Overview of NLP

a) Natural Language Processing is a discipline of computer science that focuses on the intersection of linguistics and technology. Specifically, how to develop programs that can parse and understand human language both written and spoken. More recently, NLP is using machine learning algorithms to analyze and thusly generate text.

b) NLP can be considered a branch of the study of Artificial Intelligence. Artificial Intelligence is a much broader term that encompasses enabling computers to mimic human capabilities. For example, self-driving cars or making predictions based on previous knowledge. NLP is specifically the language aspect of artificial intelligence.

c) Natural language understanding is the principle of computers being able to comprehend language while natural language generation is computers being able to write text.

d) Some applications of NLP include predictive text/emails, sentiment analysis, conversational chatbots, classifying text, and speech recognition.

e) The rules-based approach was the first NLP approach that came about in the 1960s. These rules consist of formal grammar or context free grammar rules that can be applied in order to classify texts in a structural linguistic manner. Regular expressions and basic conditional statements are also used. Examples include spell check and IBM's Eliza. However, rules-based is disadvantageous as you scale up because natural language is much more complex and nuanced than rules can cover.

Statistical and probabilistic approaches to NLP are the development of language models that use probability to predict and learn. The model requires a large corpus of data for training in order to analyze things like word frequencies and the probability of a word based on the words preceding. These approaches utilize traditional machine learning algorithms. Examples include programs like Natural Language Toolkit and spaCy. One drawback is that this approach takes a moderate amount of data and processing power.

Deep Learning in NLP is the newest approach delivering sophisticated language translation, generation, and understanding. It is based in more advanced machine learning techniques some of which include Neural Networks, Reinforcement Learning, and Deep Generative Models. These methods learn and analyze by undergoing multiple phases of processing and using complex algorithms on very large amounts of text data. This can be considered a drawback because of the amount of processing power needed in addition to the hype surrounding it. Deep learning is fast growing but it is not at a point where it can be considered Artificial General Intelligence.

f) My interest in NLP stems from my love for language. The intricacy of human language is beautiful to me and I am intrigued by the idea of a computer being able to capture that complexity. A personal project using spaCy to compare word similarity and generate text based on an input word kickstarted my drive to learn more. I would like to learn more about not just how to implement a NLP library and use its methods, but the mathematics and logic behind it. I hope to be able to use this knowledge in ways that highlight the intersection of tech and language.