Introduction. My second go-to fun fact, just behind my being an archery instructor, is that I studied classics and computer science in college. I always like telling people this fact because the combination never fails to surprise them, even though it has always made sense to me. As a college student at Princeton University, working across disciplines gave me the opportunity to connect with a variety of people and form bridges between seemingly unrelated fields. Now, as a PhD student at Carnegie Mellon's Language Technologies Institute, I am looking forward to combining my background in these disciplines with my experiences teaching and working in industry in order to solve open problems in natural language processing. I plan to focus my research on language and stylistics, particularly on developing computational approaches to combat gender bias.

My interest in studying language traces back to my first Latin class over 10 years ago. Since then, a few memorable instances exemplify what I find fascinating about implicit meaning and bias. While an undergraduate, I took a class on rhetoric during the Obama-Romney Presidential election. As part of the class, we would discuss the presidential debates. One unforgettable expression emerged from the second debate, when Mitt Romney, attempting to demonstrate his commitment to gender equality, referred to female job candidates as "binders full of women". This phrase became so infamous that there is an entire Wikipedia page dedicated to it, and my professor later speculated it marked the moment Romney lost the election. I find the power of language fascinating, particularly how easy it is to utter a comment without fully realizing the effect it may have on the listener. I believe computer science has the ability to shape how we use language by instantaneously showing us our comments in a larger social context before we say them. Not everyone has the luxury of a communications team dedicated to crafting a statement that perfectly expresses our meaning. Instead, we can build NLP tools to analyze large volumes of text more accurately and efficiently than we can do by hand. While such tools may not have been able to save Mitt Romney in the middle of a live debate, they can affect our everyday activities by making us more conscientious users of language.

Research Experience. I first sought to combine my interest in language and computer science when I worked on analyzing the authorship of Xenophon's *Cyropaedia*, an ancient Greek text. Some scholars suspect that the last chapter of the *Cyropaedia* was not written by Xenophon but instead was appended to his work later on. The project involved counting frequencies of N-grams in order to compare the last chapter to the rest of the work. This method is fairly well established for authorship attribution, but, to my surprise, it had not previously been applied to classical texts. Arvind Narayanan, a professor in the Computer Science Department, primarily advised my research, but I also worked closely with Michael Flower, a professor in the Classics Department who was considering arguing that Xenophon did not write the last chapter. My results, which suggest that the chapter has a similar style to the rest of the work, caused Professor Flower to reevaluate his approach to analyzing the text.

As a second project during junior year, I worked with Ed Felten on analyzing the traces that programs leave behind in a computer's main memory. My goal was to evaluate the damage potential of an attack that reads main memory. The project allowed me to explore a new field and become more comfortable with Linux tools, but ultimately I returned to the study of language for my undergraduate senior thesis.

My undergraduate thesis focused on an automated analysis of grammar usage in Latin. The driving premise behind the project was that when authors have a choice of grammar constructions they show personal preferences, making syntax at least as good an indicator of style as lexical metrics. My primary advisor was Christiane Fellbaum, an experienced

computational linguist who helped me formulize my ideas into concrete approaches. I also consulted with Yelena Baraz, a classics professor who specializes in Latin prose. One of the main problems I encountered was a lack of annotated training data, which made it difficult to use machine learning algorithms typical of stylistic analysis. Instead, my background in Latin allowed me to target specific grammar constructions by combining probabilistic tagging with a rule-based approach. I was then able to confirm my results by comparing them with hand-analyses performed by classicists over the last several centuries. Some of the trends I observed include similar styles between histories and military reports, and inconsistencies between the last chapter of Caesar's *Gallic Wars* and the rest of the work, supporting the widely accepted theory that the last chapter was not written by Caesar.

 Although I enjoyed conducting research, following graduation, I began working as a software developer at a finance company; however, it was not long before I found myself turning back to research. Towards the end of my first year in industry, I began revisiting my undergraduate thesis and expanding some of the analyses. I would look forward to the end of the workday not merely because it signified dinnertime, but because I could spend the rest of the evening doing research. Eventually, I submitted a revised version of my thesis to *Digital Classics Online*, where it was accepted for publication in December 2016. The process taught me to formalize my writing and revise my work in accordance with feedback from reviewers.

Industry Experience. I worked as a software developer at a trading firm for two years after college primarily to gain experience and improve my programming skills. My role there also taught me to think quickly about unfamiliar topics. Although I had a background in computer science, I had no background in finance, so understanding the company's operations was a constant balance between knowing when to ask for help and when to study topics on my own. I developed the ability to self-teach new approaches and to determine the right questions to ask, essential skills for conducting research.

 The experience also showed me how theoretical social issues play out in practice. When I first started, I was one of two women on a team of 25 developers. Unsurprisingly, my company sent me as a representative to numerous college career fairs, and within a few weeks I began conducting half of all first and second round interviews. My role as an interviewer led to numerous conversations with the head of the software team about how to interview effectively and why the company had so few women. One point we emphasized was the need to interview more female candidates than male candidates in order to combat the unintentional bias inherent in the interview process. I had similar conversations with friends at other companies, where HR would discount interviewers' reports if they contained flagged language, such as using "aggressive" to describe women. As my company hired more developers, I began serving as a mentor for the new women, talking them through the company's systems and planning social events. By the time I left, we had hired 6 new female software developers, but the question of how to combat unintentional bias remained with me.

 After leaving my company, for several months I worked full time at an archery range where I had been working part time for the prior two years. As an archery instructor, I met people from a variety of backgrounds, not only in the customers I coached, but also in my fellow staff members. My past archery experience helped me rebuild the range's youth programs and create a web application to track my trainees' progress. I had not exactly imagined spending my last summer before graduate school teaching myself JavaScript, but I enjoyed the chance to direct the entire project, while coordinating with other coaches. I believe my time spent working

as a software developer and an archery instructor has helped me broaden my horizons and develop multi-faceted problem solving skills, which directly translate to conducting research. **Intellectual Merit.** While working in industry has helped me improve my skill set, my combined education in classics and computer science provides me with a unique background for studying NLP. I can approach research questions with knowledge of driving language theories and also an understanding of the abilities and limitations of technical approaches. My research as an undergraduate has given me experience in carrying out self-directed projects, while consulting frequently with computer scientists and linguists. Now as a PhD student at CMU, I will be working with Yulia Tsvetkov, who has broad experience in NLP and machine learning, including applying computer science to social issues. I expect to interact frequently with numerous experts in the field, allowing me to learn rapidly and further my research goals. **Broader Impact/Future Goals.** My own experience working in industry has encouraged me to focus my research on questions that have real world applications. I believe that NLP research can be used to address social issues, of which gender bias is only one example. Furthermore, I view sharing findings not just with experts in the field, but also with students and non-experts, an essential part of a successful project. As someone who has worked across disciplines I am used to having technical conversations with various communities. Additionally, my background in teaching and working one-on-one with individuals will help me present my ideas at conferences, work with students, and share findings with policymakers and others outside the field.

Outreach has always been very important to me, not just in the course of my research but also in other venues. As an undergraduate, I participated in the Petey Green prison-tutoring program, which involved going to local prisons once a week to help inmates study for the GED. The program could be challenging: some of the students spoke little English and my high school Spanish classes did not exactly cover fractions and decimals, but I found the experience incredibly rewarding. By junior year I became a Leader Trainer for Princeton's Outdoor Action program, and I taught classes to underclassmen on wilderness survival skills, including leading "training trips", during which trainees would practice their newly acquired skills out in the wilderness.

Finally, I have spent many years coaching archery, not only as an instructor in New York, but beginning back in high school. While in college, I founded the Princeton Archery Team. I considered simply practicing on my own without forming an entire team, but I quickly realized that for me, one of the most important aspects of the sport is sharing it with others. I spent much of my sophomore year writing numerous proposals for funding and coming up with creative solutions to allow the team to exist within university regulations. For example, since the university would not allow the team to own bows or arrows, I structured our grant for equipment so that all of the gear was technically owned by the US Collegiate Archery Association, but permanently loaned to our team. Ultimately, my desire to share the sport with other students motivated me to create a sanctioned team and help teammates train for competitions.

After finishing graduate school, I hope to pursue a career in academia, where I can work closely with students. One of my main motivations in leaving industry for pursuing a PhD is to work on higher impact projects, which can not only advance the field of NLP, but also inspire future students, especially young women in computer science. I plan to also engage in community outreach, helping to bring an understanding of computer science and language processing to populations that may not have access to training in these areas. I look forward to graduate school as a chance to draw on my past experience, learn from my present surroundings, and pursue meaningful projects.