Maria Berger, née Moritz

Ruhr University Bochum, Germany | maria.berger-a2l@rub.de | http://staff.germanistik.rub.de/digitale-forensische-linguistik/

https://bitbucket.org/mariamoritz/reuse modification analysis

https://github.com/Goldstein-Berger/law-dmr21-sentence-meaning-annotation

Research Interests

The application of AI/NLP methods to research questions in the humanities: Formal analysis of lexical and morphological text modifications in historical texts using thesauri-based and statistical techniques

Java, Python, (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 of €550 & travel bursary €300 for the paper '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)

Grant of ϵ 20,000 I received to study the balance between noisy OCR'd data and automatic authorship and reuse analysis. Wrote the proposal collaboratively.

Experience

Academic

since 2021 RUHR UNIVERSITY BOCHUM

Bochum, Germany

- Postdoctoral researcher in computational linguistics and digital humanities
- Epistemic functions and paraphrase prediction, metaphor recognition
- Teaching in Digital Humanities

2019-2021 GERMAN RESEARCH CENTER FOR ARTIFICIAL INTELLIGENCE

Berlin, Germany

Postdoctoral researcher in language and speech technology

- 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 University of Goettingen

Göttingen, Germany

Research associate and PhD student in natural language processing & digital humanities

- 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 University of Leipzig

Leipzig, Germany

Research associate in digital humanities, project "Reinventing Humanities Publication"

- Early career research group by European Social Fund (€1.2 mio.)
- Modeling and development of an Ancient Greek e-learning application
- Data cleansing, API development, Graph databases

2011-2013 University of Leipzig

Leipzig, Germany

Research associate in computer science, project 'eTRACES – Recherche und Analyse von Zitationsspuren und Wissenstransfer in sozialwissenschaftlichen Texten und deutsprachinger Literatur'

- Research project by Federal Ministry of Education and Research (€1.2 mio.)
- Collaboration with the Google Ancient Places¹ team referring named entity enrichment
- Annotation environment for work-specific text reuse diversity annotation

2011 University of Leipzig

Leipzig, Germany

Student research assistant, Department of Natural Language Processing

Non-symmetric citation extraction on Ancient Greek texts

Business

2011 SELFEMPLOYED, LEIPZIG, GERMANY

Development and research work 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 infrequent seed words (Pearl)

2009-2010 G.LABS GMBH, LEIPZIG, GERMANY

Development of browser games (PHP, SQL, JavaScript, Symphony)

Education

2015-2018 University of Goettingen

Göttingen, Germany

PhD student in computer science

- Historical text reuse analysis and reuse transformation analysis
- Thesauri-based and machine learning methodologies for reuse style analysis
- Passed final exam (defense) on May 2 2019, certificate since Oct 2019

2008-2011 University of Leipzig

Leipzig, Germany

Master's student in computer science

- Information recognition and extraction on Ancient Greek texts
- · Natural language processing, information retrieval and text mining, databases, embedded systems

2004-2008 University of Leipzig

Leipzig, Germany

Candidate for bachelor's degree in computer science

- Content management systems and ontologies Ontowiki plugin for machine learning environment²
- Databases, algorithms, computer networks, software engineering, signal processing, logics, math

Publications

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.

2020

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

¹ http://googleancientplaces.wordpress.com/

 $^{2\} http://dl\text{-learner.org/wiki/OntoWikiPlugin}$

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, Orians Anvari, Andis Lagzdiņš, 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.

2019

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 Paraphrasticality. 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.

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 Weß, **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.

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)*. Reykjavík. 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	W1 Professorship in Machine Learning with Tenure Option on W2, University of Passau, Germany (rejected the interview due to child bed)
Feb. 25/26 2020	Assistant Professorship in Artificial intelligence with Tenure Option, University of Gothenburg, Sweden (top 5 of 63 applicants)

Expert Talks & Presentations Given

Lybert lair	S & Fresentations diven
Aug. 25, 2018	LATECH-CLFL Workshop 2018 at COLING Conference, Santa Fe, USA Poster Presentation 'A Human-Interpretable Method to Predict Paraphrasticality'
June 29, 2018	DIGITAL HUMANITIES CONFERENCE, MEXICO CITY, MEXICO Short Paper Talk 'Towards a Metric for Paraphrastic Modification'
May 10, 2018	LREC Conference, Miyazaki, Japan Poster Presentation 'Lexical and Semantic Features for Cross-lingual Text Reuse Classification: an Experiment in English and Latin Paraphrases'
Jan. 26, 2018	Workshop Corpus-based Research in the Humanities, Vienna, Austria Paper Talk 'On the Impact of Time Proximity on the Alignment of Spelling Variants in old English Bibles: A Case Study'
Dec. 5, 2017	Invited Talk, University of Jena, Germany Invited Talk at the DH Seminar of the CL Group JULIELab 'Ähnlichkeit in Historischen Texten'
Aug. 11, 2017	DIGITAL HUMANITIES CONFERENCE, MONTREAL, CANADA Short Paper Talk 'An Automated Approach to Model the Transformation Process of the Reuse in Bernard de Clairvaux: How Do Lexical Resources help?'
June 1, 2017	Invited Poster at DATECH Conference, Göttingen, Germany Invited Poster Presentation 'Non-Literal Text Reuse in Historical Texts' at DATECH 2017
May 22, 2017	Workshop on Processing Historcal Langiages at 21st Nordic Conference on Computational Linguistics (NoDaLiDa), Gothenburg, Sweden Paper Talk 'Ambiguity in Semantically Related Word Substitutions: an investigation in historical Bible translations'
Feb. 17, 2017	Invited Talk, University of Groningen, Netherlands 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'
Nov. 3, 2016	CONFERENCES ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP) 2016, AUSTIN, USA

Poster Presentation 'Non-Literal Text Reuse in Historical Texts: An Approach to Identify Reuse Transformations and its Application to Bible Reuse'

Oct. 14, 2016 Invited Talk, University of Hildesheim, Germany

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 DIGITAL HUMANITIES IN ESTONIA 2015, TARTU, ESTONIA

Invited Talk with Marco Büchler: 'eTRAP Overview: An early career research group on historical text reuse – Methodologies & research fields'

Jun. 30, 2015 CLARNI-D FACH-AG WORKSHOP, LEIPZIG, GERMANY

Invited Talk with Marco Büchler at CLARIN-D Fach-AG Workshop 2015: 'eTRAP: electronic Text Re-use Acquisition Project'

March 3, 2015 DH-SUMMIT, BERLIN, GERMANY

Poster Presentation at DH-Summit Berlin '15: 'eTRAP: A research group implementing the A.C.I.D. eHumanities paradigm.'

Nov. 2014 E-LEARN WOLRD CONFERENCE, NEW ORLEANS, USA

Paper Talk 'AncientGeek: Primary Sources Powering Historical Language Learning'

Dec. 6, 2013 GERMAN ARCHEOLOGICAL INSTITUTE, BERLIN, GERMANY

Invited Talk at Digital Classicist Seminar 'Dynamic Syllabi for Historical Language Instruction'

May 6, 2013 University of Leipzig, Leipzig, Germany

Project Report Talk at eTRACES Sponsor Meeting, presentation of project work

Visited Workshops & Trainings

Oct. 20 2017	TIME MANAGEMENT IN	DOCTORAL DESEARCH	CÖTTINGEN CE	DMANIX
OCL 20 2017	TIME WIANAGEMENT IN	DOCTORAL KESEARCH.	GOTTINGEN, GE	RMANY

Training offered by the Graduate School of Computer Science in Göttingen

Sept.-Nov. 2016 Introduction to Natural Language Processing, The University of Michigan, USA

Coursera Online Course, taught by Dragomir R. Radev, intermediate level (95%)

2015/2016 University of Goettingen, Göttingen, Germany

eTRAP team coaching

Understanding personalities, leadership, decision making, stress management in teams

Nov. 9-10 2015 University of Goettingen, Göttingen, Germany

An introduction to experimental methods and design for social studies and humanities: from theory to practice

Oct. 14 2015 UNIVERSITY OF GOETTINGEN, GÖTTINGEN, GERMANY

Managing your Risks

Sphere of influence, rational and intuitive planning, definition of and dealing with failure

Critical factor analysis, scenario planning, Mini SWOT analysis, self-coaching tools, Peer coaching groups

May 19-21 2014 Furman University, Furman, South Carolina, USA

Andrew W. Mellon Workshop on Canonical Citation

Philological principles of canonical citations and technologies for implementing those principles

July 23-31, 2012 European Summer University (ESU) on Culture & Technology, Leipzig, Germany

Participating workshop on stylometry

Research Visits & Collaborations

Nov. 2017 and Research stay at the Computational Linguistics Research Group of Jena University, Jena, Germany Feb. 2019 Collaboration on a measure for paraphrasticality and machine translation

RESEARCH MEETING AT THE CENTRE FOR LINGUISTIC THEORY AND STUDIES IN PROBABILITY OF THE UNIVERSTIY OF GOTHENBURG, SWEDEN

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 Büchler in which I conducted my PhD research. Currently, as a postdoctoral fellow, I am obliged to work full-time in funded research projects. However, whenever there was the opportunity to be involved with teaching before, I enjoyed participating as a teaching assistant, supervisor or mentor. In addition, not listed below, I assisted with various tutorials on NLP techniques, and I supervised an interdisciplinary student project consisting of a programming part and a paper (published at DATECH 2017). Throughout these activities, I obtained valuable inter-disciplinary experiences interacting with two very different research fields: Computer Science and Humanities.

June 18, 2018	Substitute Lecturer, Leibniz Institute for European History, Mainz, Germany Gave a lecture on similarity measures for NLP in the course 'Natural Language Processing' substituting the main course lecturer
Summer 2017	Co-Supervisor of a Master's Thesis, the FACULTY OF ARTS, UNIVERSITY OF GRONINGEN, NETHERLANDS Master's Thesis of Peter Sprenger in Digital Humanities on the Evaluation of Topic Models for Arabic, co-supervised with Johan Bos and Barbara Plank
July 11, 2016	Tutorial Lecturer, DIGITAL HUMANITIES CONFERENCE, KRAKOW, POLAND Taught tutorial with colleagues within our full-day workshop 'Historical Text Reuse Tutorial'
2015/2016	Cybermentor at cybermentor de Online, Germany Mentored a high school student interested in computer science, supporting her interest in academia
2015/2016	Teaching Assistant, University of Goettingen, Göttingen, Germany Supervised a student group in 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:

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

(Co-)Reviewer of EMNLP-IJCNLP 2019, 2021: 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 56th Annual Meeting of the Association for Computational Linguistics

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

Program Committee member of IWCS 2019: 13th International Conference on Computational Semantics

Effective member of Scientific Committee of LREC 2016-2020: Language Resources and Evaluation Conference

Digital Humanities:

Reviewer for DH 2017-2020: 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: 31st and 33rd IEEE/ACM International Conference on Automated Software Engineering

Co-reviewer of **ICSE NIER 2018**: 40th International Conference on Software Engineering, New Ideas and Emerging Results Co-reviewer of **WAMA 2016**: 1st 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
Tatjana Scheffler	Jun. Professor for Digital Forensic Linguistics (Università Cattolica del Sacro Cuore)	MANAGER	Tatjana.scheffler@rub.de
Marco Passarotti	ASSOC. PROFESSOR FOR COMPUTATIONAL LINGUISTICS	COLLABORATOR	marco.passarotti@unicatt.it
	(Università Cattolica del Sacro Cuore)		+39 02 72342380
Marco Büchler	ETRAP RESEARCH GROUP LEADER	FORMER MANAGER	mbuechler@etrap.eu
	(University of Goettingen, Göttingen, Germany)		+40 (0)551 20 20470
			+49 (0)551 39-20479
			jan.rybicki@uj.edu.pl
Jan Rybicki	Assist. Professor for English Studies	COLLABORATOR	
	(Jagiellonian University, Krakow, Poland)		+48 12 663 4426