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 improve annotation across the org.

- Developed analytical metrics to report consistent annotation accuracy and efficiency measures to stakeholders.
- Innovated methodology to optimize the annotation process. Accuracy improved by 10%, throughput by 30%.
- Spearheaded saving of ∼25% of work volume via targeted data sampling and model assisted annotation in >40 locales.

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 linguistic research, shepherding projects from ideation to dissemination, resulting in: 1 monograph, 1 edited book, >20 papers, and >60 presentations.
- Obtained >$120k in research support through +10 grants & awards.
- Developed and taught 14 courses with >500 enrolled students; designed curricula, lecture notes, assignments, and assessments. Mentored and supervised 14 teaching assistants.
- Supervised 6 mentees. Provided feedback on research directions & writing. Aided in professionalization & career planning.

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

- Hired a team of 24 research assistants to create a corpus of >22k English sentences. Provided training and ongoing feedback to trainees, managed and approved budgets, and supervised quality assurance of the results.

Graduate Student Researcher

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

- Synthesized +2000 datapoints from high- & low-resource languages and the results of 30+ behavioral studies and semi-structured interviews with consultants to develop linguistic theories, leading to >15 papers, >35 presentations, 11 grants & awards.
- Recruited >2500 experiment participants. Implemented methods to detect cheating or distracted behavior to ensure data quality and reliability. Analyzed results in R.
- Co-created ***Turktools***, a free toolkit for implementing and analyzing behavioral language experiments on Amazon Mechanical Turk, including HTML and R scripts for data visualization and analysis. The tools have powered several dozen studies world wide and are 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 data.

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.

Hadas Kotek, hkotek@gmail.com