CONCLUSION AND FUTURE PROSPECTS
CONCLUSION AND FUTURE PROSPECTS

In the introductory chapter we have seen the details of computational lexicography and related fields. It gives various types of lexicons and uses of COLEX. Different types of knowledge required in a lexicon, the applications, and the limitations of available lexical data, etc. are discussed. The second chapter is about the phonological and morphological aspects in a lexicon. The classification of word and various word formation rules, techniques, etc. are discussed. Inflection, derivation and compounding are the main issues examined in detail, particularly about those in Malayalam. The third chapter is on the syntactic and the related issues and it is a brief discussion on lexical structures in some important grammar formalisms. Syntactic characteristics of lexical items and semantic aspects for a lexicon - argument structure, thematic roles/theta-grid, case/karaka relations, semantics of compounds and phrases and lastly, the pragmatic and other deep level lexical information are examined. The fourth chapter is on the structure and design of lexicons for different purposes, especially for machine consumption - syntactic or semantic parsing for MT, and for human uses like language teaching and learning, and dictionary look-up. The fifth chapter is the implementation and discussion of the model of the system proposed.

The system developed for Malayalam-Hindi is a prototype meant for lexicon look-up and it can be used as an on-line aid for human translation or language learning. That's why all the relevant lexical information and possible paradigms are included. Its final aim is a multipurpose dictionary database structure irrespective of grammar formalisms. Or one can say it has incorporated all the relevant information of some major and important grammatical formalisms/ theories. It is as designed extendible and reusable. Though it entertains only partial morphological analysis and generation for limited number of constructions, the same could be made complete for the construction of all the word formations and all the possible paradigm forms from both the languages.

The design and implementation discussed here are about a lexicon for the specific purpose of an on-line COLEX for human use. Lexical analysis of only Malayalam language is done. Since the source/target conversion facility can be availed, it can be used as a bi-way lexicon
also after simple modifications. Through the lexical editor any number of lexical items can be included. The same is the case of lexical knowledge.

As the purpose for which and the way a dictionary is used affect its contents, it is not necessary to, or one cannot, include all the information in a single system. Because of this, a number of views have been expressed, by different groups, regarding the type of entries and contents that should be included in a lexicon. A consensus is still hard to be foreseen in the near future. As the machine dictionary is a must for any natural language based system there is no other option but to use a source dictionary to extract necessary parts and edit them through the addition, modification or deletion of information to suit one’s individual purpose. A particular grammatical formalism can be adopted in whatever levels it is possible, and the shortcomings of that theory can be adjusted by cross checking with other related theories. For example, the concepts like argument structure and case relations can be found equivalent to the aakaanksa and karaka relations of Panini. In that case, the system can be said to have incorporated all the required information of the three important grammar formalisms (LFG, GB and PG).

As the present system is a simple lexical system for lexical look-up it can be modified and used for other purposes. The database has a few words only but since it is dynamic and open ended any number of lexical items can be included.

It is genuinely expected that the communication and cultural gaps of the Indian geography and demography can find a possible way-out, if the present prototype developed is extended into a multilingual one to include various other Indian languages as well as upgraded and updated to accommodate and carry the voluminous information needed. A serious computational application of Indian languages requires a detailed lexical database. In this connection, interacting with the various attempts and projects going on in the country to develop colexes of Indian languages (including English) that are mentioned already in the course of the present work is extremely relevant. As it is clear from the contrastive studies on Indian languages, the difference between the languages can be solved mostly at the lexical level. As pan Indian features exist among Indian languages, the goal of realising a single colex for Indian languages may not be very distant.
It is painstaking to produce a lexicon and so it should be reusable. By using the computer memory, large amount of lexical knowledge can be included in an on-line dictionary for the user at any level. Once the structure is ready of a lexical system the required data of the language can be put to a better use of the lexicon.

The system allows a linguist or language scholar to study a language (source) in terms of another (target) language. Since the analyser is provided, one can see the construction of words, formation of words, the context of usage, etc. Certain kinds of data require more interpretation and study, particularly the derivative forms in Malayalam. Further attention is needed on certain language phenomena and constructions.

This system provides just a structure and a beginning of an evolution system. It can be developed to a full fledged lexical system with the further input of entire lexical entries of the language. Information at any level can be incorporated by using all the new technological devices including multimedia. Information at any level can be retrieved by using required module.

By adding suitable modules it can be enriched for purposes like: i. a device to understand the source language, ii. aid for human translation, iii. fully automatic domain specific or free machine translation system, and iv. to initiate language enterprise and commercialize language research.

By changing the respective fields, the present one-way prototype (Malayalam to Hindi) can be modified into a bi-way dictionary. This has the potential to develop an intermediate language (interlingua) for a group of languages.

This endeavor opens up many avenues for future research and development of usable working systems. It can be developed to a fully automated lexical system for all NLP applications related to these two languages.

Improvements to the system can be made after getting feedbacks and suggestion from the users. The best way to get proper and realistic feedback is through useful working systems.
So one should try to keep in touch with working systems at every stage and proceed according to the users’ needs. After making use of this system for any application, one can find out the shortcomings and rectify them. The system can be refined or revised further.

Morphological analyser can be developed for all types of lexical formation present in the languages. Derivation process of generating words from the root can also be developed for all words and categories. And same is the case for lexical parsing. All these can be used to extend the system.

With further improvements in the user interface or with suitable interfaces, the system can come to the aid of a number of applications. This is important because the translator uses the system only if it can increase the productivity. So is the case with people involved in language studies.

The system developed has been named **KOSAKAM** (Computer sabda kosham of Malayalam).