

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

## Publications

Monographs – Journals – Book Chapters – Conference Proceedings – Edited Volumes – Miscellaneous

### Monographs

Evert, Stefan (2004, published 2005). *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart, URN urn:nbn:de:bsz:93-opus-23714. [<http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/>>]

Hoffmann, Sebastian; Evert, Stefan; Smith, Nicholas; Lee, David; Berglund Prytz, Ylva (2008). *Corpus Linguistics with BNCweb – a Practical Guide*, volume 6 of *English Corpus Linguistics*. Peter Lang, Frankfurt am Main. [<http://www.peterlang.com/index.cfm?vID=56315>>]

### Journal Papers

Michelbacher, Lukas; Evert, Stefan; Schütze, Hinrich (in press). **Asymmetry in corpus-derived and human word associations**. To appear in *Corpus Linguistics and Linguistic Theory*.

Evert, Stefan (2006). **How random is a corpus? The library metaphor**. *Zeitschrift für Anglistik und Amerikanistik*, **54**(2), 177 – 190. [manuscript: <http://purl.org/stefan.evert/PUB/Evert2006.pdf>>, journal homepage: <http://www.zaa.koenigshausen-neumann.de/aktuell.htm>>]

Carletta, Jean, Evert, Stefan, Heid, Ulrich, Kilgour, Jonathan, and Chen, Yiya (2005). **The NITE XML Toolkit: data model and query language**. *Language Resources and Evaluation*, **39**(4), 313 – 334. [NXT homepage: <http://nite.sourceforge.net/>>]

Evert, Stefan and Krenn, Brigitte (2005). **Using small random samples for the manual evaluation of statistical association measures**. *Computer Speech & Language* **19**(4), 450 – 466. [manuscript: <http://purl.org/stefan.evert/PUB/EvertKrenn2005.pdf>>]

### Book Chapters

Evert, Stefan, Frötschl, Bernhard, and Lindstrot, Wolf (2009). **Statistische Grundlagen**. In K.–U. Carstensen, C. Ebert, C. Ebert, S. Jekat, R. Klabunde, and H. Langer, editors, *Computerlinguistik und Sprachtechnologie: Eine Einführung*, pages 114 – 158. Spektrum Akademischer Verlag, Heidelberg, 3rd edition.

Ebert, Christian; Schiehlen, Michael; Klabunde, Ralf; Evert, Stefan (2009). **Semantik**. In K.–U. Carstensen, C. Ebert, C. Ebert, S. Jekat, R. Klabunde, and H. Langer, editors, *Computerlinguistik und Sprachtechnologie: Eine Einführung*, pages 330 – 393. Spektrum Akademischer Verlag, Heidelberg, 3rd edition.

Evert, Stefan (2008). **Corpora and collocations**. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, article 58. Mouton de Gruyter, Berlin. [extended manuscript: [http://purl.org/stefan.evert/PUB/Evert2007HSK\\_extended\\_manuscript.pdf](http://purl.org/stefan.evert/PUB/Evert2007HSK_extended_manuscript.pdf)>]

Baroni, Marco and Evert, Stefan (2008). **Statistical methods for corpus exploitation**. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, article 36. Mouton de Gruyter, Berlin. [manuscript: [http://purl.org/stefan.evert/PUB/BaroniEvertHSK38\\_manuscript.pdf](http://purl.org/stefan.evert/PUB/BaroniEvertHSK38_manuscript.pdf)>]

Evert, Stefan and Fitschen, Arne (2001). **Textkorpora**. In K.–U. Carstensen, C. Ebert, C. Endriss, S. Jekat, R. Klabunde, and H. Langer (eds.), *Computerlinguistik und Sprachtechnologie – Eine Einführung*, pages 369 – 376. Spektrum Akademischer Verlag, Heidelberg, Berlin.

### Conference Proceedings and Collections

2010 – 2009 – 2008 – 2007 – 2006 – 2005 – 2004 – 2003 – 2002 – 2001 – 2000

2011

Ebert, Cornelia; Evert, Stefan; Wilmes, Katharina (2011). Focus marking via gestures. In I. Reich *et al.* (eds.), *Proceedings of Sinn & Bedeutung 15*, Saarbrücken, Germany. Universaar – Saarland University

Press. To appear. [<http://purl.org/stefan.evert/PUB/EbertEvertWilmes2011.pdf>]

2010

Evert, Stefan (2010). **Google Web 1T5 n-grams made easy (but not for the computer)**. In *Proceedings of the 6th Web as Corpus Workshop (WAC-6)*, Los Angeles, CA. [[http://purl.org/stefan.evert/PUB/Evert2010\\_WAC6.pdf](http://purl.org/stefan.evert/PUB/Evert2010_WAC6.pdf)]

2009

Giesbrecht, Eugenie and Evert, Stefan (2009). **Part-of-speech tagging – a solved task? An evaluation of POS taggers for the Web as corpus**. In I. Alegria, I. Leturia, and S. Sharoff, editors, *Proceedings of the 5th Web as Corpus Workshop (WAC5)*, San Sebastian, Spain. [[http://purl.org/stefan.evert/PUB/GiesbrechtEvert2009\\_Tagging.pdf](http://purl.org/stefan.evert/PUB/GiesbrechtEvert2009_Tagging.pdf)]

2008

Evert, Stefan (2008). **A lightweight and efficient tool for cleaning Web pages**. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco. To appear. [[http://purl.org/stefan.evert/PUB/Evert2008\\_NCleaner.pdf](http://purl.org/stefan.evert/PUB/Evert2008_NCleaner.pdf)]

Evert, Stefan (2008). **A lexicographic evaluation of German adjective-noun collocations**. In *Proceedings of the LREC Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, Marrakech, Morocco. To appear. [[http://purl.org/stefan.evert/PUB/Evert2008\\_MWE\\_Resource.pdf](http://purl.org/stefan.evert/PUB/Evert2008_MWE_Resource.pdf)]

2007

Baroni, Marco and Evert, Stefan (2007). **Words and echoes: Assessing and mitigating the non-randomness problem in word frequency distribution modeling**. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 904 – 911, Prague, Czech Republic. [<http://purl.org/stefan.evert/PUB/BaroniEvert2007.pdf>]

[talk slides: <http://purl.org/stefan.evert/PUB/BaroniEvert2007.handout.pdf>]

Bauer, Daniel; Degen, Judith; Deng, Xiaoye; Herger, Priska; Gasthaus, Jan; Giesbrecht, Eugenie; Jansen, Lina; Kalina, Christin; Krüger, Thorben; Martin, Robert; Schmidt, Martin; Scholler, Simon; Steger, Johannes; Stemle, Egon and Evert, Stefan (2007). **FIASCO: Filtering the Internet by automatic subtree classification, Osnabrück**. In C. Fairon, H. Naets, A. Kilgarriff, and G.-M. de Schryver (eds.), *Building and Exploring Web Corpora: Proceedings of the 3rd Web as Corpus Workshop (WAC3), incorporating CLEANVAL*, pages 111 – 121, Louvain-la-Neuve, Belgium. [[http://purl.org/stefan.evert/PUB/BauerEtc2007\\_FIASCO.pdf](http://purl.org/stefan.evert/PUB/BauerEtc2007_FIASCO.pdf)]

Evert, Stefan (2007). **StupidOS: A high-precision approach to boilerplate removal**. In C. Fairon, H. Naets, A. Kilgarriff, and G.-M. de Schryver (eds.), *Building and Exploring Web Corpora: Proceedings of the 3rd Web as Corpus Workshop (WAC3), incorporating CLEANVAL*, pages 123 – 133, Louvain-la-Neuve, Belgium. [[http://purl.org/stefan.evert/PUB/Evert2007\\_StupidOS.pdf](http://purl.org/stefan.evert/PUB/Evert2007_StupidOS.pdf)]

Evert, Stefan and Baroni, Marco (2007). **zipfR: Word frequency distributions in R**. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Session*, pages 29 – 32, Prague, Czech Republic. [<http://purl.org/stefan.evert/PUB/EvertBaroni2007.pdf>]

Lüdeling, Anke, Evert, Stefan, and Baroni, Marco (2007). **Using Web data for linguistic purposes**. In M. Hundt, N. Nesselhauf, and C. Biewer, editors, *Corpus Linguistics and the Web*, volume 59 of *Language and Computers – Studies in Practical Linguistics*, pages 7 – 24. Rodopi, Amsterdam, New York. [manuscript: <http://purl.org/stefan.evert/PUB/LuedelingEvertBaroni2005.pdf>]

Michelbacher, Lukas, Evert, Stefan, and Schütze, Hinrich (2007). **Asymmetric association measures**. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2007)*, Borovets, Bulgaria. [<http://purl.org/stefan.evert/PUB/MichelbacherEtc2007.pdf>]

2006

Bernardini, Silvia, Baroni, Marco, and Evert, Stefan (2006). **A WaCky introduction**. In M. Baroni and S. Bernardini, editors, *Wacky! Working papers on the Web as Corpus*, pages 9 – 40. GEDIT, Bologna. [<http://wackybook.sslmit.unibo.it/>]

Hoffmann, Sebastian and Evert, Stefan (2006). **BNCweb (CQP-edition): The marriage of two corpus tools.** In S. Braun, K. Kohn, and J. Mukherjee (eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, volume 3 of *English Corpus Linguistics*, pages 177 – 195. Peter Lang, Frankfurt am Main. [<http://purl.org/stefan.evert/PUB/HoffmannEvert2006.pdf>]

2005

Baroni, Marco and Evert, Stefan (2005). **Testing the extrapolation quality of word frequency models.** In P. Danielsson and M. Wagenmakers (eds.), *Proceedings of Corpus Linguistics 2005*, volume 1 of *The Corpus Linguistics Conference Series*. ISSN 1747–9398. [<http://purl.org/stefan.evert/PUB/EvertBaroni2005.pdf>]

Evert, Stefan and Schönenberger, Manuela (2005). **Separating the sheep from the goats: Clarifying corpus content using XML.** In P. Danielsson and M. Wagenmakers (eds.), *Proceedings of Corpus Linguistics 2005*, volume 1 of *The Corpus Linguistics Conference Series*. ISSN 1747–9398. [<http://purl.org/stefan.evert/PUB/EvertSchoenenberger2005.pdf>]

Krenn, Brigitte and Evert, Stefan (2005). **Separating the wheat from the chaff: Corpus-driven evaluation of statistical association measures for collocation extraction.** In B. Fisseni, H.–C. Schmitz, B. Schröder, and P. Wagner (eds.), *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen. Beiträge zur GLDV-Tagung 2005 in Bonn*, volume 8 of *Computer Studies in Language and Speech*, pages 104 – 117. Peter Lang, Frankfurt am Main. [<http://purl.org/stefan.evert/PUB/KrennEvert2005.pdf>]

Lüdeling, Anke and Evert, Stefan (2005). **The emergence of productive non-medical –itis. Corpus Evidence and qualitative analysis.** In: Kepser, Stephan and Reis, Marga (eds.), *Linguistic Evidence. Empirical, Theoretical, and Computational Perspectives*, pages 351 – 370, Mouton de Gruyter, Berlin, New York. [manuscript: <http://purl.org/stefan.evert/PUB/LuedelingEvert2005.pdf>]

2004

Evert, Stefan (2004a). **A simple LNRE model for random character sequences.** In *Proceedings of the 7èmes Journées Internationales d'Analyse Statistique des Données Textuelles*, pages 411 – 422, Louvain-la-Neuve, Belgium. [<http://purl.org/stefan.evert/PUB/Evert2004a.pdf>]

Evert, Stefan (2004b). **The statistical analysis of morphosyntactic distributions.** In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1539 – 1542, Lisbon, Portugal. [<http://purl.org/stefan.evert/PUB/Evert2004b.pdf>]

Evert, Stefan (2004c). **Significance tests for the evaluation of ranking methods.** In *Proceedings of the 20th International Conference on Computational Linguistics (Coling 2004)*, pages 945 – 951, Geneva, Switzerland. [<http://purl.org/stefan.evert/PUB/Evert2004c.pdf>]

Evert, Stefan; Heid, Ulrich; Spranger, Kristina (2004). **Identifying morphosyntactic preferences in collocations.** In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 907 – 910, Lisbon, Portugal. [<http://purl.org/stefan.evert/PUB/EvertHeidSpranger2004.pdf>]

Evert, Stefan; Heid, Ulrich; Säuberlich, Bettina; Debus-Gregor, Esther; Scholze-Stubenrecht, Werner (2004). **Supporting corpus-based dictionary updating.** In *Proceedings of the 11th Euralex International Congress*, pages 255 – 264, Lorient, France. [<http://purl.org/stefan.evert/PUB/EvertEtc2004.pdf>]

Krenn, Brigitte; Evert, Stefan; Zinsmeister, Heike (2004). **Determining intercoder agreement for a collocation identification task.** In *Proceedings of KONVENS 2004*, pages 89 – 96, Vienna, Austria. [<http://purl.org/stefan.evert/PUB/KrennEvertZinsmeister2004.pdf>]

Lüdeling, Anke and Evert, Stefan (2004). **The emergence of productive non-medical –itis: corpus evidence and qualitative analysis.** In *Proceedings of the First International Conference on Linguistic Evidence*, pages 91 – 95, Tübingen, Germany. [<http://purl.org/stefan.evert/PUB/LuedelingEvert2004.pdf>]

2003

Evert, Stefan and Kermes, Hannah (2003a). **Experiments on candidate data for collocation extraction.**

In *Companion Volume to the Proceedings of the 10th Conference of The European Chapter of the Association for Computational Linguistics*, pages 83 – 86. [<http://purl.org/stefan.evert/PUB/EvertKermes2003a.pdf>]

Evert, Stefan and Kermes, Hannah (2003b). **Annotation, storage, and retrieval of mildly recursive structures**. In K. Simov and P. Osenova (eds.), *Proceedings of the Workshop on Shallow Processing of Large Corpora (SProLaC 2003)*, pages 23 – 33, Lancaster, UK. [<http://purl.org/stefan.evert/PUB/EvertKermes2003b.pdf>]

Carletta, Jean; Kilgour, Jonathan; O'Donnell, Timothy; Evert, Stefan; Voormann, Holger (2003). **The NITE object model library for handling structured linguistic annotation on multimodal data sets**. In *Proceedings of the EACL Workshop on Language Technology and the Semantic Web (3rd Workshop on NLP and XML, NLPXML-2003)*, pages 17 – 24, Budapest, Hungary. [<http://purl.org/stefan.evert/PUB/CarlettaEtc2003.pdf>]

Kermes, Hannah and Evert, Stefan (2003). **Text analysis meets corpus linguistics**. In D. Archer, P. Rayson, A. Wilson, and T. McEnery (eds.), *Proceedings of the Corpus Linguistics 2003 Conference*, pages 402 – 411. UCREL. [<http://purl.org/stefan.evert/PUB/KermesEvert2003.pdf>]

Lüdeling, Anke and Evert, Stefan (2003). **Linguistic experience and productivity: corpus evidence for fine-grained distinctions**. In D. Archer, P. Rayson, A. Wilson, and T. McEnery (eds.), *Proceedings of the Corpus Linguistics 2003 Conference*, pages 475 – 483. UCREL. [<http://purl.org/stefan.evert/PUB/LuedelingEvert2003.pdf>]

2002

Kermes, Hannah and Evert, Stefan (2002). **YAC – a recursive chunker for unrestricted German text**. In M. G. Rodriguez and C. P. Araujo (eds.), *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, volume V, pages 1805 – 1812, Las Palmas, Spain.

2001

Evert, Stefan and Krenn, Brigitte (2001). **Methods for the qualitative evaluation of lexical association measures**. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 188 – 195, Toulouse, France. [<http://purl.org/stefan.evert/PUB/EvertKrenn2001.pdf>]

Evert, Stefan and Lüdeling, Anke (2001). **Measuring morphological productivity: Is automatic preprocessing sufficient?** In P. Rayson, A. Wilson, T. McEnery, A. Hardie, and S. Khoja (eds.), *Proceedings of the Corpus Linguistics 2001 Conference*, pages 167 – 175, Lancaster. UCREL. [<http://purl.org/stefan.evert/PUB/EvertLuedeling2001.pdf>]

Kermes, Hannah and Evert, Stefan (2001). **Exploiting large corpora: A circular process of partial syntactic analysis, corpus query and extraction of lexicographic information**. In P. Rayson, A. Wilson, T. McEnery, A. Hardie, and S. Khoja (eds.), *Proceedings of the Corpus Linguistics 2001 Conference*, pages 332 – 340, Lancaster. UCREL. [<http://purl.org/stefan.evert/PUB/KermesEvert2001.pdf>]

Krenn, Brigitte and Evert, Stefan (2001). **Can we do better than frequency? A case study on extracting PP-verb collocations**. In *Proceedings of the ACL Workshop on Collocations*, pages 39 – 46, Toulouse, France. [<http://purl.org/stefan.evert/PUB/KrennEvert2001.pdf>]

2000

Evert, Stefan; Heid, Ulrich; Lezius, Wolfgang (2000). **Methoden zum Vergleich von Signifikanzmaßen zur Kollokationsidentifikation**. In W. Zühlke and E. G. Schukat-Talamazzini (eds.), *KONVENS-2000 Sprachkommunikation*, pages 215 – 220. VDE-Verlag. [<http://purl.org/stefan.evert/PUB/EvertHeidLezius2000.pdf>]

Berman, Steve; Evert, Stefan; Heid, Ulrich (2000). **Searchable metaspaces**. In *Proceedings of the EAGLES/ISLE Workshop on Metadata*, Athens, Greece.

Heid, Ulrich; Evert, Stefan; Docherty, Vincent; Worsch, Wolfgang; Wermke, Matthias (2000). **A data collection for semi-automatic corpus-based updating of dictionaries**. In U. Heid, S. Evert, E. Lehmann, and C. Rohrer (eds.), *Proceedings of the 9th EURALEX International Congress*, pages 183 – 195.

Lüdeling, Anke; Evert, Stefan; Heid, Ulrich (2000). **On measuring morphological productivity**. In **Wkz** and E. G. Schukat-Talamazzini (eds.), *KONVENS-2000 Sprachkommunikation*, pages 57 – 61. VDE-Verlag. [<http://purl.org/stefan.evert/PUB/LuedelingEvertHeid2000.pdf>]

### Edited Volumes

Rayson, Paul; Villada Moirón, Begoña; Sharoff, Serge; Piao, Scott; Evert, Stefan (2009). **Multiword expressions: hard going or plain sailing?** Special issue of the *International Journal of Language Resources and Evaluation*. [Call for papers: <http://ucrel.lancs.ac.uk/publications/LREJournalMWEcfp.pdf>]

Heid, Ulrich; Evert, Stefan; Lehmann, Egbert; Rohrer, Christian (eds.) (2000). **Proceedings of the 9th EURALEX International Congress**, Stuttgart, Germany.

### Other Publications

Evert, Stefan and Pipa, Gordon (2010). **Probability estimation of rare events in linguistics and computational neuroscience**. Presentation at *KogWis 2010*, Potsdam, Germany.

Pipa, Gordon and Evert, Stefan (2010). **Statistical models of non-randomness in natural language**. Presentation at *KogWis 2010*, Potsdam, Germany.

Ebert, Cornelia; Evert, Stefan; Wilmes, Katharina (2010). **Focus marking via gestures**. Presentation at *Sinn und Bedeutung 15*, Saarbrücken, Germany. [abstract: <http://purl.org/stefan.evert/PUB/EbertEvertWilmes2010.abstract.pdf>]

Evert, Stefan (2009). **Rethinking corpus frequencies**. Presentation at the *ICAME 30 Conference*, Lancaster, UK. [handout: [http://purl.org/stefan.evert/PUB/Evert2009\\_ICAME.handout.pdf](http://purl.org/stefan.evert/PUB/Evert2009_ICAME.handout.pdf)]

Evert, Stefan (2007). **Room for improvement? Upper limits on collocation extraction with statistical association measures**. Poster presentation at the *Computational Linguistics Poster Session at the Annual Meeting of the German Linguistics Association (DGfS 2007)*. [poster: [http://purl.org/stefan.evert/PUB/Evert2007\\_DGfS\\_poster.pdf](http://purl.org/stefan.evert/PUB/Evert2007_DGfS_poster.pdf)]

Evert, Stefan and Baroni, Marco (2006). **ZipfR: Working with words and other rare events in R**. Presentation at the *useR! 2006 Conference*, Vienna, Austria. [handout: [http://purl.org/stefan.evert/PUB/EvertBaroni2006\\_useR.pdf](http://purl.org/stefan.evert/PUB/EvertBaroni2006_useR.pdf)]

Evert, Stefan (2005). **Empirical research on association measures: The UCS toolkit**. Software demonstration at the *Phraseology 2005 Conference*, Louvain-la-Neuve, Belgium. [abstract: <http://purl.org/stefan.evert/PUB/Evert2005phraseology.pdf>]

Evert, Stefan and Krenn, Brigitte (2005). **Exploratory collocation extraction**. Presentation at the *Phraseology 2005 Conference*, Louvain-la-Neuve, Belgium. [abstract: <http://purl.org/stefan.evert/PUB/EvertKrenn2005phraseology.pdf>]

Evert, Stefan and Hoffmann, Sebastian (2005). **BNCweb (CQP edition): The marriage of two corpus tools**. Presentation at the *Corpus Linguistics 2005 Conference*, Birmingham, UK.

Evert, Stefan (2004d). **An on-line repository of association measures**. <http://www.collocations.de/AM/>.

Evert, Stefan; Carletta, Jean; O'Donnell, Timothy J.; Kilgour, Jonathan; Vögele, Andreas; Voormann, Holger (2003). **The NXT object model**. Technical report, IMS, University of Stuttgart. Version 2.1. [<http://purl.org/stefan.evert/PUB/NITE-NOM.pdf>]

Evert, Stefan and Voormann, Holger (2003). **NQL – a query language for multi-modal language data**. Technical report, IMS, University of Stuttgart. Version 2.1. [<http://purl.org/stefan.evert/PUB/NITE-NQL.pdf>]

Evert, Stefan and Kermes, Hannah (2002). **The influence of linguistic pre-processing on candidate data**. In *Workshop on Computational Approaches to Collocations*, Vienna, Austria.

Schönenberger, Manuela and Evert, Stefan (2002). **The benefit of doubt**. Presentation at the *Workshop on*

*Quantitative Investigations in Theoretical Linguistics (QITL)*, Osnabrück, Germany, October 2002. Slides can be downloaded from <http://www.cogsci.uni-osnabrueck.de/qitl/>.

Evert, Stefan (1999). *Das Verhalten von Lösungen der vektoriellen Helmholtzgleichung in Außenräumen für kleine Frequenzen unter elektrischen Randbedingungen*. Unpublished Diplom thesis, University of Stuttgart. [<http://purl.org/stefan.evert/PUB/Evert1999diplom.pdf>]