egon w. stemle

Eurac Research
Viale Druso, 1
I-39100 Bolzano (BZ)
✉ *egon.stemle@eurac.edu*

Publications

2009 **Stemle, Egon W.** "Hybrid Sweeping: Streamlined Perceptual Structured-Text Refinement". Master's thesis. University of Osnabrück, Apr. 2009. Unpublished

2019 **Stemle, Egon W.; Boyd, Adriane; Janssen, Maarten; Lindström Tiedemann, Therese; Mikelić Preradović, Nives; Rosen, Alexandr; Rosén, Dan, and Volodina, Elena**. Working together towards an ideal infrastructure for language learner corpora. In: *Widening the Scope of Learner Corpus Research. Selected Papers from the 4th Learner Corpus Research Conference 2017* (2019), ed. by Andrea Abel et al. ISSN: 20346417. "In press"

Frey, Jennifer-Carmen; Stemle, Egon W., and Doğruöz, A. Seza. Comparison of Automatic vs. Manual Language Identification in Multilingual Social Media Texts. In: *Building Computer-Mediated Communication Corpora for sociolinguistic Analysis*, ed. by Ciara R. Wigham et al. June 13, 2019. ISBN: 978-2-84516-860-2. [Link]

Wigham, Ciara R. and Stemle, Egon W., eds. *Building Computer-Mediated Communication Corpora for sociolinguistic Analysis*. Clermont-Ferrand, FR: Presses Universitaires Blaise Pascal, June 13, 2019. 155 pp. ISBN: 978-2-84516-860-2. [Link]

2018 **Abel, Andrea and Stemle, Egon W.** On the Detection of Neologism Candidates as Basis for Language Observation and Lexicographic Endeavours: The STyrLogism Project. In: *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, ed. by Jaka Čibej et al. Aug. 2018, pp. 535–544. ISBN: 978-961-06-0097-8. [Link]

Stemle, Egon and Onysko, Alexander. Using Language Learner Data for Metaphor Detection. In: *Proceedings of the Workshop on Figurative Language Processing*. Workshop on Figurative Language Processing. June 2018, pp. 133–138. DOI: *10.18653/v1/W18-0918*. [Link]

Baumgartner, Barbara and Angler, Martin. Was darf Forschung mit Social Media Daten? In: *Academia-Interview Titelthema* (May 2018), in collab. with Lukas Egarter Vigl and Egon Stemle, pp. 10–11. [Link]

2017 **Beißwenger, Michael; Wigham, Ciara R., et al.** Connecting Resources: Which Issues have to be Solved to Integrate CMC Corpora from Heterogeneous Sources and for Different Languages? In: *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities*, ed. by Egon W. Stemle et al. Oct. 2017, pp. 52–55. DOI: *10.5281/zenodo.1041877*. [Link]

Stemle, Egon W. and Wigham, Ciara R., eds. *Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities*. Bolzano, Italy, Oct. 2017. DOI: *10.5281/zenodo.1040875*. [Link]

Beißwenger, Michael; Chanier, Thierry; Erjavec, Tomaž; Fišer, Darja; Herold, Axel; Lubešić, Nikola; Lüngen, Harald; Poudat, Céline; Stemle, Egon; Storrer, Angelika, and Wigham, Ciara. 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*. May 2017, pp. 1–18. [Link]

2016 **Abel, Andrea; Glaznieks, Aivars; Nicolas, Lionel, and Stemle, Egon**. An extended version of the KoKo German L1 Learner corpus. In: *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)*, ed. by Pierpaolo Basile et al. Dec. 2016. DOI: *10.4000/books.aaccademia.1743*. [Link]

Frey, Jennifer-Carmen; Glaznieks, Aivars, and Stemle, Egon W. The DiDi Corpus of South Tyrolean CMC Data: A multilingual corpus of Facebook texts. In: *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)*, ed. by Pierpaolo Basile et al. Dec. 2016. [Link]

Stemle, Egon W. bot.zen @ EVALITA 2016 - A minimally-deep learning PoS-tagger (trained for Italian Tweets). In: *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)*. EVALITA, ed. by Pierpaolo Basile et al. Dec. 2016. [Link]

Beißwenger, Michael; Chanier, Thierry; Chiari, Isabella; Erjavec, Tomaž; Fišer, Darja; Herold, Axel; Lubešić, Nikola; Lüngen, Harald; Poudat, Céline; Stemle, Egon; Storrer, Angelika, and Wigham, Ciara. 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*. Oct. 2016. [Link]

Stemle, Egon W. *Scientific Report of Short Term Scientific Mission COST-STSM-IS1305-34353*. EURAC (Bolzano, IT) and Centre for Language Resources and Technologies (Ljubljana, SI), Sept. 2016. [Link]

Cook, Paul; Evert, Stefan; Schäfer, Roland, and Stemle, Egon, eds. *Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task*. Berlin: Association for Computational Linguistics, Aug. 2016. ISBN: 978-1-945626-15-9. [Link]

Stemle, Egon W. 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*. WAC-X, ed. by Paul Cook et al. Aug. 2016, pp. 115–119. [Link]

2015 **Frey, Jennifer-Carmen; Glaznieks, Aivars, and Stemle, Egon W.** 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)* (Sept. 29, 2015), pp. 1–6. [Link]

Nicolas, Lionel; Stemle, Egon; Glaznieks, Aivars, and Abel, Andrea. A Generic Data Workflow for Building Annotated Text Corpora. In: *Studies in Learner Corpus Linguistics: Research and Applications for Foreign Language Teaching and Assessment*, ed. by Erik Castello et al. June 2015, pp. 337–351. ISBN: 978-3-0351-0736-4. DOI: *10.3726/978-3-0351-0736-4*

Stemle, Egon and Onysko, Alexander. Automated L1 identification in English learner essays and its implications for language transfer. In: *Transfer Effects in Multilingual Language Development* (Apr. 2015), ed. by Hagen Peukert, pp. 297–321. DOI: *10.1075/hsld.4.13ste*. [Link]

2014 **Glaznieks, Aivars and Stemle, Egon**. Challenges of building a CMC corpus for analyzing writer's style by age: The DiDi project. In: *Journal for Language Technology and Computational Linguistics (JLCL)* (Dec. 2014), ed. by Michael Beißwenger et al., pp. 31–57. ISSN: 2190-6858. [Link]

Glaznieks, Aivars; Abel, Andrea; Lyding, Verena; Nicolas, Lionel, and Stemle, Egon. Establishing a Standardised Procedure for Building Learner Corpora. In: *Apples - Journal of Applied Language Studies* (Dec. 2014), ed. by Tarja Nikula et al., pp. 5–20. ISSN: 1457-9863. [Link]

Généreux, Michel; Stemle, Egon W.; Nicolas, Lionel, and Lyding, Verena. Correcting OCR errors for German in Fraktur font. In: *Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it 2014)*, ed. by Roberto Basili et al. Dec. 2014. [Link]

Frey, Jennifer-Carmen; Stemle, Egon W., and Glaznieks, Aivars. Collecting language data of non-public social media profiles. In: *Workshop Proceedings of the 12th Edition of the KONVENS Conference*, ed. by Gertrud Faaß et al. Oct. 2014, pp. 11–15. [Link]

Abel, Andrea; Glaznieks, Aivars; Nicolas, Lionel, and Stemle, Egon. KoKo: An L1 Learner Corpus for German. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ed. by Nicoletta Calzolari et al. May 2014, pp. 2414–2421. ISBN: 978-2-9517408-8-4. [Link]

Lyding, Verena; Nicolas, Lionel, and Stemle, Egon. 'interHist' - an interactive visual interface for corpus exploration. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, ed. by Nicoletta Calzolari et al. May 2014, pp. 635–641. ISBN: 978-2-9517408-8-4. [Link]

Lyding, Verena; Stemle, Egon; Borghetti, Claudia; Brunello, Marco; Castagnoli, Sara; Orletta, Felice Dell; Dittmann, Henrik; Lenci, Alessandro, and Pirrelli, Vito. The PAISÀ Corpus of Italian Web Texts. In: *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*. Apr. 26, 2014, pp. 36–43. [Link]

2013   **Stemle, Egon W. and Onysko, Alexander**. Language as a Detective Story. In: *Academia* (Dec. 2013), pp. 24–25

Lyding, Verena; Borghetti, Claudia; Dittmann, Henrik; Nicolas, Lionel, and Stemle, Egon. Open Corpus Interface for Italian Language Learning. In: *Proceedings of the International Conference ICT for Language Learning, 6th edition*. Nov. 14, 2013. ISBN: 978-88-6292-423-8. [Link]

Nicolas, Lionel; Stemle, Egon W.; Kranebitter, Klara, and Lyding, Verena. High-Accuracy Phrase Translation Acquisition Through Battle-Royale Selection. In: *Proceedings of Recent Advances in Natural Language Processing, RANLP 2013*, ed. by Galia Angelova et al. Sept. 2013, pp. 516–524. [Link]

Evert, Stefan; Stemle, Egon, and Rayson, Paul, eds. *Proceedings of the 8th Web as Corpus Workshop (WAC-8)*. Lancaster, UK: WAC-8 Organising Committee, July 22, 2013. [Link]

Kranebitter, Klara and Stemle, Egon W. Constructing concept relation maps to support building concept systems in comparative legal terminology. In: *Terminologie & Ontologie: Théories et Applications. Actes de la septième conférence TOTh 2013*, ed. by Christophe Roche et al. June 2013, pp. 97–116

2012   **Nicolas, Lionel; Stemle, Egon W., and Kranebitter, Klara**. Towards high-accuracy bilingual phrase acquisition from parallel corpora. In: *11th Conference on Natural Language Processing, KONVENS 2012, Empirical Methods in Natural Language Processing*, ed. by Jeremy Jancsary. Sept. 2012, pp. 471–479. [Link]

Bonin, Francesca; Cavulli, Fabio; Noriller, Aronne; Poesio, Massimo, and Stemle, Egon W. Annotating Archaeological Texts: An Example of Domain-Specific Annotation in the Humanities. In: *Proceedings of the Sixth Linguistic Annotation Workshop (LAW 2012)*. Series Title: LAW VI '12. July 2012, pp. 134–138. [Link]

2011   **Ekbal, Asif; Bonin, Francesca; Saha, Sriparna; Stemle, Egon; Barbu, Eduard; Cavulli, Fabio; Girardi, Christian, and Poesio, Massimo**. Rapid Adaptation of NE Resolvers for Humanities Domains using Active Annotation. In: *Journal for Language Technology and Computational Linguistics (JLCL)* (Nov. 2011), pp. 39–51. [Link]

Poesio, Massimo; Barbu, Eduard; Bonin, Francesca; Cavulli, Fabio; Ekbal, Asif; Stemle, Egon, and Girardi, Christian. The Humanities Research Portal: Human Language Technology Meets Humanities Publication Archives. In: *Proceedings of Supporting Digital Humanities (SDH2011): Answering the unaskable*, ed. by Bente Maegaard. Nov. 2011

Murphy, Brian and Stemle, Egon W. 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*. July 2011, pp. 22–29. [Link]

Poesio, Massimo; Barbu, Eduard; Stemle, Egon W., and Girardi, Christian. 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)*. June 2011, pp. 54–62. [Link]

2010   **Rodríguez, Kepa Joseba; Delogu, Francesca; Versley, Jannick; Stemle, Egon W., and Poesio, Massimo**. Anaphoric Annotation of Wikipedia and Blogs in the Live Memories Corpus. In: *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, ed. by Nicoletta Calzolari et al. May 2010. ISBN: 2-9517408-6-7. [Link]

2009   **Steger, Johannes and Stemle, Egon**. KrdWrd: Architecture for Unified Processing of Web Content. In: *Proceedings of the Fifth Web as Corpus Workshop (WAC5)*, ed. by Iñaki Alegria et al. Sept. 2009, pp. 63–70. [Link]

2007   **Bauer, Daniel; Degen, Judith; Deng, Xiaoye; Herger, Priska; Gasthaus, Jan; Giesbrecht, Eugenie; Jansen, Lina; Kalina, Christin; Krüger, Thorben; Märtin, Robert; Schmidt, Martin; Scholler, Simon; Steger, Johannes; Stemle, Egon, and Evert, Stefan**. FIASCO: Filtering the Internet by Automatic Subtree Classification, Osnabrück. In: *Proceedings of the Third Web as Corpus Workshop (WAC3)*, ed. by Cédrick Fairon et al. Sept. 2007. [Link]

Blohm, Sebastian; Cimiano, Philipp, and Stemle, Egon. Harvesting Relations from the Web - Quantifiying the Impact of Filtering Functions. In: *Proceedings of the 22nd Conference on Artificial Intelligence (AAAI-07)*. July 2007, pp. 1316–1323. ISBN: 978-1-57735-323-2. [Link]

2005   **Bleichner, Martin; Giesbrecht, Eugenie; Gust, Helmar; Leicht, Eva-Maria; Ludewig, Petra; Möller, Sabine; Müller, Wiebke; Schmidt, Martin; Stefaner, Moritz; Stemle, Egon, and Wilke, Katja**. *ASADO: The Analysis and Structuring of Aviation Documents - Final Report*. Institute of Cognitive Science at the University of Osnabrück and Institute of Applied Linguistics at the University of Hildesheim, Nov. 2005

Workshops, Conferences and Review Activities

10.2019   **Member of the Programme Committee**, *Conference*, 6th biennial conference on electronic lexicography (eLex 2019).

09.2019   **Member of the Programme Committee**, *Conference*, 7th conference on CMC and Social Media Corpora for the Humanities (cmccorpora19).

07.2019   **Reviewer**, *Conference Tracks on Linguistic Theories, Cognitive Modeling and Psycholinguistics; Machine Learning; Resources and Evaluation; and Textual Inference and Other Areas of Semantics*, 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019).

09.2018   **Member of the Programme Committee**, *Conference*, 6th conference on CMC and Social Media Corpora for the Humanities (cmccorpora18).

07.2018   **Member of the Programme Committee**, *Congress*, XVIII EURALEX International Congress.

05.2018   **Member of the Programme Committee**, *Workshop*, GLOBALEX 2018: Lexicography & Word-Nets.

12.2017   **Assessor for the University of Innsbruck**, *Anniversary Fund*, University of Innsbruck and the Medical University of Innsbruck.

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).

09.2017   **Member of the Programme Committee**, *Conference*, 5th biennial conference on electronic lexicography (eLex 2017).

07.2017   **Member of the Programme Committee**, *Workshop*, 11th Web as Corpus Workshop (WAC-XI).

09.2016   **Member of the Programme 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).