

Stefan Evert – Research – Teaching – CV – Publications – Software – Private Life

## Curriculum Vitae

Overview – Events – Professional Service – Invited Talks

### Overview

#### Since September 2012

Professor of Corpus Linguistics <http://www.linguistik.uni-erlangen.de/> at Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany.

#### September 2011 – September 2012

Professor of English Computational Corpus Linguistics at Technische Universität Darmstadt <http://www.linglit.tu-darmstadt.de/>, Germany.

#### October 2005 – September 2011

*Juniorprofessor* for Computational Linguistics at the Institute of Cognitive Science <http://www.cogsci.uos.de/>, University of Osnabrück, Germany.

#### March 2005 – September 2005

Researcher and lecturer at the Institute of Cognitive Science, University of Osnabrück, Germany.

#### 30 August 2004

Dr. phil. in Computational Linguistics (grade: *mit Auszeichnung bestanden*), IMS, University of Stuttgart, Germany. Thesis: *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. [PDF <http://purl.org/stefan.evert/PUB/Evert2004phd.pdf>] Supervisor: PD Dr. Ulrich Heid <http://www.ims.uni-stuttgart.de/~uli>.

#### September 1999 – February 2005

Research assistant in computational linguistics at the IMS <http://www.ims.uni-stuttgart.de/> (Institut für maschinelle Sprachverarbeitung), University of Stuttgart, Germany.

#### 6 July 1999

Diplom in Mathematics (grade: *mit Auszeichnung bestanden*), University of Stuttgart, Germany. Thesis: *Das Verhalten von Lösungen der vektorialen Helmholtzgleichung in Außenräumen für kleine Frequenzen unter elektrischen Randbedingungen* (no joke :-). [PDF <http://purl.org/stefan.evert/PUB/Evert1999diplom.pdf>] Supervisor: Prof. Dr. Peter Werner <http://www.iadm.uni-stuttgart.de/LstAnaMPhy/Werner/>.

#### October 1991 – March 2000

Studies in Mathematics (major), English (major), and Physics (minor), University of Stuttgart <http://www.uni-stuttgart.de/>, Germany.

#### December 1991

Cambridge Certificate of Proficiency in English (grade A pass).

#### 16 May 1990

Abitur (grade: 1,0), Friedrich-Schiller-Gymnasium <http://www.fsg-marbach.de/> Marbach am Neckar, Germany.

#### 13 October 1970

Born in Ludwigsburg, Germany.

### Events

- *5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)* <https://sites.google.com/site/cogalex2016/> and shared task on *Corpus-Based Identification of Semantic Relations* <https://sites.google.com/site/cogalex2016/home/shared-task> at the COLING 2016 Conference in Osaka, Japan, 12 December 2016 (with Alessandro Lenci, Enrico Santus and Michael Zock).
- Shared task on *Automatic Linguistic Annotation of Computer-Mediated Communication (EmpiriST 2015)* <http://sites.google.com/site/empirist2015/> (with Sabine Bartsch, Michael Beißwenger and Kay-Michael Würzner).

- The *10th Workshop on Multiword Expressions* <<http://multiword.sourceforge.net/mwe2014/>> (MWE 2014) at the EACL Conference in Gothenburg, Sweden, 26–27 April 2014 (with Valia Kordoni, Markus Egg, Agata Savary and Eric Wehrli).
- Workshop on *Modelling Non-Standardized Writing* <<http://www.empirikom.net/bin/view/Aktivitaeten/DgfsNonstandard>> at the Annual Conference of the German Linguistic Society (DGfS 2013), Potsdam, Germany, 13–15 March 2013 (with Michael Beißwenger, Stefanie Dipper and Bianka Trevisan).
- Workshop on *Semantic Theory and Empirical Evidence* <<http://www.cogsci.uni-osnabrueck.de/~empiricalEvidence/>>, 18–19 September 2009, Osnabrück, Germany (with Carla Umbach and Cornelia Ebert).
- ESSLLI 2008 Workshop on *Distributional Lexical Semantics: Bridging the gap between semantic theory and computational simulations* <<http://wordspace.collocations.de/doku.php/workshop:esslli:start>>, 4–9 August 2008, Hamburg, Germany (with Marco Baroni and Alessandro Lenci).
- The 4th *Web as Corpus Workshop: Can we beat Google?* <<http://WebAsCorpus.sf.net/WAC4/>> (WAC-4) at the LREC Conference in Marrakech, Morocco, 1 June 2008 (with Adam Kilgarriff and Serge Sharoff).
- LREC workshop: *Towards a Shared Task for Multiword Expressions* <<http://multiword.sf.net/mwe2008/>> (MWE 2008) at the LREC Conference in Marrakech, Morocco, 1 June 2008 (with Nicole Grégoire and Brigitte Krenn).
- ACL workshop: *A Broader Perspective on Multiword Expressions* <<http://multiword.sourceforge.net/mwe2007/>> (MWE 2007) at the ACL Conference in Prague, Czech Republic, 28 June 2007 (with Nicole Grégoire and Su Nam Kim).
- Conference on *Quantitative Investigations in Theoretical Linguistics 2* <<http://www.cogsci.uni-osnabrueck.de/~qitl/>> (QITL-2) in Osnabrück, Germany, 1–2 July 2006 (with Peter Bosch, Graham Katz, Philip Reuter, Bettina Schrader and Carla Umbach).
- ACL workshop on *Multiword Expressions: Identifying and Exploiting Underlying Properties* (MWE-2006) at the ACL/COLING conference in Sydney, Australia, 23 July 2006 (with Diana McCarthy, Suzanne Stevenson, Aline Villavicencio, Begona Villada Moiron). [proceedings <<http://acl.ldc.upenn.edu/W/W06/#W06-1200>>]

### Professional Service

- Referee for various journals (*Computers and the Humanities, Language Resource and Evaluation, Computational Linguistics, Computer Speech and Language, Research on Language and Computation, Natural Language Engineering, International Journal of Corpus Linguistics*, etc.), conferences (*ACL, EACL, COLING, EMNLP, LREC, EURALEX, Corpus Linguistics*, etc.), and workshops (*Multiword Expressions, Web as Corpus*, etc.).
- Reviewer for book publishers (*Mouton de Gruyter, Springer, CUP*) and funding organisations (DFG, BMBF, ISF, NWO)
- Chair of the ACL Special Interest Group on the Web as Corpus <<http://sigwac.org.uk/>> (SIGWAC, 2012–2015)
- Member of the Council of the Faculty of Philosophy <<http://www.phil.fau.de/>> and the Board of the Department Germanistik & Komparatistik <<http://www.germanistik.phil.fau.de/>> (2015–)
- Board member of the Interdisciplinary Centres for *Digital Humanities and Social Sciences* <<http://www.izdigital.fau.de/>> and *Lexicography, Valency and Collocation* <<http://www.lexi.uni-erlangen.de/>> (FAU Erlangen–Nürnberg, 2014–)
- Member of the Challenge and Innovation Panel of the *ESRC Centre for Corpus Approaches to Social Science* <<http://cass.lancs.ac.uk/>>, Lancaster University, UK (2013–)

- Member of the ACL SIGLEX Executive Board (2004–2007), the Editorial Board of the *International Journal of Corpus Linguistics* <<http://benjamins.com/#catalog/journals/ijcl/board>> (2011–), and the Editorial Board of the *International Journal of Learner Corpus Research* <<http://benjamins.com/#catalog/journals/ijlcr/board>> (2015–)
- Member of the DIN Working Group 105–00–06 AA “Sprachressourcen” (*language resources*) and ISO TC37 SC4 (2014X)
- Member of the Association for Computational Linguistics (ACL), the Cognitive Science Society, the German Linguistics Society (DGfS) and the German Society for Computational Linguistics & Language Technology (GSCL)
- ERASMUS & international affairs coordinator for the Institute of Cognitive Science, University of Osnabrück with Sven Walter, 2009 – 2011)
- Local coordinator for the ERASMUS student exchange programme at the IMS, University of Stuttgart (March 2001 – February 2005)

### Invited Talks

- Talk on *Modelling Free Associations with Co-Occurrence Data* in the symposium *From computational modelling to behavior via multimodal corpus data: integrative approaches to the study of semantic representation and processing*, *International Convention of Psychological Science (ICPS 2017; Vienna, Austria, March 2017; with Gabriella Lapesa)*
- Talk on *Naïve Discriminative Learning: From Rescorla-Wagner to Machine Learning in Natural Language Processing* at the Institute of Cognitive Science, University of Osnabrück, Germany (June 2017)
- Talk series on *Evaluating the corpus-based identification of lexical collocations*
  - latest slides (PDF) <[http://purl.org/stefan.evert/PUB/Measure\\_for\\_Measure\\_2016.pdf](http://purl.org/stefan.evert/PUB/Measure_for_Measure_2016.pdf)>
  - Centre for Translation Studies, University of Leeds, UK (June 2016)
  - FU Berlin, Germany (February 2016)
  - Laboratoire CLLE-ERSS, Université de Toulouse 2 (September 2015)
- Plenary talk on *Multivariate Ansätze zur Modellierung von Sprachvariation* in the workshop *Digitale Forschungsmethoden in der Romanistik* <<http://www.romanistik.de/aktuelles/1491>>, *Forum Junge Romanistik* (Würzburg, Germany, March 2016)
- Statistical consultant for research project OPTACH <<http://www.eurac.edu/de/research/projects/Pages/projectdetails.aspx?pid=11263>>, EURAC Research (Bolzano, Italy, December 2015)
- Workshop tutorial on *Introduction to statistics for corpus linguistics*, *Symposium on Methods and Linguistic Theories* <<http://www.uni-bamberg.de/bagl/malt-2015/>> (Bamberg, Germany, November 2015)
- Talk on *A multivariate approach to linguistic variation and distribution* at the workshop on *Data mining and its use and usability for linguistic analysis* (Saarbrücken, Germany, March 2015) and in the GECCo colloquium (Saarbrücken, Germany, September 2015)
- Talk on *An NLP Approach to the Evaluation of Web Corpora* at QLVL, KU Leuven (Belgium, February 2015) [slides (PDF) <[http://purl.org/stefan.evert/PUB/Evert2015\\_WebEval\\_slides.pdf](http://purl.org/stefan.evert/PUB/Evert2015_WebEval_slides.pdf)>, handout (PDF) <[http://purl.org/stefan.evert/PUB/Evert2015\\_WebEval\\_handout.pdf](http://purl.org/stefan.evert/PUB/Evert2015_WebEval_handout.pdf)>]
- Talk on *Current approaches to linguistic querying and quantitative analysis of large corpora* at e-cosmos Workshop *Computational Perspectives on Multimodal Data* (Aachen, Germany, November 2014) [handout (PDF) <[http://purl.org/stefan.evert/PUB/eCosmos2014\\_Evert\\_handout.pdf](http://purl.org/stefan.evert/PUB/eCosmos2014_Evert_handout.pdf)>]
- Tutorial *Statistics for Linguists* at DGKL 7 Conference (Erlangen, Germany, September 2014)
- Evening lecture on *A multivariate approach to linguistic variation and distribution* at the 26th European Summer School in Logic, Language and Information (ESSLLI 2014, Tübingen, Germany,

- August 2014 [slides (PDF)  
 <[http://purl.org/stefan.evert/PUB/Evert2014\\_ESSLLI\\_evening\\_lecture.pdf](http://purl.org/stefan.evert/PUB/Evert2014_ESSLLI_evening_lecture.pdf)>, video 1 (p. 21)  
 <[http://purl.org/stefan.evert/media/esslli\\_2014/bnc\\_correl\\_example.mp4](http://purl.org/stefan.evert/media/esslli_2014/bnc_correl_example.mp4)>, video 2 (p. 29)  
 <[http://purl.org/stefan.evert/media/esslli\\_2014/en\\_pca3\\_register\\_sides.mp4](http://purl.org/stefan.evert/media/esslli_2014/en_pca3_register_sides.mp4)>, video 3 (p. 39)  
 <[http://purl.org/stefan.evert/media/esslli\\_2014/en\\_lda\\_trans\\_register.mp4](http://purl.org/stefan.evert/media/esslli_2014/en_lda_trans_register.mp4)>, video 4 (p. 40)  
 <[http://purl.org/stefan.evert/media/esslli\\_2014/en\\_lda\\_trans\\_status.mp4](http://purl.org/stefan.evert/media/esslli_2014/en_lda_trans_status.mp4)>]
- Talk on *An NLP Approach to the Evaluation of Web Corpora* in the workshop *AG 11: Web Data as a Challenge for Theoretical Linguistics and Corpus Design* at the DGfS 2014 Conference (Marburg, Germany, February 2014) [handout (PDF)  
 <[http://purl.org/stefan.evert/PUB/Evert2014\\_WebEval\\_handout.pdf](http://purl.org/stefan.evert/PUB/Evert2014_WebEval_handout.pdf)>]
  - Talk series on *Rethinking corpus frequencies: Understanding and managing non-randomness in corpus data*
    - latest slides (PDF) <[http://purl.org/stefan.evert/PUB/On\\_Corpus\\_Frequencies\\_2013.pdf](http://purl.org/stefan.evert/PUB/On_Corpus_Frequencies_2013.pdf)>
    - Summer school *Documentary Linguistics and Variationist Sociolinguistics*, University of Bamberg, Germany (July 2014)
    - Workshop *The relative frequencies of nouns, pronouns, and verbs in discourse* at the *10th Biennial Conference of the Association for Linguistic Typology*, Leipzig, Germany (August 2013)
    - Linguistics Department, University of Zurich, Switzerland (March 2012)
    - Interdisciplinary Centre for Research on Lexicography, Valency and Collocation, University of Erlangen, Germany (December 2010)
    - Computational Linguistics Group, LORIA INRIA Nancy, France (June 2010)
    - Workshop *Ontologien und Wortnetze in Lexikographie und Textwissenschaft*, LexicoLux & University of Trier, Luxembourg (December 2009)
    - Conference on *Diversity in Language Corpora*, University of Basel, Switzerland (April 2009)
    - Linguistics Department, University of Konstanz, Germany (January 2009)
    - Linguistics Department, Lancaster University, UK (December 2008)
    - Linguistics Department, University of Potsdam, Germany (May 2008)
  - Participation in Dagstuhl Seminar 13462, *Computational Models of Language Meaning in Context* <<http://www.dagstuhl.de/13462>> (Schloß Dagstuhl, Germany, November 2013)
  - Tutorial “Statistik für Linguisten”, DGfS 2012 Conference (Frankfurt, Germany, March 2012) and DGfS 2013 Conference (Potsdam, Germany, March 2013); with Amir Zeldes [material <[http://wordspace.collocations.de/doku.php/corpus\\_tutorial](http://wordspace.collocations.de/doku.php/corpus_tutorial)>]
  - Talk on *Historische Korpora und distributionelle Semantik* at *Varianz und Vielfalt interdisziplinär: Wörter und Strukturen* (Darmstadt, Germany, November 2012; with Cristina Sánchez Marco)
  - Talk on *The Role of Dimensionality Reduction in Distributional Semantics* as special contributor at *Leuven Statistics Days 2012* (Leuven, Belgium, June 2012) [handout (PDF)  
 <[http://purl.org/stefan.evert/PUB/lrd\\_2012\\_dimensionality\\_reduction.handout.pdf](http://purl.org/stefan.evert/PUB/lrd_2012_dimensionality_reduction.handout.pdf)>]
  - Talk on *Web N-Grams as a Resource for Corpus Linguistics* at the Computational Linguistics Group, University of Zurich (Zurich, Switzerland, March 2012)
  - Talks on *Making Sense of Distributional Semantic Models* at the University of Amsterdam (Amsterdam, The Netherlands, September 2010) and the Tandem Workshop on *Optimality in Language and Geometric Approaches to Cognitions* (Berlin, December 2010)
  - Talks on *Association & Word Meaning* at the University of Vienna (Vienna, Austria, January 2010) and the University of Potsdam (Potsdam, Germany, March 2010)
  - Tutorial “Korpuslinguistik mit Online-Ressource – eine interaktive Einführung für Linguisten”, DGfS 2010 Conference (Berlin, Germany, February 2010), LIPP (LMU München, Germany, January 2011)

- and DGfS 2011 Conference (Göttingen, Germany, February 2011); with Stefanie Dipper and Heike Zinsmeister [material <[http://wordspace.collocations.de/doku.php/corpus\\_tutorial](http://wordspace.collocations.de/doku.php/corpus_tutorial)>]
- Talk *Inside Word Skteches* at Berlin-Brandenburgische Akademie der Wissenschaften (BBAW, Berlin, Germany, January 2010)
  - Talk *Inside the IMS Corpus Workbench* at the Institut Universitari de Lingüística Aplicada (IULA), Universitat Pompeu Fabra (Barcelona, Spain, September 2008)
  - Talk on *The Significance of Corpus Frequencies* at the Colloquium of the Chair of Applied Linguistics and Translation Studies, Saarland University (Saarbrücken, Germany, January 2008)
  - Invitation to *Expert Meeting on the Dutch Web as Corpus* (Instituut voor Nederlandse Lexicologie & Koninklijke Bibliotheek, Den Haag, The Netherlands April 2007)
  - ERASMUS visiting lecturer at CIMEC <<http://www.cimec.unitn.it/>>, University of Trento (Rovereto/Trento, Italy, March 2007)
  - Participation in Dagstuhl Seminar 06491, *Digital Historical Corpora – Architecture, Annotation, and Retrieval* <<http://www.dagstuhl.de/06491>> (Schloß Dagstuhl, Germany, December 2006)
  - Presentation *Towards better association measures for collocation extraction* at the Collocations and Idioms 2006 Conference, Berlin-Brandenburg Academy of Sciences and Humanities (Berlin, Germany, November 2006)
  - Talk on *Statistical Association and Natural Language* in the Computational Linguistics Colloquium, Saarland University (Saarbrücken, Germany, May 2006)
  - Talk on *The Web as Corpus* at the Department of Linguistics, Stockholm University (Stockholm, Sweden, November 2005)
  - Course on *Statistical Methods for Corpus Linguistics* at the European Academy (Eurac, Bolzano, Italy, September 2005)
  - Talk on *Indexing and Interfaces for the Web as Corpus* at the Workshop on the Web as Corpus (Corpus Linguistics 2005 Conference, Birmingham, UK, July 2005)
  - Teacher at TWC seminar on *The Case for a Corpus* (Tuscan Word Centre, Siena, Italy, June 2005)
  - Talk on *How Random is a Corpus? – The Library Metaphor* at the Workshop on The Scope and Limits of Corpus Linguistics (FU Berlin, Germany, June 2005)
  - Talk on *Using the Web as a Corpus: Procedures, Pitfalls, and my Pet Peeve* at the LingLit Special Event 2005: Corpus applications in literary studies, linguistics and information technology (TU Darmstadt, Germany, February 2005)
  - Talk on *Sketch of a WWW Interface for Very Large Corpora and Lexical Databases* at the Symposium on the Web as Corpus (SITLeC, University of Bologna, Italy, January 2005)
  - Talk on *Intuitive Association Measures* (HU Berlin, Germany, December 2004)
  - Talk on *Design Patterns for Richly Annotated Corpora* at the Antragskonferenz DDD (HU Berlin, Germany, December 2003)
  - Consultant on tools for text corpora (GLiCom, UPF Barcelona, Spain, October 2001)
  - Consultant for the research project *Computational Processing of Portuguese* (SINTEF, Oslo, Norway, March 2001)
  - EURALEX tutorial on *Tools for Corpus Lexicography* with Arne Fitschen and Ulrich Heid (Stuttgart, Germany, August 2000)