Hadas Kotek, Ph.D.

Linguistics | data science | project management | HCOMP and annotation

Research Affiliate, MIT

hkotek@gmail.com

hkotek.com

linguistics.mit.edu/
user/hkotek/

Experience

Senior Data Scientist

Apple, Siri and Language Technologies, NL team   2022–present

Provided data-driven solutions to measure and improve annotation across the AI|ML org.

- ▶ Innovated annotation tools and tasks to improve efficiency, consistency, and accuracy in 41 locales.
- ▶ Led annotation efficiency improvement of ∼25% worldwide via targeted data sampling and model assisted annotation.
- ▶ Designed and implemented methodology to optimize annotation quality. Accuracy improved by 10%, throughput by 30%.
- ▶ Developed analytical metrics to report consistent measurement of annotation accuracy to stakeholders and annotators.

Senior NL Annotation Lead

Apple, Siri and Language Technologies, NL team   2019–2022

Drove annotation projects to build and test large-scale ML models.

- ▶ Facilitated cross-functional collaboration with data scientists, modeling engineers, platform developers, admins, multi-vendor project managers, and annotators globally.
- ▶ Set and maintained weekly and quarterly schedules for multiple top priority projects to align with organizational goals.
- ▶ Reviewed R&D annotation projects to implement and improve new Siri features, including many currently in production.
- ▶ Onboarded and trained dozens of annotators worldwide.
- ▶ Created guidelines, trainings, and new feedback procedures.

Professor of Linguistics

Yale University; New York University; McGill University   2014–2019

- ▶ Led cutting-edge collaborative linguistic research at R1 universities, shepparding projects from ideation to dissemination, resulting in: 1 monograph, 1 edited volume, >20 published papers, >60 presentations, 10 grants and awards.
- ▶ Obtained and managed over $100,000 in research funding.

Skills

Data science

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

Project management

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

Ontology and annotation

Ontology design
Guideline creation
Data labeling
Linguistic data analysis

NLP

Named Entity Recognition
Natural Question Generation
Query rewrite
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

- ▶ Designed curricula; developed teaching materials; taught lectures and seminars; managed teaching assistant teams (14 new courses in syntax, semantics, & experimental linguistics; 15 teaching assistants; over 500 enrolled students.)
- ▶ Advised PhD dissertations, graduate Qualifying Papers, and undergraduate theses.

Graduate Student Researcher

Massachusetts Institute of Technology; Tel-Aviv University; Center for General Linguistics, Berlin   2007–2014

- ▶ Synthesized large amounts of natural language data from high- and low-resource languages to develop linguistic theories, resulting in: >15 papers, >35 presentations, 11 grants and awards.
- ▶ Designed and implemented dozens of behavioral studies and semi-structured interviews.

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.
- ▶ Powered >60 studies world wide and still actively used today.

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.

Education

PhD, Linguistics   2014
Massachusetts Institute of Technology

BA, Linguistics and Political Science   2007
Tel-Aviv University

Other relevant coursework:

- ○ SQL for Data Science
- ○ Python
- ○ Infinitesimal Calculus 1+2, Discrete Mathematics
- ○ Statistical Analysis, Research Methods

Languages

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

Selected publications

2021   Patel, Alkesh, Joel Moniz, Roman Nguyen, **Hadas Kotek**, Nick Tzou, Vincent Renkens. MMIU: Dataset for Intent Understanding in Multimodal Assistant". *WeCNLP*.

2021   Patel, Alkesh, Akanksha Bindal, **Hadas Kotek**, Christopher Klein, and Jason D. Williams. Generating Natural Questions from Images for Multimodal Assistants. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

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)*.

See additional details on my academic CV.