MIS: 111708049

Corpus used : Brown Corpus
Total Tagged sentences: 10000 (Due to CPU constraints)
Tagset: ['.', 'ADJ', 'ADP', 'ADV', 'CONJ', 'DET', 'NOUN', 'NUM', 'PRON', 'PRT', 'VERB', 'X']
Number of training sentences: 7500
Number of testing sentences: 2500

Features for each word in sentence:

1. word
2. is_numeric(is a number)
3. is_joint (contains hyphen)
4. is_first(is start of sentence)
5. is_last(is last of sentence)
6. is_capitalized(is first letter capital - mostly Proper Nouns)
7. is_all_caps(is an abbreviation)
8. is_all_lower(No capital letter)
9. capitals_inside(has capital letters in the word)
10. prefix1(first letter of word)
11. prefix2(first two letters of word)
12. prefix3(first three letters of word)
13. prefix4(first four letters of word)
14. suffix1(last one letter of word)
15. suffix2(last two letters of word)
16. suffix3(last three letters of word)
17. suffix4(last four letters of word)
18. prev_word(Previous word)
19. prev_prev_word(second previous word)
20. next_word(Next word)
21. next_next_word(second next word)
22. prevToprev_prev(Bigram of second previous word and previous word)
23. prev_curr(Bigram of previous and current word)
24. curr_next(Bigram of current and next word)
25. next_nextTonext(Bigram of next and second next word)
26. prevToprev_prev_curr(Trigram of second previous, previous, current word)
27. prev_curr_next(Trigram of previous, current, next word)
28. curr_next_nextTonext(Trigram of current, next, second next word)
29. prevToprev_prev_curr_next(4gram of second previous, previous, current, next word)
30. prev_curr_next_nextTonext(4gram of previous, current, next, second next word)
31. context(5gram of second previous, previous, current, next, second next word)
32. word_prev3(third previous word)
33. word_prev4(fourth previous word)
34. word_prev5(fifth previous word)

35. word_next3(third next word)
36. word_next4(fourth next word)
37. word_next5(fifth next word)

Accuracy : 0.9552549601026171

Format : Tag, Precision, Recall, Fscore

1. NUM Precision: 0.935251798561151 Recall: 0.9643916913946587 Fscore: 0.9495982468955441
2. . Precision: 0.9998701130016885 Recall: 1.0 Fscore: 0.9999350522829122
3. NOUN Precision: 0.9194740629292864 Recall: 0.9645896239363162 Fscore: 0.9414916775511571
4. PRON Precision: 0.9902370990237099 Recall: 0.9598918431726002 Fscore: 0.9748283752860412
5. VERB Precision: 0.9478466076696165 Recall: 0.943061751584879 Fscore: 0.9454481256988171
6. DET Precision: 0.9930118798043326 Recall: 0.99231843575419 Fscore: 0.9926650366748165
7. ADJ Precision: 0.9053189141469874 Recall: 0.8368013055895552 Fscore: 0.869712710696491
8. CONJ Precision: 0.9966714217784118 Recall: 0.9896128423040604 Fscore: 0.9931295901445155
9. ADV Precision: 0.9452459016393443 Recall: 0.8956197576887233 Fscore: 0.9197639176902217
10. ADP Precision: 0.9732379438261791 Recall: 0.9803816895769385 Fscore: 0.9767967555348712
11. PRT Precision: 0.9330323551542513 Recall: 0.9323308270676691 Fscore: 0.9326814591951861