1. Ellipsis (…) is tagged as comma (,) because there are only 9 punctuations that are well tagged in collection of pos tags, and among them are #, $, (,), closing quote, opening quote, period, colon, and comma. The lack of tags for special characters may affect the pipeline’s functioning, for example: adding ellipsis in a sentence to emphasize the hesitation of the subject can cause the position change of dependency parse of punctuation; however, the extent of severeness seems to be trivial in case we don’t concentrate on punctuations much.

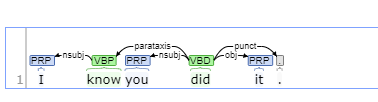


Figure 1. dependencies of sentence: “I know you did it” (https://corenlp.run/)

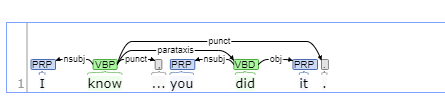


Figure 2. dependencies of sentence: “I know … you did it” (https://corenlp.run/)

1. NLP pipeline can identify an adverb as a determiner plus an adjective in some cases due to irregular structure, e.g., “a little clearer” has a structure of adverb plus an adjective, but after it goes through our NLP pipeline, it is tagged as a determiner followed by two adjectives. This situation is caused by the mandatory execution order of splitting and tagging. That is, we have to split our sentence first before we can assign pos tags to tokens when using CoreNLP, but once the words are separated, the original meaning may distort in the NLP pipeline.