

Damian Yukio | Romero Diaz

Tucson, Arizona, USA, 85745 | <https://spanish.arizona.edu/people/damianiji> | <https://d-romero.com>
damianiji@email.arizona.edu

Ph.D. candidate at The University of Arizona. Analytical linguist with a focus on corpus linguistics, computational linguistics, and natural language processing. Interested in improving communication outcomes among humans and with computers. Experience in English, Spanish, and Portuguese projects for academia and industry. Keen, hardworking, and eager to learn. Able to work under high-demand situations.

SKILLS

- **Collection and annotation of language data:** Project management for the creation and implementation for the collection of language corpora | Copyright and legal sanity checks for language datasets | Creation of annotation workflows and guidelines | XML annotation | Use of regular expressions and other tools for semi-automation of language annotation | Data anonymization | Annotation tools such as Label Studio and Prodigy
- **Linguistic data analysis:** Use of APA and IPA conventions for spoken data transcription and analysis | Experience in morpho-syntactic, phonological, and pragmatic analyses | Familiarity with WordNet and semantic categories | Basic use of R software for statistics | Corpus tools such as LancesBox and AntConc
- **Scripting languages & version control:** Python & Perl (mainly) | Bash | MATLAB | IBM's Annotation Query Language | SQL (basic prior knowledge) | Git (& git flow) & GitHub
- **Natural language processing and machine learning:** Python libraries for ML and NLP: spaCy, Natural Language Toolkit (NLTK), scikit-learn, PyTorch | Kaldi ASR (basic knowledge)
- **Natural language proficiency:** Spanish (Native and Language Expert for Professional Purposes) | English (C2/High proficiency in technical, academic, and everyday language) | Portuguese (B1/Intermediate conversational and written proficiency) | French (A2/Basic, studying at the moment).
- **Pedagogical expertise:** Collaborative learning | Online teaching and learning | Classroom diversity and inclusion | Community-based learning | Task-based learning (learn by doing) | Student-oriented facilitation

WORK & RESEARCH EXPERIENCE

Dissertation work, The University of Arizona, U.S.A.

Dr. Carmen King de Ramirez, *Dissertation committee chair*,

07/2019 - today

- Currently working on a theoretical and computational study of people's (instructors' and crowdsourced) perceptions of student emails.
- Creation of corpora, datasets, workflows, annotation conventions, experiments and evaluation (precision, recall, F1) for the supervised multilabel classification of documents using Natural Language Processing, as well as conventional linguistic analysis methods.

ExplosionAI, GmbH, Berlin, Germany (Remote)

Machine Learning Intern

05/2021 - today

- Creation of annotation guidelines and resources for natural language processing
- Data annotation, design and implementation of machine learning architectures for English and Spanish

School of Information, The University of Arizona, U.S.A.

NLP Research Assistant for Dr. Tracy Crane and Dr. Steven Bethard

08/2020 - today

- Theorization of causal models for Motivational Interviewing in diverse cultural settings
- Development of resources for speech data annotation and analysis
- Creation of NLP models to help improve outcomes of cancer prevention interventions

University Libraries - Office of Digital Innovation and Stewardship, The University of Arizona, U.S.A.
Research Data Repository Graduate Assistant, Supervisor: Fernando Rios 08/2020 - 05/2021

- Currently working on the creation and launching of a research data repository service for open data that will enable open and reproducible science at The University of Arizona.
- Software development (backend), data curation, and campus outreach.

Department of Spanish and Portuguese, The University of Arizona, U.S.A.

Research assistant for Dr. Julieta Fernandez 08/2017 - 12/2017

- Designed workflows and conventions of two oral corpora of learners of Spanish.
- Transcribed and annotated oral language interviews using some of these conventions.

IBM Deutschland - Research and Development GmbH, Böblingen, Germany

Intern, Watson Analytics for Social Media 05/2017 - 08/2017

- Creation of vocabularies, grammars and other linguistic resources for Spanish and Portuguese.
- Computational analysis of Brazilian Portuguese data using rule-based approaches.

Management Information Systems Department, The University of Arizona, U.S.A.

Research assistant for Dr. David Kauchak, Pomona College 08/2015 - 07/2016

- Worked with a team of researchers from the University of Arizona and Pomona College.
- Performed language analysis and processing through rule-based approaches and database (MySQL) queries for text simplification on a project funded by the National Institute of Health (U.S.A.)

EDUCATION

Ph.D. in Hispanic Linguistics

The University of Arizona, Tucson, Arizona, U.S.A.

Expected graduation: Spring, 2022, GPA: 3.889/4.0

Dissertation: Instructor Perceptions of College Student Emails: Success Factors in Spanish and English

M.S. in Human Language Technology

The University of Arizona, Tucson, Arizona, U.S.A.

Expected graduation date: Fall, 2021, GPA: 3.889/4.0

M.A. in Hispanic Linguistics

The University of Arizona, Tucson, Arizona, U.S.A.

Graduation date: Spring, 2014, GPA: 3.909/4.0

Outstanding Graduate Teaching Assistant Award, College of Humanities

Outstanding Graduate Student Award, Hispanic Alumni Association

B.A. in Linguistics and Spanish Teaching as a Second Language

Autonomous University of Querétaro (UAQ), Querétaro, México

Graduation date: Summer, 2012, Cumulative grade: 9.359/10

Sponsored study abroad program at Concordia University (Montreal, Canada) in Fall 2010 & Spring 2011

FEATURED PUBLIC PROJECTS ON GITHUB

(English) MCEC-EmG: <https://github.com/MCECorpus/>

I continuously develop Python exercises in Jupyter Notebooks for myself and other graduate students.

(Spanish) Cuarentena 2020: https://github.com/damian-romero/cuarentena_2020

Code and live coding video tutorial for writing a video game with Pygame.