


# Joshua Tanner

NLP Engineer

 [linkedin.com/in/joshuatanner2](https://www.linkedin.com/in/joshuatanner2)

 [mindful.jt@gmail.com](mailto:mindful.jt@gmail.com)

 <https://github.com/mindful/>

## Experience

### Principal Applied Scientist

Health Data Intelligence

Oracle

2024/7 - Current

- Research into training medicine-specialized LLMs and medical entity linking

### NLP Lead

Development Team

Mantra Inc

2021/12 - 2024/5

- Lead NLP engineering (5 engineers) across projects including:
  - Apply LLMs to manga translation, allowing terminology and story aware translation of >30k manga chapters
  - Explore training methods to improve LLM translation, pruning noisy OCR data for monolingual and bilingual fine-tuning
  - Train multilingual and terminology-aware translation models with Fairseq, improving BLEU on multiple languages by >10
  - Work with professional translators to improve evaluation
- Develop and maintain application backend using Django and AWS

### NLP Consultant

Freelance

Multiple Companies

2020/1 - Current

- Led development of an interactive translation system allowing users to translate by repeatedly selecting from model-suggested words
- Implemented optimized character-level transformer architecture in PyTorch and pretrained it with masked language modeling, outperforming state-of-the-art in Japanese topic classification [\[link\]](#)
- Rewrote scientific information extraction pipeline using spaCy, increasing micro f1 by more than 20 points

### Research Intern

Institute of Industrial Science

University of Tokyo

2020/12 - 2021/12

- Research into to retrieval-augmented and explainable grammatical error correction

### Software Engineer

New Product Development & Global Growth

eBay

2016/11 - 2019/9

- Lead development of, tested, and maintained sole ownership of Scala distributed scheduler that powered proactive messaging for eBay's chat bot and Chinese iPhone app
- Research into bootstrapping NLP process to extract informative noun/adjective pairs from product reviews was accepted for presentation at eBay's internal research conference (25% acceptance rate)
- Developed Ruby on Rails web service that allowed non-technical staff to write and schedule messages from chat bot to users

### Software Engineer

Global Shipping Program

eBay

2015/9 - 2016/11

- Rearchitected the pipeline used to import shipping charge data from shipping partners, increasing the accuracy of international shipping prices on eBay's search results page by up to 40%
- Developed new features in Java/Spring services at scale (several million requests per hour), wrote tests using JUnit and Mockito
- Drove an effort to automate excessive manual operations work, saving team tens of hours of work per week

### Intern

Global Data Infrastructure

eBay

2013/6 - 2014/9

- Owned and expanded functionality of Ruby on Rails application automating Hadoop permissions changes and other operational tasks

## Research

### Project MOSLA: Recording Every Moment of Second Language Acquisition

M. Hagiwara, J Tanner

LREC-COLING 2024 Oral Presentation

[\[link\]](#)

### Utilizing Longer Context than Speech Bubbles in Automated Manga Translation

H Kaino, S Sugihara, T Kajiwar, T

Ninomiya, J Tanner and S Ishiwatari

LREC-COLING 2024 Poster [\[link\]](#)

### MWE as WSD: Solving Multiword Expression Identification with Word Sense Disambiguation

J Tanner, J Hoffman

EMNLP 2023 Findings & Poster [\[link\]](#)

### GrammarTagger: A Multilingual, Minimally-Supervised Grammar Profiler for Language Education

M Hagiwara, J Tanner, K Sakaguchi

Preprint 2021 [\[link\]](#)

## Awards

- LREC-COLING 2024 Outstanding Paper Award
- 2018 eBay Critical Talent award (top 1% performance)
- 2016 eBay Innovation Expo Presenter

## Education

### M.S. Comp. Linguistics

University of Washington / 2021

### B.S. Computer Science

The Evergreen State College / 2015

### Exchange Program

University of Hyogo / 2015

## Projects

### MWE Dictionary

2023

Multiword-expression aware dictionary for language learners using models trained with Pytorch Lightning and a gRPC backend

### Parasite

2009

Lead development of StarCraft 2 mod reaching two million plays per week

## Languages

English Native

Japanese Fluent (JLPT N1)