Chapter 6

Experimental Results and Evaluation

6.1 Experimental Results

Experiment was done in four phases according to the size of the lexicon. The four phases were based on lexicon sizes: 5000, 10000, 15000, and 20000. Lexicons were compiled based on different domains viz., government/politics, sports, tourism, etc.

Test data were randomly selected from different domains. Same test data were manually tagged in order to compare the accuracies of tagger. An application was built, which takes automatically tagged test data and manually tagged data as input. In order to see the percentage of error over test corpus, tag of a word in test corpus was compared against the tag of manually tagged corpus.

We have taken four different test sets with similar corpus sizes, and the tagging results obtained for each corpus are given below:

Experiment Set 1 based on lexicon size: 5000 words

Exp1: Total number of words in the test corpus is 1000. Total number of words which are correctly tagged is 506. Hence, accuracy of tagging is 506/1000*100=50.60%.

Exp 2: Total number of words in the test corpus is 1500. Total number of words which are correctly tagged is 755. Hence, accuracy of tagging is 755/1500*100=50.60%.

Exp 3: Total number of words in the test corpus is 2000. Total number of words which are correctly tagged is 997. Hence, accuracy of tagging is 997/2000*100=49.85%.

Exp 4: Total number of words in the test corpus is 2500. Total number of words which are correctly tagged is 1269. Hence, accuracy of tagging is 1269/2500*100=50.60%.

Exp 5: Total number of words in the test corpus is 3000. Total number of words which are correctly tagged is 1538. Hence, accuracy of tagging is 1538/3000*100=51.26%.

Exp 6: Total number of words in the test corpus is 3500. Total number of words which are correctly tagged is 1829. Hence, accuracy of tagging is 1829/3500*100=52.25%.

Exp 7: Total number of words in the test corpus is 4000. Total number of words which are correctly tagged is 2210. Hence, accuracy of tagging is 2210/4000*100=50.60%.

Exp 8: Total number of words in the test corpus is 4500. Total number of words which are correctly tagged is 2512. Hence, accuracy of tagging is 2512/4500*100=55.82%.

Exp 9: Total number of words in the test corpus is 5000. Total number of words which are correctly tagged is 2903. Hence, accuracy of tagging is 2903/5000*100=58.06%.

Experiment-1				
Size of lexicon (No of words)	Exp	Total words	Correctly Tagged	Accuracy
5000	1	1000	506	50.6
5000	2	1500	755	50.333
5000	3	2000	997	49.85
5000	4	2500	1269	50.76
5000	5	3000	1538	51.26
5000	6	3500	1829	52.25
5000	7	4000	2210	55.25
5000	8	4500	2512	55.82
5000	9	5000	2903	58.06
Average accuracy				52.66

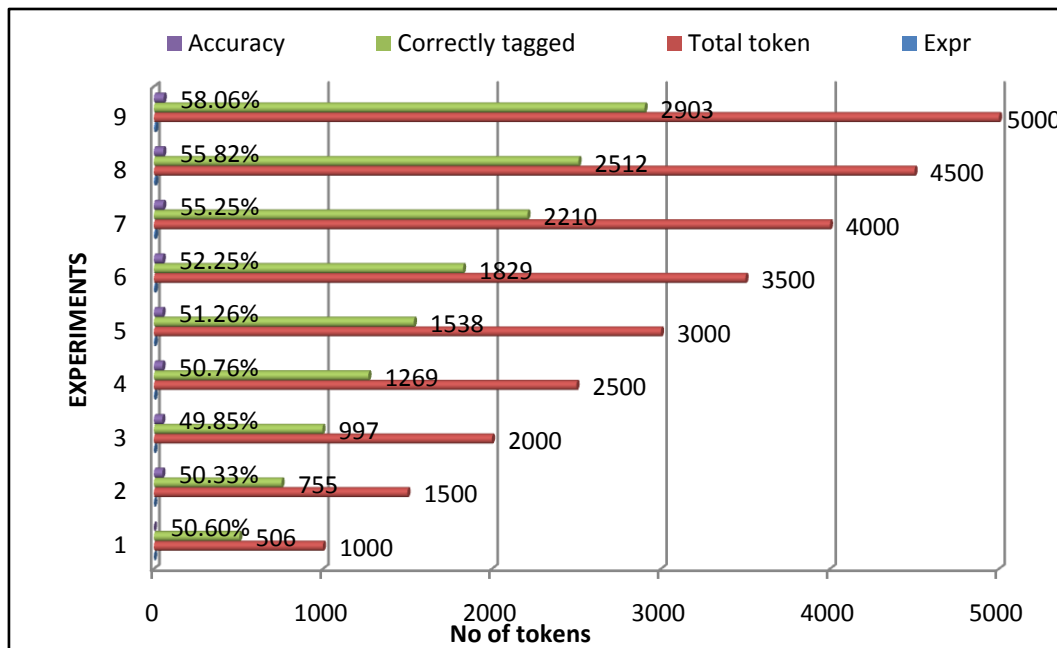**Table 6.1 Experiment Set 1 based on lexicon size: 5000 words**

Fig 6.1: Experiment Set 1 based on lexicon size: 5000 words

Experiment Set 2 based on lexicon size: 10000 words

Exp 1: Total number of words in the test corpus is 1000. Total number of words which are correctly tagged is 506. Hence, accuracy of tagging is 682/1000*100=68.20%.

Exp 2: Total number of words in the test corpus is 1500. Total number of words which are correctly tagged is 1036. Hence, accuracy of tagging is 1036/1500*100=69.06%.

Exp 3: Total number of words in the test corpus is 2000. Total number of words which are correctly tagged is 997. Hence, accuracy of tagging is 1420/2000*100=71%.

Exp 4: Total number of words in the test corpus is 2500. Total number of words which are correctly tagged is 1269. Hence, accuracy of tagging is 1708/2500*100=68.32%.

Exp 5: Total number of words in the test corpus is 3000. Total number of words which are correctly tagged is 1538. Hence, accuracy of tagging is 2078/3000*100=69.26%.

Exp 6: Total number of words in the test corpus is 3500. Total number of words which are correctly tagged is 1829. Hence, accuracy of tagging is 2445/3500*100=69.85%.

Exp 7: Total number of words in the test corpus is 4000. Total number of words which are correctly tagged is 2734. Hence, accuracy of tagging is 2734/4000*100=68.35%.

Exp 8: Total number of words in the test corpus is 4500. Total number of words which are correctly tagged is 3151. Hence, accuracy of tagging is 3151/4500*100=70.02%.

Exp 9: Total number of words in the test corpus is 5000. Total number of words which are correctly tagged is 2903. Hence, accuracy of tagging is 3523/5000*100=70.46%.

Experiment-2				
Size of lexicon (No of words)	**Exp**	**Total words**	**Correctly Tagged**	**Accuracy**
10000	1	1000	682	68.2
10000	2	1500	1036	69.066
10000	3	2000	1420	71
10000	4	2500	1708	68.32
10000	5	3000	2078	69.266
10000	6	3500	2445	69.85
10000	7	4000	2734	68.35
10000	8	4500	3151	70.022
10000	9	5000	3523	70.46
Average accuracy				**69.39**

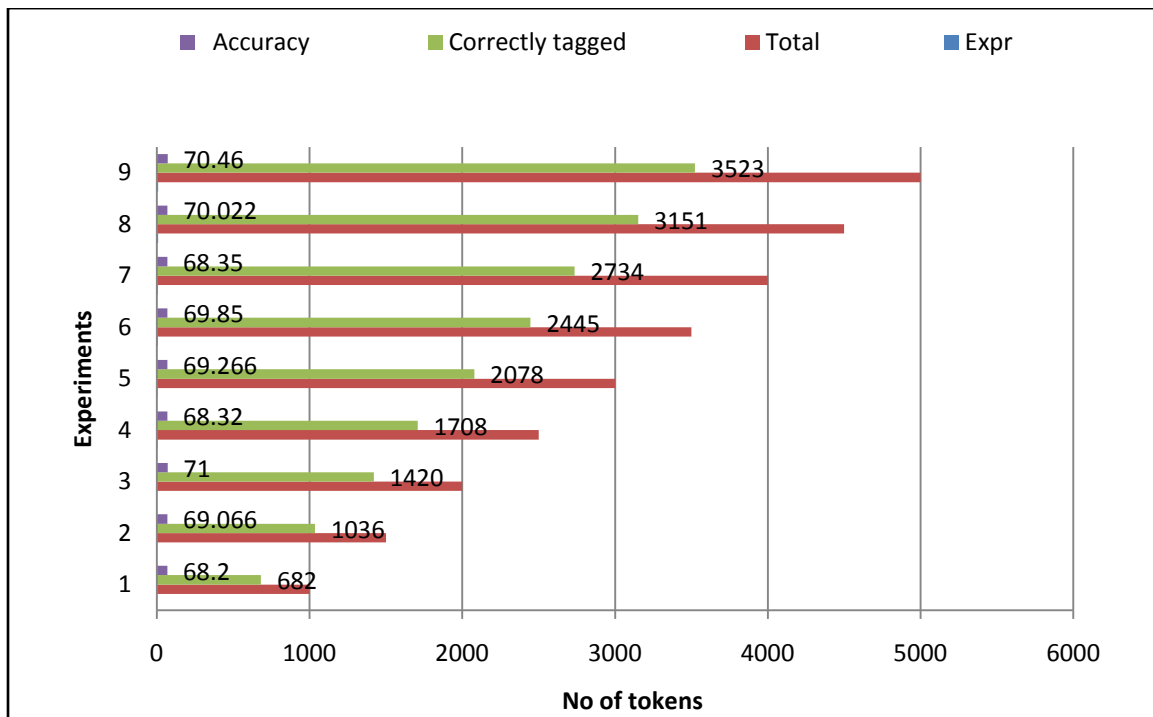**Table 6.2: Experiment Set 2 based on lexicon size: 10000 words**

Fig 6.2: Experiment Set 2 based on lexicon size: 10000 words

Experiment Set 3 based on lexicon size: 15000 words

Exp 1: Total number of words in the test corpus is 1000. Total number of words which are correctly tagged is 845. Hence, accuracy of tagging is 845/1000*100=84.50%.

Exp 2: Total number of words in the test corpus is 1500. Total number of words which are correctly tagged is 1292. Hence, accuracy of tagging is 1292/1500*100=86.13%.

Exp 3: Total number of words in the test corpus is 1000. Total number of words which are correctly tagged is 1688. Hence, accuracy of tagging is 1688/2000*100=84.40%.

Exp 4: Total number of words in the test corpus is 2500. Total number of words which are correctly tagged is 2170. Hence, accuracy of tagging is 2170/2500*100=86.80%.

Exp 5: Total number of words in the test corpus is 3000. Total number of words which are correctly tagged is 2603. Hence, accuracy of tagging is 2603/3000*100=86.76%.

Exp 6: Total number of words in the test corpus is 3500. Total number of words which are correctly tagged is 3075. Hence, accuracy of tagging is 3075/3500*100=87.85%.

Exp 7: Total number of words in the test corpus is 4000. Total number of words which are correctly tagged is 3594. Hence, accuracy of tagging is 3594/4000*100=89.85%.

Exp 8: Total number of words in the test corpus is 4500. Total number of words which are correctly tagged is 3921. Hence, accuracy of tagging is 3921/4500*100=87.13%.

Exp 9: Total number of words in the test corpus is 5000. Total number of words which are correctly tagged is 5000. Hence, accuracy of tagging is 4450/1000*100=89.00%.

Experiment-3				
Size of lexicon (No of words)	**Exp**	**Total words**	**Correctly Tagged**	**Accuracy**
15000	1	1000	845	84.50
15000	2	1500	1292	86.13

15000	3	2000	1688	84.40
15000	4	2500	2170	86.80
15000	5	3000	2603	86.76
10000	6	3500	3075	87.85
15000	7	4000	3594	89.85
15000	8	4500	3921	87.13
15000	9	5000	4450	89.00
Average accuracy				**86.93**

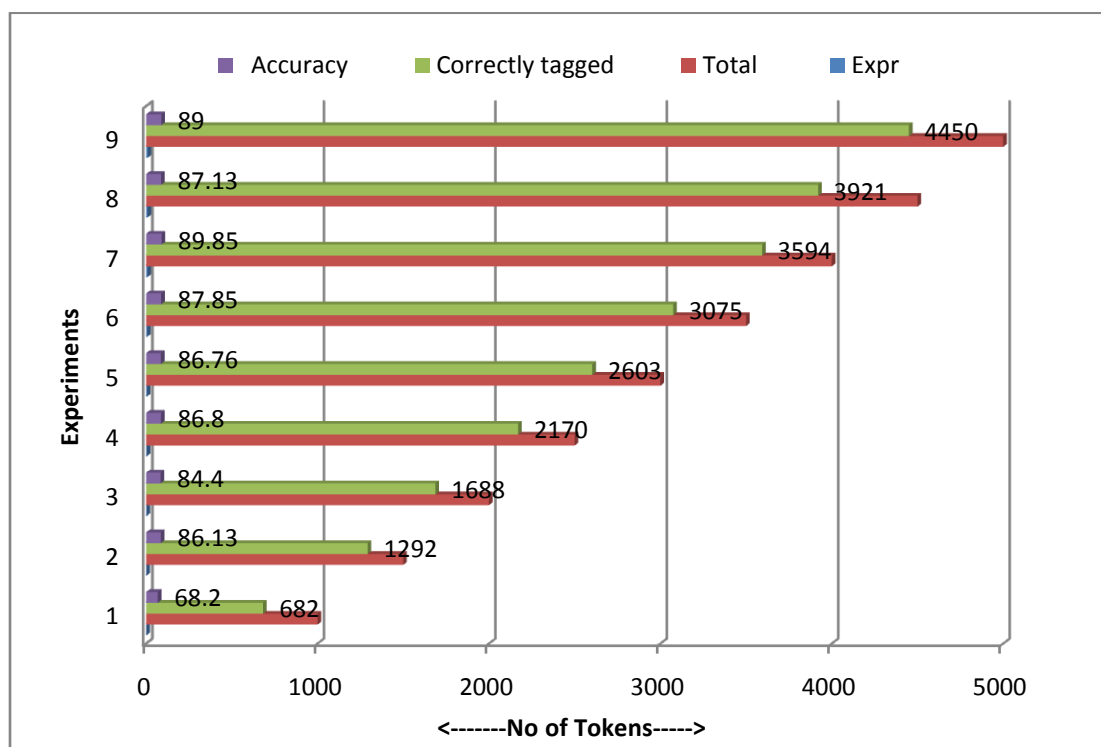**Table 6.3Experiment Set 3 based on lexicon size: 15000 words**

Fig 6.3: Experiment Set 3 based on lexicon size: 15000 words

Experiment Set 4 based on lexicon size: 20000 words

Exp 1: Total number of words in the test corpus is 1000. Total number of words which are correctly tagged is 922. Hence, accuracy of tagging is 922/1000*100=92.20%.

Exp 2: Total number of words in the test corpus is 1500. Total number of words which are correctly tagged is 1392. Hence, accuracy of tagging is 1392/1500*100=92.80%.

Exp 3: Total number of words in the test corpus is 2000. Total number of words which are correctly tagged is 1837. Hence, accuracy of tagging is 1837/2000*100=91.85%.

Exp 4: Total number of words in the test corpus is 2500. Total number of words which are correctly tagged is 2301. Hence, accuracy of tagging is 2301/2500*100=92.04%.

Exp 5: Total number of words in the test corpus is 3000. Total number of words which are correctly tagged is 2803. Hence, accuracy of tagging is 2803/3000*100=93.43%.

Exp 6: Total number of words in the test corpus is 3500. Total number of words which are correctly tagged is 3275. Hence, accuracy of tagging is 3275/3500*100=93.57%.

Exp 7: Total number of words in the test corpus is 4000. Total number of words which are correctly tagged is 3790. Hence, accuracy of tagging is 3594/4000*100=94.75%.

Exp 8: Total number of words in the test corpus is 4500. Total number of words which are correctly tagged is 4211. Hence, accuracy of tagging is 4211/4500*100=93.57%.

Exp 9: Total number of words in the test corpus is 5000. Total number of words which are correctly tagged is 4706. Hence, accuracy of tagging is 4706/1000*100=94.12%.

Experiment-4				
Size of lexicon	**Exp**	**Total words**	**Correctly Tagged**	**Accuracy**
20000	1	1000	922	92.2
20000	2	1500	1392	92.8
20000	3	2000	1837	91.85
20000	4	2500	2301	92.04
20000	5	3000	2803	93.43
20000	6	3500	3275	93.57
20000	7	4000	3790	94.75
20000	8	4500	4211	93.57
20000	9	5000	4706	94.12
			Average accuracy	**93.14**

Table 6.4   Experiment Set 4 based on lexicon size: 20000 words
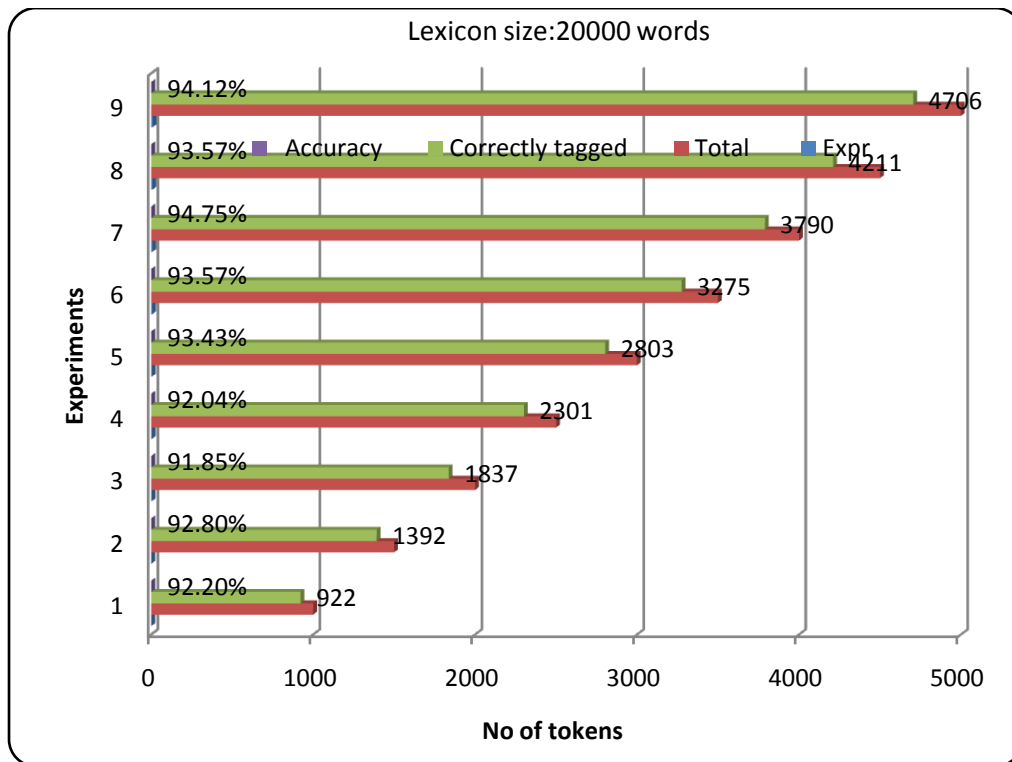
Fig 6.4: Experiment Set 4 based on lexicon size: 20000 words

6.2 Performance of different number of words:

We have tested many experiments using rule based approach and HMM with rule based approach on different corpus till we get the best accuracy. Then, we have seen that POS Tag using HMM with rule based approach get the better accuracy than using rule based approach only. Table II shows the performances of POS tagging according to the different approaches on different number of words in corpus. Figure 2 also, shows the comparison of these improvements in accuracy along with the increase in the size of annotated training data on different methods.

Number of words	Accuracy using Rule Based Approach	Accuracy using HMM with Rule Based Approach
5000	47.11	52.66
7500	61.34	65.43
10000	61.54	69.33
12500	75.42	82.13
15000	77.87	88.11
17500	80.56	90.33
20000	84.89	93.50

Table 6.5   Performance of different number of words
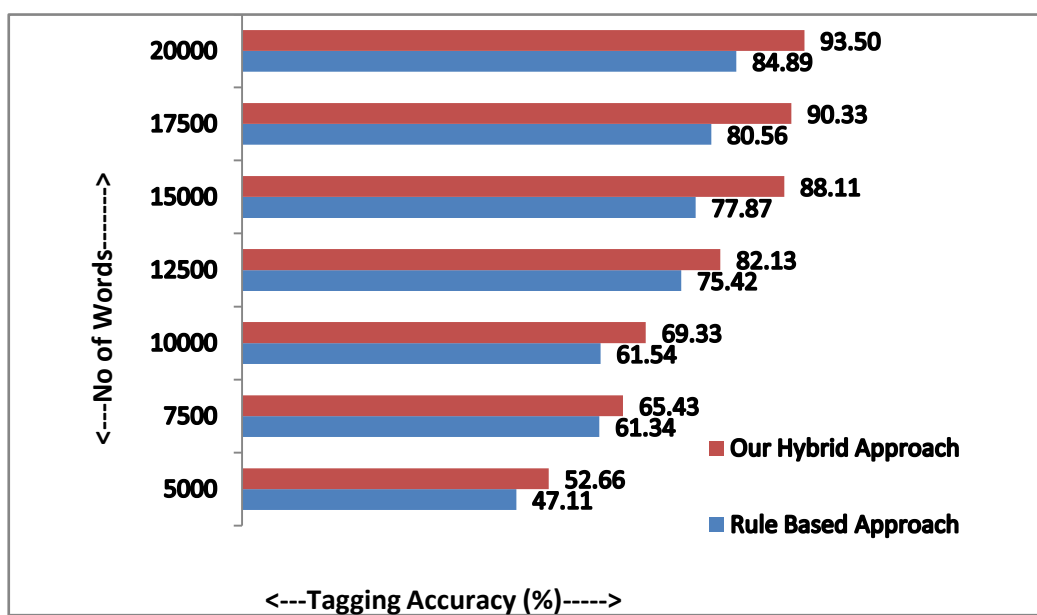

Fig 6.5   Performance of different number of words

6.4 Input/Output of the Tagger

Input

६१ वर्षीय   छेत्री अप्रेल   भिन्केन नोभेम्बर २९   बाट सल्लाहकार को रूप मा सञ्चालक समिति मा

आउनुहुनेछ ।श्री भिन्केन डच प्रकाशन समूह एल्सेभियर एन.भी. को अध्यक्ष हुनुहुन्छ ।

81

कन्सोलिडेटिड गोल्ड फिल्ड्स पीएलसीका पूर्व सभापति ५५ वर्षीय रूडोल्फ अग्न्युलाई यस ब्रिटिस औद्योगिक समूहको सल्लाहकारको रूपमा मनोनयन गरिएको थियो ।एकताका केन्ट चुरोटको फिल्टर बनाउन प्रयोग भएको एक प्रकारको अस्बेस्टोस तीस वर्षभन्दा अगाडि यसको सम्पर्कमा आएका कामदारहरूको समूहमा क्यान्सरबाट मृत्यु हुनेको उच्च प्रतिशतको कारण बनेको छ , अनुसन्धाताहरूले जानकारी दिए ।

Output

६१/CD वर्षीय/JJ छेत्री/NN अप्रेल/ NN भिन्केन/ NNP नोभेम्बर/NNP २९/CD बाट/POP सल्लाहकार/NN को/ PKO रूप/NN मा/POP सञ्चालक/NN समिति/NN मा/POP आउनुहुनेछ/VBX श्री/NN भिन्केन/NNP डच/NNP प्रकाशन/NN समूह/NN एल्सेभियर/ NNP एन.भी./FB को/PKO अध्यक्ष/NN हुनुहुन्छ/VBF कन्सोलिडेटिड/NNP गोल्ड/NN फिल्ड्स/NN पीएलसीका/NN पूर्व/JJ सभापति/NN ५५/CD वर्षीय/JJ रूडोल्फ/NNP अग्न्युलाई/NN यस/DUM ब्रिटिस/NNP औद्योगिक/JJ समूहको/NN सल्लाहकारको/NN रूपमा/NN मनोनयन/NN गरिएको/VBKO थियो/VBX एकताका/RBO केन्ट/NNP चुरोटको/NN फिल्टर/NN बनाउन/VBI प्रयोग/NN भएको/VBKO एक/CD प्रकारको/NN अस्बेस्टोस/NNP तीस/ CD वर्षभन्दा/NN अगाडि/RBO यसको/NN सम्पर्कमा/NN आएका/VBKO कामदारहरूको/NN समूहमा/NN क्यान्सरबाट/NN मृत्यु/NN हुनेको/VBKO उच्च/दर प्रतिशतको/ NN कारण/NN बनेको/VBKO छ/VBF ,/YM अनुसन्धाताहरूले/NN जानकारी/NN दिए/VBF
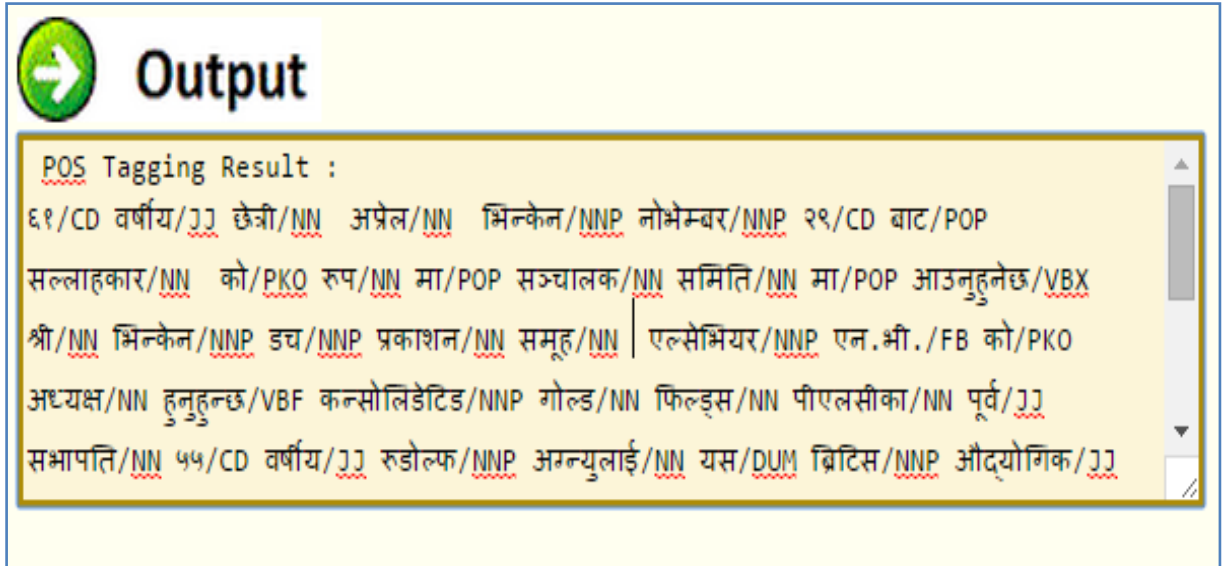
Fig 6.6: Output of the tagger

6.5 System Performance

For testing the performance of our system, we measured different corpus with different lengths and we calculated the precision, recall and f-measure as they are considered to be standard performance indicators of a system.

PRECISION:

Precision: Precision is the number of relevant records retrieved divided by the total number of irrelevant and relevant records retrieved. In other words, Precision is the ratio of the number of correct POS tags assigned by the system to the total number of POS tags assigned by the system. Generally it is expressed as a percentage. These can be calculated using the following equation.

$$\text{Precision (P)} = \frac{\text{No of correct POS tags assigned by the system}}{No of POS tags assigned by t \quad he system}$$

.

RECALL:

Precision: Recall is the number of relevant records retrieved divided by the total number relevant records retrieved. In other words, Recall is the ratio of the number of correct POS tags assigned by the system to the total number of POS tags in the annotated corpus. Generally it is expressed as a percentage. These can be calculated using the following equations

$$R = \frac{\text{Number of items tagged by the system}}{\text{Total number of items to be tagged in the annotated test corpus}}$$

Recall is also considered to be the accuracy score of the system as it assigns correct tags against the total no. of tags in the corpus.

F-measure:

F-measure, sometimes known as F-Score is used to measure of a test's accuracy [70]. It is combination of both the precision P and the recall R to compute the score or performance. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. It combines Recall (R) and Precision (P) using the formula. F-measure or balanced F-score (**F1 score**) is the mean of precision and recall and in simplified form it can be written as

$$F = \frac{2RP}{R+P}$$

Measurements of different corpus based on the above mentioned indicators are summarized in the Table 6.6.

Test Corpus	Total No of words	Precision			Recall			f-measure F= 2RP/R+P
		Unknown	Known	All	Unknown	Known	All	
1	7889	88.98 %	96.86 %	92.87%	88.98 %	96.86 %	92.87%	92.92%
2	8765	90.45 %	97.26 %	93.75%	90.45 %	97.26 %	93.75%	93.85%
3	6550	89.43 %	96.98 %	93.20%	89.43 %	96.98 %	93.20%	93.20%
4	9743	87.87 %	96.98 %	92.42%	87.87 %	96.98 %	92.42%	92.42%
5	4324	89.88 %	97.90 %	93.89%	89.88 %	97.90 %	93.89%	93.89%
6	3253	91.43 %	96.76 %	94.99%	93.43 %	96.56 %	94.99%	94.09%
7	2064	90.78 %	97.67 %	94.22%	90.78 %	97.67 %	94.22%	94.22%
8	1087	88.67 %	96.75 %	92.71%	88.67 %	96.75 %	92.71%	92.7%1
9	6675	90.76 %	97.64 %	94.60%	90.76 %	97.64 %	94.60%	94.20%
Average →		89.80 %	97.20 %	93.50%	89.80 %	97.20%	93.50%	93.50%

Table6.6: Precision and Recall accuracy on different corpus

Test scores of our system are as follows:

Average precision for unknown words   = 89.80%

Average precision for known words   = 97.20%

Average precision for all the words   = (89.80+97.20)/2

=93.50%

Average Recall for unknown words   = 89.80%

Average Recall for known words   = 97.20%

Average precision for all the words   = (89.80+97.20)/2

$$=93.50\%$$

Since the scores of precision and recall are same, the f-measure would also be same. Thus accuracy of the system is **93.50%.**

From tests made for various sizes of training data, it had been shown that the performance of the tagger depends upon the size of the training data, as well as number of tokens that are present and absent in the training data. **Here, in this dissertation, the average overall accuracy of this tagger for morphologically rich and order free language ‑Nepali is 93.50%.**