

# Sept 14 Prep: Parts of Speech and Tagging

CC And PRP I VBP have TO to VB tell PRP you PRP I VBP 'm RB here , , CC and RB very RB strongly RB here

PRP\$ Our NNS allies VBP are RB not VBG paying PRP\$ their JJ fair NN share , , CC and PRP I VBP have VBN been VBG talking IN about DT this RB recently DT a NN lot . .

DT The JJ amazing NN thing VBZ is IN that DT the NNS politicians IN in DT this NN country VBP live IN by DT a JJ whole JJ different NN standard WRB when PRP it VBZ comes TO to NN health NN care . .

The above 3 sentences have been chosen from the collection to test the part of speech tagger. All the above have been selected to test different lexical and syntactical structures.

After analyzing the result, I was able to identify a few incorrectly tagged tokens. However, it seems that the issue occurs when certain words can be identified and categorized under multiple parts of speech. I have highlighted above the tokens, I believe have been incorrectly tagged and have provided the correct parts of speech below:

Token	Incorrect Tag	Correct Tag
To tell	To	Infinitive verb
This	Determiner	Pronoun
A lot	Determiner	Adverb
That	Conjunction	Pronoun
This	Determiner	Adjective
To	To	Preposition

It was interesting to note that during my process I found the text to contain multiple tokens of the form I've, I'm, I'll, etc. These tokens need to be handled properly based on the root verb form.

