Meghana Kambhampati
MXK190048
CS4395

Overview of NLP

a) Natural language processing (NLP) describes a field of study in which computer programs are written to allow for the processing of human languages. With the help of various algorithms, computers can recognize human languages, spoken or written, and use that input to complete a task.

b) NLP is a subset of artificial intelligence (AI). Many language processing projects will involve other aspects of AI that are not purely part of NLP. Machine learning is another branch of AI.

c) Natural language understanding involves interpreting human language inputs, while natural language generation is the creation and production of human language responses. Both require the processing of natural language through some kind of algorithm.

d) Language translators like Google Translate, email filters, autocomplete, and voice assistants are all modern examples of NLP. These tools use NLP concepts to complete specific tasks.

e) NLP is an evolving field with many techniques. There are three main approaches to NLP: rules-based, statistical and probabilistic based, and deep learning. These techniques are often used together for NLP projects.

Rules-based techniques, the oldest approach to NLP, focuses mainly on a set of regular expressions and grammar rules. This includes context free grammars and spell-checking software. Another example of the rules-based approach is Eliza, an early chatbot. This approach is hard to scale up, however, as human languages are extremely complex.

Statistical and probabilistic approaches, popularized in the 1980s, involved a more mathematical approach to NLP. With this approach, words and sequences of words are counted. This data is then used to work out the probability of word combinations. This model is often used for language models, predictive text, and is the basis for traditional machine learning algorithms. For the statistical and probabilistic approach to be useful, a moderate amount of data and processing power is required.

Deep learning is a product of increased access to data and high processing power. Deep learning is an iteration of neural networks. The growth of this technique saw advancements in language translation, generation, and understanding. There is a lot of hype surrounding this approach that may or may not be warranted. In addition, deep learning requires massive amounts of data and processing power to produce results.

f) I am interested in both computer science and languages, so taking this course seemed like a good choice. I have also always wondered how computers process human languages. Google Translate and autocomplete seemed like magic to me when I was younger. NLP makes up so much of our modern day conveniences, and I really want to understand it better. Once I

get a basic understanding of how NLP works, I'll be able to do my own research for personal projects.