|  |  |  |
| --- | --- | --- |
| **Hadas Kotek, Ph.D.**  **Linguistics | data science | project management | HCOMP and annotation**  **Research Affiliate, MIT**  **Experience** | | hkotek@gmail.com  hkotek.com  linguistics.mit.edu/ user/hkotek/  **Skills** |
| **Senior Data Scientist** | | **Data science**  Experiment design  Data sampling, processing, and visualization |
| **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**

Statistical analysis   
Error analysis

**Project management**

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

|  |  |  |
| --- | --- | --- |
| **Apple, Siri and Language Technologies, NL team** | **2019–2022** | **Ontology and annotation** |

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.

Ontology design   
Guideline creation   
Data labeling   
Linguistic data analysis

▶Set and maintained weekly and quarterly schedules for multi- **NLP**

ple 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**

Named Entity Recognition Natural Question Generation Query rewrite   
Multi-turn conversations

**Technology**

Python (pandas, numpy, NLTK)

|  |  |  |
| --- | --- | --- |
| **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. | R (tidyverse, dplyr, ggplot2, lme4, nlme) | |
| Markdown | Git |
| LATEX | HTML (basic) |
| regex | SQL (basic) |

▶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 direc-tions & writing. Aided in professionalization & career planning.

**Research**   
See [academic CV](https://hkotek.com/KotekCV.pdf)

▶Hired a team of 24 research assistants to create a corpus of >22k English sentences. Provided training and ongoing feed-

**Education**

|  |  |  |
| --- | --- | --- |
| back to trainees, managed and approved budgets, and super- | **PhD, Linguistics** | 2014 |
| vised quality assurance of the results. | Massachusetts Institute of |

Technology

|  |  |  |  |
| --- | --- | --- | --- |
| **Graduate Student Researcher** | | **BA, Linguistics and Political** | 2007 |
| **Massachusetts Institute of Technology; Tel-Aviv University;** | | **Science** |
| **Center for General Linguistics, Berlin** | **2007–2014** | Tel-Aviv University |

▶Synthesized +2000 datapoints from high- & low-resource lan-guages and the results of 30+ behavioral studies and semi-structured interviews with consultants to develop linguistic the-ories, leading to >15 papers, >35 presentations, 11 grants & awards.

▶Recruited >2500 experiment participants. Implemented meth-ods to detect cheating or distracted behavior to ensure data quality and reliability. Analyzed results in R.

▶Co-created ***[Turktools](http://turktools.net/)***, a free toolkit for implementing and an-alyzing behavioral language experiments on Amazon Mechan-ical 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.

Other relevant coursework:

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

**Languages**

|  |  |
| --- | --- |
| Hebrew | native |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Machine Learning Coordinator** | | English | native-like | |
| German | fluent, ZOP certificate | |
| **Ontology Ltd., Tel-Aviv, Israel** | **2005–2006** | Arabic | conversational | |
| Built a knowledge graph for machine learning software based on a multilingual classification algorithm.  ▶Contributed to ontology design.  ▶Collected, processed, and annotated Spanish data. | | 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](https://arxiv.org/abs/2012.03678)  [Natural Questions from Images for Multimodal Assistants](https://arxiv.org/abs/2012.03678). *IEEE International Conference on Acous-* *tics, 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](https://arxiv.org/abs/2012.04169). *The 28th Interna-tional Conference on Computational Linguistics (COLING)*.

See additional details on my [academic CV](https://hkotek.com/KotekCV.pdf).

Hadas Kotek, hkotek@gmail.com