Graduate Research Plan Statement

I want to teach machines to learn to read poetry. Poetry highlights technical difficulties in natural language processing as it relies on sophisticated language (e.g. connotation, metaphor, and reference) and less structured syntax. But what truly differs in the approach of having machines read literary art rather than newspaper articles is that art is not an attempt to communicate information but rather an attempt to communicate emotion. Leo Tolstoy explains this better:

"To evoke in oneself a feeling one has once experienced, and having evoked it in oneself, then by means of movements, lines, colors, sounds, or forms expressed in words, so to transmit that feeling that others may experience the same feeling - this is the activity of art." [1]

Models that structure information about the world are very different from models that could structure affect. Current sentiment analysis focuses on interpretting how a person felt when they wrote a sentence: 'My room is as cold as the South Pole!' should be labeled as having negative affect because that's the typical human response to the fact. In contrast, the emotion that poetry expresses cannot be determined by correctly interpreting factual statements. Instead it is the result of cognitive leaps between subtle observations. I propose that thematic emotions can be modeled using poetry as a starting point.

In addition to its complex use of language and goal of communicating emotion, a poem has multiple interpretations; its meaning varies not only among its audience members but even individual readers' interpretations change with subsequent readings. Some of these interpretations may be difficult to access without a sympathetic experience or literary training, but no one, not even the poet, accesses all possible interpretations. All natural language has this feature of multiple accessible meanings and for machines to build language models that can provide useful tools and interpretations back to people we must build algorithms that not only account for variability in interpretation but utilize it.

Work done on connotation frames at the University of Washington begins to explore complex language use and variable emotional response. They introduce connotation frames as a formalism to organize how a single predicate verb can have different implied sentiments based on the perspectives, value, effects and mental state[2]. By breaking out interpretation of a single word into multiple frames they have begun to address the necessity of context, not the kind of context that common sense knowledge of a world would provide but rather the context a unique perspective brings. Currently this formalism relies on an a priori set of possible connotation frames. The next step could be to learn connotation frames from a specific corpus, similar to how humans bring their own context to reading a text based on their unique experiences. Another direction would be to apply it to more parts of speech or work on a model that could combine connotation frames across words. Eventually a connotation frame for a sentence should be influenced by the context of the document in which it occurs. Modeling how frames interact is a large, but necessary, step.

Work on generating creative language, such as that from the University of California at Berkeley on using intersecting word vectors to generate figurative language[3], also moves towards a subtle reading of language. Machine intepretation of metaphor especially has long

relied on a human-generated corpus of metaphors which will have to be constantly updated to keep up with the progression of language[4] and inherently cannot read texts that use novel language. Generative systems, though still in their infancy for figurative language, allow us to create a figurative language corpus based on current natural language usage, as opposed to requiring human annotation. With a figurative language corpus in hand, we can better access sophisticated language use and expand work on contextual frames.

As a first step towards consolidating these concepts I plan to develop a poetry recommendation engine. A recommendation engine would require machines to read poems and generate meaningful models that relate poems in the corpus. It would then provide a way to evaluate the effectiveness of the model through human interaction.

This work would not only require highly technical understanding and implementation of current artificial intelligence techniques but would have an impact on many avenues of natural language processing. Sentiment analysis would be furthered by a focus on the direct communication of emotions, as opposed to emotions as the reaction to facts; research on sarcasm[4] has demonstrated that more work on context is necessary to accurately interpret sentiment. Similarly machine translation would benefit from better interpretation of figurative language. Speech-based human computer interfaces will require more contextually-aware interpretation of natural language to move beyond basic commands.

The impact of using poetry as a model example of machine interpretation of natural language is wide-reaching. The immediate benefit is to make poetry more accessible to people: sophisticated recommendation engines should also be applied to the arts. But considering how machines will interpret art also encourages us to ask larger questions about how robust intelligence should interact with people in emotionally complex situations. What kind of models do we want machines to build of poetry? The answer to this question, which ultimately is the model we create machines to build and is the work I want to pursue, is the starting point for understanding how machines will exist in this world.

[1] Tolstoy, L. 1897. What is Art?

[2] H. Rashkin, S. Singh and Y. Choi. 2016. Connotation Frames: A Data-Driven Investigation. *Association for Computational Linguistics.*

[3] A. Gagliano, E. Paul, K. Booten and M. A. Hearst. 2016. Intersecting Word Vectors to Take Figurative Language to New Heights. *Workshop on Computational Linguistics for Literature, NAACL.*

[4] E. Shutova. 2010. Models of Metaphor in NLP. *Association for Computational Linguistics.*

[5] D. Bamman and N. A. Smith. 2015. Contextualized Sarcasm Detection on Twitter. *AAAI Conference on Web and Social Media.*