# egon w. stemle

*Researcher*

*Eurac Research*
*Viale Druso, 1*
*I-39100 Bolzano (BZ)*
☎ *+39 0471 055.129*
✉ *egon.stemle@eurac.edu*
🖱 *iiegn.eu/work*

Egon Stemle is a researcher in the Institute for Applied Linguistics at Eurac Research, Bolzano, Italy. As a cognitive scientist, he passionately combines the humanistic and analytical methods of the arts and formal sciences, focusing on the area where computational linguistics and artificial intelligence converge.

He works on the creation, standardisation, and interoperability of tools for editing, processing, and annotating linguistic data and enjoys working together with other scientists on their data but also collects or helps to collect new data from the Web, from computer-mediated communication and social media, and from language learners. He is an advocate of open science to make research and data available for others to consult or reuse in new research.

His curiosity in research is driven by the question why humans handle incomplete and – more often than not – inconsistent structured concepts just fine, whereas computational processes are often of little avail or fail completely.

## Research and Professional Experience

**since 02.2012** — **Researcher**, *Institute for Applied Linguistics at Eurac Research, Bolzano, Italy*.

**04.2009 – – 01.2012** — **Research Fellow**, *LiveMemories sponsored grant, financed by the Provincia Autonoma of Trento, for research activities at the Center for Mind/Brain Sciences (CIMeC) of the University of Trento, Italy*.

**01.2009 – – 07.2018** — **Technology Consultant and Shareholder**, *Whitematter Labs*. Provider of neuroscientific technologies for marketers.

**04.2008 – – 2010** — **Project Member**, *GoodGaze*, Institute of Cognitive Science, University of Osnabrück, Germany. Research project to predict where people will look on Web pages.

**2008** — **Freelancer**, *ontoprise GmbH, Karlsruhe, Germany*. Integrating in-house technology with IBM's OmniFind; Feasibility Study: evaluate to which extent very specific information extraction from large web data collections combined with semi-automatic cleaning is technically feasible, what benefits and efforts are to be expected – while also considering external alternatives.

**10.2006 – – 03.2008** — **Research Participant**, *DAAD funded research project (Project-linked exchange of academics and scientists, PPP)*, Collaboration between the Institute of Cognitive Science, U. Osnabrück, Germany, the CNRS UMR 8163 (Savoirs, Textes, Langages), U. of Lille (III), France, and the Departament de Traducció i Filologia, U. Barcelona (Pompeu Fabra), Spain. Reference to Abstract Objects in Natural Language (OntoRef).

**04.2006 – – 03.2008** — **Research Assistant**, Institute of Applied Informatics and Formal Description Methods, University of Karlsruhe, Germany, Knowledge Management Research Group. Maintenance and development of software and scientific counsel; e.g. evaluating CL tools for OmniFind+UIMA and hooking them together.

**08.2004 – – 09.2005** — **Research Participant**, *ASADO*, Cooperation between the Universities of Osnabrück and Hildesheim, and the aircraft manufacturer AIRBUS. Project to research methodologies and technologies to analyze and structure the huge amount of documentation produced during aircraft construction.

**08.2003 – – 12.2003** — **Student Assistant**, *Institute of Cognitive Science, University of Osnabrück, Germany*, Artificial Intelligence Research Group. Maintenance and development of software for the MiLCA (Media intensive learning units for courses in computational linguistics) project.

| | |
|---|---|
| 10.2001 – – 03.2009 | **System Administrator**, *Institute of Cognitive Science, University of Osnabrück, Germany*. Maintenance of computer hardware and software, server administration (Web Services, Databases, File Sharing, Printing, Mail, Calendaring, (D)VCS, CMS), user support; assisting work groups in making strategic decisions about hardware and software. |

## Education

| | |
|---|---|
| 04.2006 – – 03.2009 | **MSc in Cognitive Science**, *with distinction*, University of Osnabrück, Germany. Majors: *Linguistics and Computational Linguistics*, and *Artificial Intelligence*, One Year Study Project: Analysis and Structure of Aviation Documents (ASADO). |
| Thesis | **Hybrid Sweeping: Streamlined Perceptual Structured-Text Refinement**. Development of a perceptually driven content extraction architecture for Web pages – exemplified by a Web page cleaning system. Supervisors: Prof. Dr. Stefan Evert and Prof. Dr. Peter König. |
| 03.2002 – – 09.2002 | **Semester abroad**, *ILLC (Institute for Logic, Language and Computation)*, University of Amsterdam, The Netherlands. |
| 10.2000 – – 01.2006 | **BSc in Cognitive Science**, University of Osnabrück, Germany. |
| 12.1998 – – 07.2000 | **Apprenticeship as Travel Agent**, WORLDWIDE Touristik, Nürnberg, Germany. |
| 06.1997 | **Abitur**, Sigmund-Schuckert-Gymnasium, Nürnberg, Germany. |

## Teaching Experience

| | |
|---|---|
| 05.2016 | **Session on Corpus Building and Practical Corpus Building**, *COST ENeL Training School 2016:Tools and methods for creating innovative e-dictionaries*, Central Hotel, Ljubljana, Slovenia, [link]. |
| 2009 – 2011 (yearly) | **Lecture on Web Corpora**, *Course in Text Processing class*, Coordinated by Marco Baroni, University of Trento, Italy, Offered to students of the School of Humanities and Philosophy, of the International Master in Cognitive Science and of the Master in Human Language Technologies and Interfaces. |
| 10.2003 – – 02.2006 (every other semester) | **Artificial Intelligence Tutor**, *Methods of Artificial Intelligence class*, Institute of Cognitive Science, University of Osnabrück, Germany, Second year Cognitive Science class. Weekly tutoring of the lecture material, assessment of programming assignments, homework corrections, and exam preparation. |
| 04.2003 – – 07.2006 (every other semester) | **Computational Linguistics Tutor**, *Introduction to Computational Linguistics class*, Institute of Cognitive Science, University of Osnabrück, Germany, First year Cognitive Science class. Weekly tutoring of the lecture material, preparation and assessment of programming assignments, homework corrections, and exam preparation. |
| 10.2002 – – 02.2003 | **Theoretical Neuroscience Tutor**, *Introduction to Theoretical Neuroscience class*, Institute of Cognitive Science, University of Osnabrück, Germany, Second year Cognitive Science class. Weekly tutoring of the lecture material, homework corrections. |
| 10.2001 – – 02.2002 | **Computer Science Tutor**, *Algorithms class*, Computer Science Department, University of Osnabrück, Germany, First year Computer Science class. Weekly oral assessment of 10 2-people groups and homework corrections. |

## Publications

| | |
|---|---|
| 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" |
| | **Stemle, Egon W.; Abel, Andrea, and Lyding, Verena**. "Language varieties meet One-Click Dictionary". Accepted paper. Accepted paper. 2019. "Accepted at eLex" |

**König, Alexander and Stemle, Egon W.** "Technical Solutions for Reproducible Research". Accepted paper. Accepted paper. 2019. "Accepted at CLARIN"

**Frey, Jennifer-Carmen; König, Alexander, and Stemle, Egon W.** How FAIR are CMC Corpora? In: *Proceedings of the 7th Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora2019)*. Sept. 9, 2019, pp. 25–30. [Link]

**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. DOI: *10.4000/books.aaccademia.1782*. [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. DOI: *10.4000/books.aaccademia.1968*. [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* (Nov. 2011), pp. 39–51. ISSN: 2190-6858. [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. [Link]

**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]

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

## Presentations (talks, posters, etc. without proceedings)

2019 **Stemle, Egon W.** "Towards an infrastructure for FAIR language learner corpora". Talk. 8th Natural Language Processing for Computer-Assisted Language Learning (NLP4CALL) Workshop (Turku, FI). Sept. 30, 2019. Upcoming talk

**Stemle, Egon W.** "News from the STyrLogism Project: Language varieties meet One-Click Dictionary". Talk. Neologismen: Korpuslinguistische Ermittlung und lexikographische Bearbeitung (Leibniz-Institut für Deutsche Sprache, Mannheim, DE). June 3, 2019. [Link]

2018 **Abel, Andrea and Stemle, Egon W.** "On the Detection of Neologism Candidates as Basis for Language Observation and Lexicographic Endeavours: the STyrLogisms Project". Talk. XVIII EURALEX International Congress (Ljubljana, SI). July 19, 2018. [Link]

**Stemle, Egon W. and Onysko, Alexander**. "Using Language Learner Data for Metaphor Detection". Talk. Work in Progress Series (Kompetenzzentrum Sprachen, Freie Universität Bozen, Bolzano, IT). June 5, 2018. [Link]

**Melzer, Tine; Oosterhof, Frans; Franck, Dorothea, and Stemle, Egon**. "Bilder kippen! Aspektsehen in künstlerischer Praxis und Lehre". Panel talk. Forschungs-Mittwoch (Bern University of the Arts, Bern, CH). May 30, 2018

2017 **Stemle, Egon W.** "Learner Corpus Infrastructure (LCI) @ Eurac Research". Talk. SWE-CLARIN Workshop on Interoperability of Second Language Resources and Tools (University of Gothenburg, Gothenburg, SE). Dec. 2017. [Link]

**Stemle, Egon W.** "DiDi Corpus". Talk. Integrating a new type of language resource into the Digital Humanities landscape: French-German colloquium on standards for corpora of computer-mediated communication (University of Duisburg-Essen, Duisburg, DE). June 19–20, 2017. [Link]

2016 **Lyding, Verena; Nicolas, Lionel, and Stemle, Egon W.** "Cross-institutional cooperation initiatives in the Digital Humanities - challenges and infrastructures". Discussion. International Workshop:"Initiativen und Innovationen in den Digital Humanities" (Akademie Deutsch-Italienischer Studien, Merano, IT). Nov. 2016

2015 **Stemle, Egon W.** "The DiDi Project: Collecting, Annotating, and Analysing South Tyrolean Data of Computer-mediated Communication". Invited talk. First international research days on Social Media and CMC Corpora for the eHumanities (ird-cmc-rennes) (Rennes 2 University, Rennes, FR). Oct. 23, 2015. [Link]

**Stemle, Egon W.** "South Tyrolian Neologisms Project". Short talk. COST ENel WG3 meeting "Automatic Knowledge Acquisition for Lexicography" (Herstmonceux Castle, Sussex, UK). Aug. 13, 2015

**Stemle, Egon W.** "Web Corpus Creation, Cleaning and Evaluation". Lecture. Web as Corpus Meeting@eLex 2015 (Herstmonceux Castle, UK). Aug. 10, 2015. [Link]

2014 **Stemle, Egon W. and Onysko, Alexander**. "Automated L1 identification in English learner essays and its implications for language transfer". Talk. Work in Progrss Series (Kompetenzzentrum Sprachen, Freie Universität Bozen, Bolzano, IT). Apr. 2014

2013 **Stemle, Egon W.** "Automatische Annotation von Schülertexten - Herausforderungen und Lösungsvorschläge am Beispiel des Projekts KoKo". Talk. Workshop from the "Arbeitsgruppe: Korpusbasierte Linguistik" at the 40. Österreichische Linguistiktagung (Universität Salzburg, Salzburg, AT). Nov. 22, 2013. [Link]

**Stemle, Egon W.** "Herausforderungen bei der automatischen Verarbeitung von dialektalen IBK-Daten". Talk. Publication Title: Workshop on "Verabeitung und Annotation von Sprachdaten aus Genres internetbasierter Kommunikation" at the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2013) (TU Darmstadt, Darmstadt, DE). Sept. 23, 2013. [Link]

**Stemle, Egon W.** "The future of BootCaT: A Creative Commons License filter". Talk. BootCaTters of the world unite! (BOTWU), A workshop (and a survey) on the BootCaT toolkit (Department of Interpreting and Translation, University of Bologna, Forlì, IT). June 24, 2013. [Link]

**Stemle, Egon W.** "(Technical Aspects of) Harvesting Data from Social Network Sites". Talk. International workshop "Building Corpora of Computer-Mediated Communication: Issues, Challenges, and Perspectives" (Dortmund, DE). Feb. 14, 2013. [Link]

2012   **Stemle, Egon W.** "Web Corpus Creation and Cleaning". Plenary talk. Student Research Workshop:Computer Applications in Linguistics (CSRW2012) (English Corpus Linguistics Group at the Institute of Linguistics and Literary Studies, Technische Universität Darmstadt, Darmstadt, DE). July 13, 2012. [Link]

**Stemle, Egon W.** "On visual Approaches towards Corpus Exploration". Short talk. 3rd workshop of the academic network on "Internet Lexicography" (EURAC research, Bolzano, IT). May 4, 2012. [Link]

2011   **Poesio, Massimo; Barbu, Eduard; Stemle, Egon, and Girardi, Christian**. "Portale Ricerca Umanistica". Live demo and poster. LiveMemories Final Event - Internet, Memoria e Futuro and The Semantic Way (FBK, Povo di Trento, IT). Nov. 24, 2011. [Link]

## Workshops, Conferences and Review Activities

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

09.2019   **Member of the Programme Committee**, *Conference*, 7th conference on CMC and Social Media Corpora for the Humanities (cmccorpora19), Cergy-Pontoise, France, [link].

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), Florence, Italy, [link].

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

07.2018   **Member of the Programme Committee**, *Congress*, XVIII EURALEX International Congress, Ljubljana, Slovenia, [link].

05.2018   **Member of the Programme Committee**, *Workshop*, GLOBALEX 2018: Lexicography & WordNets, Workshop at the 11th Language Resources and Evaluation Conference (LREC 2018), Miyazaki, Japan.

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

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), Bolzano, Italy, [link].

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

07.2017   **Member of the Programme Committee**, *Workshop*, 11th Web as Corpus Workshop (WAC-XI), Workshop at ninth international Corpus Linguistics conference (CL2017), Birmingham, UK, [link].

09.2016   **Member of the Programme Committee**, *Conference*, 4th conference on CMC and Social Media Corpora for the Humanities, Ljubljana, Slovenia, [link].

09.2016 **Member of the Programme Committee**, *Workshop*, 3rd Workshop on Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC 2016), Bochum, Germany, [link].

08.2016 **Member of the Organising Committee and Programme Committee**, *Workshop*, 10th Web as Corpus Workshop (WAC-X), Workshop at the annual meeting of the Association for Computational Linguistics (ACL), Berlin, Germany, [link].

08.2016 **Member of the Reviewing Board**, *Special issue of the open-access journal Slovenščina 2.0 on Computer-Mediated Communication*, [link].

09.2015 **Member of the Programme Committee**, *Workshop*, 2nd Workshop on Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC 2015), Pre-conference workshop at the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL-2015), Essen, Germany, [link].

10.2014 **Member of the Programme Committee**, *Workshop*, Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC), Pre-conference workshop at KONVENS 2014, Hildesheim, Germany, [link].

04.2014 **Member of the Programme Committee**, *Workshop*, 9th Web as Corpus Workshop (WAC-9), Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014), Gothenburg, Sweden, [link].

07.2013 **Member of the Organising Committee and Programme Committee**, *Workshop*, 8th Web as Corpus Workshop (WAC-8), Workshop at the seventh international Corpus Linguistics conference (CL2013), Lancaster University, UK, [link].

## Administrative Experience

since 08.2012 **Secretary of the Special Interest Group on the Web as Corpus (SIGWAC)**, SIG of the Association for Computational Linguistics.

04.2006 – **Member of the Examination Board (Prüfungsausschuss)**, Cognitive Science Study Programme,
– 03.2007 University of Osnabrück, Germany.

12.2005 – **Member of the Search Committee (Besetzungkommission)**, Institute of Cognitive Science,
– 02.2006 University of Osnabrück, Germany.
BAT IIa research associate in Artificial Intelligence.

04.2005 – **Member of the Academic Studies Commission (Studienkommission)**, Cognitive Science Study
– 03.2008 Programme, University of Osnabrück, Germany.

02.2005 – **Member of the Search Committee (Berufungskommission)**, Faculty of Humanities, University
– 01.2006 of Osnabrück, Germany.
W3 professorship Artificial Intelligence and Cognitive Science.

10.2003 – **Member of the Steering Committee (Vorstand)**, Institute of Cognitive Science, University of
– 03.2008 Osnabrück, Germany.

## Other Activities

04.2006 **A for Alibi Symposium**, Uqbar Foundation, Utrecht University Museum [link].

12.2005 **Amsterdam 2.0 Exhibition**, Mediamatic, Amsterdam [link].
Technical counsel for Kasper Andreasen & Tine Melzer

03.2003 **Spring School**, Interdisciplinary College 2003 (IK2003), Günne at Lake Möhne.
Focus Theme: Applications, Brains and Computers

2003 **The Complete Dictionary**, Tine Melzer, 26 volumes, A–Z.
Programming and Processing [link]

## Languages

German Native

English Excellent                                                          *Main education language at university.*

| | | | |
|---|---|---|---|
| Italian | Basic | | |
| French | Basic | | *5 years training during high-school.* |

## Computer Skills

| | | | |
|---|---|---|---|
| programming | Python, JavaScript, Java, Prolog | scripting | Bash, Perl, PHP |
| scientific | Keras, Theano, R, CQP, SkE, Oracle/Sun Grid Engine, Matlab, UIMA, PDP++, Octave, Protégé | design | LaTeX, gnuplot, Inkscape, Scribus, InDesign, Illustrator, Photoshop, GIMP |
| OS | Linux, OS X | services | Kubernetes, Docker, LAMP Stack, PostgreSQL, Apache Tomcat, Mailing Stack, Trac, ISC Bind, OpenVPN, Shorewall, OpenLDAP, NIS/YP, Kerberos+NFSv4 |