Huda Khayrallah

Hackerman 226
3400 N Charles Street
Baltimore, MD 21218

phone: +1 (510) 545-6782
email: huda@jhu.edu
web: cs.jhu.edu/~huda

Education

Johns Hopkins University — August 2015 - present
Ph.D. in Computer Science — *Baltimore, MD*
Affiliation: Center for Language and Speech Processing
Advisor: Philipp Koehn

Johns Hopkins University — August 2015 - May 2017
M.S.E. in Computer Science — *Baltimore, MD*

University of California, Berkeley — August 2011 - May 2015
B.A. in Computer Science — *Berkeley, CA*
Minor in Applied Language Studies

Awards

Best Paper Nomination — 2020
West Coast NLP Summit (WeCNLP)
for: Simulated Multiple Reference Training Improves Low-Resource Machine Translation
by: **Huda Khayrallah**, Brian Thompson, Matt Post, and Philipp Koehn.

Jelinek Fellowship — 2019
Johns Hopkins University Center for Language and Speech Processing

Outstanding Contribution Award — 2018
Workshop on Neural Machine Translation and Generation (WNMT)
for: On the Impact of Various Types of Noise on Neural Machine Translation
by: **Huda Khayrallah** and Philipp Koehn.

Graduate Research Fellowship Program (GRFP) Honorable Mention — 2015, 2016, 2017
National Science Foundation

Upsilon Pi Epsilon - Computer Science Honors Society — Inducted December 2012
UC Berkeley Computer Science

Publications

◇ indicates a student I mentored. * indicates authors contributed equally.
Corresponding presentations for publications can be found at cs.jhu.edu/~huda.
More information can also be found on my Google Scholar page.

Huda Khayrallah, Brian Thompson, Matt Post, and Philipp Koehn. 2020. Simulated Multiple Reference Training Improves Low-Resource Machine Translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, pages 82–89. https://doi.org/10.18653/v1/2020.emnlp-main.7
[Nominated for best paper at the WeCNLP summit]

Huda Khayrallah and João Sedoc. 2020. SMRT Chatbots: Improving Non-Task-Oriented Dialog with Simulated Multi-Reference Training. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, pages 4489–4505. https://doi.org/10.18653/v1/2020.findings-emnlp.403

Jacob Bremerman,◇ **Huda Khayrallah**, Douglas Oard, and Matt Post. 2020. On the Evalua-

tion of Machine Translation n-best Lists. In *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*. Association for Computational Linguistics, Online, pages 60–68. https://doi.org/10.18653/v1/2020.eval4nlp-1.7

Huda Khayrallah, Jacob Bremerman,◇ Arya D. McCarthy, Kenton Murray, Winston Wu, and Matt Post. 2020. The JHU Submission to the 2020 Duolingo Shared Task on Simultaneous Translation and Paraphrase for Language Education. In *Proceedings of the Fourth Workshop on Neural Generation and Translation*. Association for Computational Linguistics, Online, pages 188–197. https://www.aclweb.org/anthology/2020.ngt-1.22
[Highest scoring submission in all 5 language pairs.]

Brian Thompson,* Rebecca Knowles,* Xuan Zhang,* **Huda Khayrallah**, Kevin Duh, and Philipp Koehn. 2019. HABLex: Human Annotated Bilingual Lexicons for Experiments in Machine Translation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, pages 1382–1387. https://doi.org/10.18653/v1/D19-1142

Adrian Benton, **Huda Khayrallah**, Biman Gujral, Dee Ann Reisinger, Sheng Zhang, and Raman Arora. 2019. Deep Generalized Canonical Correlation Analysis. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*. Association for Computational Linguistics, Florence, Italy, pages 1–6. https://doi.org/10.18653/v1/W19-4301

Brian Thompson, Jeremy Gwinnup, **Huda Khayrallah**, Kevin Duh, and Philipp Koehn. 2019. Overcoming Catastrophic Forgetting During Domain Adaptation of Neural Machine Translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, pages 2062–2068. https://doi.org/10.18653/v1/N19-1209

J. Edward Hu, **Huda Khayrallah**, Ryan Culkin, Patrick Xia, Tongfei Chen, Matt Post, and Benjamin Van Durme. 2019. Improved Lexically Constrained Decoding for Translation and Monolingual Rewriting. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, pages 839–850. https://doi.org/10.18653/v1/N19-1090

Huda Khayrallah, Rebecca Knowles, Kevin Duh, and Matt Post. 2019. An Interactive Teaching Tool for Introducing Novices to Machine Translation. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. ACM, New York, NY, USA, SIGCSE '19, pages 1276–1276. https://doi.org/10.1145/3287324.3293840

Brian Thompson, **Huda Khayrallah**, Antonios Anastasopoulos, Arya D. McCarthy, Kevin Duh, Rebecca Marvin, Paul McNamee, Jeremy Gwinnup, Tim Anderson, and Philipp Koehn. 2018. Freezing Subnetworks to Analyze Domain Adaptation in Neural Machine Translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers*. Association for Computational Linguistics, Brussels, Belgium, pages 124–132. https://doi.org/10.18653/v1/W18-6313

Philipp Koehn, **Huda Khayrallah**, Kenneth Heafield, and Mikel L. Forcada. 2018. Findings of the WMT 2018 Shared Task on Parallel Corpus Filtering. In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*. Association for Computational Linguistics, Belgium, Brussels, pages 726–739. https://doi.org/10.18653/v1/W18-6453

Huda Khayrallah, Hainan Xu, and Philipp Koehn. 2018. The JHU Parallel Corpus Filter-

ing Systems for WMT 2018. In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*. Association for Computational Linguistics, Belgium, Brussels, pages 896–899. https://doi.org/10.18653/v1/W18-6479

Huda Khayrallah and Philipp Koehn. 2018. On the Impact of Various Types of Noise on Neural Machine Translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*. Association for Computational Linguistics, Melbourne, Australia, pages 74–83. https://doi.org/10.18653/v1/W18-2709
[Outstanding Contribution Award] [syllabus]

Huda Khayrallah, Brian Thompson, Kevin Duh, and Philipp Koehn. 2018. Regularized Training Objective for Continued Training for Domain Adaptation in Neural Machine Translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*. Association for Computational Linguistics, Melbourne, Australia, pages 36–44. https://doi.org/10.18653/v1/W18-2705

Steven Shearing,[◇] Christo Kirov, **Huda Khayrallah**, and David Yarowsky. 2018. Improving Low Resource Machine Translation using Morphological Glosses. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)*. Association for Machine Translation in the Americas, pages 132–139. http://aclweb.org/anthology/W18-1813

Huda Khayrallah, Gaurav Kumar, Kevin Duh, Matt Post, and Philipp Koehn. 2017. Neural Lattice Search for Domain Adaptation in Machine Translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Asian Federation of Natural Language Processing, Taipei, Taiwan, pages 20–25. https://www.aclweb.org/anthology/I17-2004

Ryan Cotterell, Ekaterina Vylomova, **Huda Khayrallah**, Christo Kirov, and David Yarowsky. 2017. Paradigm Completion for Derivational Morphology. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, pages 714–720. https://doi.org/10.18653/v1/D17-1074

Shuoyang Ding, **Huda Khayrallah**, Philipp Koehn, Matt Post, Gaurav Kumar, and Kevin Duh. 2017. The JHU Machine Translation Systems for WMT 2017. In *Proceedings of the Second Conference on Machine Translation*. Association for Computational Linguistics, Copenhagen, Denmark, pages 276–282. https://doi.org/10.18653/v1/W17-4724

Biman Gujral, **Huda Khayrallah**, and Philipp Koehn. 2016. Translation of Unknown Words in Low Resource Languages. In *Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA)*

Shuoyang Ding, Kevin Duh, **Huda Khayrallah**, Philipp Koehn, and Matt Post. 2016. The JHU Machine Translation Systems for WMT 2016. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*. Association for Computational Linguistics, Berlin, Germany, pages 272–280. https://doi.org/10.18653/v1/W16-2310

Huda Khayrallah, Sean Trott, and Jerome Feldman. 2015. Natural Language For Human Robot Interaction. In *Proceedings of the Workshop on Human-Robot Teaming at the ACM/IEEE conference on Human-Robot Interaction (HRI)*. http://www.bradhayes.info/hri15/papers/2.pdf

Colleen M. Lewis, **Huda Khayrallah**, and Amy Tsai. 2013. Mining Data from the AP CS a Exam: Patterns, Non-patterns, and Replication Failure. In *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research*. ACM, New York, NY, USA, ICER '13, pages 115–122. https://doi.org/10.1145/2493394.2493415

Preprints

Huda Khayrallah and João Sedoc. 2020. Measuring the 'I don't know' Problem through the Lens of Gricean Quantity. arXiv preprint arXiv:2010.12786v1. https://arxiv.org/abs/2010.12786

Xuan Zhang, Gaurav Kumar, **Huda Khayrallah**, Kenton Murray, Jeremy Gwinnup, Marianna J Martindale, Paul McNamee, Kevin Duh, and Marine Carpuat. 2018. An Empirical Exploration of Curriculum Learning for Neural Machine Translation. arXiv preprint arXiv:1811.00739. http://arxiv.org/abs/1811.00739

Research Experience

Research Assistant August 2015 - present
Johns Hopkins University, Department of Computer Science *Baltimore, MD*
Advisor: Philipp Koehn

- Improving methods for low-resource & domain adapted machine translation:
- Developed Simulated Multiple Reference Training (SMRT), a novel MT training method that improves translation performance by approximating the full space of possible translations for each training example using a paraphraser (Khayrallah et al., 2020).
- Conducted analysis on training machine translation on noisy text (Khayrallah and Koehn, 2018) and co-organized a shared task on parallel corpus filtering (Koehn et al., 2018).
- Lead participant in yearly DARPA LORELEI low resource MT evaluation (2016-2018).
- Participate in WMT Machine Translation evaluations (Ding et al., 2016, 2017; Khayrallah et al., 2018).

Project Advisor: João Sedoc December 2019 - Present

- Improving dialog systems with insights from machine translation:
- Applied SMRT paraphrastic data augmentation for dialog (Khayrallah and Sedoc, 2020).
- Conducting analysis on the evaluation of dialog systems (Khayrallah and Sedoc, 2020).

Project Advisor: Matt Post January 2020 - May 2020

- Lead participant in JHU submission to the Duolingo STAPLE shared task on Simultaneous Translation And Paraphrase for Language Education at the WNGT workshop at ACL 2020.
- Achieved the highest scoring submission in all 5 language pairs (Khayrallah et al., 2020).

Project Advisor: Kevin Duh September 2017 - December 2018

- Improving methods for domain adaptation in Machine Translation:
- Improved continued training for domain adaptation using a regularized training objective (Khayrallah et al., 2018).
- Novel analysis of continued training for domain adaptation (Thompson et al., 2018).
- Curriculum learning for neural machine translation (Zhang et al., 2018).
- Overcoming catastrophic forgetting in continued training using a method for parameter-specific regularization (Thompson et al., 2019).

Project Advisors: Kevin Duh & Matt Post January 2017 - September 2017

- Built a hybrid MT system that leveraged adequacy and fluency from SMT and NMT to improve performance in domain adaptation (Khayrallah et al., 2017).

Project Advisor: Raman Arora October 2015 - November 2016

- Developed Deep Generalized Canonical Correlation Analysis, a multiview learning technique that extends prior work beyond two views of data (Benton et al., 2019).

Research Intern June 2019 - August 2019
Lilt *Berlin, Germany*
Supervisors: Patrick Simianer, Joern Wuebker & John DeNero

- Meta-Learning to improve adaptation for translator-in-the-loop machine translation.

Undergraduate Research Assistant May 2014 - January 2015

International Computer Science Institute *Berkeley, CA*
Supervisor: Jerome Feldman
- Created a natural language interface for a robot simulator based on Embodied Construction Grammar (Khayrallah et al., 2015).

Research Programmer January 2013 - September 2013
International Computer Science Institute *Berkeley, CA*
Supervisor: Andreas Stolcke
- Extended a tool for conversational speech processing & analyzed multi-speaker recordings.

Undergraduate Research Assistant May 2012 - September 2012
UC Berkeley Group in Science & Math Education *Berkeley, CA*
Supervisor: Colleen M. Lewis
- Data mining and analysis of student performance on the AP CS exam (Lewis et al., 2013).

Teaching & Mentoring

Mentor, CS MS student March 2020 - Present
Johns Hopkins University
Professor: Matt Post
- Mentored a University of Maryland, College Park masters student intern (Jacob Bremerman) working on the Duolingo STAPLE shared task on Simultaneous Translation And Paraphrase for Language Education shared task (Khayrallah et al., 2020).
- Mentored the student on his first-author followup workshop paper (Bremerman et al., 2020).

Guest Lecturer, Artificial Intelligence Fall 2019
Johns Hopkins University
Computer Science 601.464/664
Professors: Benjamin Van Durme, João Sedoc
- Gave guest lecture: Bayes Nets Independence.

Teaching Assistant, Machine Translation Fall 2018
Johns Hopkins University
Computer Science 601.468/668
Professor: Philipp Koehn
- Teaching Assistant for ∼60 student graduate/undergraduate course on Machine Translation.
- Gave guest lectures: EM algorithm for IBM model 1; Neural Networks Language Models; Domain Adaptation; Non-Parallel Corpora.
- Held office hours.
- Designed homework assignments.

Guest Lecturer, Artificial Intelligence Fall 2018
Johns Hopkins University
Computer Science 601.464/664
Professor: Benjamin Van Durme
- Gave guest lectures: Bayes Nets Independence; Bayes Nets Inference.

Teaching Assistant, Artificial Intelligence Spring 2018
Johns Hopkins University
Computer Science 601.464/664
Professor: Benjamin Van Durme
- Teaching Assistant for ∼70 student graduate/undergraduate course on Artificial Intelligence.
- Gave guest lectures: Bayes Nets Independence; Bayes Nets Inference.
- Held office hours.
- Designed & graded exams.

Mentor, CS BS/MS student — August 2017 - July 2018
Johns Hopkins University
Professors: David Yarowsky & Philipp Koehn
- Mentored a undergraduate/masters student (Steven Shearing) working on very low resource machine translation.
- Mentored him during two DARPA LORELEI evaluations and during his research work.
- Resulted in the student's first-author publication (Shearing et al., 2018).

Teaching Assistant, Machine Translation — June 2016
Johns Hopkins Summer School on Human Language Technology
Professors: Matt Post & Kevin Duh
- Designed interactive lesson on rule-based machine translation (MT) to teach ∼30 graduate/undergraduate students about MT challenges.

Co-Instructor, Explorations in Robotics — Summer 2013
UC Berkeley Academic Talent Development Program
- Designed and taught 5 weekly 3-hour lessons to 10 middle and high school students about programming and CS Principles in a graphical programming language, along with real world application of computer science.

Grader, Discrete Mathematics and Probability Theory — Summer 2013
UC Berkeley
Computer Science 70
Instructor: Tom Watson
- Graded weekly problem sets.

Lab Assistant, Structure and Interpretation of Computer Programs (CS 1) — Fall 2012
UC Berkeley
Computer Science 61A
Professor: John DeNero
- Assisted students with programming lab assignments.

Lab Assistant, Structure and Interpretation of Computer Programs (CS 1) — Spring 2012
UC Berkeley
Computer Science 61A
Professor: Paul Hilfinger
- Assisted students with programming lab assignments.

Talks

Machine Translation with Diverse Data Sources — September 1, 2019
Venue: NYU Abu Dhabi Computer Science Seminar
Host: Nizar Habash

Machine Translation with Diverse Data Sources — April 8, 2019
Venue: University of Pennsylvania Computational Linguistics Seminar
Host: Chris Callison-Burch

Machine Translation with Diverse Data Sources — March 29, 2019
Venue: Johns Hopkins University Center for Language and Speech Processing Seminar

Continued Training Algorithms — August 9, 2018
Venue: SCALE 2018 Workshop at Johns Hopkins University

Neural Lattice Search for Domain Adaptation in Machine Translation — November 17, 2017
(+ An overview of Machine Translation)

Venue: Johns Hopkins University Center for Language and Speech Processing Seminar

Panel

Machine Translation for LoRes: December 4, 2020
Where do we stand and what the future holds?
Venue: Workshop on Technologies for MT of Low Resource Languages (LoResMT 2020) at AACL

Invited Workshops

EECS Rising Stars 2020
- Invited to attend a professional development workshop for senior female PhD students and post-docs in preparation for pursuing academic careers in EE & CS.
- Hosted by UC Berkeley.

CRA Grad Cohort 2018
- Invited to attend a sponsored professional development workshop for junior female graduate students in computer science.
- Hosted by the Computing Research Associations Committee on Widening Participation in Computing Research (CRA-WP).

Academic Service

Program Committees
ACL 2018, 2020; EMNLP 2018, 2020; WMT 2018-20; IWSLT 2018; NAACL 2019; MT Summit 2019; WiNLP 2019-20; ACL SRW 2020

Local Service

Ph.D. Admissions Committee December 2017 - Present
Johns Hopkins Center For Language and Speech Processing

Organizer May 2016 - Present
Johns Hopkins Machine Translation Reading group
- Coordinate presenters and scheduling for ∼25 faculty and students.

Outreach

Mentor October 2018 - Present
Black in AI
- Run a yearly session on the graduate application process.
- Mentor researcher applying to Ph.D. programs. Suggest programs to consider & provide feedback on application.

Invited Speaker November 12, 2020
Harvey Mudd undergraduate NLP course
- Spoke to students at a liberal arts college about being a graduate student in NLP, and applying to graduate school.

Co-Founder, Webmaster May 2016 - Present
Johns Hopkins Graduate Women in CS & ECE
- Organize a weekly lunch to promote cross-departmental community.
- Plan mentoring events for undergraduate students.

Organizing Committee November 2015 - March 2017
North American Computational Linguistics Olympiad (NACLO) at Johns Hopkins
- Recruited Baltimore area middle and high school students for an international computational linguistics competition.

Advisor, Coordinator September 2011 - May 2015
UC Berkeley CS KickStart

- CS KickStart is a one week workshop for incoming female freshman that creates a community and build students interest and confidence to pursue a CS major.
- Led team of students in planning and fundraising for the $18,000 summer program.
- Mentored new leadership and co-author grant proposals (awarded $10,000 NCWIT grant).

Staff Program Director, Event Coordinator January 2012 - May 2014
UC Berkeley Pioneers in Engineering (PiE)

- PiE runs a robotics competition, provides a robotics kit that's fully designed and developed by college students, and trains college student mentors for teams of students from underserved Bay Area high schools.
- Managed non-profit incorporation, IT systems, resume book and workshop, staff recruitment and training.
- Coordinated 5 all-day events for a $40,000 robotics competition for 300 high school students, 100 college student mentors and 50 college student staff.