My research interests are in natural language processing (NLP), particularly in its intersections with machine learning and linguistics. These interests have drawn me toward two main research directions:

- **Linguistics for NLP**: How can we best use linguistic domain knowledge to improve the interpretability and generalizability of NLP systems?
- **Computational Linguistics**: What can learning systems tell us about human language?

I have been working with Professor Noah Smith from the onset of my first year at the University of Washington, and I have been fortunate to conduct some initial work in these directions. These initial steps and my future plans are outlined below.

Linguistics for NLP The lack of interpretability and generalizability in state-of-the-art models hampers our ability to ethically deploy equitable NLP systems for broader impact upon society.

I am excited about overcoming these limitations by using linguistic structure to augment neural models with useful inductive biases. Linguistic structure underlies all language data, but the majority of neural NLP models do not explicitly consider this unifying feature. On the contrary, modern neural architectures are unconstrained by design; these overparameterized models overfit to their training datasets. Reining in our models with appropriate linguistic biases will aid generalization across inputs and may also improve interpretability, perhaps through the generation of intermediary structures.

My impetus for pursuing this research direction comes from experiences during an research internship at the Allen Institute for Artificial Intelligence, where I worked with Dr. Matt Gardner on improving science exam question answering systems by transferring knowledge from a larger, out-of-domain question answering dataset (SQuAD). This did not work very well; neural models are brittle, and models with "superhuman" performance on a particular data domain fail to generalize to others. Furthermore, it was difficult to decipher what knowledge these neural models were missing, let alone find any principled ways of improving them.

My firsthand experiences with the opacity of modern NLP models motivated my subsequent research on better understanding recurrent neural networks (RNNs), which form the backbone of almost all state-of-the-art NLP systems. In our paper at the ACL 2018 Workshop on Representation Learning for NLP (Liu et al., 2018b), which received a best paper award, we noted that although RNNs have seen great success on language-based tasks, they are inherently models of arbitrary sequential data—perhaps properties of natural language in particular augment them in some way? Through a carefully-controlled synthetic data experiment, I found that RNNs exploit linguistic attributes of data, even when trained on inherently non-linguistic tasks, and that these features improve model generalization.

To further understand how RNNs encode information, and as a necessary initial step in my proposed research in integrating linguistic structure and model structure, I am working to identify what linguistic knowledge is implicitly captured by current unstructured NLP systems. In a paper to be submitted to NAACL 2019 (Liu et al., 2019), I led a project that seeks to broadly characterize what linguistic information is encoded by several recent models for contextualized word representation. Our findings indicate that contextualized word vectors fail to capture discourse phenomena or coreference information, which suggests that explicit discourse structures or notions of entities could be promising avenues for augmentation. I plan to continue this line of analysis to further understand the empirical abilities and limitations of other core components of neural NLP models, such as RNNs and the recently popular self-attentive encoders (Vaswani et al., 2017).

When adding structure and domain knowledge to models, I intend to build on initial work that augments model architectures with explicit notions of linguistic syntactic structure (recurrent neural network grammars; Dyer et al., 2016) and compositionality (neural module networks; Andreas et al., 2016). I am also interested in developing new ways of evaluating how well models generalize to better understand their strengths and weaknesses when deployed in the real world. Through this direction, I hope to further explore my nascent interests in algorithmic fairness. For instance, I will develop evaluations that assess the ability of models (typically trained on newswire text) to generalize to dialects such as African-American English.

Computational Linguistics I am also interested in exploring how learning models can help us test linguistic hypotheses and learn more about language itself, a research direction sparked by my linguistics coursework. In my introductory linguistics course, I learned about the arbitrariness of the sign, which is the notion that a word's phonetic and orthographic forms have no relation to its meaning. Phonesthemes, which are noncompositional, submorphemic phonetic units that consistently occur in words with similar meanings, were mentioned as an interesting exception to this principle.

To better understand this exception to the arbitrariness of language, I developed a method for inducing phonesthemes from text (Liu et al., 2018a). I framed this as a feature selection problem by training a sparsely-regularized linear model to identify character n-grams that are most predictive of word vectors—these are likely phonesthemes. Phonesthemes had previously only been proposed by linguists, and our method discovered many of these previously proposed clusters and even novel ones.

I believe that neural models can be a powerful tool for testing linguistic hypotheses (e.g., Kuncoro et al., 2017). One direction that I am excited about is investigating the Poverty of the Stimulus Argument, which argues that the linguistic input received by young children is insufficient for acquisition of every feature of their native language. The Poverty of the Stimulus Argument is often used to reinforce the theory of a strong Universal Grammar, which claims that all humans share particular language universals. Examining how performance on a grammaticality judgment task changes as we add stronger linguistic inductive biases to our neural models may lead to new insights into the Poverty of the Stimulus Argument.

Future Plans My career aspiration is to become a professor, since an academic career offers unique opportunities to mentor students through teaching and research advising. This choice is particularly informed by my positive experiences as an undergraduate teaching assistant. As a TA for the NLP course, I led a discussion section and held office hours, while my duties as a TA for the NLP capstone involved advising student teams on the design and implementation of original NLP projects; I find that I enjoy both teaching and research advising. Pursuing a Ph.D. will enable me to continue my research while also gaining further teaching experience.

At Stanford, I am especially interested in the work of Professors Christopher Manning, Percy Liang, and Dan Jurafsky. I also appreciate that the Stanford NLP group includes students in both computer science and linguistics—this tight integration is unique strength. Following the work of these groups has led me to see a clear fit for my skills and interests at Stanford, and I am confident that it is a great place for me to pursue a Ph.D.

Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. Learning to compose neural networks for question answering. In *Proc. of NAACL*, 2016.

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. Recurrent neural network grammars. In *Proc. of NAACL*, 2016.

Adhiguna Kuncoro, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, and Noah A. Smith. What do recurrent neural network grammars learn about syntax? In *Proc. of EACL*, 2017.

Nelson F. Liu, Gina-Anne Levow, and Noah A. Smith. Discovering phonesthemes with sparse regularization. In *Proc. of the Second Workshop on Subword and Character Level Models in NLP*, 2018a.

Nelson F. Liu, Omer Levy, Roy Schwartz, Chenhao Tan, and Noah A. Smith. LSTMs exploit linguistic attributes of data. In *Proc. of the 3rd Workshop on Representation Learning for NLP*, 2018b.

Nelson F. Liu, Yonatan Belinkov, Matt Gardner, and Noah A. Smith. On the transferability of language model representations. In *Proc. of NAACL*, 2019. **Currently Under Review**.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proc. of NIPS*, 2017.