

# Dr. Roland Schäfer | Lebenslauf

roland.schaefer@fu-berlin.de | Email  
http://rolandschaefer.net | Web

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

## Forschungsschwerpunkte

- ◇ Morphologie und Syntax des Gegenwartsdeutschen
- ◇ gebrauchsbasierte und probabilistische linguistische Theorie
- ◇ Grammatikvermittlung in Schulen und linguistische Ausbildung für das Lehramt
- ◇ empirische Methoden und statistische Verfahren
- ◇ Korpuserstellung und Korpusevaluation

## Akademische Ausbildung

2013 – 2019	<b>Habilitation</b> <i>Venia Legendi</i> in Allgemeiner Sprachwissenschaft und Germanistischer Linguistik, Humboldt-Universität zu Berlin, Thema <i>Probabilistic German Morphosyntax</i> , Gutachter: Anke Lüdeling, Stefan Müller, Matthias Hüning
2003 – 2008	<b>Promotion</b> <i>magna cum laude</i> in Englischer Sprachwissenschaft, Georg-August-Universität Göttingen, Thema <i>Arguments and Adjuncts at the Syntax-Semantics Interface</i> , Gutachter*innen: Gert Webelhuth, Regine Eckardt
2002	<b>Magister Artium</b> (1,0) in Allgemeiner, Vergleichender und Japanologischer Sprachwissenschaft, Philipps-Universität Marburg, Thema <i>Typologie des Wandels von Konsonantenverbindungen</i> , Gutachter: Michael Job, Götz Keydana
1993 – 2002*	<b>Studium</b> der Allgemeinen Sprachwissenschaft, Japanologie und Vergleichenden Sprachwissenschaft an der Philipps-Universität Marburg

\* Das Studium wurde 2000–2001 wegen Berufstätigkeit unterbrochen.

## Akademische Tätigkeiten (einschließlich Gastprofessuren)

ab April 2020	<b>Privatdozent</b> an der Sprach- und Literaturwissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
ab April 2020	<b>Wissenschaftlicher Mitarbeiter</b> im Projekt A04 <i>Situated Syntax</i> des SFB 1412 <i>Register</i> an der Humboldt-Universität zu Berlin
2018–2020	<b>Vertretung der W3-Professur</b> für Deutsche Grammatik an der Freien Universität Berlin
2016 SS	<b>Vertretung der W3-Professur</b> für Deutsche Grammatik an der Freien Universität Berlin
2015 – 2018	<b>Projektleiter</b> im selbst erworbenen DFG-Projekt <i>Linguistische Webcharakterisierung und Webkorpuserstellung</i> (DFG SCHA1916/1–1) an der Freien Universität Berlin
2014	<b>Wissenschaftlicher Mitarbeiter</b> bei Språkbanken an der Universität Göteborg
2011	<b>Vertretung der W3-Professur</b> für Englische Syntax an der Georg-August-Universität Göttingen
2007 – 2014	<b>Wissenschaftlicher Mitarbeiter</b> am Lehrstuhl Prof. Dr. Stefan Müller für Deutsche Grammatik an der Freien Universität Berlin
2002 – 2007	<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

### Projekt *Linguistische Webcharakterisierung und Webkorpuserstellung*

Drittmittelgeber	Deutsche Forschungsgemeinschaft (DFG)
Typ	Sachmittel/Eigene Stelle
Fördersumme	286.100€
Laufzeit	3 Jahre
DFG-Kennzeichen	SCHA1916/1-1
Kooperationen	Friedrich-Alexander Universität Erlangen-Nürnberg, Korpuslinguistik (Prof. Dr. Stefan Evert)  Institut für Deutsche Sprache, Mannheim, Abteilung Grammatik (Prof. Dr. Angelika Wöllstein)  Institut für Maschinelle Sprachverarbeitung, Stuttgart (PD Dr. Sabine Schulte im Walde)

## Drittmittelinwerbung im Antrags- und Entscheidungsprozess

2019	Eigenverantwortliche Antragserstellung für eine DFG-Förderung <i>Korpora und Individualgrammatik</i> , Einreichung 2019
seit 2018	Eigenverantwortliche Antragserstellung für ein ERC-Consolidator-Projekt <i>Individual language and Individual Input (ILII)</i> , zweite Einreichung geplant für 2020/2021
seit 2017	Mitarbeit in einer Arbeitsgruppe zur Antragserstellung eines Sonderforschungsbereichs (DFG), an der Humboldt-Universität zu Berlin und dem Leibniz-Zentrum Allgemeine Sprachwissenschaft zum Thema <i>Register: Understanding Situational-Functional Variation in Language</i>

## Betreute Abschlussarbeiten (Auswahl)

Hinweis: Mein langjähriger Arbeitgeber, die Freie Universität Berlin, erlaubt die Betreuung von Abschlussarbeiten durch nicht habilitiertes Personal nur im Ausnahmefall.

- 2019 | Zweitprüfer BA: *A probabilistic analysis of diachronic change in the productivity of nominalizing morphological processes in Early New High German*, Germanistische Linguistik, Humboldt Universität zu Berlin (Kandidatin: Elizabeth Pankratz).
- Erstprüfer: *Evaluative Morphologie in Neuen Medien: Ein Vergleich der Suffixe -ismus und -itis im Deutschen*, Deutsche Philologie, Freie Universität Berlin (Kandidatin: Geraldine Schulz).
- Erstprüfer BA: *Flexionsschwund in internetbasierter Kommunikation am Fall der schwachen Maskulina*, Deutsche Philologie, Freie Universität Berlin (Kandidatin: Sarah Schultze).
- Erstprüfer BA: *Substantiv-Verb-Verbindungen als graphematische Zweifelsfälle*, Deutsche Philologie, Freie Universität Berlin (Kandidat: Hannes Marggraf).
- 2016 | Erstprüfer: BA *Einflussfaktoren der Parallel- und Wechselflexion bei Nominalphrasen mit zwei attributiven Adjektiven*, Deutsche Philologie, Freie Universität Berlin (Kandidatin: Lisa Krack).
- 2015 | Erstprüfer: BA *Das Passiv in der Transformationsgrammatik*, Deutsche Philologie, Freie Universität Berlin (Kandidatin: Chiara Jancke).
- 2013 | Erstprüfer: BA *Kasuzuweisung zur Artangabe pseudopartitiver Konstruktionen*, Deutsche Philologie, Freie Universität Berlin (Kandidat: Samuel Reichert).
- Zweitprüfer: BA *Eigennamenerkennung in Web-Korpora des Deutschen. Eine Herausforderung für die (Computer)linguistik*, Germanistische Linguistik, Humboldt Universität zu Berlin (Kandidatin: Lea Arianna Helmers).
- 2011 | Zweitprüfer BA: *Die Flexionstypen des deutschen Adjektivs*, Deutsche Philologie, Freie Universität Berlin (Kandidat: Nicolai Sinn).

## Akademische Gremien- und Verwaltungstätigkeit

2013 – 2018	Mittelbauvertretung in mehreren Promotionskommissionen
2011 – 2014	Mitglied im Mentoringausschuss Deutsche und niederländische Philologie Freie Universität Berlin
2009 – 2014	Modulbeauftragter für Basis- und Aufbauphase Linguistik, Deutsche und niederländische Philologie Freie Universität Berlin
2004 – 2005	Mitarbeit an der Ordnung für die BA/MA-Studiengänge in Allgemeiner und Indogermanischer Sprachwissenschaft an der Georg-August Universität Göttingen
2001	Studentischer Vertreter in der Auswahlkommission zur Besetzung der W3-Professur für Allgemeine und Vergleichende Sprachwissenschaft an der Philipps-Universität Marburg

## Mitgliedschaften in Expertengremien und Netzwerken

seit 2017	Mitarbeit in der Lehramtsinitiative der DGfS
2013 – 2014	Mitglied der Expertenrunde der DFG zur Erarbeitung der Richtlinien zu technischen und rechtlichen Aspekten der Handhabung von Sprachkorpora

## Ämter in Vereinen und Verbänden

seit 2016	Mitwirkung im Interdisziplinären Forschungsverbund <i>Digital Humanities in Berlin</i> (if DH b)
seit 2015	Vorsitzender der <i>Special Interest Group for Web as Corpus</i> der Association for Computational Linguistics

## **Gutachtertätigkeit: Zeitschriften und Verlage (Auswahl)**

seit 2014	Language Science Press
2013	Journal of Germanic Linguistics
2012	Lingua

## **Gutachtertätigkeit: Konferenzen (Auswahl)**

2019	KONVENS, Nürnberg Challenges in the Management of Large Corpora, Cardiff
2018	KONVENS, Wien
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 HPSG, Paris

## **Organisierte Tagungen**

2020	12th Web as Corpus Workshop (WAC-XII) bei LREC, Marseille, mit Adrien Barbaresi, Felix Bildhauer, Egon Stemle
2017	11th Web as Corpus Workshop (WAC-XI) bei Corpus Linguistics, Birmingham, mit Adrien Barbaresi, Felix Bildhauer

- |      |                                                                                                                          |
|------|--------------------------------------------------------------------------------------------------------------------------|
| 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 und Teilnahme an Podien (Auswahl)

- |      |                                                                                                                                                                                                                                                                                 |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2020 | Roland Schäfer: <i>Assessing compatibility and stability of individual grammars through replication</i> . Jahrestagung der DGfS 2020, Hamburg.                                                                                                                                  |
|      | Felix Bildhauer & Roland Schäfer: <i>Beyond Multidimensional Analysis: probabilistic register induction for large corpora</i> . Jahrestagung der DGfS 2020, Hamburg.                                                                                                            |
|      | Roland Schäfer: <i>Are incompatibility and instability of individual grammars reflected in experimental and corpus data?</i> Vortrag auf Einladung des Sprachwissenschaftlichen Seminars der Georg-August-Universität Göttingen im Rahmen der LinG Colloquium Series 2019/2020. |
| 2019 | Felix Bildhauer & Roland Schäfer: <i>Assessing and comparing corpus composition using lexico-grammatical features</i> . Jahrestagung der DGfS 2019, Bremen.                                                                                                                     |
| 2018 | Podiumsmitglied auf Einladung bei der Podiumsdiskussion zur Vernetzung von Fachwissenschaft und Fachdidaktik im Lehramtsstudium, Tagung des Projekts <i>ProfJL</i> , Jena.                                                                                                      |
| 2017 | Roland Schäfer & Ulrike Sayatz: <i>Wie viel Grammatik braucht das Germanistikstudium?</i> Vortrag auf Einladung beim 2. Treffen <i>Grammatik im Lehramtsstudium</i> (DFG Grafüs), Leipzig.                                                                                      |

Roland Schäfer: *Parallel- und Wechselflexion in obliquen Adjektivreihen – Eine empirische Vervollständigung*. Vortrag auf Einladung der Abteilung Grammatik des Instituts für Deutsche Sprache (IDS), Mannheim.

Felix Bildhauer & Roland Schäfer: *Automatic register annotation for linguistic research?* DGfS Jahrestagung, AG Register in linguistic theory: Modeling functional variation, Saarbrücken.

2016 Roland Schäfer: *The Register Casino – Should you risk your grammar in an outside bet?* Grammar and Corpora 2016, Mannheim.

Felix Bildhauer & Roland Schäfer: *Modelling alternation phenomena with aggregated and non-aggregated predictors*. Grammar and Corpora 2016, Mannheim.

Felix Bildhauer & Roland Schäfer: *Induktive Topikmodellierung und extrinsische Topikdomänen*. Jahrestagung des Instituts für Deutsche Sprache 2016, Mannheim.

Roland Schäfer: *On bias-free crawling and representative web corpora*. 10th Web as Corpus workshop (WAC-X) bei ACL 2016, Berlin.

Roland Schäfer & Felix Bildhauer: *Automatic classification by topic domain for meta data generation, web corpus evaluation, and corpus comparison*. 10th Web as Corpus workshop (WAC-X) bei ACL 2016, Berlin.

Lars Borin, Markus Forsberg, Martin Hammarstedt, Dan Rosén, Roland Schäfer & Anne Schumacher: *Sparv: Språkbanken's corpus annotation pipeline infrastructure*. Swedish Language Technology Conference (SLTC), Umeå.

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.
- Felix Bildhauer & Roland Schäfer: *Design decisions in web corpus construction and their impact on distributional semantic models*. Gastvortrag auf Einladung der University of Cambridge, Computer Science Lab.
- 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: <i>On exceptional Vorfeldbesetzung</i> . DGfS Jahrestagung, AG <i>Quirky Vorfeld Phenomena</i> , Berlin. |
| 2007 | Jakob Maché & Roland Schäfer: <i>Linking gestures</i> . Generative Grammatik des Südens, Leipzig.                        |
| 2006 | Roland Schäfer: <i>Frequency adjectives force NPs to have event readings</i> . Sinn und Bedeutung 11, Barcelona.         |

## Veröffentlichte Ressourcen und Software

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

- |           |                                                                                                                                                                                                                                                                    |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| seit 2018 | Entwicklung der SeaCOW-Schnittstelle zur skriptbasierten Abfrage der COW-Korpora mit NoSketchEngine und Python                                                                                                                                                     |
| seit 2016 | Entwicklung der COREX und CORECO 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 COW-Online-Korpusplattform <i>webcorpora.org</i>                                                                                                                                                                                                   |
| seit 2011 | Erstellung der COW 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) ( <i>corporafromtheweb.org</i> ) |
| seit 2011 | Entwicklung der quelloffenen Software <i>texrex</i> zur vollautomatischen Webkorpus-Erstellung                                                                                                                                                                     |
| seit 2011 | Entwicklung der quelloffenen COW Toolchain zur linguistischen Annotation von Korpora                                                                                                                                                                               |