

VALENTIN HOFMANN

7 Keble Rd, Oxford OX1 3QG, UK

valentinhofmann.github.io ◊ valentin.hofmann@ling-phil.ox.ac.uk

EDUCATION

University of Oxford DPhil, Linguistics <i>Advisors: Janet Pierrehumbert, Hinrich Schütze</i>	<i>10/2018 – Present</i>
LMU Munich MSc, Computational Linguistics & Computer Science <i>Advisor: Hinrich Schütze</i>	<i>10/2018 – 09/2020</i> GPA 1.1
University of Oxford MSt, Linguistics <i>Advisor: Janet Pierrehumbert</i>	<i>10/2017 – 07/2018</i> Distinction
LMU Munich BA, Linguistics BA, Slavic Studies	<i>04/2013 – 07/2017</i> GPA 1.0 GPA 1.0
Lomonosov Moscow State University Visiting Student, Linguistics	<i>08/2015 – 12/2015</i>
University of Zagreb Visiting Student, Linguistics	<i>02/2015 – 06/2015</i>

RESEARCH

Natural Language Processing Group, Stanford University Visiting Student Researcher <i>Host: Dan Jurafsky</i>	<i>09/2022 – Present</i>
Center for Information and Language Processing, LMU Munich Research Assistant <i>Supervisor: Hinrich Schütze</i>	<i>06/2019 – Present</i>
Language Team, DeepMind Research Scientist Intern <i>Hosts: Angeliki Lazaridou, Elena Gribovskaya, Jordan Hoffmann</i>	<i>03/2022 – 07/2022</i>
Institute of Historical Linguistics, LMU Munich Student Assistant <i>Supervisor: Olav Hackstein</i>	<i>04/2013 – 12/2014</i>

PUBLICATIONS

Geographic adaptation of pretrained language models

Valentin Hofmann, Goran Glavaš, Nikola Ljubešić, Janet Pierrehumbert, Hinrich Schütze
arXiv:2203.08565

Unsupervised detection of contextualized embedding bias with application to ideology

Valentin Hofmann, Janet Pierrehumbert, Hinrich Schütze
ICML 2022

Modeling ideological salience and framing in polarized online groups with graph neural networks and structured sparsity

Valentin Hofmann, Xiaowen Dong, Janet Pierrehumbert, Hinrich Schütze
NAACL 2022 (Findings)

The Reddit Politosphere: A large-scale text and network resource of online political discourse

Valentin Hofmann, Hinrich Schütze, Janet Pierrehumbert
ICWSM 2022

An embarrassingly simple method to mitigate undesirable properties of pretrained language model tokenizers

Valentin Hofmann, Hinrich Schütze, Janet Pierrehumbert
ACL 2022

CaMEL: Case marker extraction without labels

Leonie Weissweiler, **Valentin Hofmann**, Masoud Jalili Sabet, Hinrich Schütze
ACL 2022

Superbizarre is not superb: Derivational morphology improves BERT's interpretation of complex words

Valentin Hofmann, Janet Pierrehumbert, Hinrich Schütze
ACL 2021

Dynamic contextualized word embeddings

Valentin Hofmann, Janet Pierrehumbert, Hinrich Schütze
ACL 2021

DagoBERT: Generating derivational morphology with a pretrained language model

Valentin Hofmann, Janet Pierrehumbert, Hinrich Schütze
EMNLP 2020

Predicting the growth of morphological families from social and linguistic factors

Valentin Hofmann, Janet Pierrehumbert, Hinrich Schütze
ACL 2020

A graph auto-encoder model of derivational morphology

Valentin Hofmann, Hinrich Schütze, Janet Pierrehumbert
ACL 2020

INVITED TALKS

Natural Language Processing Group, University of Würzburg 08/2022
Geographic adaptation of pretrained language models

MilaNLP Lab, Bocconi University 07/2022
What role could and should sociolinguistic variables play in modern-day NLP?

Data and Web Science Group, University of Mannheim 05/2022
Modeling ideological language with graph neural networks and structured sparsity

LLab, Carnegie Mellon University 09/2020
Predicting the growth of morphological families from social and linguistic factors

TEACHING

Oxford Internet Institute, University of Oxford 01/2021 – 03/2021
Teaching Assistant (*NLP for the Social Sciences*)

Oxford Internet Institute, University of Oxford 01/2020 – 03/2020
Lab Instructor (*NLP for the Social Sciences*)

Faculty of Linguistics, University of Oxford
Tutor (*Phonetics and Phonology*)

10/2019 – 02/2020

Institute of Slavic Philology, LMU Munich
Tutor (*Russian*)

10/2016 – 03/2017

AWARDS

British Arts and Humanities Research Council

10/2018 – Present

German Academic Scholarship Foundation

10/2012 – Present

Scholarship awarded to top 0.5% of high school graduates in Germany

Maximilianeum Foundation

10/2012 – 07/2019

Scholarship awarded to top five high school graduates in Bavaria

Max Weber Program of the State of Bavaria

10/2012 – 07/2019

Three LMU Munich Student Research Grants

10/2016 – 04/2017

Alfried Krupp von Bohlen und Halbach Foundation

02/2015 – 04/2016

Erasmus Program

02/2015 – 06/2015

PROGRAMMING SKILLS

Python: machine learning, deep learning (HuggingFace, JAX, Keras, PyTorch)

R: data analysis, statistical modeling

C++: basic programming

LANGUAGE SKILLS

Native proficiency: German

Professional working proficiency: Croatian, English, Polish, Russian

Limited working proficiency: Farsi, French, Urdu

Reading knowledge: Ancient Greek (Graecum), Latin (Latinum), Old Church Slavonic

PERSONAL INTERESTS

Juggling: ball juggling, ring juggling, siteswaps

Classical music: viola, choir, chamber music