editor, *11th Conference on Natural Language Processing, KONVENS 2012, Empirical Methods in Natural Language Processing*, pages 471–479.
- 07.2012 **Francesca Bonin, Fabio Cavulli, Aronne Noriller, Massimo Poesio, and Egon W. Stemle**, *Annotating Archaeological Texts: An Example of Domain-Specific Annotation in the Humanities*, In *Proceedings of the Sixth Linguistic Annotation Workshop*, number July in LAW VI '12, pages 134–138.
- 11.2011 **Massimo Poesio, Eduard Barbu, Francesca Bonin, Fabio Cavulli, Asif Ekbal, Egon Stemle, and Christian Girardi**, *The Humanities Research Portal: Human Language Technology Meets Humanities Publication Archives*, In Maegaard, editor, *Proceedings of Supporting Digital Humanities (SDH2011): Answering the unaskable*.
- 11.2011 **Asif Ekbal, Francesca Bonin, Sriparna Saha, Egon Stemle, Eduard Barbu, Fabio Cavulli, Christian Girardi, and Massimo Poesio**, *Rapid Adaptation of NE Resolvers for Humanities Domains using Active Annotation*, volume 26.
- 07.2011 **Brian Murphy and Egon W. Stemle**, *PaddyWaC: A Minimally-Supervised Web-Corpus of Hiberno-English*, In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, pages 22–29.
- 06.2011 **Massimo Poesio, Eduard Barbu, Egon W. Stemle, and Christian Girardi**, *Structure-Preserving Pipelines for Digital Libraries*, In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2011)*, pages 54–62.
- 05.2010 **Kepa Joseba Rodríguez, Francesca Delogu, Jannick Versley, Egon W. Stemle, and Massimo Poesio**, *Anaphoric Annotation of Wikipedia and Blogs in the Live Memories Corpus*, In Calzolari et al., editors, *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*.
- 09.2009 **Johannes Steger and Egon Stemle**, *KrdWrd: Architecture for Unified Processing of Web Content*, In Alegria et al., editors, *Proceedings of the Fifth Web as Corpus Workshop (WAC5)*, pages 63–70.
- 09.2007 **Daniel Bauer, Judith Degen, Xiaoye Deng, Priska Herger, Jan Gasthaus, Eugenie Giesbrecht, Lina Jansen, Christin Kalina, Thorben Krüger, Robert Martin, Martin Schmidt, Simon Scholler, Johannes Steger, Egon Stemle, and Stefan Evert**, *FIASCO: Filtering the Internet by Automatic Subtree Classification*, Osnabrück, In Fairon et al., editors, *Proceedings of the Third Web as Corpus Workshop (WAC3)*.
- 07.2007 **Sebastian Blohm, Philipp Cimiano, and Egon Stemle**, *Harvesting Relations from the Web - Quantifying the Impact of Filtering Functions*, In *Proceedings of the 22nd Conference on Artificial Intelligence (AAAI-07)*, pages 1316–1323.
- 11.2005 **Martin Bleichner, Eugenie Giesbrecht, Helmar Gust, Eva-Maria Leicht, Petra Ludewig, Sabine Möller, Wiebke Müller, Martin Schmidt, Moritz Stefaner, Egon Stemle, and Katja Wilke**, *ASADO: The Analysis and Structuring of Aviation Documents - Final Report*, Technical report, Institute of Cognitive Science at the University of Osnabrück and Institute of Applied Linguistics at the University of Hildesheim.

---

## Workshops, Conferences and Review Activities

- 10.2017 **Member of the Programme Committee and Head of the Local Organising Committee**, *Conference*, 5th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora17).
- 09.2016 **Member of the Scientific Committee**, *Conference*, 4th conference on CMC and Social Media Corpora for the Humanities.
- 09.2016 **Member of the Programme Committee**, *Workshop*, 3rd Workshop on Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC 2016).
- 08.2016 **Member of the Organising Committee and Programme Committee**, *Workshop*, 10th Web as Corpus Workshop (WAC-X).

- 08.2016 **Member of the Reviewing Board**, *Special issue of the open-access journal Slovenščina 2.0 on Computer-Mediated Communication*.
- 09.2015 **Member of the Programme Committee**, *Workshop*, 2nd Workshop on Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC 2015).
- 10.2014 **Member of the Programme Committee**, *Workshop*, Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC).
- 04.2014 **Member of the Programme Committee**, *Workshop*, 9th Web as Corpus Workshop (WAC-9).
- 07.2013 **Member of the Organising Committee and Programme Committee**, *Workshop*, 8th Web as Corpus Workshop (WAC-8).