

# Maria Berger, née Moritz

Ruhr University Bochum, Germany | [maria.berger-a2l@rub.de](mailto:maria.berger-a2l@rub.de) | +49 (0) 176 9915 2387

<https://scholar.google.de/citations?user=nbGHbD4AAAAJ>



## Research Interests

Application of AI/NLP methods to interdisciplinary research questions:  
e.g., in the humanities: analysis of text modifications in historical parallel texts, epistemic functions, knowledge modeling, transfer-learning techniques

Python, Java (further: Octave, PHP, Neo4j, SQL, some: Bash/AWK)  
German (native), English (professional), French (elementary)

## Grants & Awards

- 2017 **Award at the Digital Humanities Conference, Montréal, Canada**  
ADHO Bursary Cash Award (€550) & travel bursary (€300) for 'An Automated Approach to Model the Transformation Process of the Reuse in Bernard de Clairvaux: How Do Lexical Resources help?'. DH 2017.
- 2016 **Grant from the University of Göttingen: TRAIN (Tracing Authorship in Noise, €20,000)**  
To study the balance between OCR'd data and automatic authorship. Wrote the proposal collaboratively.

## Experience

### Academic

- since 08/2021 **Postdoctoral researcher in computational linguistics, Ruhr University Bochum, Germany**  
Epistemic functions and paraphrase prediction, metaphor recognition  
Teaching in Computational Linguistics and Humanities
- 2019-2021 **Postdoc in language & speech technology, German Research Center for Artificial Intelligence (DFKI), Berlin, Germany**  
(parental leave 11/2019 – 09/2021)  
Project 'European Language Grid (ELG)' by the European Commission & Cordis – Horizon2020 (€7 mio.  
Integrated platform for European language resources and tools; member of the research & development team
- 2015-2019 **Research associate and PhD student in natural language processing & digital humanities, University of Göttingen, Germany**  
eTRAP Early career research group funded by the Federal Ministry of Education and Research (€1.6 mio)  
Historical text re-use analysis & re-use transformation analysis
- 2013-2014 **Research associate in digital humanities, project "Reinventing Humanities Publication", University of Leipzig, Germany**  
Early career research group by European Social Fund (€1.2 mio.)  
Modeling and development of an Ancient Greek e-learning application (esp. cleansing, API, Graph databases)
- 2011-2013 **Research associate in computer science, University of Leipzig, Germany**  
project 'eTRACES – Recherche und Analyse von Zitationsspuren und Wissenstransfer in sozialwissenschaftlichen Texten und deutschsprachiger Literatur'  
Research project by Federal Ministry of Education and Research (€1.2 mio.)  
Collaboration with the Google Ancient Places<sup>1</sup> team referring named entity enrichment  
Annotation environment for work-specific text reuse diversity annotation
- 2011 **Student research assistant, Department of Natural Language Processing, University of Leipzig, Germany**  
Non-symmetric citation extraction on Ancient Greek texts

### Business

- 2011 **Selfemployed, Leipzig, Germany**  
Development and research as a freelancer
- 2007, 2011 **Toxido Web & IT-Lösungen (intern, freelance), Altenbach, Germany**  
Web development (PHP, JS, MySQL), development of an SSH-Client (Java), development of a document similarity retrieval service using less frequent seed words (Pearl)

<sup>1</sup> <http://googleancientplaces.wordpress.com/>

2009-2010      **g.Labs GmbH, Leipzig, Germany**  
Development of browser games (PHP, SQL, JavaScript, Symphony)

## Education

---

2015-2018      **PhD student in computer science, University of Göttingen, Germany**  
Historical text reuse analysis and reuse transformation analysis  
Thesauri-based and machine learning-based methodologies for reuse style analysis  
Passed final exam (defense) on **May 2 2019**, certificate since Oct 2019

2008-2011      **Master's student in computer science, University of Leipzig, Germany**  
Information recognition and extraction in Ancient Greek texts  
NLP, information retrieval & text mining, databases, embedded systems

2004-2008      **Candidate for bachelor's degree in computer science, University of Leipzig, Germany**  
Content management systems and ontologies – Ontowiki plugin system for DL Learner<sup>2</sup>  
Databases, algorithms, computer networks, software engineering, signal processing, logics, math

## Publications

---

### 2022

**Maria Berger** (forthcoming). Transfer Learning Parallel Metaphor. In: (Proceedings) *The Third Workshop on Figurative Language Processing (FigLang2022)*. Abu Dhabi, United Arab Emirates. ACL.

### 2019 – 2021

**Maria Berger** and Elizabeth Goldstein (2021). Increasing Sentence-Level Comprehension Through Text Classification of Epistemic Functions. In: (Proceedings) *The Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*. Punta Cana, Dominican Republic. ACL.

**Maria Berger** (2020). Evaluating Semantic Relation Alignment in Historical Bibles using Historical Embeddings. In: (Proceedings) *Digital Humanities 2020 (DH 2020)*. Ottawa/online. ADHO.

Penny Labropoulou, Katerina Gkirtzou, Maria Gavriilidou, Miltos Deligiannis, Dimitris Galanis, Stelios Piperidis, Georg Rehm, **Maria Berger**, Valérie Mapelli, Michael Rigault, Victoria Arranz, Khalid Choukri, Gerhard Backfried, Jose Manuel Gomez Perez, and Andres Garcia-Silva (2020). Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid. In: (Proceedings) *12th International Conference on Language Resources and Evaluation (LREC 2020)*. Marseille. ELRA.

Georg Rehm, **Maria Berger**, Ela Elsholz, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Stelios Piperidis, Miltos Deligiannis, Dimitrios Galanis, Katerina Gkirtzou, Penny Labropoulou, Kalina Bontcheva, David Jones, Ian Roberts, Jan Hajič, Jana Hamrlová, Lukáš Kačena, Khalid Choukri, Victoria Arranz, Andrejs Vasiljevs, Oriens Anvari, Andis Lagzdīņš, Jūlija Meļņika, Gerhard Backfried, Erinç Dikici, Miroslav Janosik, Katja Prinz, Christoph Prinz, Severin Stampler, Dorothea Thomas-Aniola, José Manuel Gómez Pérez, Andres Garcia Silva, Cristian Berrío, Ulrich Germann, Steve Renals, and Ondrej Klejch (2020). European Language Grid: An Overview. In (Proceedings) *12th International Conference on Language Resources and Evaluation (LREC 2020)*. Marseille. ELRA.

Georg Rehm, Karolina Zaczynska, Julián Moreno Schneider, Malte Ostendorff, Peter Bourgonje, **Maria Berger**, Jens Rauenbusch, André Schmidt and Mikka Wild (2020). Towards Discourse Parsing-inspired Semantic Storytelling. In: (CEUR Workshop Proceedings) *Conference on Digital Curation Technologies – Exploring the Future of Digital Communication (QURATOR 2020)*. Berlin. January 2020.

Malte Ostendorff, Peter Bourgonje, **Maria Berger**, Julián Moreno Schneider, Georg Rehm., and Bela Gipp (2019). Enriching BERT with Knowledge Graph Embeddings for Document Classification. In: (Proceedings) *Preliminary proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)*. Erlangen, Germany. German Society for Computational Linguistics & Language Technology.

### 2016 – 2018

Greta Franzini, Marco Passarotti, **Maria Moritz**, and Marco Büchler (2018). Using and evaluating TRACER for an Index fontium computatus of the Summa contra Gentiles of Thomas Aquinas. In: (Proceedings) *The Fifth Italian Conference on Computational Linguistics (CliC-it 2018)*. Turin. AILC.

**Maria Moritz**, Johannes Hellrich, and Sven Büchel (2018). A Human-Interpretable Method to Predict Paraphrasticity. In: (Proceedings) *Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-ClfL 2018)*. Santa Fe. ACL.

**Maria Moritz**, Johannes Hellrich, and Sven Büchel (2018). Towards a Metric for Paraphrastic Modification. In: (Proceedings) *Digital Humanities 2018 (DH 2018)*. Mexico City. ADHO.

**Maria Moritz** and David Steding. (2018). Lexical and Semantic Features for Cross-lingual Text Reuse Classification: an Experiment in English and Latin Paraphrases. In: (Proceedings) *11th International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki. ELRA.

---

2 <https://dl-learner.org/>

**Maria Moritz.** (2018). On the Impact of Time Proximity on the Alignment of Spelling Variants in old English Bibles: A Case Study. In: (Proceedings) *Corpus-based Research in the Humanities (CRH 2018)*. Vienna. Gerastree.

**Maria Moritz** and Marco Büchler. (2017). An Automated Approach to Model the Transformation Process of the Reuse in Bernard de Clairvaux: How Do Lexical Resources help?. In: (Proceedings) *Digital Humanities 2017 (DH 2017)*. Montreal. ADHO.

Markus Paluch, Gabriela Rotari, David Steding, Maximilian Weiß, **Maria Moritz**, and Marco Büchler. (2017). Analysis of Part-Of-Speech Tagging of Historical German Texts. In: (Proceedings) *The International Conference on Digital Access to Textual Cultural Heritage (DATECH 2017)*. Göttingen. ACM.

**Maria Moritz** and Marco Büchler. (2017). Ambiguity in Semantically Related Word Substitutions: an investigation in historical Bible translations. In: (Proceedings) *Workshop on Processing Historical Language at NODALIDA 2017 (ProcHistLang 2017)*. Gothenburg. Linköping University Electronic Press.

**Maria Moritz**, Andreas Wiederhold, Barbara Pavlek, Yuri Bizzoni and Marco Büchler. (2016). Non-Literal Text Reuse in Historical Texts: An Approach to Identify Reuse Transformations and its Application to Bible Reuse. In: (Proceedings) *The 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*. Austin. ACL.

**Maria Moritz**, Barbara Pavlek, Greta Franzini and Gregory Crane. (2016). Sentence Shortening via Morpho-Syntactic Annotated Data in Historical Language Learning. *Journal on Computing and Cultural Heritage (JOCCH)*. 9, 1, Article 3 (February 2016), ACM.

## 2012 – 2014

Marco Büchler, Greta Franzini, Emily Franzini, and **Maria Moritz.** (2014). Scaling Historical Text Re-Use. In: (Proceedings) *IEEE International Conference on Big Data (IEEE BigData 2014)*. Washington DC. IEEE Computer Society Press.

**Maria Moritz**, Monica Lent, Thomas Köntges, Emily Franzini, Maryam Foradi, and Gregory Crane. (2014). AncientGeek: Primary Sources Powering Historical Language Learning. In: (Proceedings) *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (ELEARN 2014)*. New Orleans. AACE.

Frederik Baumgardt, Monica Berti, Giuseppe Celano, Gregory R. Crane, Stella Dee, Maryam Foradi, Emily Franzini, Greta Franzini, Simon Frazier, Monica Lent, **Maria Moritz** and Simona Stoyanova. (2014). Open Philology at the University of Leipzig. In: (Proceedings) *9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavik. ELRA.

Marco Büchler, Gregory Crane, **Maria Moritz** and Alison Babeu. (2012). Increasing Recall for Text Reuse in Historical Documents to Support Research in Humanities. In: (Proceedings) *Theory and Practice of Digital Libraries (TPDL 2012)*. Cyprus. Springer.

## Interview Invitations

Feb. 7., 2020	<b>W1 Professorship in Machine Learning with Tenure Option on W2, University of Passau, Germany</b> (declined the interview due to childbed)
Feb 25/26 2020	<b>Assistant Professorship in Artificial Intelligence with Tenure Option, University of Gothenburg, Sweden</b> (top 5 of 63 applicants)

## Expert Talks & Presentations Given (w/o Paper Presentations)

Dec. 5, 2017	<b>Invited Talk, University of Jena, Germany</b> Invited Talk at the DH Seminar of the CL Group JULIELab 'Ähnlichkeit in Historischen Texten'
June 1, 2017	<b>Invited Poster at DATECH Conference, Göttingen, Germany</b> Invited Poster Presentation 'Non-Literal Text Reuse in Historical Texts' at DATECH 2017
Feb. 17, 2017	<b>Invited Talk, University of Groningen, Netherlands</b> Invited Talk at the research seminar of the Computational Linguistics Group: 'Non-Literal Text Reuse in Historical Texts: An Approach to Identify Reuse Transformations, its Application and Future Work'
Oct. 14, 2016	<b>Invited Talk, University of Hildesheim, Germany</b> Invited Talk at the 2nd Göttingen-Hildesheim Workshop on Computational Linguistics and Digital Humanities 'Non-Literal Text Reuse in Historical Texts: An Approach to Identify Reuse Transformations and its Application to Bible Reuse'
Oct. 20, 2015	<b>Digital Humanities in Estonia 2015, Tartu, Estonia</b> Invited Talk with Marco Büchler: 'eTRAP Overview: An early career research group on historical text reuse – Methodologies & research fields'
Jun. 30, 2015	<b>CLARIN-D-Fach-AG Workshop, Leipzig, Germany</b> Invited Talk with Marco Büchler at CLARIN-D Fach-AG Workshop 2015: 'eTRAP: electronic Text Re-use Acquisition Project'
March 3, 2015	<b>DH-Summit, Berlin, Germany</b> Poster Presentation at DH-Summit Berlin '15: 'eTRAP: A research group implementing the A.C.I.D. eHumanities paradigm.'

Dec. 6, 2013	<b>German Archeological Institute, Berlin, Germany</b> Invited Talk at Digital Classicist Seminar 'Dynamic Syllabi for Historical Language Instruction'
May 6, 2013	<b>University of Leipzig, Germany</b> Project Report Talk at eTRACES Sponsor Meeting, presentation of project work

## Visited Workshops & Trainings

---

Oct. 20 2017	<b>Time Management in Doctoral Research, Göttingen, Germany</b> Training offered by the Graduate School of Computer Science in Göttingen
Sept.-Nov. 2016	<b>Introduction to Natural Language Processing, The University of Michigan, USA, online</b> Coursera Online Course, taught by Dragomir R. Radev, intermediate level (95%)
2015/2016	<b>eTRAP team coaching, University of Göttingen, Germany</b> Understanding personalities, leadership, decision making, stress management in teams
Nov. 9-10 2015	<b>Experimental Methods, University of Göttingen, Germany</b> An introduction to experimental methods and design for social studies and humanities: from theory to practice
Oct. 14 2015	<b>Managing your Risks, University of Göttingen, Germany</b> Rational and intuitive planning, dealing with failure, critical factor analysis, scenario planning, mini SWOT
May 19-21 2014	<b>Andrew W. Mellon Workshop on Canonical Citation, Furman University, SC, USA</b> Philological principles of canonical citations and technologies for implementing those principles
July 23-31, 2012	<b>European Summer University on Culture &amp; Technology (ESU), Leipzig, Germany</b> Participating workshop on stylometry

## Research Visits & Collaborations

---

Nov. 2017	<b>Research Stay at the Computational Linguistics Research Group of Jena University, Germany</b>
Feb. 2019	Collaboration on a measure for paraphrasticity and machine translation
May 2017	<b>Research Visit, Centre for Linguistic Theory &amp; Studies in Probability, University of Gothenburg, Sweden</b> Follow-up collaboration of the EMNLP paper on measuring modification in Latin and Ancient Greek

## Teaching Experiences

---

I spent the large majority of my academic career in research groups without teaching obligations, such as the research group of Marco Böhler in which I conducted my PhD research. However, whenever there was a possibility to assist in teaching duties, I enjoyed it. In addition, I supervised an interdisciplinary student project consisting of a programming part and a paper (published at DATECH 2017). I also supervised a small number of final thesis (see below). Throughout these activities, I obtained valuable inter-disciplinary experiences interacting with two very different research fields: Computer Science and Humanities.

Since summer 2022	<b>Lecturer, Ruhr University Bochum, Germany</b> Text Similarity Measures and their Application in Computational Humanities
Since winter 21/22	<b>Lecturer, Ruhr University Bochum, Germany</b> Computational Humanities Methods for Historical Texts
June 18, 2018	<b>Substitute Lecturer, Leibniz Institute for European History, Mainz, Germany</b> Gave a lecture on similarity measures in the course 'Natural Language Processing' substituting the main course lecturer
Summer 2017	<b>Co-Supervisor of a Master's Thesis, the Faculty of Arts, University of Groningen, Netherlands</b> Master's Thesis in Digital Humanities on the Evaluation of Topic Models for Arabic co-supervised with Johan Bos and Barbara Plank
July 11, 2016	<b>Tutorial Lecturer, Digital Humanities Conference, Krakow, Poland</b> Taught tutorial with colleagues within our full-day workshop 'Historical Text Reuse Tutorial'
2015/2016	<b>Teaching Assistant, University of Göttingen, Germany</b> Co-supervised the course 'From written manuscripts to big data: Cultural Heritage Programming' (M. Inf.1904) The inter-disciplinary project consisted of a programming part and a paper (published at DATECH 2017)

Oct. 21, 2015

Tutorial Lecturer, Digital Humanities Conference, Tartu, Estonia

Taught tutorial with colleagues within our full-day workshop 'Historical Text Reuse'

## Service

### Natural Language Processing & Computational Linguistics

Reviewer of **ACL Rolling Reviews** since **Oct. 2021**: reviewing procedure of ACL, EMNLP, NAACL (a.o.)

Program Committee member of **ACL 2019, 2021**: Annual Meeting of the Association for Computational Linguistics

(Co-)Reviewer of **EMNLP-(IJCNLP) 2019, 2021, 2022**: Conference on Empirical Methods in Natural Language Processing & International Joint Conference on Natural Language Processing

Program Committee member of **ACL SRW 2018**: Student Research Workshop at the 56<sup>th</sup> Annual Meeting of the ACL

Program Committee member of **BEA 2017-2020**: WS on Innovative Use of NLP for Building Educational Applications

Program Committee member of **IWCS 2019**: 13<sup>th</sup> International Conference on Computational Semantics

Effective member of Scientific Committee of **LREC 2016-2022**: Language Resources and Evaluation Conference

Program Committee member of **KONVENS 2022**: (Konferenz zur Verarbeitung natürlicher Sprache/Conference on Natural Language Processing)

### Digital Humanities

Reviewer for **DH 2017-2022**: Digital Humanities

Board member of **GDDH 2015, 2016**: Göttingen Dialogue on Digital Humanities

Reviewer for **JDMDH**: Journal of Data Mining and Digital Humanities Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages (forthcoming)

### Natural Language Processing in Software Engineering

Co-reviewer of **ASE 2016, 2018**: 31<sup>st</sup> and 33<sup>rd</sup> IEEE/ACM International Conference on Automated Software Engineering

Co-reviewer of **ICSE NIER 2018**: 40<sup>th</sup> International Conference on Software Engineering, New Ideas and Emerging Results

Co-reviewer of **WAMA 2016**: 1<sup>st</sup> Workshop on App Market Analytics at the ACM SIGSOFT International Symposium on the Foundations of Software Engineering

## Organized Workshops

Göttingen, Germany (conference workshop), May 30, 2017: TRACER Tutorial. At the DATeCH Conference.

Krakow, Poland (conference workshop), July 11, 2016: Historical Text Reuse Tutorial. At the Digital Humanities Conference.

Tartu, Estonia (conference workshop), Oct. 21, 2015: Workshop on Text Reuse. At the Conference of Digital Humanities in Estonia.

Göttingen, Germany (talk series), Summer 2015, 2016: Göttingen Dialog on Digital Humanities.

Göttingen, Germany (hackathon), July 2015: Digital Humanities Hackathon on Text Re-Use: 'Don't leave your data problems at home!'

Leipzig, Germany (meeting), May 2012: annual eTRACES Sponsor Meeting: *Introduction of the ASV Subproject - Report on recent state of work and other activities*. Leipzig, 2012.

Leipzig (conference workshop), Oct. 2010: eHumanities - How does computer science benefit?. At 40. Jahrestag der Gesellschaft für Informatik e.V.: Service Science - Neue Perspektiven für die Informatik.

## References

Name	Position & Affiliation	Relation	Contact
Marco Passarotti	<b>Assist. Prof. for Computational Linguistics</b> (Università Cattolica del Sacro Cuore, Milan, Italy)	Collaborator	marco.passarotti@unicatt.it +39 02 72342380
Georg Rehm	<b>Research Group leader &amp; Honorary Professor</b> (DFKI & Humboldt University, Berlin, Germany)	Former manager	georg.rehm@hu-berlin.de +49 (0)30 23895-1833
Jan Rybicki	<b>Assist. Prof. for English Studies</b> (Jagiellonian University, Krakow, Poland)	Collaborator	jan.rybicki@uj.edu.pl +48 12 663 4426