Chapter 8 :CONCLUSION

Our research work described a language model for morphological analysis of Manipuri written texts minus its tonal description, using finite- state techniques. The analysis of the words is performed using two main components: a lexicon and a set of rules, in the framework of finite state network based algorithm. The present underlying model has over 3,00 lexicon entries of roots. The average number of inflected forms of a lexicon entry is 47 and for derivation it is 44. By virtue of the characteristics of finite-state transducers which allows look up and look down routines in the network, our analyzer can also be used for generation of word forms. And also for the same reason the time complexity of analysis and generation is linear in the length of the string. In practice, our current analyzer can analyze about 1000 words and take less than a second.

RESOLVED RESEARCH QUESTIONS

1) ***Linguistic insights in the light of computational morphology***

Manipuri morphotactics have been described in detail only for specific local varieties, yet the finite-state approach requires a level of generalization and abstraction that cannot be found in any conventional grammar. The partitioning into slots of the verbal and nominal root structure and their particular morphotactical restrictions represent new knowledge about the linguistic features of the language.

2) ***Efficiency and sufficiency of the technique***

If appropriate linguistic facts are established, finite-state approach for implementation of morphological analysis and synthesis is very straightforward and simple. In addition finite-state systems are known for their speedy performance.

3) ***Coverage***

The experiments with the xfst tool have shown that the finite-state techniques are ideally suited to cover a wide range of possible word forms in a morphologically rich agglutinative language like Manipuri. Nevertheless, the major challenge to every morphology system for Manipuri are the numerous loan words, and other unearthed morphological features like reduplication, etc. which make a 100% coverage appear highly unrealistic.

CONTRIBUTION OF THE RESEARCH WOK

Literature has shown that finite state techniques and its algorithms are successfully utilized in solving many NLP problems like Part-Of-Speech (POS) tagging, chunking, parsing, etc. The present reaserch study has, no doubt, opened up and paved the way to use finite state algorithms and techniques in these NLP applications for Manipuri language. No such attempts has been made for Manipuri language as of today.

The developed model for morphological analysis of Manipuri work can be used for the purpose of spelling correction and word form checking by increasing the size of each category lexicons. Further, it can be enhanced for other NLP problems like word sense disambiguation (WSD), machine translation (MT), information Retrieval (IR), etc. by adding more features on the existing system.

OPEN RESEARCH QUESTION(S)

Effort

The amount of effort put in the development of Morphological Analysis for Manipuri language with finite-state techniques is comparable to other agglutinative languages like Aymara, Turkish, and Finnish, which also have rich and complex morphological word structure. The disentanglement of the exact morphotactics is a major challenge for such languages. Turkish handles this problem grammatically and the only challenge it has is in correct handling of the complex morphophonology.

To achieve successful and satisfactory output from a Morphological Analyzer in an NLP application, it is essentially required to have substantial amount of grammatical,

linguistic and vocabulary insights of the language in question. The main bottleneck during the course of the journey of the investigation is the lack of a standard documentation of Manipuri language grammar and the incomplete written representation of the language with Bengali script for tones and the monophthong and diphthong sound patterns available in the language. As for example, certain morphemes are missing their lexical/ grammatical meaning in a given context.

FUTURE DIRECTIONS OF RESEARCH

One of the complex issues in agglutinative and complex morphological word structured languages is reduplication. Manipuri language exhibits reduplication-partial as well as complete. Work in this direction to capture these morphological phenomena for computational morphology will be challenging as well as interesting one.

Pure finite-state networks have no stack or other "memory" to store information about what morphemes or features have been accumulated; each transition from one state to the next depends only on the current input symbol. Manipuri exhibits a kind of dependency, called long distance dependency/ separated morphotactics where the occurrence of a morpheme is constrained by another morpheme in the word structure not adjacent to the said morpheme. Any feasible method to allow or restrict the morpheme will definitely solve the over-generation or under generation problem in the networks.