ITAI 2373 – Natural Language Processing

Student: Natalia Solorzano W207818526

Professor: Narges DeBary

Spring 2025 Course Number: 16579

A07: NLP field is the part-of-speech (POS) tagging topic.

Part-of-Speech (POS) tagging is a foundational technique in Natural Language Processing (NLP) used to assign grammatical labels—such as noun, verb, adjective, or adverb—to each word in a sentence. The purpose of POS tagging is to help computers understand the syntactic role of each word in context, enabling more advanced tasks like parsing, semantic analysis, question answering, and machine translation.

There are several types of POS tagging approaches. Rule-based POS taggers apply a set of hand-written rules to identify the appropriate tags. Statistical models such as Hidden Markov Models (HMM) and Conditional Random Fields (CRF) analyze the probability of sequences to determine tags. More recent approaches use machine learning and deep learning to tag text accurately by learning patterns from annotated corpora.

One of the most used tools for POS tagging in Python is the Natural Language Toolkit (NLTK). It provides access to pre-trained POS taggers and corpora like the Penn Treebank. The `nltk.pos_tag()` function, for example, takes a tokenized sentence and returns a list of word-tag pairs. This function uses the Averaged Perceptron Tagger by default.

POS tagging is especially valuable in applications such as sentiment analysis and machine translation, where the role of a word influences meaning. For instance, in the sentence 'They can fish,' the word 'can' can be a verb or a noun depending on context, and POS tagging helps in disambiguating such usage.

While POS tagging is an essential building block in NLP, it does have limitations. Errors may occur with ambiguous words, and taggers may struggle with informal or ungrammatical text. Despite these challenges, POS tagging remains a critical component in the NLP pipeline.

References

- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.

- Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing (3rd ed.). Stanford University.

- NLTK Documentation: https://www.nltk.org/