

Assignment#6 Write- up
Yuqian Zhang
N19945556

Conclusion:

After trying 12 feature combinations, I found the following combination gives the best score for F-measure.

curToken,firstUpper, preToken, postToken, curPOS, prePOS, postPOS, curTag, preTag, preName

```
49792 out of 51578 tags correct
accuracy: 96.54
5917 groups in key
5768 groups in response
4744 correct groups
precision: 82.25
recall:    80.18
F1:        81.20
```

I expected to have higher F-measure when adding external name list source (lastName, femaleFirstName, maleFirstName). However, it got 80.38 for the F-measure, not that high.

```
49728 out of 51578 tags correct
accuracy: 96.41
5917 groups in key
5783 groups in response
4702 correct groups
precision: 81.31
recall:    79.47
F1:        80.38
```

Thus, my CONLL_test.name is run by the higher F-measure feature combination stated above.

Considering that the external name list feature might not be obvious through the development corpus, I tested the CONLL_test.pos-chunk with external name list source as well. Please refer to CONLL_test(extra).name for the result with external name list feature added.

Here are all feature testing results:

1. curToken,curPOS, prePOS, postPOS, preCur, curPost, curTag, preTag, postTag, preName

```
49148 out of 51578 tags correct
accuracy: 95.29
5917 groups in key
5729 groups in response
4367 correct groups
precision: 76.23
recall:    73.80
F1:        75.00
```

2. curToken,curPOS, prePOS, postPOS, preCur, curPost, curTag, preName

```
49180 out of 51578 tags correct
accuracy: 95.35
5917 groups in key
5733 groups in response
4388 correct groups
precision: 76.54
recall:    74.16
F1:        75.33
```

3. curToken,preToken,postToken,curPOS, prePOS, postPOS, preCur, curPost, curTag, preName

```
49537 out of 51578 tags correct
accuracy: 96.04
5917 groups in key
5653 groups in response
4529 correct groups
precision: 80.12
recall:    76.54
F1:        78.29
```

4. curToken,firstUpper,preToken,postToken,curPOS, prePOS, postPOS, preCur, curPost, curTag, preName

```
49777 out of 51578 tags correct
  accuracy: 96.51
5917 groups in key
5819 groups in response
4729 correct groups
  precision: 81.27
  recall:    79.92
  F1:        80.59
```

5. curToken,firstUpper,preToken,postToken,curPOS, prePOS, postPOS, curTag, preName

```
49788 out of 51578 tags correct
  accuracy: 96.53
5917 groups in key
5770 groups in response
4743 correct groups
  precision: 82.20
  recall:    80.16
  F1:        81.17
```

6. curToken,firstUpper,allUpper, preToken,postToken,curPOS, prePOS, postPOS, curTag, preName

```
49801 out of 51578 tags correct
  accuracy: 96.55
5917 groups in key
5789 groups in response
4741 correct groups
  precision: 81.90
  recall:    80.13
  F1:        81.00
```

7. curToken,firstUpper, preToken,hasTHE, postToken,curPOS, prePOS, postPOS, curTag, preName

```
49786 out of 51578 tags correct
  accuracy: 96.53
5917 groups in key
5778 groups in response
4739 correct groups
  precision: 82.02
  recall:    80.09
  F1:        81.04
```

8. . curToken,firstUpper, preToken, postToken,hasComma, curPOS, prePOS, postPOS, curTag, prename

```
49780 out of 51578 tags correct
accuracy: 96.51
5917 groups in key
5779 groups in response
4735 correct groups
precision: 81.93
recall:    80.02
F1:        80.97
```

9. curToken,firstUpper, preToken, postToken,hasComma, curPOS, prePOS, postPOS, curTag, prename

```
49796 out of 51578 tags correct
accuracy: 96.55
5917 groups in key
5771 groups in response
4743 correct groups
precision: 82.19
recall:    80.16
F1:        81.16
```

10. curToken,firstUpper, preToken, postToken, curPOS, prePOS, postPOS, curTag, preTag, preName

```
49792 out of 51578 tags correct
accuracy: 96.54
5917 groups in key
5768 groups in response
4744 correct groups
precision: 82.25
recall:    80.18
F1:        81.20
```

11. curToken,firstUpper, preToken, postToken, curPOS, prePOS, postPOS, curTag, postTag, preName

```
49780 out of 51578 tags correct
accuracy: 96.51
5917 groups in key
5753 groups in response
4731 correct groups
precision: 82.24
recall:    79.96
F1:        81.08
```

12. curToken,firstUpper, inNamelist, preToken, postToken, curPOS, prePOS, postPOS, curTag, preTag, preName

```
49728 out of 51578 tags correct
  accuracy: 96.41
5917 groups in key
5783 groups in response
4702 correct groups
  precision: 81.31
  recall:    79.47
  F1:        80.38
```