Personal information

Name	egon w. stemle
Address	European Academy of Bozen/Bolzano (EURAC)
	Institute for Specialised Communication and Multilingualism
	Viale Druso, 1
	I-39100 Bolzano (BZ)
Telephone	+39 0471 055.129
Nationality	German
Date of birth	20.Sept.1977
E-mail	egon.stemle@eurac.edu
Website	iiegn.eu/work

Main areas of research

Computational Linguistics, Corpus Linguistics, Web Corpora
Artificial Intelligence, Machine Learning, Ontologies
Cognitive Science

Academic experience

since 02.2012	**Researcher**, *Institute for Specialised Communication and Multilingualism at the European Academy of Bozen/Bolzano (EURAC)*.
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.*
04.2008 – – 2010	**Project Member**, *GoodGaze*, Institute of Cognitive Science, University of Osnabrück, Germany.
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.
04.2006 – – 03.2008	**Research Assistant**, Institute of Applied Informatics and Formal Description Methods, University of Karlsruhe, Germany.
08.2004 – – 09.2005	**Research Participant**, *ASADO*, Cooperation between the Universities of Osnabrück and Hildesheim, and the aircraft manufacturer AIRBUS.
08.2003 – – 12.2003	**Student Assistant**, *Institute of Cognitive Science, University of Osnabrück, Germany*, Artificial Intelligence Research Group.
02.2006 – – 03.2009	**MSc in Cognitive Science**, *with distinction*, University of Osnabrück, Germany.
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.

Invited accademic talks

10.2015	**Egon W. Stemle**, *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).

08.2015	**Egon W. Stemle**, *Web Corpus Creation, Cleaning and Evaluation*, lecture, Web as Corpus Meeting@eLex 2015.
02.2013	**Egon W. Stemle and Aivars Glaznieks**, *(Technical Aspects of) Harvesting Data from Social Network Sites*, talk, International workshop "Building Corpora of Computer-Mediated Communication: Issues, Challenges, and Perspectives".
07.2012	**Egon W. Stemle**, *Web Corpus Creation and Cleaning*, plenary talk, Student Research Workshop:Computer Applications in Linguistics (CSRW2012).
05.2012	**Egon W. Stemle, Verena Lyding, and Lionel Nicolas**, *On visual Approaches towards Corpus Exploration*, short talk, 3rd workshop of the academic network on "Internet Lexicography".

Five most important academic prizes/awards

09.2013	**RANLP-2013 Young Researcher Award**, *best paper*, High-Accuracy Phrase Translation Acquisition through Battle-Royale Selection.
04.2009 – – 01.2012	**Research Fellow**, *Grant for research activities at the Center for Mind/Brain Sciences (CIMeC) of the University of Trento*.

Five most important peer review activities, editorships and/or memberships in academic organisations

08.2016	**Member of the reviewing board for the special issue of the open-access journal Slovenščina 2.0 on Computer-Mediated Communication**, [www.trojina.org/slovenscina2.0/en/].
2016,2014	**Member of the Organising Committee and Programme Committee**, *Workshops*, 10th & 8th Web as Corpus Workshops.
2015,2014	**Member of the Programme Committee**, *Workshops*, 1st & 2nd Workshops on Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC).
since 08.2012	**Secretary of the Special Interest Group on the Web as Corpus (SIGWAC)**, *SIG of the Association for Computational Linguistics*, [www.sigwac.org.uk].
since 11.2002	**Member of the German Cognitive Science Society (GK e.V.)**, [www.gk-ev.de].

List of the 10 most important scientific / scholarly publication in the entire career to date

09.2015	**Egon Stemle and Alexander Onysko**, *Automated L1 identification in English learner essays and its implications for language transfer*.
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*, pages 337–351.
12.2014	**Aivars Glaznieks, Andrea Abel, Verena Lyding, Lionel Nicolas, and Egon Stemle**, *Establishing a Standardised Procedure for Building Learner Corpora*.
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.
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.
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*.

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)*.

07.2007 **Sebastian Blohm, Philipp Cimiano, and Egon Stemle**, *Harvesting Relations from the Web - Quantifiying the Impact of Filtering Functions*, In *Proceedings of the 22nd Conference on Artificial Intelligence (AAAI-07)*, pages 1316–1323.

List of scholarly publications in the last 5 years

Journal Articles

09.2015 **Egon Stemle and Alexander Onysko**, *Automated L1 identification in English learner essays and its implications for language transfer*.

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*, pages 337–351.

12.2014 **Aivars Glaznieks and Egon Stemle**, *Challenges of building a CMC corpus for analyzing writer's style by age: The DiDi project*.

12.2014 **Aivars Glaznieks, Andrea Abel, Verena Lyding, Lionel Nicolas, and Egon Stemle**, *Establishing a Standardised Procedure for Building Learner Corpora*.

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.

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*.

Conference Proceedings

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)*.

12.2014 **Michel Généreux, 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.

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.

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.*

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.