

# Dr. Roland Schäfer | Lebenslauf

roland.schaefer@fu-berlin.de | Email  
http://hpsg.fu-berlin.de/~rsling/ | Web

6. Januar 1974 | geboren in Düsseldorf, Nordrhein-Westfalen  
1993 | Abitur in Bad Arolsen, Hessen

## Forschungsschwerpunkte

- ◇ Morphosyntax und Graphematik des Gegenwartsdeutschen
- ◇ empirische Methoden und statistische Verfahren in der Linguistik
- ◇ kognitiv ausgerichtete linguistische Theoriebildung
- ◇ Korpuserstellung und Korpusevaluation

## Akademische Ausbildung

2003–2008	<b>Promotion</b> in Englischer Sprachwissenschaft an der Georg-August-Universität Göttingen, <i>magna cum laude</i> , Arbeit <i>Arguments and Adjuncts at the Syntax-Semantics Interface</i> bei Prof. Dr. Gert Webelhuth und Prof. Dr. Regine Eckardt
2002	<b>Magister Artium</b> in Vergleichender Sprachwissenschaft an der Pilipps-Universität Marburg, Note 1.0, Arbeit <i>Typologie des Wandels von Konsonantenverbindungen</i> bei Prof. Dr. Michael Job
1993–2002 (unterbrochen 2000 SS–2001 WS)	<b>Studium</b> der Vergleichenden Sprachwissenschaft, Allgemeinen Sprachwissenschaft und Japanologie an der Philipps-Universität Marburg

## Akademische Tätigkeiten

2016 SS	<b>Vertretung der W3-Professur</b> für Deutsche Grammatik an der Freien Universität Berlin
seit Jan. 2015 (beurlaubt 2016 SS)	<b>Projektleiter</b> im selbst eingeworbenen DFG-Projekt <i>Linguistische Webcharakterisierung und Webkorpuserstellung</i> (DFG SCHA1916/1-1) an der Freien Universität Berlin
Sep.–Nov. 2014	<b>Forschungsingenieur</b> bei Språkbanken an der Universität Göteborg
2011 WS	<b>Vertretung der W3-Professur</b> für Englische Syntax an der Georg-August-Universität Göttingen
2007 WS–2014 SS (beurlaubt 2011 WS)	<b>Wissenschaftlicher Mitarbeiter</b> am Lehrstuhl Prof. Dr. Stefan Müller für Deutsche Grammatik an der Freien Universität Berlin
2002 WS–2007 SS	<b>Wissenschaftlicher Mitarbeiter</b> am Lehrstuhl Prof. Dr. Michael Job für Allgemeine und Vergleichende Sprachwissenschaft an der Georg-August-Universität Göttingen

## Nichtakademische Tätigkeiten

2000–2001 | Leiter Kundenkommunikation bei PrimusAvitos AG

## Eingeworbene Drittmittel

Im Jahr 2014 wurde mein korpuslinguistisches Projekt *Linguistische Webcharakterisierung und Webkorpuserstellung* von der Deutschen Forschungsgemeinschaft (DFG) bewilligt. In dem Projekt wird die Zusammensetzung großer Korpora aus Webdaten untersucht, insbesondere durch Entwicklung und Evaluation automatischer Verfahren zur Klassifikation der in den Korpora enthaltenen Dokumente nach linguistisch relevanten inhaltlichen und grammatischen Kriterien.

Typ	Sachmittel/Eigene Stelle
Fördersumme	286.100€
Laufzeit	3 Jahre
DFG-Kennzeichen	SCHA1916/1-1

In dem Projekt bestehen Kooperationen mit folgenden Einrichtungen:

- ◇ Friedrich-Alexander Universität Erlangen-Nürnberg,  
Korpuslinguistik (Prof. Dr. Stefan Evert)
- ◇ Georgetown University, Washington, DC,  
Korpuslinguistik (Ass. Prof. Dr. Amir Zeldes)
- ◇ Institut für Deutsche Sprache, Mannheim,  
Abteilung Lexik (Prof. Dr. Stefan Engelberg) und  
Abteilung Grammatik (Prof. Dr. Angelika Wöllstein)
- ◇ Institut für Maschinelle Sprachverarbeitung, Stuttgart  
(PD Dr. Sabine Schulte im Walde)

## Betreute Abschlussarbeiten

Hinweis: Der Fachbereich Philosophie und Geisteswissenschaften der Freien Universität Berlin erlaubt die Betreuung von Abschlussarbeiten durch wissenschaftliche Mitarbeiter nur mit Sondergenehmigung und nur bis zum Bachelor of Arts.

2016	Erstprüfer: <i>Einflussfaktoren der Parallel- und Wechselflexion bei Nominalphrasen mit zwei attributiven Adjektiven</i> im Fach Deutsche Philologie, Freie Universität Berlin, von Lisa Krack.
2015	Erstprüfer: <i>Das Passiv in der Transformationsgrammatik</i> im Fach Deutsche Philologie, Freie Universität Berlin, von Chiara Jancke.
2013	Erstprüfer: <i>Kasuzuweisung zur Artangabe pseudopartitiver Konstruktionen</i> im Fach Deutsche Philologie an der Freien Universität Berlin, von Samuel Reichert.

- |        |   |
|--------|---|
| (2013) | Zweitprüfer: <i>Eigennamenerkennung in Web-Korpora des Deutschen. Eine Herausforderung für die (Computer)linguistik</i> im Fach Germanistische Linguistik an der Humboldt Universität zu Berlin, von Lea Arianna Helmers. |
|        | Erstprüfer: <i>Mögliche Prädiktoren für die Pluralintegration von Fremdwörtern – eine korpuslinguistische Untersuchung</i> im Fach Deutsche Philologie an der Freien Universität Berlin, von Elisabeth Metz.              |
| 2011   | Zweitprüfer: <i>Die Flexionstypen des deutschen Adjektivs</i> im Fach Deutsche Philologie an der Freien Universität Berlin, von Nicolai Sinn.   |

## Akademische Gremien- und Verwaltungstätigkeit

- |           |  |
|-----------|--|
| 2011–2014 | Mitglied im Mentoringausschuss<br>Deutsche und niederländische Philologie<br>Freie Universität Berlin  |
| 2009–2014 | Modulbeauftragter für Basis- und Aufbauphase<br>Linguistik, Deutsche und niederländische Philologie<br>Freie Universität Berlin  |
| 2009–2014 | Koordinator für das System zur Verwaltung<br>von Studienleistungen ( <i>Campus Management</i> )<br>Linguistik, Deutsche und niederländische Philologie<br>Freie Universität Berlin |
| 2004–2005 | Mitarbeit an der Ordnung für die BA/MA-Studiengänge<br>in Allgemeiner und Indogermanischer Sprachwissenschaft<br>an der Georg-August Universität Göttingen                         |
| 2000–2001 | Studentischer Vertreter in der Auswahlkommission zur<br>Besetzung der W3-Professur für Allgemeine und Vergleichende<br>Sprachwissenschaft an der Philipps-Universität Marburg      |

## Mitgliedschaften in Expertengremien

2013–2014	Mitglied der Expertenrunde der DFG zur Erarbeitung der Richtlinien zu technischen und rechtlichen Aspekten der Handhabung von Sprachkorpora
-----------	---

## Gutachtertätigkeit: Zeitschriften und Verlage

2014	Language Science Press (anonyme Buchrezension)
2013	Journal of Germanic Linguistics
2012	Lingua

## Gutachtertätigkeit: Konferenzen

2016	KONVENS, Bochum Linguistic Evidence, Tübingen Language Resources and Evaluation, Portorož
2014	Linguistic Evidence, Tübingen Language Resources and Evaluation, Reykjavík Challenges in the Management of Large Corpora, Lancaster
2010	The 17th International Conference on Head-Driven Phrase Structure Grammar, Paris

## Veröffentlichte Ressourcen und Software

Alle meine Softwareprojekte sind quelloffen und öffentlich auf der Plattform *GitHub* verfügbar: <https://github.com/rsling>.

seit 2016	Entwicklung der <i>COReX</i> und <i>COReCO</i> Software zur automatischen Annotation großer Korpora nach Textthema und Texttyp/Register, in Kooperation mit dem Institut für Deutsche Sprache (Felix Bildhauer)
seit 2014	Entwicklung der Online-Korpusplattform <i>Colibri</i> <sup>2</sup> <a href="https://webcorpora.org">https://webcorpora.org</a>
seit 2011	Erstellung der <i>COW</i> Webkorpora (mit Felix Bildhauer) für Deutsch (20 Mrd. Tokens), Englisch (17 Mrd. Tokens), Französisch (11 Mrd. Tokens), Niederländisch (9 Mrd. Tokens), Schwedisch (8,5 Mrd. Tokens), Spanisch (7 Mrd. Tokens) <a href="http://corporafromtheweb.org">http://corporafromtheweb.org</a>
seit 2011	Entwicklung der quelloffenen Software <i>texrex</i> zur vollautomatischen Webkorpus-Erstellung
seit 2011	Entwicklung der quelloffenen <i>COW Toolchain</i> zur linguistischen Annotation von Korpora

## Ämter in Vereinen und Verbänden

seit 2015	Vorsitzender der <i>Special Interest Group for Web as Corpus</i> der Association for Computational Linguistics
seit 2015	Organisator des Arbeitskreises <i>Digital Humanities</i> der Gesellschaft für Computerlinguistik und Sprachtechnologie

## Organisierte Tagungen

- |      |  |
|------|--|
| 2016 | 10th Web as Corpus Workshop (WAC-X) bei ACL, Berlin, mit Paul C. Cook, Stefan Evert, Egon Stemle                         |
| 2014 | DGfS Jahrestagung, AG <i>Web Data as a Challenge for Theoretical Linguistics and Corpus Design</i> , mit Felix Bildhauer |
|      | 9th Web as Corpus Workshop (WAC-9) bei EACL, Göteborg, mit Felix Bildhauer   |

## Vorträge (Auswahl)

- |      |   |
|------|---|
| 2017 | Felix Bildhauer & Roland Schäfer: <i>Automatic register annotation for linguistic research?</i> DGfS Jahrestagung, AG <i>Register in linguistic theory: Modeling functional variation</i> , Saarbrücken.            |
| 2016 | Roland Schäfer: <i>The Register Casino – Should you risk your grammar in an outside bet?</i> Grammar and Corpora 2016, Mannheim.  |
|      | Felix Bildhauer & Roland Schäfer: <i>Modelling alternation phenomena with aggregated and non-aggregated predictors.</i> Grammar and Corpora 2016, Mannheim.   |
|      | Felix Bildhauer & Roland Schäfer: <i>Induktive Topikmodellierung und extrinsische Topikdomänen.</i> Jahrestagung des Instituts für Deutsche Sprache 2016, Mannheim.   |
|      | Roland Schäfer: <i>On bias-free crawling and representative web corpora.</i> 10th Web as Corpus workshop (WAC-X) bei ACL 2016, Berlin.  |
|      | Roland Schäfer & Felix Bildhauer: <i>Automatic classification by topic domain for meta data generation, web corpus evaluation, and corpus comparison.</i> 10th Web as Corpus workshop (WAC-X) bei ACL 2016, Berlin. |

- 2015 Roland Schäfer: *Competing constructions for German measure NPs*. Corpus Linguistics 2015, Lancaster.
- Roland Schäfer: *Processing and querying large web corpora with the COW14 architecture*. Challenges in the Management of Large Corpora (CMLC-3) bei Corpus Linguistics 2015, Lancaster.
- 2014 Felix Bildhauer & Roland Schäfer: *Sehr große Webkorpora – Aufbau, Zusammensetzung und Anwendung*. Gastvortrag auf Einladung am Institut für Deutsche Sprache, Mannheim.
- Roland Schäfer, Adrien Barbaresi & Felix Bildhauer: *Focused web corpus crawling*. 9th Web as Corpus workshop (WAC-9) bei EACL 2014, Göteborg.
- 2013 Roland Schäfer, Adrien Barbaresi & Felix Bildhauer: *The good, the bad, and the hazy: design decisions in web corpus construction*. 8th Web as Corpus workshop (WAC-8) bei Corpus Linguistics 2013, Lancaster.
- Felix Bildhauer & Roland Schäfer: *Token-level noise in large Web corpora and non-destructive normalization for linguistic applications*. Corpus Analysis with Noise in the Signal (CANS 2013) bei Corpus Linguistics 2013, Lancaster.
- Roland Schäfer: *Inflectional alternations in German weak nouns*. DGfS Jahrestagung, AG Usage-Based Approaches to Morphology, Potsdam.
- Ulrike Sayatz & Roland Schäfer: *Klitika und Apostrophschreibung in Webkorpora: Graphematik und Registerklassifikation*. DGfS Jahrestagung, AG Modelling Non-Standardized Writing, Potsdam.
- 2012 Roland Schäfer: *Linguistic constructions related to the distinction between emotion-denoting and affect-denoting German nouns*. Jahrestagung der SLE 2012, Stockholm.
- Roland Schäfer & Felix Bildhauer: *Why only large web corpora provide answers to certain linguistic questions*. Jahrestagung der SLE 2012, Stockholm.



- 2010 | Jakob Maché & Roland Schäfer: *From parts of speech to parts of body*. Conference of the International Society for Gesture Studies 2010, Frankfurt/Oder.
- Roland Schäfer: *On exceptional Vorfeldbesetzung*. DGfS Jahrestagung, AG *Quirky Vorfeld Phenomena*, Berlin.
- 2007 | Jakob Maché & Roland Schäfer: *Linking Gestures*. Generative Grammatik des Südens, Leipzig.
- 2006 | Roland Schäfer: *Frequency Adjectives Force NPs to Have Event Readings*. Sinn und Bedeutung 11, Barcelona.