

# Joachim Daiber

Karel du Jardinstraat 29-3  
1072SE, Amsterdam  
(+31) 62501 2205  
daiber.joachim@gmail.com

## Personal Details

Born: April 23, 1987 in Bad Waldsee, Germany. Nationality: German.

## Education

### 2013–2017 **PhD Student in Machine Translation and Natural Language Processing**

*Statistical Language Processing and Learning Lab.*

ILLC, University of Amsterdam, The Netherlands.

Supervisors: Prof. Khalil Sima'an and Dr. Wilker Aziz.

### 2011–2013 **European Masters Program in Language and Communication Technologies**

Grade: *Summa cum laude*

Charles University in Prague, Czech Republic (Master of Science)

University of Groningen, The Netherlands (Master of Arts)

Thesis: Robust parsing of noisy content (with Prof. Gertjan van Noord, Dr. Dan Zeman)

- statistical methods in NLP
- statistical machine translation
- machine learning
- morphology, syntax and semantics

### 2007–2011 **Bachelor of Science in Computer Science, English**

Free University of Berlin, Germany

Thesis: Candidate selection and evaluation in DBpedia Spotlight

- data structures and algorithms
- parallel programming
- theoretical computer science
- computer and software engineering
- semantic web
- English and general linguistics

## Scholarships and Awards

### 2013–2016 **Marie Skłodowska-Curie Fellowship**

European Commission

### 2011–2013 **Erasmus Mundus Scholarship**

European Commission

### 2011–2012 **Merit Scholarship (prospěchové stipendium)**

Charles University in Prague

## Languages

German (native), English (fluent), Spanish (conversational), Dutch (basic).

---

## Experience

### Teaching

- 2014–2015 **Teaching Assistant, Natural Language Processing 1**  
Prof. Ivan Titov.
- 2014–2016 **Teaching Assistant, Natural Language Processing 2**  
Prof. Khalil Sima'an and Dr. Wilker Aziz.
- 2016 **Co-supervision of Master thesis (Master of AI)**  
Title: *Tree-Structured Markov Models for Word Alignment*. With Prof. Khalil Sima'an.
- 2015 **Supervision of student project project in AI**  
Project title: *Splitting German Compounds with Word Embeddings*. With Dr. Stella Frank.

### Research and Development

- 2016 **Research Intern, Unbabel**  
Lisbon, Portugal. 1 month, full time.
- 2015 **Machine Translation Marathon**  
Proposed and executed project: *Better unsupervised processing of compound words*.  
Prague, Czech Republic.
- 2014–2015 **Mentor: Google Summer of Code**  
Organization: DBpedia Spotlight.
- 2014 **European Summer School in Logic, Language and Information**  
Tübingen, Germany.
- 2012 **Lisbon Machine Learning Summer School**  
Lisbon, Portugal.
- 2012 **Participant: Google Summer of Code 2012**  
Organization: DBpedia Spotlight.  
Project: *Implementation of an efficient probabilistic model for entity linking*.
- 2012 **Visiting Research Intern**  
Web-based Systems Group (DBpedia project), Free University of Berlin.
- 2010–2011 **Research Assistant**  
Language Technology Lab.  
German Research Center for Artificial Intelligence (DFKI), Berlin. 16 months, part time.
- 2010 **Intern in Computational Linguistics and NLP**  
vionto GmbH, Berlin. 3 months, full time.

---

## Computer and Programming Skills

- Advanced Scala, Java, Python/NumPy, Bash,  $\LaTeX$ , JavaScript, CSS.
- Intermediate Perl, Haskell, PHP, TensorFlow.
- Basic C, C++, Objective-C.

---

## Publications

### Machine Translation

- Joachim Daiber, Miloš Stanojević, and Khalil Sima'an. Universal reordering via linguistic typology. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*, Osaka, Japan, December 2016b. Association for Computational Linguistics
- Joachim Daiber, Miloš Stanojević, Wilker Aziz, and Khalil Sima'an. Examining the relationship between preordering and word order freedom in machine translation. In *Proceedings of the First Conference on Machine Translation, Volume 1: Research Papers*, pages 118–130, Berlin, Germany, August 2016a. Association for Computational Linguistics
- Joachim Daiber and Khalil Simaan. Machine translation with source-predicted target morphology. In *Proceedings of the 15th Machine Translation Summit (MT Summit 2015)*, pages 283–296, Miami, USA, 2015
- Joachim Daiber, Lautaro Quiroz, Roger Wechsler, and Stella Frank. *Proceedings of the 1st Deep Machine Translation Workshop*, chapter Splitting Compounds by Semantic Analogy, pages 20–28. ÚFAL MFF UK, 2015. URL <http://aclweb.org/anthology/W15-5703>
- Joachim Daiber and Khalil Sima'an. *Proceedings of the 1st Deep Machine Translation Workshop*, chapter Delimiting Morphosyntactic Search Space with Source-Side Reordering Models, pages 29–38. ÚFAL MFF UK, 2015. URL <http://aclweb.org/anthology/W15-5704>

### Entity Extraction

- Joachim Daiber, Max Jakob, Chris Hokamp, and Pablo N. Mendes. Improving efficiency and accuracy in multilingual entity extraction. In *Proceedings of the 9th International Conference on Semantic Systems, I-SEMANTICS '13*, pages 121–124, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1972-0. doi: 10.1145/2506182.2506198. URL <http://doi.acm.org/10.1145/2506182.2506198>
- Pablo Mendes, Joachim Daiber, Rohana Rajapakse, Felix Sasaki, and Christian Bizer. Evaluating the impact of phrase recognition on concept tagging. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7

### Dependency Parsing

- Joachim Daiber and Rob van der Goot. The Denoised Web Treebank: Evaluating dependency parsing under noisy input conditions. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, May 2016. European Language Resources Association (ELRA). ISBN 978-2-9517408-9-1