Personal statement. Before starting at the University of Massachusetts Amherst, academia was entirely foreign to me. No one had once suggested to me that laboratories were not just confined to television, and lab reports were more than uninspiring high school assignments. What I did begin with is an interest in Latin and an inclination towards understanding how systems worked. In retrospect, these two qualities inspired my college career. My interest in Latin developed into a broader interest in Linguistics, and research gave my an outlet for inquiry. I pursued a path of research experiences in information extraction, natural language processing (NLP), and psycholinguistics; my undergraduate career has helped me achieve significant personal growth and intellectual development. Three years after starting at UMass, as a culmination of research opportunities, mentorship, and encouragement I received, I will present my *first first-author paper at an EMNLP 2018 workshop*.

After writing the acknowledgements section to thank all my previous mentors, I wrote correspondence to each to reflect on how my experience with them helped me produce my work today. Almost every previous research experience contributed to my technical skills or my understanding of the literature, which I will continue to carry forward in producing high quality research papers towards a graduate degree. This work also serves as the basis of my graduate research direction, outlined in the research proposal, on evaluating language generation systems using parsing techniques. A brief outline of my work and how it required my training is outlined below, which is the theme my relevant background section assumes.

Technical skills and outline. The core of my research trains a seq2seq model from sentences in a language like HPSG-grammar, and parses the output with respect to the same grammar to observe the syntactic phenomenon exhibited. This work is among many in the literature with the goal of interpreting neural network models. I gained technical skills required to carry out this large scale parsing in a project described in §1. My interest and exposure to parsing technologies was developed in projects described in §2. The problem of understanding neural language models was first introduced to me during summer research described in §3. The research was conducted during an NSF-REU described in §4.

Intellectual merit. The approach my work takes in evaluating the linguistic ability of neural networks has several advantages over existing methods. Existing methods either evaluate models in limited, artificial language settings to deeply understand the neural model, or evaluate models with real language on constrained linguistic phenomena to gain a shallow understanding. My methodology strikes a balance on this spectrum that is less extreme - we can have a non-shallow understanding of our model on nearly real language. Equally important, the methodology applies to sequence-to-sequence (seq2seq) models, widespread in NLP, while other works do not. Both reviewers of my EMNLP workshop submission mentioned that this was a "nice contribution" to analyzing neural networks for NLP. Followup work outlined in my research proposal aims to translate our analysis insights into a metric to directly facilitate research on seq2seq models applied to language generation.

Broader impact: Research. For a natural language generation (NLG) system, producing grammatical and coherent text is a key requirement to deployment in real world settings. Better understanding of linguistic properties of the output of our natural language generation (NLG) systems can help us improve the quality of output text. Followup work will develop a metric based on our analysis methods that will advance the pace of NLG research and ensure reliable results of automatic evaluations. NLG encompasses tasks such as dialogue, which can be readily applied to educational chatbots. Educational chatbots

performs at the level of traditional online learning, but can reach broader audiences, and be optimized with bandit algorithms to improve student learning outcomes.

Relevant background: §1. Information extraction, UMass (2015-2016). The methodologies developed in my paper requires large scale parsing of seq2seq output. The technical skills to run large scale parsing jobs were acquired in my freshman year, where I worked in Prof. Andrew McCallum's lab in information extraction (IESL). I assisted two graduate students by standardizing bibliographic datasets and implementing baseline disambiguation algorithms. In dealing with 200GB+ bibliography datasets, they introduced me to the research project's codebase and taught me how to utilize the computing clusters to parallelize data processing. I'll never forget seeing how their code compiled and organized results for different configurations of each run. While the work was never published, the technical skills gained allowed me to work on complex, computationally heavy projects.

In my time at IESL, I also worked on two other projects to hone my technical skills. I led the development of crowdsourcing methodology to create a dataset of over 8000+ academics with over 90%+ accuracy, which was *presented at the UMass undergraduate research conference*. This dataset will be integrated with OpenReview conference management system to reduce conflict of interests in peer reviewing. Our lab also participated in the 2016 TAC-KBP shared task, where I built a distributed system to collect search results at scale to supervise a relation extraction system. Our system gains 2.4% in F1 (11.7% performance increase from 2015) *placing as a top team without labor-intensive active learning methods*.

§2. NLP for dialetical English, UMass (2016-2017). Central to my methodology is parsing. My interest in parsing was developed through research on parsing technology for dialectal English with Prof. Brendan O'Connor. I annotated 500+ tweets, 250 of which exhibited the African American English (AAE) dialect, with their dependency parse trees. Next, I conducted experiments on inter-annotator agreement and communicated annotation challenges and guidelines. This treebank forms the basis of a *paper published in ACL 2018*. With this dataset we saw a 5.9% to 15.7% labeled attachment score disparity between Mainstream American English tweets and AAE tweets using off-the-shelf parsers; with training on half of our dataset and the use of Twitter embeddings we saw a jump in 11.7% LAS. This work highlights a shortcoming of a technology performing disparately on languages associated with minority groups. Seeing tweets and their dependency structures struck a lasting impression on me that deeper structure underlies the surface forms of all valid language. In my work, this deep structure is of interest, and parsing continues to play a large role.

In work leading up to parsing, I also worked in language identification, a common first step in pipelines analyzing social media data. I annotated and double-checked a corpus of 10,000 tweets of whether they were English, accounting for code switching and other phenomenon. This dataset forms the basis of a *paper published in a EMNLP 2017 workshop, which won a best paper award*. We found that off-the-shelf language identification models have lackluster performance on Twitter data (88% recall). This work hopes to highlight concerns in using Twitter for English tasks, as dialectical English strongly associated with demographic groups may be ignored in tasks such as opinion mining.

§3. Interpreting neural networks, UMass and UW (2017). Literature on analyzing black box models was exposed to me through my research experience. My exposure to techniques in this research direction was first developed in studying psycholinguistics. As a research assistant under Prof. Brian Dillon's lab, eyetracking was the primary method used

to gather evidence for theories on human processing of reflexives. I ran over 30+ hours of eyetracking studies and attended weekly lab meetings. Independently, I formulated research on the Chinese *ziji* reflexive and collected human judgments. The results were inconclusive, but my abstract was *accepted to an undergraduate Linguistics conference in Montreal*.

Later, when I was a visiting undergraduate at Prof. Noah Smith's ARK lab at the University of Washington, my experience in psycholinguistics prompted me to work with several postdocs on understanding representations learned from neural networks. I was given freedom to research anything, and I decided to begin with an exploration of how LSTMs could resolve postfix expressions (e.g. "4 5 3 * +" evaluates to 19). This task was particularly interesting to me because we could manipulate expressions to examine the model capability in different aspects. Several papers that I cited in my recent work were exposed to me during this summer. While we did not produce any publishable work, the ideas on using grammars to evaluate grammars stuck with me, and was truly the inspiration for my current research.

§4. Current research, Stanford CSLI (2018). Attending an NSF-REU at the Center for Study of Language and Information (CSLI) at Stanford facilitated the environment to conduct the research in my EMNLP workshop submission. My roommate, another REU intern, worked alongside me day and night, and is the second author of the work. Coincidentally, Dan Flickinger, the developer of the HPSG-grammar used in my work, is faculty at CSLI. His comments on my work inspired my followup research direction. This summer, by writing my own paper, I experienced all that comes with leading my own research project that I didn't experience when I was an assistant. Mostly on my own, I learned how to formulate a research question, apply methodology, analyze data, and communicate results. This work demanded many skills from my experience in relevant research projects prior. I am grateful for all those who have led me here today. During this time, I also applied bandits to educational chatbots, and *a paper is currently under submission to CHI'19*.

Broader impact: Outreach. Outside of my direct research goals, I have worked in jobs helping other undergraduates with research and academics. My motivation is in part due to an undergraduate I met at UW who encouraged me to write my own paper, when I was having trouble beginning independent research in NLP. I've personally seen many fellow undergraduates capable of making non-trivial contributions, but do not end up pursuing research because getting involved is hard. However, with the right mentorship, these students could fulfill part of the rising demands for computer science researchers. These reasons lead me to believe that encouraging undergraduates research is crucial, and can be effective.

To support undergraduates, I was involved with several campus jobs throughout my time at UMass. My first experience was as a residential peer mentor, to support first year students transition into collegiate academic life in the dorms. My second job was as a teaching assistant for an undergraduate course in programming methodology. I hosted weekly discussion sections and office hours. My third and current job is as a research mentor in our campus's office of undergraduate research (OURS). I have advised 10+ students primarily on beginning a research career. My goal is for CS undergraduates to begin independent research projects. Future goals I have with OURS include surveying CS undergraduate research involvement and writing comprehensive tutorials to independent research projects.

Career goals. My immediate goal is to pursue a graduate degree, to eventually become faculty. As faculty, I hope to focus on scientific outreach to undergraduates, both through involvement in my own research and through university outreach programs.