Hadas Kotek, Ph.D.

Linguistics, project management, ontology, data science

Research Affiliate at MIT

hkotek@gmail.com

hkotek.com

linguistics.mit.edu/user/hkotek/

Experience

Apple

NL Annotation Lead   **2019–present**

AI|ML Siri Understanding, Natural Language, Data Science team

Project management

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.
- ▶ Reviewed and provided feedback on Engineering efforts, contributing to the development of Siri Understanding R&D projects.
- ▶ Facilitated cross-functional collaboration with data scientists, modeling engineers, platform developers, annotation operations admins, multi-vendor project managers and annotators.

Data science and ontology

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

- ▶ Led annotation efficiency improvement of 25% worldwide, through smart sampling and model assisted annotation.
- ▶ Designed methodology to optimize quality of annotation projects, improving accuracy by 10% and throughput by 30%.
- ▶ Led NEU ontology development for new Siri products.

Faculty member in Linguistics

Yale University   **2016–2017, 2018–2019**

New York University   **2017–2018**

McGill University   **2014–2016**

Led cutting-edge linguistic research at R1 universities. Engaged in teaching, advising, and service to the profession.

- ▶ 1 monograph (MIT Press); 1 edited volume; >20 published papers; >60 presentations; 10 grants and awards.
- ▶ Created teaching materials; taught lectures and seminars; managed Teaching Assistant teams (14 courses in syntax, semantics, & experimental linguistics; over 500 enrolled students.)

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 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

- ▶ Co-organized 6 intl conferences with hundreds of attendees.
- ▶ Advised PhD dissertations, graduate Qualifying Papers, and un-dergraduate theses.

Graduate student researcher

Massachusetts Institute of Technology	**2009–2014**
Tel-Aviv University	**2008–2009**
Center for General Linguistics, Berlin	**2007–2008**

Developed linguistic theories based on large amounts of natural language data using behavioral, experimental, and fieldwork methods.

- ▶ >15 papers, >35 presentations, 11 grants and awards.
- ▶ Designed and implemented psycholinguistic, neurolinguistic, processing, acquisition, & grammaticality judgment experiments.
- ▶ Created data visualization and statistical analysis scripts.

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.

Led outreach activities, teaching linguistics to a general audience.

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

Publications

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).*