Chapter 7
Conclusion and Future Scope

7.1 Conclusion
This thesis explains the development of a UNL based MT system for Punjabi language. UNL based MT systems follow the interlingua approach of MT. The approach in UNL revolves around the development of an EnConverter and a DeConverter for a natural language. The EnConverter is used to convert a given sentence in natural language to an equivalent UNL expression; and the DeConverter is used to convert a given UNL expression to an equivalent natural language sentence. UNL system has the potential to bridge the language barriers across the world with the development of $2n$ components, while traditional approaches require the $n^*(n-1)$ components, where $n$ is the number of languages. UNL represents the information sentence by sentence in the form of Universal Words (UWs), UNL relations and UNL attributes. The concepts are expressed by UWs and UNL relations are used to specify the role of each word in a sentence. The subjective meanings intended by the author are expressed through UNL attributes. UNL system makes use of word dictionaries in the form of Language-UW lexicon of respective languages for its processing. A major outcome of this research work is the development of Punjabi EnConverter, Punjabi DeConverter and a web interface for online EnConversion and DeConversion process. Punjabi EnConverter uses Punjabi shallow parser for processing the input Punjabi sentence. This parser performs the tasks of tokenizer, morph analyzer, part-of-speech tagger and chunker for the processing of input sentence. Architecture of Punjabi EnConverter has seven phases, namely, parser phase (to parse the input sentence with Punjabi shallow parser), linked list creation phase, universal word lookup phase, case marker lookup phase, unknown word handling phase, user interaction phase (this phase is optional) and UNL generation phase. Punjabi EnConverter developed in this work can convert simple and complex Punjabi sentences to equivalent UNL expressions.

Punjabi DeConverter involves UNL parser, lexeme selection, morphology generation,
function word insertion and syntax planning phases. First stage of a DeConverter is UNL parser which parses the input UNL expression to build a node-net from the input UNL expression. During lexeme selection phase, Punjabi root words and their dictionary attributes are selected for the given UWs in the input UNL expression from the Punjabi-UW dictionary. After that, the nodes are ready for generation of morphology according to the target language in the morphology phase. The proposed system makes use of morphology rule base for Punjabi language to handle attribute label resolution morphology; relation label resolution morphology; and noun, adjective, pronoun and verb morphology. In function word insertion phase, the function words are inserted to the morphed words. Finally, the syntax planning phase is used to define the word order in the generated sentence so that output matches with a natural language sentence. Algorithms and pseudocodes have been implemented for syntax planning for simple sentences, syntax planning of UNL graph with a scope node, untraversed parent, multiple parents, special cases in syntax planning and clausal sentences. All these algorithms and pseudocodes used in different phases of Punjabi EnConverter and DeConverter have been implemented in Java to develop the proposed system.

The proposed system has been tested for one thousand sentences. These sentences have been considered in such a way that generation of all possible UNL relations and attributes can be tested. These sentences include the sentences taken from Spanish UNL Language Server and also from agricultural domain threads developed by IIT Bombay, India. The proposed system has achieved a fluency score of 3.61 (on a 4-point scale) and an adequacy score of 3.70 (on a 4-point scale). The proposed system is able to achieve a BLEU score of 0.72. The proposed system has a word error rate of 5.43% and sentence error rate of 20.8%.

7.2 Limitations of the proposed system

In this section, some of the limitations of the developed MT system are presented.

- The accuracy of Punjabi EnConverter is highly dependent on the accuracy of Punjabi shallow parser used for the parsing of input Punjabi sentence. If it does not parse the input sentence correctly, then the MT system will also not be able to generate its equivalent UNL expression correctly.
• Firing of an accurate EnConversion and DeConversion rule sometime requires semantic information of verbs. The non availability of this information for some verbs in lexicon results into incorrect resolution of UNL relation between UWs.
• The system is not able to resolve ambiguous words; it altogether depends upon Punjabi shallow parser for resolving their part of speech information.
• The system is not able to handle very large and complex sentences.
• The system is not able to perform automatic EnConversion and DeConversion of web pages.

7.3 Future scope of the proposed work
In future, this work can further be enriched by incorporating following components into the proposed system. Some of the work that can be carried out in future includes following.
• In case of incorrect parsing of input sentence by Punjabi shallow parser, an interactive system can be designed to override the part of speech information given by the parser.
• The coverage and accuracy of system can further be improved by expanding the Punjabi-UW dictionary and enriching it with more semantic information.
• The rule base can further be improved linguistically to increase the accuracy of the system and to handle very large and complex sentences.
• The ambiguous words can be handled by adding a word sense disambiguation module to the proposed Punjabi EnConverter.
• A UNL based Punjabi Language Server can be developed that may host the EnConverter and DeConverter of Punjabi language. It will start conversions after receiving requests from any web application, and provide the results when the conversions are completed.
• A UNL Proxy Server can be developed to serve as a filter to allow Internet browsers to recognize web pages written in UNL and then selecting an appropriate Language Server on the Internet so that the document can be read in a natural language.
• The proposed system can further be tested on other languages by using corresponding language’s rule base and lexicon.
Chapter Summary

This chapter concludes the work done during this thesis. This thesis explains the development of a Punjabi-UNL EnConverter and UNL-Punjabi DeConverter. All the algorithms and pseudocodes used in different phases of Punjabi EnConverter and DeConverter have been implemented in Java to develop the proposed system. A web interface for online EnConversion and online DeConversion tasks has also been developed. The proposed system has been tested for one thousand sentences. The proposed system has achieved a fluency score of 3.61 (on a 4-point scale) and an adequacy score of 3.70 (on a 4-point scale). The proposed system has a BLEU score of 0.72, a word error rate of 5.43% and sentence error rate of 20.8%.

There are some limitations of the proposed system that include dependence of Punjabi EnConverter on the accuracy of Punjabi shallow parser, the non availability of semantic information for some verbs in lexicon, dependence on Punjabi shallow parser to resolve ambiguous words and the inability to handle very large and complex sentences.

In future this work can further be enriched by incorporating an interactive system to override incorrect information given by Punjabi shallow parser in case of incorrect parsing of input sentence. Punjabi-UW dictionary can further be enhanced with addition of more semantic information. The proposed system can further be improved by adding a word sense disambiguation module to resolve ambiguities. Setting up a UNL based Punjabi Language Server for automatic EnConversion and DeConversion of web pages can further increase its usage.