Chapter 10

Conclusion

There is a radical shift proposed in the architecture of the anusaaraka from its earlier version. The earlier version, very well demonstrated its usefulness as an aid to overcome the language barrier. But the output of the system involved a learning component. In this thesis, we showed, how the system can be improved further, taking it towards Machine Translation, still retaining the fidelity to the original text. We have also shown how the various concepts from IGT are relevant in developing a Machine Translation system. We discuss below some experimental feedback we got on the present system which provides directions for future development.

10.1 Present: some experimental feedback

Anusaaraka developed with an aim to provide a faithful access to the SL text, was found to be a good tool to assist the students learning English as a second language too. The usefulness and effectiveness of the tool was tested at various Hindi medium schools of Madhya Pradesh, India. Madhya Pradesh Sanskrit Board at Bhopal provided help from time to time to get feedback on the anusaaraka at various stages of its development.
The first three experiments were carried out with only first layer of anusaaraka output \[13\]. Thus students had to undergo some training in the śabdasūtra and also the lessons describing the structural differences in English and Hindi. The students who, prior to the experiment, were not able to comprehend simple sentences were at the end of a week’s period able to comprehend stories of 10-15 sentences, with the help of anusaaraka output. Most of these stories were from Children’s story books. During these experiments need was felt to provide more help in the form of word groupings, phrase boundaries, likely POS tags, etc., which resulted in the present architecture.

The main emphasis of anusaaraka being the faithfulness, scientific texts and other texts carrying ‘information’ are more relevant for the use of anusaaraka rather than texts meant for amusement, entertainment, etc. To find out how useful the anusaaraka in current state (as of 2005) is, it was decided to conduct some more experiments.

With the improved architecture, again an experiment was planned, but this time with the science texts. The students of 10th class who have studied 9th standard science texts through Hindi medium were given the anusaaraka outputs of science texts, in order to know how much this tool helps a student to understand the scientific English texts with which s/he is already familiar with.

The following are our observations:

- The major problem was with the scientific concepts. Though the students had studied the scientific texts through Hindi medium, they found it difficult to comprehend the anusaaraka output of the same text from English into Hindi because, they were not thorough with the scientific concepts involved.

- Second problem was the domain specific constructions. It was necessary to develop domain specific modules to handle the specific syntactic constructions and also use domain specific dictionary to get better results.
Finally it was noticed that adding a WSD module will improve the quality and reduce the burden on the user further to a great extent.

This then defined the course of our future work.

### 10.2 Future

The first version of the anusaaraka demonstrated the technological feasibility of anusaaraka to overcome the language barrier. The second version further demonstrated the relevance of Indian Grammatical Tradition in the development of this technology. The design of architecture showed various ways to add comfort to the end product and make the end product more user friendly and comfortable reducing the burden on the user.

As it was noticed during the experiments in the schools, now, in the third version a strong need is felt about the Word Sense Disambiguation module. Use of WASP workbench helped us in developing the WSD rules semi-automatically. But the rules were so large in number that, it failed to provide any linguistic insight. There is a need to keep the rules in human readable form, so that, if need arises, one can improve upon them. Secondly, the task being voluminous, with always scope for further improvement, there is a need to involve large number of people. Work is in progress towards developing a user friendly interface for developing WSD rules. The importance of involvement of the society in the development of Oxford dictionary, and online knowledge repositories such as Wikipedia are well-known. The goal of the next version of anusaaraka then should be to make the system self sustainable or in other words, provide an opportunity to the users for participation in the development makes a consumer a contributor.

The close interaction of Sanskrit scholars with the system, should provide better in-
sights in developing parser interfaces for various parsers, by the way of designing an
intermediate parser representation based on Pāñinian system.

The architecture together with an expert shell for decision making should provide
a good base for Linguist’s Workbench, where a user can observe various language
behaviours, play with various parsers, understand different means of coding informa-
tion in language strings, understand the information dynamics within a language, and
understand the information flow from one language into the other in an MT system.