

Damian Yukio Romero Diaz

Charlotte, North Carolina, U.S.A., 28262 | <https://d-romero.com> | damianiji@email.arizona.edu

Analytical linguist with a focus on data collection, management of digital assets, 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 passionate for constant learning. Able to work under high-demand situations but believes in life balance and joy at work.ⁱ

SKILLS

- **Language data collection:** Creation and management of annotation workflows and guidelines for language corpora | Use of regular expressions and other tools for semi-automation of language annotation | Experience with annotation tools like Prodigy and LabelStudio | Data de-identification and PII law compliance
- **Stewardship of digital assets:** Project management for the creation, distribution, and reuse of digital language datasets | Copyright and legal sanity checks for language datasets | Creation of data management plans | Familiarity and practical experience with Open Access and FAIR principles | Familiarity and eagerness to practice CARE principles
- **Scripting languages & version control:** Python (mainly) | Bash | Perl (rusty) | MATLAB (rusty) | IBM's Annotation Query Language (rusty) | 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), Transformers, scikit-learn, PyTorch | Kaldi ASR (basic knowledge)
- **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 | R software for basic statistics | Corpus tools such as LancesBox and AntConc
- **Natural language proficiency:** Spanish (Native) | English (C2/High proficiency in technical, academic, and everyday language) | Portuguese (A2/Basic conversational and written proficiency) | French (A1/Basic)
- **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

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

Dissertation work, Committee chair: Dr. Carmen King de Ramirez

Current

- Using qualitative analyses and machine learning techniques, I study how linguistic and contextual factors in emails trigger receiver's perceptions of, for example, appropriateness and formality.
- Creation of corpora, workflows, annotation conventions, and evaluation (precision, recall, F1) for supervised multilabel classification of machine learned models for people's email perceptions.
- Emphasis on data stewardship topics like FAIR principles of digital assets, management of personally identifiable information, compliance with data regulations, and collaboration with ethics committees.

Explosion AI, 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 models for English and Spanish.
- Social media community engagement, internal team building, and team engagement.

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

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

08/2021 - today

- [NIH Grant Project #1R21CA256680-01](#): Using Natural Language Processing to Determine Predictors of Healthy Diet and Physical Activity Behavior Change in Ovarian Cancer Survivors.
- [Created an award-winning data management plan](#) focused on machine learned model sharing and reuse, and I planned future data-centric ML research via [actionable data management foresight](#).
- Theorization of causal models for Motivational Interviewing counseling in diverse cultural settings.
- Creation of NLP models to help improve outcomes of cancer prevention interventions.
- Developed resources for speech data annotation and managed the work of seven part-time annotators.

University Libraries - Office of Digital Innovation and Stewardship, The University of Arizona, U.S.A.
DevOps and outreach, University Libraries, Supervisor: Fernando Rios **08/2020 - 05/2021**

- Worked on the creation and implementation of a [research data repository](#) for hosting and enabling diverse, open, and reproducible archival datasets at The University of Arizona. Designed [documentation website](#).
- Data curation, data stewardship, community outreach, and basic backend software development.
- [Published over 50 alternative research outcomes](#), including datasets, conference presentations, and posters.

Department of Spanish and Portuguese, The University of Arizona, U.S.A.
Teaching, outreach, research **08/2017 - 07/2020**

- Taught most Spanish levels for in-person and online programs (102 - 251, 325, 333, and 425).
- Designed workflows and conventions of two (private) oral language corpora for Dr. Julieta Fernandez.
- Worked on [outreach and community building](#) for Dr. Benjamin Fraser, [Journal of Urban Cultural Studies](#)

IBM Deutschland - Research and Development GmbH, Böblingen, Germany
Intern at Watson Analytics for Social Media, Supervisor: Dr. Thilo Götz **05/2017 - 08/2017**

- Creation of vocabularies, grammars, and other linguistic resources for Spanish and Portuguese.
- Computational analysis of Brazilian Portuguese social media 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**

- [Evidence-based Strategy and Tool to Simplify Text for Patients and Consumers](#).
- Analyzed the role of [semantic and grammatical](#) features on text simplification.
- Contributed to the creation of a [negation parser](#) for text simplification.

EDUCATION

Ph.D. in Hispanic Linguistics (Candidate)
The University of Arizona, Tucson, Arizona, U.S.A.
 Expected graduation: Spring 2023, GPA: 3.889/4.0

M.S. in Human Language Technology
The University of Arizona, Tucson, Arizona, U.S.A.
 Expected graduation date: Summer 2022, 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 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

ⁱ Full CV and references available upon request