

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. [official version <<http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/>>, PDF <<http://purl.org/stefan.evert/PUB/Evert2004phd.pdf>>, companion website <<http://www.collocations.de/>>]

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. [ordering information <<http://www.peterlang.com/index.cfm?vID=56315>>]

Journal Papers

Schäfer, Fabian; Evert, Stefan; Heinrich, Philipp (2017). **Japan’s 2014 general election: Political bots, right-wing internet activism and PM Abe Shinzō’s hidden nationalist agenda**. *Big Data*, 5(4). To appear. [manuscript <<http://purl.org/stefan.evert/PUB/SchaeferEvertHeinrich2017.pdf>>]

Evert, Stefan; Proisl, Thomas; Jannidis, Fotis; Reger, Isabella; Pielström, Steffen; Schöch, Christof; Vitt, Thorsten (2017). **Understanding and explaining Delta measures for authorship attribution**. *Digital Scholarship in the Humanities*. Advanced access. [free access (PDF) <<https://academic.oup.com/dsh/article/doi/10.1093/lc/fqx023/3865676/Understanding-and-explaining-Delta-measures-for?guestAccessKey=54b7daa4-be40-4687-880a-543d2b41254e>>]

Evert, Stefan; Greiner, Paul; Baigger, João Filipe; Lang, Bastian (2016). **A distributional approach to open questions in market research**. *Computers in Industry*, 78, 16X28. [accepted manuscript (PDF) <http://purl.org/stefan.evert/PUB/EvertGreinerEtc2016_COMIND.pdf>, journal homepage <<http://dx.doi.org/10.1016/j.compind.2015.10.008>>]

Lapesa, Gabriella and Evert, Stefan (2014). **A large scale evaluation of distributional semantic models: Parameters, interactions and model selection**. *Transactions of the Association for Computational Linguistics*, 2, 531–545. [PDF <<http://purl.org/stefan.evert/PUB/LapesaEvert2014tacl.pdf>>, supplementary material <<http://www.linguistik.fau.de/dsmeval/>>]

Biemann, Chris; Bildhauer, Felix; Evert, Stefan; Goldhahn, Dirk; Quasthoff, Uwe; Schäfer, Roland; Simon, Johannes; Swiezinski, Leonard; Zesch, Torsten (2013). **Scalable construction of high-quality Web corpora**. *Journal for Language Technology and Computational Linguistics (JLCL)*, 28(2), 23–59. [PDF <http://www.jlcl.org/2013_Heft2/2Biemann.pdf>]

Ansorge, Ulrich; Reynvoet, Bert; Hendler, Jessica; Oettl, Lennart; Evert, Stefan (2013). **Conditional automaticity in subliminal morphosyntactic priming**. *Psychological Research*, 77, 399–421.

Michelbacher, Lukas, Evert, Stefan, and Schütze, Hinrich (2011). **Asymmetry in corpus-derived and human word associations**. *Corpus Linguistics and Linguistic Theory*, 7(2), 245–276.

Evert, Stefan (2006). **How random is a corpus? The library metaphor**. *Zeitschrift für Anglistik und Amerikanistik*, 54(2), 177–190. [manuscript (PDF) <<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 (PDF) <<http://purl.org/stefan.evert/PUB/EvertKrenn2005.pdf>>]

Book Chapters

Evert, Stefan (2013). **Tools for the acquisition of lexical combinatorics**. In R. H. Gouws, U. Heid, W. Schweickard, and H. E. Wiegand (eds.), *Dictionaries. An International Encyclopedia of Lexicography*.

Supplementary volume: Recent Developments with Focus on Electronic and Computational Lexicography (HSK 5.4), chapter 104, pages 1415–1432. Mouton de Gruyter, Berlin, New York.

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, pages 1212–1248. Mouton de Gruyter, Berlin. [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, pages 777–803. Mouton de Gruyter, Berlin. [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

2017 – 2016 – 2015 – 2014 – 2013 – 2012 – 2011 – 2010 – 2009 – 2008 – 2007 – 2006 – 2005 – 2004 – 2003 – 2002 – 2001 – 2000

2017

Evert, Stefan; Heinrich, Philipp; Henselmann, Klaus; Rabenstein, Ulrich; Scherr, Elisabeth; Schröder, Lutz (2017). **Combining machine learning and semantic features in the classification of corporate disclosures**. In *Proceedings of the Workshop on Logic and Algorithms in Computational Linguistics (LACompLing 2017)*, pages 47X62, Stockholm, Sweden. [PDF <<http://purl.org/stefan.evert/PUB/EvertHeinrichEtc2017.pdf>>]

Evert, Stefan and Neumann, Stella (2017). **The impact of translation direction on characteristics of translated texts. A multivariate analysis for English and German**. In G. De Sutter, M.–A. Lefer, and I. Delaere (eds.), *Empirical Translation Studies. New Theoretical and Methodological Traditions*, number 300 in Trends in Linguistics. Studies and Monographs (TiLSM). Mouton de Gruyter, Berlin. Online supplement: <<http://www.stefan-evert.de/PUB/EvertNeumann2017/>>.

Evert, Stefan; Uhrig, Peter; Bartsch, Sabine; Proisl, Thomas (2017). **E–VIEW-alation X a large-scale evaluation study of association measures for collocation identification**. In *Electronic lexicography in the 21st century. Proceedings of the eLex 2017 conference*, pages 531X549, Leiden, The Netherlands. [PDF <<http://purl.org/stefan.evert/PUB/EvertUhrigEtc2017.pdf>>, slides <http://purl.org/stefan.evert/PUB/EvertUhrigEtc2017_slides.pdf>, video <<https://www.youtube.com/watch?v=xYo3wTRx8F8>>, E–VIEW-alation <<http://www.collocations.de/eviewalation/>>]

Evert, Stefan; Wankerl, Sebastian; Nöth, Elmar (2017). **Reliable measures of syntactic and lexical complexity: The case of Iris Murdoch**. In *Proceedings of the Corpus Linguistics 2017 Conference*, Birmingham, UK. [PDF <<http://purl.org/stefan.evert/PUB/EvertWankerlNoeth2017.pdf>>, Slides <http://purl.org/stefan.evert/PUB/EvertWankerlNoeth2017_Slides.pdf>]

Lapesa, Gabriella and Evert, Stefan (2017). **Large-scale evaluation of dependency-based DSMS: Are they worth the effort?** In *Proceedings of the 15th Annual Meeting of the European Association for Computational Linguistics (EACL 2017)*, pages 394X400, Valencia, Spain. [PDF <<http://purl.org/stefan.evert/PUB/LapesaEvert2017.pdf>>, supplementary material <<http://www.linguistik.fau.de/dsmeval/>>]

Wankerl, Sebastian; Nöth, Elmar; Evert, Stefan (2017). **An n-gram based approach to the automatic diagnosis of alzheimer's disease from spoken language.** In *Proceeding of INTERSPEECH 2017*, pages 3162X3166, Stockholm, Sweden. [PDF <<http://purl.org/stefan.evert/PUB/WankerlNoethEvert2017.pdf>>]

2016

Beißwenger, Michael; Bartsch, Sabine; Evert, Stefan; Würzner, Kay-Michael (2016). **EmpiriST 2015: A shared task on the automatic linguistic annotation of computer-mediated communication and web corpora.** In *Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task*, pages 44–56, Berlin, Germany. [PDF <<http://purl.org/stefan.evert/PUB/BeisswengerEtc2016.pdf>>, task homepage <<https://sites.google.com/site/empirist2015/>>]

Evert, Stefan (2016). **CogALex-V shared task: Mach5 X a traditional DSM approach to semantic relatedness.** In *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*, pages 92X97, Osaka, Japan. [PDF <http://purl.org/stefan.evert/PUB/Evert2016_Mach5.pdf>, system & data <<http://www.collocations.de/data/#mach5>>]

Evert, Stefan; Jannidis, Fotis; Dimpel, Friedrich Michael; Schöch, Christof; Pielström, Steffen; Vitt, Thorsten; Reger, Isabella; Büttner, Andreas; Proisl, Thomas (2016). **XDeltaX in der stilometrischen Autorschafts Attribution.** In *Proceedings of DHD 2016*, pages 61X74, Leipzig, Germany. [HTML <<http://www.dhd2016.de/abstracts/sektionen-002.html>>]

Santus, Enrico; Gladkova, Anna; Evert, Stefan; Lenci, Alessandro (2016). **The CogALex-V shared task on the corpus-based identification of semantic relations.** In *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*, pages 69X79, Osaka, Japan. [PDF <<http://purl.org/stefan.evert/PUB/SantusEtc2016.pdf>>, task homepage <<https://sites.google.com/site/cogalex2016/home/shared-task>>]

Wankerl, Sebastian; Nöth, Elmar; Evert, Stefan (2016). **An analysis of perplexity to reveal the effects of Alzheimer's disease on language.** In *ITG-Fachbericht 267: Speech Communication*, pages 254X259, Paderborn, Germany. [PDF <<http://purl.org/stefan.evert/PUB/WankerlNoethEvert2016.pdf>>]

2015

Evert, Stefan and Arppe, Antti (2015). **Some theoretical and experimental observations on naïve discriminative learning.** In *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics (QITL-6)*, Tübingen, Germany. [PDF <<http://purl.org/stefan.evert/PUB/EvertArppe2015.pdf>>, handout (PDF) <http://purl.org/stefan.evert/PUB/EvertArppe2015_handout.pdf>, slides (PDF) <http://purl.org/stefan.evert/PUB/EvertArppe2015_slides.pdf>]

Evert, Stefan and Hardie, Andrew (2015). **Zigurat: A new data model and indexing format for large annotated text corpora.** In *Proceedings of the 3rd Workshop on the Challenges in the Management of Large Corpora (CMLC-3)*, pages 21–27, Lancaster, UK. [PDF <<http://purl.org/stefan.evert/PUB/EvertHardie2015.pdf>>]

Evert, Stefan; Proisl, Thomas; Jannidis, Fotis; Pielström, Steffen; Schöch, Christof; Vitt, Thorsten (2015). **Towards a better understanding of Burrows's Delta in literary authorship attribution.** In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, Denver, CO. Co-located with NAACL-HLT 2015. [PDF <http://purl.org/stefan.evert/PUB/EvertEtc2015_CL4Lit.pdf>]

Plotnikova, Nataliia; Kohl, Micha; Volkert, Kevin; Lerner, Andreas; Dykes, Natalie; Ermer, Heiko; Evert, Stefan (2015). **KLUEless: Polarity classification and association.** In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 619–625, Denver, Colorado. [PDF <http://purl.org/stefan.evert/PUB/PlotnikovaEtc2015_KLUEless.pdf>, extended report <http://purl.org/stefan.evert/PUB/PlotnikovaEtc2015_KLUEless_report.pdf> with qualitative evaluation]

Plotnikova, Nataliia; Lapesa, Gabriella; Proisl, Thomas; Evert, Stefan (2015). **SemantiKLUE: Semantic textual similarity with maximum weight matching.** In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 111–116, Denver, Colorado. [PDF <http://purl.org/stefan.evert/PUB/PlotnikovaEtc2015_SemantiKLUE.pdf>]

2014

Bartsch, Sabine and Evert, Stefan (2014). **Towards a Firthian notion of collocation**. In A. Abel and L. Lemnitzer (eds.), *Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetbüchern*, number 2/2014 in OPAL – Online publizierte Arbeiten zur Linguistik, pages 48–61. Institut für Deutsche Sprache, Mannheim. [PDF <<http://purl.org/stefan.evert/PUB/BartschEvert2014.pdf>>]

Diwersy, Sascha, Evert, Stefan, and Neumann, Stella (2014). **A weakly supervised multivariate approach to the study of language variation**. In B. Szmrecsanyi and B. Wälchli, editors, *Aggregating Dialectology, Typology, and Register Analysis. Linguistic Variation in Text and Speech*, *Linguae et Litterae: Publications of the School of Language and Literature, Freiburg Institute for Advanced Studies*. De Gruyter, Berlin. [online access <<http://www.degruyter.com/viewbooktoc/product/207699>>, earlier manuscript (PDF) <<http://purl.org/stefan.evert/PUB/DiwersyEvertNeumann2012.pdf>>]

Evert, Stefan (2014). **Distributional semantics in R with the wordspace package**. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, pages 110–114, Dublin, Ireland. [PDF <http://purl.org/stefan.evert/PUB/Evert2014_wordspace.pdf>, wordspace homepage <<http://wordspace.r-forge.r-project.org/>>]

Evert, Stefan; Proisl, Thomas; Greiner, Paul; Kabashi, Besim (2014). **SentiKLUE: Updating a polarity classifier in 48 hours**. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval–2014)*, Dublin, Ireland. [PDF <http://purl.org/stefan.evert/PUB/EvertEtc2014_SentiKLUE.pdf>]

Lapasa, Gabriella and Evert, Stefan (2014). **NaDiR: Naive distributional response generation**. In *Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex)*, pages 50–59, Dublin, Ireland. [PDF <<http://purl.org/stefan.evert/PUB/LapasaEvert2014.pdf>>]

Lapasa, Gabriella; Evert, Stefan; Schulte im Walde, Sabine (2014). **Contrasting syntagmatic and paradigmatic relations: Insights from distributional semantic models**. In *Proceedings of the Third Joint Conference on Lexical and Computational Semantics (*SEM 2014)*, pages 160–170, Dublin, Ireland. [PDF <<http://purl.org/stefan.evert/PUB/LapasaEtc2014.pdf>>]

Proisl, Thomas; Evert, Stefan; Greiner, Paul; Kabashi, Besim (2014). **SemantiKLUE: Robust semantic similarity at multiple levels using maximum weight matching**. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval–2014)*, Dublin, Ireland. [PDF <http://purl.org/stefan.evert/PUB/ProislEtc2014_SemantiKLUE.pdf>]

Schulze Wettendorf, Clemens; Jegan, Robin; Körner, Allan; Zerche, Julia; Plotnikova, Nataliia; Moreth, Julian; Schertl, Tamara; Obermeyer, Verena; Streil, Susanne; Willacker, Tamara; Evert, Stefan (2014). **SNAP: A multi-stage XML-pipeline for aspect based sentiment analysis**. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 578X584, Dublin, Ireland. [PDF <http://purl.org/stefan.evert/PUB/SchulzeWettendorfEtc2014_SNAP.pdf>]

2013

Greiner, Paul; Proisl, Thomas; Evert, Stefan; Kabashi, Besim (2013). **KLUE-CORE: A regression model of semantic textual similarity**. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity*, pages 181–186, Atlanta, Georgia, USA. [PDF <<http://purl.org/stefan.evert/PUB/GreinerEtc2013.pdf>>]

Lapasa, Gabriella and Evert, Stefan (2013). **Evaluating neighbor rank and distance measures as predictors of semantic priming**. In *Proceedings of the ACL Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2013)*, Sofia, Bulgaria. [PDF <<http://purl.org/stefan.evert/PUB/LapasaEvert2013a.pdf>>]

Proisl, Thomas; Greiner, Paul; Evert, Stefan; Kabashi, Besim (2013). **KLUE: Simple and robust methods for polarity classification**. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 395–401, Atlanta, Georgia, USA. [PDF <<http://purl.org/stefan.evert/PUB/ProislEtc2013.pdf>>]

2012

Boleda, Gemma; Evert, Stefan; Gehrke, Berit; McNally, Louise (2012). **Adjectives as saturators vs.**

modifiers: Statistical evidence. In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, and M. Westera (eds.), *Logic, Language and Meaning. Proceedings of the 18th Amsterdam Colloquium*, volume 7218 of *Lecture Notes in Computer Science*, pages 112–121. Springer, Berlin, Heidelberg. [PDF <<http://purl.org/stefan.evert/PUB/BoledaEtc2012.pdf>>]

2011

Evert, Stefan and Hardie, Andrew (2011). **Twenty-first century corpus workbench: Updating a query architecture for the new millennium.** In *Proceedings of the Corpus Linguistics 2011 Conference*, Birmingham, UK. [PDF <<http://purl.org/stefan.evert/PUB/EvertHardie2011.pdf>>]

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. [PDF <<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. [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. [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. [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. [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. [PDF <<http://purl.org/stefan.evert/PUB/BaroniEvert2007.pdf>>, talk slides (PDF) <<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; Märtin, 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. [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. [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. [PDF <<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 (PDF) <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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 (PDF) <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<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. [PDF <<http://purl.org/stefan.evert/PUB/EvertKrenn2001.pdf>>, colour plots <<http://www.collocations.de/EK/ACL01/figures-ek.html>>]

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. [PDF <<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. [PDF <<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. [PDF <<http://purl.org/stefan.evert/PUB/KrennEvert2001.pdf>>, colour plots <<http://www.collocations.de/EK/ACL01/figures-ke.html>>]

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. [PDF <<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 W. Zühlke and E. G. Schukat-Talamazzini (eds.), *KONVENS–2000 Sprachkommunikation*, pages 57–61. VDE-Verlag. [PDF <<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 (2017a). **Measures of productivity and lexical diversity.** Poster at the *ICAME 38 Conference*, Prague, Czech Republic. [abstract <http://purl.org/stefan.evert/PUB/Evert2017_ICAME.pdf>, poster <http://purl.org/stefan.evert/PUB/Evert2017_ICAME_poster.pdf>] X best poster award

Evert, Stefan (2017b). **Making sense of multivariate analyses of linguistic variation.** Poster at the *Corpus Linguistics 2017 Conference*, Birmingham, UK. [abstract <http://purl.org/stefan.evert/PUB/Evert2017CL/Evert2017CL_Abstract.pdf>, poster <http://purl.org/stefan.evert/PUB/Evert2017CL/Evert2017CL_Poster.pdf>, additional material <<http://purl.org/stefan.evert/PUB/Evert2017CL/>>]

Evert, Stefan; Heinrich, Philipp; Schäfer, Fabian (2017). **Social Bots in Japan’s 2014 General Election: Preliminary Results from a Corpus-Linguistic and Qualitative Study of Computational Propaganda on Twitter.** Presentation at the *International Conference on Computational Social Science (IC2S2 2017)*, Cologne, Germany. [abstract <http://purl.org/stefan.evert/PUB/EvertHeinrichSchaefer2017_abstract.pdf>]

Neumann, Stella; Evert, Stefan; De Sutter, Gert (2017). **Register-specific interference in translation.** Presentation at the *Annual Meeting of the German Linguistics Association (DGfS 2017)*, Saarbrücken, Germany. [abstract <http://purl.org/stefan.evert/PUB/NeumannEvertDeSutter2017_abstract.pdf>, slides <http://purl.org/stefan.evert/PUB/NeumannEvertDeSutter2017_slides.pdf>]

Evert, Stefan; Jannidis, Fotis; Proisl, Thomas; Vitt, Thorsten; Schöch, Christof; Pielström, Steffen; Reger, Isabella (2016). **Outliers or Key Profiles? Understanding Distance Measures for Authorship Attribution.** Presentation at *Digital Humanities 2016*, Kraków, Poland. [abstract <<http://dh2016.adho.org/abstracts/253>>]

Evert, Stefan; Schneider, Gerold; Brezina, Vaclav; Gries, Stefan Th.; Lijffijt, Jeffrey; Rayson, Paul; Wallis, Sean; Hardie, Andrew (2015). **Corpus statistics: key issues and controversies.** Panel discussion at *Corpus Linguistics 2015*, Lancaster, UK. [abstract <http://purl.org/stefan.evert/PUB/EvertSchneiderEtc2015_abstract.pdf>]

Evert, Stefan; Proisl, Thomas; Schöch, Christof; Jannidis, Fotis; Pielström, Steffen; Vitt, Thorsten (2015).

Explaining Delta, or: How do distance measures for authorship attribution work? Presentation at *Corpus Linguistics 2015*, Lancaster, UK. [abstract http://purl.org/stefan.evert/PUB/EvertProislEtc2015_abstract.pdf], slides http://purl.org/stefan.evert/PUB/EvertProislEtc2015_CL_Slides.pdf]

Bartsch, Sabine; Evert, Stefan; Proisl, Thomas; Uhrig, Peter (2015). **(Association) measure for measure.** Presentation at *ICAME 36*, Trier, Germany.

Lapesa, Gabriella; Schulte im Walde, Sabine; Evert, Stefan (2014). **Judging Paradigmatic Relations: A Collection of Ratings for English.** Poster at *Architecture and Mechanisms of Language Processing (AMLAP-2014)*, Edinburgh, UK. [poster (PDF) http://cogsci.uni-osnabrueck.de/~glapesa/materials/poster_AMLAP_14_Edinburgh_2.pdf]

Evert, Stefan and Neumann, Stella (2013). **The impact of translation direction on the characteristics of translated texts: a multivariate analysis for English and German.** Presentation at the *46th Annual Meeting of the Societas Linguistica Europaea*, Split, Croatia.

Bartsch, Sabine and Evert, Stefan (2013b). **Exploring the Firthian notion of collocation.** Presentation at *Corpus Linguistics 2013*, Lancaster, UK.

Evert, Stefan; Schneider, Gerold; Lehmann, Hans Martin (2013). **Statistical modelling of natural language for descriptive linguistics.** Presentation at *Corpus Linguistics 2013*, Lancaster, UK.

Lapesa, Gabriella and Evert, Stefan (2013b). **Thematic Roles and Semantic Space. Insights from Distributional Semantic Models.** Presentation at *Quantitative Investigations in Theoretical Linguistics (QITL-5)*, Leuven, Belgium. [abstract (PDF) http://cogsci.uni-osnabrueck.de/~glapesa/materials/QITL5-abstract_submitted.pdf], slides (PDF) http://cogsci.uni-osnabrueck.de/~glapesa/materials/presentation_qitl.pdf]

Lapesa, Gabriella and Evert, Stefan (2013a). **Item-based Prediction of Reaction Times in Priming: an Evaluation of Distributional Semantic Models.** Poster at *Architecture and Mechanisms of Language Processing (AMLAP-2013)*, Marseille, France. [poster (PDF) http://cogsci.uni-osnabrueck.de/~glapesa/materials/poster_AMLAP.pdf]

Sánchez Marco, Cristina; Marín, Rafael; Evert, Stefan (2012). **The lexical extension of *estar* + participle through psychological verbs.** Presentation at the *Annual Meeting of the German Linguistics Association (DGfS 2012)*, Frankfurt, Germany.

Sánchez Marco, Cristina; Marín, Rafael; Evert, Stefan (2012). **Measuring lexical extension: The case of Spanish *estar* + past participle.** Poster presentation at *Linguistic Evidence 2012*, Tübingen, Germany. [extended abstract (PDF) <http://purl.org/stefan.evert/PUB/SanchezEtc2012.pdf>]

Evert, Stefan (2011). **Quantitative measures of productivity and their significance.** Presentation at *Corpus Linguistics 2011*, Birmingham, UK. [handout (PDF) http://purl.org/stefan.evert/PUB/Evert2011_productivity.handout.pdf]

Sánchez Marco, Cristina and Evert, Stefan (2011). **Measuring semantic change: The case of Spanish participial constructions.** Poster presentation at *QITL-4*, Berlin, Germany.

Evert, Stefan and Pipa, Gordon (2010). **Probability estimation of rare events in linguistics and computational neuroscience.** Presentation at *KogWis 2010*, Potsdam, Germany. [abstract http://purl.org/stefan.evert/PUB/EvertPipa2010_abstract.pdf], handout http://purl.org/stefan.evert/PUB/EvertPipa2010_handout.pdf]

Pipa, Gordon and Evert, Stefan (2010). **Statistical models of non-randomness in natural language.** Presentation at *KogWis 2010*, Potsdam, Germany. [abstract (PDF) http://purl.org/stefan.evert/PUB/PipaEvert2010_abstract.pdf]

Evert, Stefan (2009). **Rethinking corpus frequencies.** Presentation at the *ICAME 30 Conference*, Lancaster, UK. [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 (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 (PDF) <http://purl.org/stefan.evert/PUB/EvertBaroni2006_useR.pdf>]
- Lüdeling, Anke; Baroni, Marco; Evert, Stefan (2006). **Need and competition in word formation and where to find data to study them**. Poster presentation at the *Second International Conference on Linguistic Evidence*, Tübingen, Germany. [abstract (PDF) <http://purl.org/stefan.evert/PUB/LuedelingBaroniEvert2006_LingEv.pdf>, poster (PDF) <http://purl.org/stefan.evert/PUB/LuedelingBaroniEvert2006_LingEv_poster.pdf>]
- Lüdeling, Anke; Baroni, Marco; Evert, Stefan (2006). **Need and competition: Deconstructing quantitative productivity**. Presentation at the *Second Conference on Quantitative Investigations in Theoretical Linguistics (QITL-2)*, Osnabrück, Germany. [abstract (PDF) <http://purl.org/stefan.evert/PUB/LuedelingBaroniEvert2006_QITL.pdf>]
- Evert, Stefan (2005). **Empirical research on association measures: The UCS toolkit**. Software demonstration at the *Phraseology 2005* Conference, Louvain-la-Neuve, Belgium. [abstract (PDF) <<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 (PDF) <<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. [PDF <<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. [PDF <<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 vektorialen Helmholtzgleichung in Außenräumen für kleine Frequenzen unter elektrischen Randbedingungen**. Unpublished Diplom thesis, University of Stuttgart. [PDF <<http://purl.org/stefan.evert/PUB/Evert1999diplom.pdf>>]