

RYAN GEORGI

PERSONAL INFORMATION

email [REDACTED]
website <http://ryan.georgi.cc>
phone [REDACTED]

PROFILE

- Researcher and Instructor in Computational Linguistics with background in resource-poor languages, syntax and parsing, and computational typology.
- Research includes work with supervised and semi-supervised machine learning techniques on noisy and resource-poor domain.
- Experience as researcher, instructor, advisor, and software engineer in industry.
- Current research interests include low-resource languages and data mining; computational semantics, particularly as related to ethics, fairness, and abusive language.

EDUCATION

Ph.D The University of Washington, Seattle 2009–2016
Ph.D, Computational Linguistics · Aug. 2016
Thesis: *From Aari to Zulu: Massively Multilingual Creation of Language Tools using Interlinear Glossed Text*
Advisors: FEI XIA & WILLIAM D. LEWIS

MA The University of Washington, Seattle 2006–2009
MA, Computational Linguistics · Jun. 2009
Thesis: *Grammar Induction with Prototypes Derived from Interlinear Text*
Advisors: FEI XIA & WILLIAM D. LEWIS

BA The University of California, Berkeley 2001–2005
BA, Major in Linguistics · Minor in Computer Science · May 2005

EMPLOYMENT – ACADEMIC

University of Washington Acting Assistant Professor Dec 2017–Present
Faculty in the UW Linguistic Department’s **Master of Science in Computational Linguistics (CLMS)** program. Duties include instruction of core NLP technologies courses (see **Teaching**), advising of students’ masters projects, theses, and internships. Also involved in general program administration, including leading informational webinars and weekly lab meetings.

University of Washington Research Asst., Postdoc. Consultant 2009–2017
Member of the *RiPLes* (information extraction and synthesis for **Resource Poor Languages**) research group. Implemented machine learning tools and user interfaces for corpus enrichment and knowledge transfer between languages. Developed code for detecting and extracting interlinear glossed text (IGT) instances from PDFs.
Supervisor: FEI XIA [REDACTED]

EMPLOYMENT – INDUSTRY

- Spoken Communications* **Delivery Engineer – Python** *Jun–Dec 2017*
Automated systems for cloud deployment of Keras speech recognition pipeline and accompanying systems on AWS. Managed OAuth frameworks for secure transactions between application components and user databases.
Supervisor: JOHN SHAO [REDACTED]
- Microsoft Research* **Linguistics Test Engineer** *Feb–Sep 2008*
Engineer on the Machine Translation Incubation Team (now Bing Translator), developed metrics and experimented with different methods of corpora selection with the aim of improving MT system performance on different target domains. Adapted new tools to existing C# codebase.
Supervisor: WILLIAM D. LEWIS [REDACTED]
- Voxify, Inc* **QA Engineer** *2004–2006*
Developed QA procedures for testing Voice user interfaces (VUIs) in customer-facing telephony systems, as well as backend database creation and testing.
Supervisor: BOB BIRSS [REDACTED]

PUBLICATIONS

- Journal Papers*
- Fei Xia, William D. Lewis, Michael W. Goodman, Glenn Slayden, Ryan Georgi, Joshua Crowgey, & Emily Bender. Enriching a Massively Multilingual Database of Interlinear Glossed Text. *Language Resources and Evaluation*, Jan. 2016.
<http://link.springer.com/article/10.1007%2Fs10579-015-9325-4>
- Ryan Georgi, Fei Xia, William D. Lewis. Capturing Divergence in Dependency Trees to Improve Syntactic Projection. *Language Resources and Evaluation*, Oct. 2014.
<http://link.springer.com/article/10.1007%2Fs10579-014-9273-4>
- Conference Papers*
- Michael W. Goodman, Ryan Georgi, and Fei Xia. PDF-to-Text Reanalysis for Linguistic Data Mining. *LREC 2018*, May 2018. (*System Demonstration*)
<http://www.lrec-conf.org/proceedings/lrec2018/pdf/947.pdf>
- Ryan Georgi, Michael W. Goodman, and Fei Xia. A Web-framework for ODIN Annotation. *ACL 2016*, Aug. 2016. (*System Demonstration*)
<https://www.aclweb.org/anthology/P/P16/P16-4006.pdf>
- Ryan Georgi, Fei Xia & William D. Lewis. Enriching Interlinear Text using Automatically Constructed Annotators. *LaTeCH-2015*, Jul. 2015.
<http://www.aclweb.org/anthology/W15-3709>
- Fei Xia, Michael W. Goodman, Ryan Georgi, Glenn Slayden, and William D. Lewis. Enriching, Editing, and Representing Interlinear Glossed Text. *CICLing 2015*, Apr. 2015.
http://link.springer.com/chapter/10.1007%2F978-3-319-18111-0_3
- Ryan Georgi, Fei Xia, William D. Lewis. Enhanced and Portable Dependency Projection Algorithms Using Interlinear Glossed Text. *ACL 2013*, Aug. 2013. (*Poster*)
<http://www.aclweb.org/anthology/P13-2055>

Ryan Georgi, Fei Xia, William D. Lewis. Improving Dependency Parsing with Interlinear Glossed Text and Syntactic Projection. *COLING 2012*, Dec. 2012.
<http://www.aclweb.org/anthology/C12-2037>

Ryan Georgi, Fei Xia, William D. Lewis. Measuring the Divergence of Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms. *LREC 2012*, May 2012. (Oral Presentation)
<http://research.microsoft.com/pubs/168001/GeorgietallLREC2012.pdf>

Ryan Georgi, Fei Xia, William D. Lewis. Comparing Language Similarity across Genetic and Typologically-Based Groupings. *COLING 2010*, Aug. 2010. (Poster)
<http://www.aclweb.org/anthology/C10-1044>

PRESENTATIONS & INVITED TALKS

From Aari to Zulu: Massively Multilingual Creation of Language Tools Using Interlinear Glossed Text. IBM Research Almaden, San Jose, CA., Jun. 20, 2016.
https://ryan.georgi.cc/wp-content/uploads/2016/01/ryan_georgi_ibm_almaden_june_20.pdf

Using IGT with INTENT: Automatically Enriching Interlinear Glossed Text (IGT). *MS/UW Symposium*, Nov. 6, 2015.
<http://research.microsoft.com/apps/video/dl.aspx?id=258553>

Measuring the Divergence of Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms. *NW-NLP*, May 11, 2012.
<http://research.microsoft.com/apps/pubs/default.aspx?id=168001>

TEACHING

Instructor Deep Processing for NLP – (Sample Syllabus) Autumn 2018
Winter 2018

Instructor for graduate-level course in the **Master of Science in Computational Linguistics (CLMS)** program, covering parsing, computational semantics, and discourse for students from a mixture of linguistic and CS backgrounds.

Instructor NLP Systems & Applications – (Sample Syllabus) Spring 2018

Instructor for capstone course of **CLMS** program. In this course, I guided students through the process of working in teams to develop an end-to-end implementation of an extractive summarization system, based on the **TAC 2009 AESOP** task. Lectures involved examination of component pieces of the project, as well as project management and group coding skills, with deliverable milestones throughout the quarter. I also worked with students in 2013 as a TA, on a **TREC question answering track**.

Instructor Ethics in NLP – (Sample Syllabus) Spring 2018

This graduate seminar covered a variety of topics related to ethics and fairness, ranging from the philosophical underpinnings of ethics and human subjects protections to examining NLP data and applications for bias and algorithmic fairness. Also introduced students to concepts of system design that account both for direct and indirect stakeholders, and how each group may be affected by the implementation of automated systems.

Teaching Assistant Advanced Statistical Methods for NLP Winter 2014,
10 & 09

Assisted **CLMS** students in understanding concepts related to machine learning concepts including statistics, classifiers, clustering, and inductive learning in semi-supervised and unsupervised settings.

PROFESSIONAL SERVICE

General Co-chair, <i>NW-NLP</i>	2018
Program Committee, <i>AAAI</i>	2018
Reviewer, <i>COLING</i>	2018
Reviewer, <i>EMNLP</i>	2017
Reviewer, <i>Computational Linguistics</i>	2016

SOFTWARE PACKAGES

INTENT

INTENT

In *INTENT*, or the *INterlinear Text ENrichment Toolkit*, I implemented machine learning techniques for automatically enriching IGT data and extracting NLP tools including POS taggers, dependency parsers, and translation lexicons. Using *NLTK* and other dependencies, I designed this package to run the experiments for my doctoral dissertation and ensure reproducibility. The package was recently used by the *RiPLEs* project to enrich 158K IGT instances for 1.5K languages for the Online Database of Interlinear Text (*ODIN*) v2.1.

<https://github.com/rgeorgi/intent>

Links: <http://xigt.org/odin>

Classification Pipeline

Document Classification and Text Extraction for *RiPLEs*

Using Python and Scikit-Learn, I implemented a two-stage classification pipeline for the *RiPLEs* project with the goal of identifying short relevant textual elements within a large (1.3M) collection of noisy text extracted from PDF documents. My *DOC-CLASSIFY* package first identifies potential linguistic documents within the open-domain set of documents, while I designed the *IGT-DETECT* package to detect and extract two-to-three line instances of Interlinear Glossed Text (IGT) within a linguistic document. The end-product of this system aims to use a human-in-the-loop to validate and clean the extracted IGT instances for addition to the *ODIN* database. While this project is still underway, preliminary results suggest that this system may triple the current coverage of *ODIN*, which currently contains 158,000 IGT instances over 1,500+ languages.

<https://github.com/xigt/doc-classify>

Links: <https://github.com/xigt/igtdetect>

IGT Editor

IGT EDITOR

Using my experience as a freelance web developer, I created this AJAX-enabled browser-based IGT editing toolkit to aid the *RiPLEs* group's annotation efforts. This tool enables annotators to clean IGT instances, rate cleanliness, and visualize the *INTENT*-enriched output. The utility of the web editor is further strengthened by the inclusion of *INTENT* server-side for first-pass word alignment and POS tagging, facilitating rapid annotation by reducing repeated tasks for human users.

Links: <http://editor.xigt.org/user/demo>

FORMAL AND NATURAL LANGUAGE EXPERIENCE

NATURAL	English – Native German – Conversational
FORMAL	Python, JQuery/AJAX/CSS – Advanced Java, C#, C – Intermediate