

egon w. stemle

Researcher

European Academy of Bozen/Bolzano
(EURAC)

Viale Druso, 1
I-39100 Bolzano (BZ)

✉ egon.stemle@eurac.edu

<http://www.eurac.edu/staff/estemle/>



Egon Stemle is a Cognitive Scientist: he studies skills like perception, thinking, learning, motor function, and language by combining the humanistic and analytical methods of the arts and the formal sciences. His research focus within this relatively new 'inter-discipline' lies in the area where Computational Linguistics and Artificial Intelligence converge. He is working on computer aided fabrication of ontologies from large document repositories, the technological feasibility thereof and the utilization of cross-linked structured data in applications, and on tools for editing, processing, and annotating linguistic data. His driving force is 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.

Education

- 02.2006 – **MSc in Cognitive Science**, *with distinction*, University of Osnabrück, Germany.
- 03.2009 Majors: *Linguistics and Computational Linguistics* and *Artificial Intelligence*, One Year Study Project: Analysis and Structure of Aviation Documents (ASADO), [<http://www.cogsci.uni-osnabrueck.de/~ASADO>].
- 03.2002 – **Semester abroad**, *ILLC (Institute for Logic, Language and Computation)*, University of Amsterdam, The Netherlands.
- 09.2002
- 10.2000 – **BSc in Cognitive Science**, University of Osnabrück, Germany.
- 01.2006
- 12.1998 – **Apprenticeship as Travel Agent**, WORLDWIDE Touristik, Nürnberg, Germany.
- 07.2000
- 06.1997 **Abitur**, Sigmund-Schuckert-Gymnasium, Nürnberg, Germany.

Master's Thesis

- title *Hybrid Sweeping: Streamlined Perceptual Structured-Text Refinement*
- supervisors Prof. Dr. Stefan Evert and Prof. Dr. Peter König
- description Development of a perceptually driven content extraction architecture for Web pages – exemplified by a Web page cleaning system.

Research and Professional Experience

- since 02.2012 **Researcher**, *Institute for Specialised Communication and Multilingualism at the European Academy of Bozen/Bolzano (EURAC)* [[→EURAC](#)].
- 04.2009 – **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* [[→LiveMemories](#), [→CIMEC](#)].
- 01.2012
- since 2009 **Technology Consultant and Shareholder**, *Whitematter Labs* [<http://whitematterlabs.com>], provider of neuroscientific technologies for marketers.

- 04.2008 – **Project Member**, *GoodGaze* [<http://www.goodgaze.com>], research project to predict where people will look on Web pages.
– 2010
- 2008 **Freelancer**, *ontoprise GmbH* [<http://www.ontoprise.de>], 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 – **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, Langues), U. of Lille (III), France, and the Departament de Traducció i Filologia, U. Barcelona (Pompeu Fabra), Spain, title: Reference to Abstract Objects in Natural Language (OntoRef) [<http://parles.upf.es/glif/pub/abstract/index.html>].
– 03.2008
- 04.2006 – **Research Assistant**, *Institute of Applied Informatics and Formal Description Methods, University of Karlsruhe, Germany*, Knowledge Management Research Group, [<http://www.aifb.uni-karlsruhe.de/Forschungsgruppen/WBS/english>].
– 03.2008
Maintenance and development of software and scientific counsel; e.g. evaluating CL tools for OmniFind+UIMA and hooking them together.
- 08.2004 – **Research Participant**, *cooperation between the Universities of Osnabrück and Hildesheim, and the aircraft manufacturer AIRBUS*, title: ASADO – project to research methodologies and technologies to analyze and structure the huge amount of documentation produced during aircraft construction.
– 09.2005
- 08.2003 – **Student Assistant**, *Institute of Cognitive Science, University of Osnabrück, Germany*, Artificial Intelligence Research Group, [<http://www.cogsci.uni-osnabrueck.de/~ai/milca.html>].
– 12.2003
Maintenance and development of software for the MiLCA (Media intensive learning units for courses in computational linguistics) project.
- 10.2001 – **System Administrator**, *Institute of Cognitive Science, University of Osnabrück, Germany*.
– 03.2009
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.

Teaching Experience

- 2009 – 2011
(yearly) **Lecture on Web Corpora**, *Course in text processing*, 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, coordinated by Marco Baroni, University of Trento.
- 10.2003 – **Artificial Intelligence Tutor**, *Methods of Artificial Intelligence*, Second year Cognitive Science class, Institute of Cognitive Science, University of Osnabrück.
– 02.2006
(every other semester) Weekly tutoring of the lecture material, assessment of programming assignments, homework corrections, and exam preparation.
- 04.2003 – **Computational Linguistics Tutor**, *Introduction to Computational Linguistics*, First year Cognitive Science class, Institute of Cognitive Science, University of Osnabrück.
– 07.2006
(every other semester) Weekly tutoring of the lecture material, preparation and assessment of programming assignments, homework corrections, and exam preparation.
- 10.2002 – **Theoretical Neuroscience Tutor**, *Introduction to Theoretical Neuroscience*, Second year Cognitive Science class, Institute of Cognitive Science, University of Osnabrück.
– 02.2003
Weekly tutoring of the lecture material, homework corrections.
- 10.2001 – **Computer Science Tutor**, *Algorithms*, First year Computer Science class, Computer Science Department, University of Osnabrück.
– 02.2002
Weekly oral assessment of 10 2-people groups and homework corrections.

Publications

- (in Preparation) **Egon Stemle, Alexander Onysko**, *The role of transfer in automated L1 identification of English learner essays*, in *Peukert, Hagen (ed.): Transfer Effects in Multilingual Language Development*, John Benjamins, Amsterdam & Philadelphia.
- (in Preparation) **Aivars Glaznieks, Andrea Abel, Verena Lyding, Lionel Nicolas, Egon Stemle**, *Establishing a Standardised Procedure for Building Learner Corpora – a Response to Demands and Suggestions of Users*, Proceedings of Learner Language, Learner Corpora 2012 (LLLC2012), Oct 5-6, University of Oulu, Finland, Tarja Nikula, Sauli Takala and Sabine Ylönen (eds.). Apples - Journal of Applied Language Studies.
- (in Preparation) **Aivars Glaznieks, Egon Stemle**, *Challenges of building a CMC corpus for analyzing writer's style by age: The DiDi project*, Special Issue of Journal of Language Technology and Computational Linguistics: Building and annotating corpora of computer-mediated discourse. Issues and Challenges at the Interface of Corpus and Computational Linguistics, Michael Beißwenger, Nelleke Oostdijk, Angelika Storrer & Henk van den Heuvel (eds.).
- 06.2014 **Verena Lyding, Lionel Nicolas, Egon Stemle**, *'interHist' – an interactive visual interface for corpus exploration*, for *Proceedings of the 9th Conference on International Language Resources and Evaluation (LREC'14)*, May 26-31, Reykjavik, Iceland. European Language Resources Association (ELRA)
- 06.2014 **Andrea Abel, Aivars Glaznieks, Lionel Nicolas, Egon Stemle**, *KoKo: An L1 learner corpus for German*, for *Proceedings of the 9th Conference on International Language Resources and Evaluation (LREC'14)*, May 26-31, Reykjavik, Iceland. European Language Resources Association (ELRA)
- 04.2014 **Verena Lyding, Egon Stemle, Claudio Borghetti, Marco Brunello, Sara Castagnoli, Felice Dell Orletta, Henrik Dittmann, Alessandro Lenci, Vito Pirrelli**, *The PAISÀ Corpus of Italian Web Texts*, in *Proceedings of the 9th Web as Corpus Workshop (WAC-9)*, Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Apr 26, Gothenburg, Sweden
- 09.2013 **Lionel Nicolas, Egon W. Stemle, Klara Kranebitter, Verena Lyding**, *High-accuracy phrase translation acquisition through battle-royale selection*, Proceedings of Recent Advances in Natural Language Processing, RANLP 2013, Sep 7-13, Hissar, Bulgaria.
- 07.2013 **Stefan Evert, Egon Stemle, Paul Rayson (eds.)**, *Proceedings of the 8th Web as Corpus Workshop (WAC-8)*, Workshop at the seventh international Corpus Linguistics conference (CL2013), Jul 22, Lancaster, UK.
- 06.2013 **Klara Kranebitter, Egon W. Stemle**, *Constructing concept relation maps to support building concept systems in comparative legal terminology*, Terminology & Ontology: Theories and applications (TOTH'2013), Jun 6-7, Chambéry, France.
- 09.2012 **Lionel Nicolas, Egon W. Stemle, Klara Kranebitter**, *Towards high-accuracy bilingual phrase acquisition from parallel corpora*, in *Proceedings of KONVENS 2012 (LexSem 2012 workshop)*, Vienna, Austria, Österreichische Gesellschaft für Artificial Intelligence (ÖGAI).
- 07.2012 **Francesca Bonin, Fabio Cavulli, Massimo Poesio, Egon W. Stemle**, *Annotating Archaeological Texts: An Example of Domain-Specific Annotation in the Humanities*, in *Proceedings of the Sixth Linguistic Annotation Workshop, ACL2012*, Jeju, Republic of Korea, Association for Computational Linguistics.

- 11.2011 **Asif Ekbal, Francesca Bonin, Sriparna Saha, Egon Stemle, Eduard Barbu, Fabio Cavulli, Christian Girardi, Massimo Poesio**, *Rapid Adaptation of NE Resolvers for Humanities Domains using Active Annotation*, in *Journal for Language Technology and Computational Linguistics (JLCL)*, 26(2):39–51, 2011.
- 11.2011 **Massimo Poesio, Eduard Barbu, Francesca Bonin, Fabio Cavulli, Asif Ekbal, Egon Stemle, Christian Girardi**, *The Humanities Research Portal: Human Language Technology Meets Humanities Publication Archives*, in *SDH2011 - Supporting Digital Humanities: Answering the unaskable*, Copenhagen, DK.
- 07.2011 **Brian Murphy, 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 (DIALECTS2011)*, Edinburgh, Scotland, UK, Association for Computational Linguistics.
- 06.2011 **Massimo Poesio, Eduard Barbu, Egon W. Stemle, 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)*, Portland, OR, USA, Association for Computational Linguistics.
- 05.2010 **Kepa J. Rodríguez, Francesca Delogu, Jannick Versley, Egon W. Stemle, Massimo Poesio**, *Anaphoric Annotation of Wikipedia and Blogs in the Live Memories Corpus*, in *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, European Language Resources Association (ELRA).
- 09.2009 **Johannes M. Steger, Egon W. Stemle**, *KrdWrd – The Architecture for unified Processing of Web Content*, in *Proceedings of the 5th Web as Corpus Workshop (WAC5)*, Pre-SEPLN Workshop, Donostia-San Sebastián, Basque Country, Spain.
[http://www.sigwac.org.uk/raw-attachment/wiki/WAC5/WAC5_proceedings.pdf]
- 09.2007 **Daniel Bauer, Judith Degen, Xiaoye Deng, Priska Herger, Jan Gasthaus, Eugenie Giesbrecht, Lina Jansen, Christin Kalina, Thorben Krüger, Robert Märtin, Martin Schmidt, Simon Scholler, Johannes Steger, Egon Stemle, Stefan Evert**, *FIASCO: Filtering the Internet by automatic Subtree Classification*, Osnabrück, in *Building and Exploring Web Corpora. Proceedings of the 3rd Web as Corpus Workshop, incorporating CLEANVAL (WAC3-2007)*, Presses universitaires de Louvain.
[http://purl.org/stefan.evert/PUB/BauerEtc2007_FIASCO.pdf]
- 07.2007 **Sebastian Blohm, Philipp Cimiano, Egon Stemle**, *Harvesting Relations from the Web - Quantifying the Impact of Filtering Functions*, in *Proceedings of the 22nd Conference on Artificial Intelligence (AAAI-07)*, Association for the Advancement of Artificial Intelligence (AAAI).
[<http://www.aifb.uni-karlsruhe.de/WBS/seb/publications/pronto-aaai07.pdf>]
- 11.2005 **Martin Bleichner, Eugenie Giesbrecht, Helmar Gust, Eva-Maria Leicht, Petra Ludewig, Sabine Möller, Wiebke Müller, Martin Schmidt, Moritz Stefaner, Egon Stemle, 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.
[<http://iiegn.de/static/doc/bleichneretal2005.pdf>]

Talks

- 09.2014 (Accepted) **Michel Genereux, Egon Stemle, Lionel Nicolas, Verena Lyding**, *Correcting OCR errors for German in Fraktur font*, CLICit - The first Italian Computational Linguistics Conference, Dec 9-10, Pisa, Italy.
- 09.2014 (Accepted) **Jennifer-Carmen Frey, Aivars Glaznieks, Egon Stemle**, *Collecting language data of non-public social media profiles*, Natural Language Processing for Computer-Mediated Communication / Social Media (NLP4CMC), Pre-conference workshop at KONVENS2014, Hildesheim/Germany, October 6, 2014.
- 08.04.2014 **Alexander Onysko*, Egon W. Stemle***, *Automated L1 identification in English learner essays and its implications for language transfer*, Work in Progress, Kompetenzzentrum Sprachen, Freie Universität Bozen, Bozen, IT.
- 02.2014 **Aivars Glaznieks*, Egon W. Stemle, Jennifer Frey**, *Nützlichkeit der Normalisierung von Dialekt für automatische Verarbeitungsschritte. Erfahrungen aus dem Projekt DiDi*, 7th workshop of the DFG scientific network Empirikom: Social Media Corpora for the eHumanities: Standards, Challenges, and Perspectives, Feb 19–21, TU Dortmund University, Germany.
- 22.11.2013 **Andrea Abel, Aivars Glazniek, Egon W. Stemle**, *Automatische Annotation von Schülertexten – Herausforderungen und Lösungsvorschläge am Beispiel des Projekts KoKo*, Arbeitsgruppe: Korpusbasierte Linguistik, Workshop at the 40. Österreichische Linguistiktagung, Nov 22-24, Universität Salzburg, Salzburg, Austria.
- 11.2013 **Verena Lyding*, Claudia Borghetti*, Henrik Dittmann, Lionel Nicolas, Egon Stemle**, *Open Corpus Interface for Italian Language Learning*, 6th edition of the International Conference ICT for Language Learning, Nov 14-15, Florence, Italy.
- 23.09.2013 **Aivars Glazniek, Egon W. Stemle**, *Herausforderungen bei der automatischen Verarbeitung von dialektalen IBK-Daten*, GSCL-Workshop *Verarbeitung und Annotation von Sprachdaten aus Genres internetbasierter Kommunikation*, Workshop at the International Conference of the German Society for Computational Linguistics and Language Technology, Sep 23, Darmstadt, Germany.
- 24.06.2013 **Egon W. Stemle, Verena Lyding**, *The future of BootCaT: A Creative Commons License filter*, BootCaTters of the world unite! (BOTWU), A workshop (and a survey) on the BootCaT toolkit, Jun 24, Department of Interpreting and Translation, University of Bologna Forlì, Italy.
- 05.2013 **Lionel Nicolas*, Aivars Glaznieks, Egon W. Stemle, Andrea Abel**, *Compiling and using a German learner corpus: the Koko project*, Compiling and Using Learner Corpora to teach and assess productive and interactive skills in foreign languages at university level - Learner Corpora 2013, May 16-17, Padova, Italy.
- 03.2013 **Aivars Glaznieks*, Egon Stemle, Andrea Abel, Verena Lyding**, *Herausforderungen bei der Erstellung eines L1-Lernerkorpus: Lösungsvorschläge aus dem Projekt "KoKo"*, DGfS 2013 Workshop on Modellierung nicht-standardisierter Schriftlichkeit, Workshop at the 35th Annual Conference of the German Linguistic Society (DGfS 2013), Mar 12-15, Potsdam, Germany.
- 03.2013 **Verena Lyding*, Lionel Nicolas, Egon Stemle**, *interHist – an interactive visualization for statistically enhanced query structures*, DGfS 2013 Workshop on the Visualization of Linguistic Patterns, Workshop at the 35th Annual Conference of the German Linguistic Society (DGfS 2013), Mar 12-15, Potsdam, Germany.
- 13.02.2013 **Egon W. Stemle, Aivars Glaznieks**, *Technical Aspects in Harvesting Data from Social Network Sites*, International workshop: Building Corpora of Computer-Mediated Communication: Issues, Challenges, and Perspectives, Feb 13-15, Department of German Language and Literature, Faculty of Culture Studies, TU Dortmund University (Germany).

- 02.2013 **Aivars Glaznieks*, Egon W. Stemle**, *The Project DIDI. Writing on Social Network Sites – A Corpus-based Observation of the Current Language Use in South Tyrol, with Particular Consideration of the Writers' Age*, International workshop: Building Corpora of Computer-Mediated Communication: Issues, Challenges, and Perspectives, Feb 13-15, Department of German Language and Literature, Faculty of Culture Studies, TU Dortmund University (Germany).
- 13.07.2012 **Egon W. Stemle**, *Web Corpus Creation and Cleaning*, Computer Applications in Linguistics: Student Research Workshop (CSRW2012), English Corpus Linguistics Group at the Institute of Linguistics and Literary Studies, Technische Universität Darmstadt.
- 04.05.2012 **Egon W. Stemle, Verena Lyding, Lionel Nicolas**, *On visual Approaches towards Corpus Exploration*, 3rd Workshop of the Academic Network on "Internet Lexicography".
(Impulsreferat)

Workshops and Conferences

- 07.2013 **Organising Committee and Programme Committee**, *Workshop*, 8th Web as Corpus Workshop (WAC-8), [<http://sigwac.org.uk/wiki/WAC8>].
Workshop day at the seventh international Corpus Linguistics conference (CL2013), Lancaster University, UK
- 09.2006 **Student Worker at OTT06**, *Workshop*, Ontologies in Text Technology: Approaches to Extract Semantic Knowledge from Syntactic Information., [<http://www.cogsci.uni-osnabrueck.de/~ott06>].
jointly organized by the Institute of Cognitive Science at the University of Osnabrück and the project C2 of the distributed DFG-research group Text Technological Modelling of Information
- 06.2006 **Student Worker at QITL-2**, *Workshop*, Second Workshop on Quantitative Investigations in Theoretical Linguistics., [<http://www.cogsci.uni-osnabrueck.de/~qitl>].
organized by the Computational Linguistics Group at the Institute of Cognitive Science
- 09.2003 **Student Worker at EuroCogSci 2003**, *Conference*, [<http://www.eurocogsci03.uni-osnabrueck.de>], First European Cognitive Science Meeting..
jointly organized by the Cognitive Science Society and the German Cognitive Science Society

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 – 03.2007 **Member of the Examination Board (Prüfungsausschuss)**, Cognitive Science Study Programme, University of Osnabrück.
- 12.2005 – 02.2006 **Member of the Search Committee (Besetzungskommission)**, Institute of Cognitive Science, University of Osnabrück.
BAT IIa research associate in Artificial Intelligence.
- 04.2005 – 03.2008 **Member of the Academic Studies Commission (Studienkommission)**, Cognitive Science Study Programme, University of Osnabrück.
- 02.2005 – 01.2006 **Member of the Search Committee (Berufungskommission)**, Faculty of Humanities, University of Osnabrück.
W3 professor ship Artificial Intelligence and Cognitive Science.
- 10.2003 – 03.2008 **Member of the Steering Committee (Vorstand)**, Institute of Cognitive Science, University of Osnabrück.
- 07.2003 – 10.2003 **Member of the Search Committee (Besetzungskommission)**, Institute of Cognitive Science, University of Osnabrück.
BAT IIa research associate in Artificial Intelligence.

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 [link]), Günne at Lake Möhne.
Focus Theme: Applications, Brains and Computers
- 2003 **The Complete Dictionary**, Tine Melzer, 26 volumes, A–Z.
Programming and Processing [link]
- since 11.2002 **Member of the German Cognitive Science Society (GK e.V.)**, [link].

Languages

German	Native	
English	Excellent	<i>Main education language at university.</i>
French	Basic	<i>5 years training during high-school.</i>
Italian	Basic	

Computer Skills

programming	Python, Java, JavaScript (ECMA-Scrip), Prolog, ML	scripting	Bash, Perl, PHP
scientific	R, CQP, Oracle/Sun Grid Engine, Apache Hadoop, Matlab, OmniFind+UIMA, PDP++, Octave, Protégé	design	L ^A T _E X, gnuplot, Inkscape, Scribus, InDesign, Illustrator, Photoshop, GIMP
OS	Linux, OS X, Windows 2000/XP/Vista	services	Apache, MySQL, PostgreSQL, Exim, Mailman, SpamAssassin, DSpam, Trac, Subversion, ISC DHCP, ISC Bind, Cyrus, Dovecot, OpenVPN, Shorewall, Darwin Calendar Server, OpenLDAP, NIS/YP, Kerberos+NFSv4