Roland Schäfer
Academic CV
July 2014

Research interests

I am a linguist currently working on the construction of very large web corpora and their linguistic exploration and characterization. After my education in Comparative Linguistics, General Linguistics, and Japanese Linguistics I concentrated on formal linguistics, especially model-theoretic semantics and syntax. This led to my doctorate thesis, which is a model-theoretic account of arguments and adjuncts in event semantics, including a computational implementation. After I had finished the doctorate, my interest shifted to empirical linguistics, specifically corpus linguistics and statistical procedures. Working in the German Department of Freie Universität Berlin, where I taught mostly future teachers, also made me aware of the great value of more theory-neutral approaches to language description. As a result, I am now working on my *Habilitationsschrift* (second thesis as required in the German academic system), which is an extensive corpus-based account of exceptive particles (like English *except for* and *but*) in German, English, and Swedish with reference to the highly formal semantic theories proposed for such particles.

For my corpus linguistic studies, I required corpora of considerable size which also included a lot of linguistic variation, possibly even non-standard variation. Since I could not find such corpora for my languages of interest, I developed a family of web-derived corpora in the region between one and ten billion tokens with my colleague Felix Bildhauer from 2011 to 2014. A great deal of our publication output is devoted to our innovative work on data collection, processing, and corpus assessment. In fact, we have become two of the major international researchers in the field and provide the largest corpora available in Dutch, German, French, Spanish, and Swedish. Since 2013, I have resumed my original linguistic work based on these corpora.

My current object languages in linguistic research are German, English, and Swedish. However, I have studied many languages in the course of my studies and even taught some of them. Among those languages are such typologically diverse ones as Japanese, Latin, Middle Welsh, and Ossetic.

Education

since 2009	Habilitation in General and German Linguistics at Freie Universität Berlin
2003 – 2008	Doctorate in English Linguistics at Georg-August University, Göttingen Title: *Arguments and Adjuncts at the Syntax-Semantics Interface* Supervisors: Gert Webelhuth, Regine Eckardt
2001	Magister Artium in Comparative and General Linguistics at Philipps University, Marburg Title: *Typology of Changes in Consonant Clusters*
1994 – 2000	Studies in Comparative Linguistics, General Linguistics, Japanese Linguistics at Philipps University, Marburg

Employment

2014 –	Research engineer at Språkbanken, University of Gothenburg
Winter 2011	Visiting professor for English syntax at Georg-August University, Göttingen
2007 – 2014	Post-doc at Freie Universität Berlin Department of German Grammar, chair of Stefan Müller
2002 – 2007	Research assistant at Georg-August University, Göttingen Department of Comparative and General Linguistics

Conferences and workshops

2014 9th Web as Corpus Workshop workshop (WAC-9) endorsed by ACL SIGWAC, workshop at the EACL annual conference 2014 (Göteborg), with Felix Bildhauer

2014 Web Data as a Challenge for Theoretical Linguistics and Corpus Design, workshop at the annual conference of the German Linguistic Society 2014 (Marburg/Lahn), with Felix Bildhauer

2013 Linguistic research with large annotated web corpora, pre-conference tutorial at the 20th international conference on Head-Driven Phrase Structure Grammar (Berlin), with Felix Bildhauer

Refereeing

2014 Language Resources and Evaluation (conference)

Linguistic Evidence (conference)

2013 Journal of Germanic Linguistics (Cambridge University Press)

Lingua (Elsevier)

Monographs and collected volumes by Benjamins, Elsevier, Language Science Press

Talks (selection)

2014 Bildhauer, Felix ; Schäfer, Roland: Sehr große Webkorpora – Aufbau, Zusammensetzung und Anwendung ("Very large web corpora – Construction, Composition, and Application"). Invited talk at Institut für Deutsche Sprache (IDS), Mannheim.

Schäfer, Roland ; Barbaresi, Adrien ; Bildhauer, Felix: Focused Web Corpus Crawling. 9th Web as Corpus workshop (WAC-9). EACL annual conference 2014, Göteborg.

2013 Schäfer, Roland ; Barbaresi, Adrien ; Bildhauer, Felix: The Good, the Bad, and the Hazy: Design Decisions in Web Corpus Construction. 8th Web as Corpus Workshop (WAC-8). Corpus Linguistics 2013, Lancaster.

Bildhauer, Felix ; Schäfer, Roland: Token-level noise in large Web corpora and non-destructive normalization for linguistic applications. Corpus Analysis with Noise in the Signal (CANS 2013). Corpus Linguistics 2013, Lancaster.

Schäfer, Roland: Inflectional Alternations in German Weak Nouns. Workshop "Usage-Based Approaches to Morphology". Annual conference of the German Linguistic Society 2013, Potsdam.

2012 Schäfer, Roland: Linguistic constructions related to the distinction between emotion-denoting and affect-denoting German nouns. Annual conference of SLE 2012, Stockholm.

2010 Maché, Jakob ; Schäfer, Roland: From Parts of Speech to Parts of Body. Conference of the International Society for Gesture Studies 2010, Frankfurt/Oder.

Schäfer, Roland: On Exceptional Vorfeldbesetzung. Workshop "Quirky Vorfeld Phenomena". Annual conference of the German Linguistic Society 2010, Berlin.

2006 Schäfer, Roland: Frequency Adjectives Force NPs to Have Event Readings. Sinn und Bedeutung 11, Barcelona.

Teaching (selection)

Titles in *italics* denote courses taught in German (titles translated into English).

Siegen University, German Department

2014 *Graduate workshop on corpora and statistics with R*

European Summer School in Logic, Language and Information (ESSLLI)

2012 Building large corpora from the web, foundational course at ESSLLI 2012, Opole (with Felix Bildhauer)

Freie Universität Berlin

Summer 2014	*Statistics for linguists and quantitative linguistics* (BA)
Winter 2013	*Introduction to the grammatical description of German* (BA)
	Advanced German Grammar (BA)
Winter 2012	*Introduction to the grammatical description of German* (BA)
Summer 2012	*Grammatical description of contemporary German* (BA)
Winter 2010	*Introduction to linguistics* (BA)
	Introduction to the grammatical description of German (BA)
	Advanced descriptive grammar (BA)
Winter 2009	*Introduction to linguistics* (BA)
	Introduction to the grammatical description of German (BA)
Summer 2009	*Empirical methods in linguistics* (BA)
Winter 2008	*Introduction to linguistics* (BA)
	Introduction to the grammatical description of German (BA)
Summer 2008	*Non-segmental phonology: tone and intonation* (BA)

Georg-August University, Göttingen

Winter 2011	Graduate colloquium in English linguistics
	Descriptive grammar and linguistic theory (BA, MA)
	Corpus linguistics (BA)
	Construction Grammar (BA)
Summer 2007	Game-theoretical semantics and pragmatics (MA)
Winter 2006	Adverbials (MA, with Regine Eckardt)
Summer 2006	Syntax I : traditional grammar (BA)
	Minimalism (MA, with Götz Keydana)
Winter 2005	Formal approaches to the syntax-semantics interface (MA, with Manfred Sailer)
Summer 2005	*Introduction to formal and computational semantics* (BA)
Winter 2004	*Connectionism and implementation of artificial neural nets* (BA)
	Middle Welsh II (BA)

Summer 2004	*Formal semantics* (BA)
	Middle Welsh I (BA)
Winter 2003	*Optimality Theory* (BA)
	Ossetic II (BA)
Summer 2003	*Phonology* (BA)
	Ossetic I (BA)

Major software projects

2011 – 2014	*texrex* fully featured web document cleaning and web corpus creation software (20,000 lines of code)