|  |
| --- |
| **Chapter 6: Future work & Conclusion** |

* 1. **Future Work**

This thesis addresses the technique to improve the quality of POS tagging by finding the possible changes of a word. In the others direction how, a word may vary in a sentence. The main limitation of the approach presented here is that it will not result well if the corpora is too little. That means if the number of word variation of a word is few or little, the result may not come according the expected output.

There are a number of possible directions for future work, based on the findings in this thesis. Some of the directions are given bellow -

Increasing the size of parallel corpora always help to improve the accuracy of the system. Adding different sentence structures and handling the idioms and phrases externally would help to improve the system.

The reordering rules by which a subject of a sentence can be measured as well as the POS tags suggested in this thesis are relatively simple, and do not perform very well on large sentences. It would be possible to replace the handcrafted rules with automatically learned rules.

Word prediction of possible word can be detect from our research. Generating rules with stemming and lemma development can also help detect similar wordform.

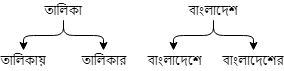


Figure 6-A: Word possibilities detection

In future, advancement of this system lies in integrating with speech recognition systems. Additionally, there is a possibility to convert this system into mobile environment for translating simple sentences.

It would be useful to perform a thorough error analysis of the translation output. Such an analysis would give improvements in future.

* 1. **Conclusion**

In this paper, we have investigated the possible way of Bengali POS tagging. Our implementation of the model was not effective due to our choice of training sets as well as the cleanliness of our data. Automatic POS tagging makes errors because many high frequency words of part-of-speech are ambiguous. Statistical tagging assigns a word its most likely tag, based on the n-set values frequencies in a training corpus.