

# Hadas Kotek, Ph.D.

Linguistics, project management, data science

Research Affiliate, MIT

hkotek@gmail.com

hkotek.com

linguistics.mit.edu/  
user/hkotek/

## Experience

---

### Senior NL Annotation Lead

Apple

2019–present

Siri Understanding, Natural Language team

Drove data annotation projects to build and test the quality of Machine Learning models.

- ▶ 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 now in production.
- ▶ Onboarded and trained several dozen annotators worldwide.
- ▶ Facilitated cross-functional collaboration with data scientists, modeling engineers, platform developers, admins, multi-vendor project managers and annotators globally.

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

- ▶ 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 NEU ontology for new Siri products.

### Professor of Linguistics

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

- ▶ Led cutting-edge collaborative linguistic research at R1 universities, resulting in: 1 monograph ([MIT Press](#)), 1 edited volume, >20 published papers, >60 presentations, 10 grants and awards.
- ▶ 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.)
- ▶ Co-organized 6 intl conferences with hundreds of attendees.
- ▶ Advised PhD dissertations, graduate Qualifying Papers, and undergraduate theses.

## Skills

---

### Project management

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

### Data science

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

### 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](#)

## 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 diverse languages and developed linguistic theories, resulting in: >15 papers, >35 presentations, 11 grants and awards.
- ▶ Designed and implemented dozens of neurolinguistic and behavioral studies (sentence processing, language acquisition, grammaticality judgments) and semi-structured interviews.
- ▶ Led outreach activities, teaching linguistics to a general audience of ~100 participants in the greater Boston area.

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.

## Selected publications

- |      |  |
|------|--|
| 2021 | Patel, Alkesh, Akanksha Bindal, <b>Hadas Kotek</b> , Christopher Klein, and Jason D. Williams. <a href="#">Generating Natural Questions from Images for Multimodal Assistants</a> . <i>IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)</i> .         |
| 2020 | Sun, David Q.*, <b>Hadas Kotek*</b> , Christopher Klein, Mayank Gupta, William Li and Jason D. Williams. <a href="#">Improving Human-Labeled Data through Dynamic Automatic Conflict Resolution</a> . <i>The 28th International Conference on Computational Linguistics (COLING)</i> . |

## 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