Natural Language Processing (NLP) enables machines to mimic the linguistics of humans as closely as possible. It is a branch of artificial intelligence (AI) that aims to communicate with humans effectively.

Components of NLP include natural language understanding (NLU) and natural language generation (NLG). NLU analyzes the grammar of a sentence and the meaning behind the sentence. Many words can be used accurately but have different meanings, so NLU is where the machine gains reading comprehension to get the correct context of the sentence. Meanwhile, NLG is where the computer generates text in the correct context. It plans which words and punctuation to use as well as follows the grammar rules to make it seem like a human is talking.

Some examples of NLP applications include voice assistants like Siri and Alexa, autocomplete text to predict what a user will type, and email filters to determine spam and what kind of email is it such as social or promotions.

The three approaches to NLP are rules-based, statistical and probabilistic, and deep learning. Rule-based is when the machine follows grammar rules in a language such as context-free grammar. Context-free grammar generates strings that can be used to form a sentence. This approach has limitations due to the complexity of human language.

The next approach is statistical and probabilistic. Statistics started being used in language models to form sentences. Examples include calculating the probability of the next word and finding the number of total words used. Other statistical approaches include Markov independence assumptions, calculating conditional probabilities, and smoothing techniques to reduce outliers in data.

The last and most recent approach is deep learning. Recurrent neural networks (RNN) are one of the approaches to deep learning, where they're useful for processing data sequentially. They have feedback loops to remember previous data and make copies to learn and transmit information to the next copy, creating a sequential order. This aligns with processing words in a humanlike sentence. Long Short-Term Memory is a type of RNN where it improves retaining data longer. A model where it predicts captions for a dataset of images is an example of utilizing deep learning in NLP.

I have previously taken CS 4375 (Machine Learning) and CS 4372 (Data Science) this past semester, and both classes covered NLP. I became interested in learning more about NLP when I had to do projects such as predicting captions from images and doing sentiment analysis on sentences to determine if it has a positive or negative tone. I would like to do more projects that cover NLP to increase my understanding of the topic.