

# Dorottya (Dora) Demszky

## EDUCATION

---

### Ph.D., Linguistics, Stanford University

September 2017 – June 2022 (exp.)

Melvin and Joan Lane Stanford Graduate Fellow

Dissertation: “Using Natural Language Processing to Support Student-Centered Education”

Committee: Dan Jurafsky (chair), Heather C. Hill, Chris Piech

### B.A., Linguistics, Princeton University

September 2013 – June 2017

Minor in Computer Science

Summa Cum Laude

## RESEARCH & TEACHING INTERESTS

---

Education Data Science; Teaching Quality; Natural Language Processing; Computational Social Science

## PUBLICATIONS

---

\* indicates equal contribution

### Under Review

1. **Demszky, D.**, Liu, J., Hill, H., Jurafsky, D., and Piech, C. (2021a). Can Automated Feedback Improve Teachers’ Uptake of Student Ideas? Evidence from a Randomized Controlled Trial in a Large-Scale Online. *Under Review at Education Researcher*

### Peer-Reviewed

2. **Demszky, D.**, Liu, J., Mancenido, Z., Cohen, J., Hill, H., Jurafsky, D., and Hashimoto, T. (2021b). Measuring Conversational Uptake: A Case Study on Student-Teacher Interactions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*
3. **Demszky, D.**, Sharma, D., Clark, J. H., Prabhakaran, V., and Eisenstein, J. (2021c). Learning to Recognize Dialect Features. In *Proceedings of the 19th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*
4. **Demszky, D.** (2021). The Role of Verb Semantics in Hungarian Verb-Object Order. In *Proceedings of the Linguistic Society of America (LSA)*
5. Lucy\*, L., **Demszky\*, D.**, Bromley, P., and Jurafsky, D. (2020). Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas US History Textbooks. *AERA Open*, 6(3)

Best Paper Award at the 2020 Educational Data Science Conference

6. **Demszky, D.**, Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., and Ravi, S. (2020). GoEmotions: A Dataset of Fine-Grained Emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*
7. Kmetty, Z., Vincze, V., **Demszky, D.**, Ring, O., Nagy, B. N., and Szabó, M. K. (2020). *Pártélet*: A Hungarian Corpus of Propaganda Texts from the Hungarian Socialist Era. In *Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC)*

8. **Demszky, D.**, Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., and Jurafsky, D. (2019). Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings. In *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*

## Public Domain

9. Bommasani, R., Hudson, D. A., ..., **Demszky, D.**, ..., and Liang, P. (2021). On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258*
10. **Demszky\*, D.**, Guu\*, K., and Liang, P. (2018). Transforming Question Answering Datasets Into Natural Language Inference Datasets. *arXiv preprint arXiv:1809.02922*

## CONFERENCE PRESENTATIONS & INVITED TALKS

---

<b>World Well-Being Project Psychology Consortium</b>	2021
“Using NLP to Analyze Participants’ Essays from Social-Behavioral Interventions”	
<b>Association for Computational Linguistics (ACL) Conference</b>	2021
“Measuring Conversational Uptake: A Case Study on Student-Teacher Interactions”	
<b>Brown University, Department of Computer Science &amp; Annenberg Institute</b>	2021
“Using NLP to Support Student-Centered Education: What We Can Learn From Teachers’ Uptake of Student Ideas”	
<b>Google, NLP Group</b>	2021
“Using NLP to Make People Feel Heard: What We Can Learn From Teachers’ Uptake of Student Ideas”	
<b>North American Association of Computational Linguistics (NAACL) Conference</b>	2021
“Learning to Recognize Dialect Features”	
<b>Women in Data Science Workshop Series</b>	2021
“Using NLP to Analyze US History Textbooks”	
<b>Rising Stars in AI for Social Good Speaker Series, Harvard CRCS</b>	2021
“Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas U.S. History Textbooks”	
<b>Cornell University, NLP Seminar</b>	2021
“Measuring Conversational Uptake: A Case Study on Student-Teacher Interactions”	
<b>Linguistic Society of America (LSA) Annual Meeting</b>	2021
“The Role of Verb Semantics in Hungarian Verb-Object Order”	
<b>Eötvös Loránd University, Hungary</b>	2020
“Content Analysis of Textbooks via Natural Language Processing”	
<b>Conference on Educational Data Science, Stanford University</b>	2020
“Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas U.S. History Textbooks”	
<b>10th Annual New Directions in Analyzing Text as Data (TADA)</b>	2019
“Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas U.S. History Textbooks”	
<b>University of San Francisco Seminar Series in Data Science</b>	2019
“Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings”	

“Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings”

## GRANTS

---

**NSF Discovery Research K-12, Design and Development Project** July 2022 – June 2026

*PI: Jing Liu (College of Education, University of Maryland)*

Co-wrote and listed as key personnel in a \$3M grant application currently under review, titled “Improving Computer Science Instruction Through Automated Feedback to Teachers: A Case Study on Teachers’ Uptake of Student Ideas”. The project builds on my research on teachers’ uptake of student contributions to support teachers’ professional development in computer science classrooms.

**Human-Centered AI Seed Grant**

2021 – 2022

*PI: Patricia Bromley (Graduate School of Education, Stanford University)*

Co-wrote and listed as key personnel on a grant application titled “Civics Education for a Just and Sustainable Future”. We received \$75,000 in funding for our project on analyzing civics and history textbooks used in the United States from the perspective of diversity and equity.

**Grassroots and Parent Activism in Education Grant**

2020 – 2021

*For Tarisznya Táborok (the NGO I co-founded)*

Co-wrote and received a \$20,000 grant (15 out of 65 organizations were awarded) as part of the Education Support Program of Open Society Foundations. Our project focuses on organizing a year-round grassroots educational afterschool program for marginalized Roma students in Litke, one of Hungary’s 300 poorest villages, by building on and hiring local human resources.

**Grassroots Activism in Education Grant**

2018 – 2019

*For Tarisznya Táborok*

Wrote and received a \$13,630 grant (25 out of 80 organizations were awarded) as part of the Education Support Program of Open Society Foundations. The grant funded our work to organize project-based educational activities for marginalized Roma students in five rural Hungarian villages by building on local resources and partnerships.

**Davis Projects for Peace Grant**

2016

*For Tarisznya Táborok*

Recipient of a \$10,000 grant (1 or 2 projects are funded per university) to organize inclusive, project-based educational summer camps in rural Hungary for marginalized Roma students in two rural Hungarian villages.

## AWARDS & HONORS

---

Rising Star in AI for Social Good (Harvard CRCS) 2021

Best Paper Award at the Conference on Education Data Science 2020

Melvin and Joan Lane Stanford Graduate Fellowship (SGF) 2017 – 2021

Senior Thesis Prize (Princeton University) 2017

Phi Beta Kappa Honor Society 2017

Linguistic Summer Institute Fellowship 2015

Dr. Elemér & Éva Kiss Scholarship 2013, 2014, 2015, 2016, 2017

Member of the Hungarian team at the 10th International Linguistics Olympiad (IOL) 2012

**TEACHING & MENTORING**

---

**CURIS Program***Research Mentor*

Summer 2021

*Stanford University*

Mentored Sterling Alic, an undergraduate student in computer science on a project to develop computational methods to study teacher questioning in elementary math classrooms.

**Natural Language Understanding (CS 224U)***Teaching Assistant*

Spring 2021

*Stanford University*

Mentored students in their course projects, helping them identify a research problem and relevant literature, learn and apply NLP methods and write a final research paper. Helped students in office hours and provided detailed feedback on assignments.

**Corpus Teaching Assistant, Stanford University**

September 2019 – September 2020

Held workshops and helped students and researchers find, access and use text corpora in their research and course projects. Managed text corpora owned by the Linguistics Department.

**Introduction to Psycholinguistics (PSYCH 140)***Teaching Assistant*

Fall 2018

*Stanford University*

Gave a guest lecture, taught sections and office hours and graded student assignments.

**Petey Greene Prison Tutoring Program***Volunteer Tutor & Course Co-Leader*

January 2015 – May 2016

*New Jersey*

Co-led an employment preparation course and tutored college level math at correctional facilities.

**Peer Academic Advisor, Princeton University**

September 2015 – May 2017

Advised a group of freshmen and sophomores on academic issues, such as adjustment to college, course selection and major choices.

**North American Computational Linguistics Olympiad**

2013 – 2017

*Member of the Problem Committee and Organizing Committee*

Wrote and helped correct problems for the contest with 1,700+ participants yearly.

**Hungarian Linguistics Olympiad**

2012 – Present

*Founder & Board Member*

Providing training and mentorship to participants of the six-round national competition. Initiated and co-organized the first Hungarian Linguistics Olympiad in 2012.

**PROFESSIONAL EXPERIENCE**

---

**Tarisznya Táborok (Knapsack Camps)***Co-Founder & Co-Director*

2016 – Present

Co-founded and co-directing an NGO that seeks to empower marginalized Roma students in Hungary through inclusive and project-based educational opportunities. Have raised more than \$60,000 since 2016 via grants and crowdfunding campaigns.

**Google***Research Intern*

Summer 2020

*Seattle, WA*

Worked on multi dialect data collection, focusing on varieties of English, Spanish and Portuguese. Developed few-shot machine learning methods to recognize dialect features, in order evaluate the robustness of NLP systems to language variation. The project led to a peer-reviewed publication.

**Google**

*Research Intern*

Summer 2019

*Mountain View, CA*

Built a dataset, conducted in-depth analyses and developed models for recognizing fine-grained emotions in text. The project led to a peer-reviewed publication.

**Center for the Study of Language and Information (CSLI)**

*Research Intern*

Summer 2016

*Stanford University*

Worked with Professor Percy Liang (Computer Science) on a project pertaining to understanding the linguistic properties of natural language inference, leading to a publication.

**ACADEMIC SERVICE**

---

Organizer, NLP for Positive Impact Workshop (NLP4PI)	2021 – 2022
Member, Association for Computational Linguistics (ACL)	2018 – Present
Reviewer, American Education Research Journal (AERJ)	2021
Reviewer, Empirical Methods in Natural Language Processing (EMNLP)	2020, 2021
Reviewer, ACL	2019, 2021
Reviewer, American Educational Research Association (AERA) Open Journal	2020
Colloquium Committee Member, Stanford Linguistics	2019 – 2020
Social Committee Member, Stanford Linguistics	2017 – 2018
Social Co-Chair, Stanford NLP Group	2017 – 2019
Co-President, Princeton Linguistics Club	2015 – 2017