

Hadas Kotek, Ph.D.

Linguistics, project management, ontology, data science

www.hkotek.com

hkotek@gmail.com

www.github.com/hkotek

Experience

Apple

NL Annotation Project Lead 2019–present

Siri Understanding, NL, Data Science team

- ▶ Drove data annotation projects to build and test the quality of Machine Learning models.
- ▶ Set and maintained schedules for multiple top priority projects to align with organizational goals and timelines.
- ▶ Designed methodology to optimize quality of annotation projects, improving accuracy by 10% and throughput by 30%.
- ▶ Led NEU ontology development for new Siri products.

Facilitated cross-functional collaboration with data scientists, ML modeling engineers, platform developers, annotation operations admins, and multi-vendor project managers and annotators.

Reviewed and provided feedback on Engineering efforts, contributing to the development of Siri Understanding R&D projects.

Drafted and maintained clear and concise grading guidelines. Created trainings for Siri and Siri International annotators.

Identified errors and regressions through data analysis, and proposed solutions to improve model accuracy and performance.

Faculty member in Linguistics

Yale University 2016–2017, 2018–2019

New York University 2017–2018

McGill University 2014–2016

- ▶ Led research and teaching activities at R1 universities.
- ▶ Advised PhD dissertations, graduate Qualifying Papers, and undergraduate theses.

Collected linguistic data and developed cutting-edge linguistic theories, resulting in: 1 monograph ([MIT Press](#)); 1 edited volume; >20 published papers; >60 presentations; 10 grants and awards.

Skills

Project management

Cross-functional collaboration
Personnel & budget management
Written & oral presentation

Ontology

Ontology design
Guideline creation

Data science

Experiment design
Data sampling, processing, and visualization
Error analysis
Statistical analysis

NLP

Named Entity Understanding
Part of Speech Tagging
Constituency Parsing
Discourse-Sensitive Parsing
Multi-turn conversations

Technology

Python (pandas, numpy, NLTK)

R (tidyverse, dplyr, ggplot2, lme4, nlme)

Markdown	Git
LaTeX	HTML (basic)
regex	SQL (basic)

Research

See [academic CV](#)

Created teaching materials; taught lectures and seminars; managed Teaching Assistant teams (14 courses in syntax, semantics, & experimental linguistics; over 500 enrolled students.)

Co-organized 6 intl conferences with hundreds of attendees.

Graduate student researcher

Massachusetts Institute of Technology 2009–2014

Tel-Aviv University 2008–2009

Center for General Linguistics, Berlin 2007–2008

- ▶ Synthesized large amounts of natural language data from several typologically unrelated languages.
- ▶ Designed and implemented psycholinguistic, neurolinguistic, processing, acquisition, & grammaticality judgment experiments.
- ▶ Created data visualization and analysis scripts.

16 papers, 37 presentations, 11 grants and awards.

Led outreach activities, teaching linguistics to a general audience.

Co-creator of [Turktools](#): Free toolkit for implementation and analysis of behavioral language experiments on Amazon Mechanical Turk.

- ▶ Coded HTML and R scripts for data visualization & analysis.
- ▶ Designed and contributed to programming of Python scripts.

Machine Learning Coordinator

Ontology Ltd, Tel-Aviv, Israel 2005–2006

Built a knowledge graph for machine learning software based on a multilingual classification algorithm.

Contributed to ontology design.

Collected, processed, and annotated Spanish datasets.

Publications

- 2020 Sun, David Q.*, **Hadas Kotek***, Christopher Klein, Mayank Gupta, William Li and Jason D. Williams. [Improving Human-Labeled Data through Dynamic Automatic Conflict Resolution](#). *The 28th International Conference on Computational Linguistics (COLING)*.
- 2020 Patel, Alkesh, Akanksha Bindal, Hadas Kotak, Christopher Klein, and Jason D. Williams. [Generating Natural Questions from Images for Multimodal Assistants](#). *West Coast NLP (WeCNLP)*.

Education

PhD, Linguistics 2014
Massachusetts Institute of Technology

BA, Linguistics 2007
Tel-Aviv University

Other relevant coursework:

- Intro to SQL for Data Science (DataCamp)
- Learn Python 2 (CodeAcademy)
- Infinitesimal Calculus 1+2, Discrete Math
- Statistical Analysis, Research Methods
- AP Computer Science, Math, Physics, English, Arabic

Languages

Hebrew	native
English	native
German	fluent, ZOP certificate
Arabic	conversational
Spanish	conversational
Japanese	beginner
French	beginner
Kaqchikel (Mayan)	fieldwork
Chuj (Mayan)	fieldwork
Tibetan	fieldwork