Analysis of the Impact of Ambiguity in Question Paper Translation from English to Hindi

ABSTRACT

of

THESIS

SUBMITTED TO

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

LUCKNOW

BABASAHEB
BHIMRAO
AMBEDKAR
UNIVERSITY

LUCKNOW

प्रज्ञा शील करुणा

ESTABLISHED 1996

FOR THE DEGREE OF

Doctor of Philosophy

IN

COMPUTER SCIENCE

Submitted by

Shweta Vikram

Enrolment No: 965/14

Under the supervision of

Prof. Sanjay Kumar Dwivedi

DEPARTMENT OF COMPUTER SCIENCE

SCHOOL FOR INFORMATION SCIENCE & TECHNOLOGY

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

(A CENTRAL UNIVERSITY; NAAC- 'A' GRADE)

VIDYA VIHAR, RAEBARELI ROAD, LUCKNOW-226 025 (U.P.), INDIA

2019

ABSTRACT

Word Ambiguity is a challenging task in almost all Natural Language Processing (NLP) based application and Word Sense Disambiguation (WSD) is a research area which helps inappropriately handling the ambiguity issue. WSD aims to automatically identify the correct sense of a word in a particular context by applying a suitable technique. This problem persists since a long in NLP which lead to many researchers to make Machine Translation (MT) projects development meaningful.

The ambiguity has been identified as one of the key issues and challenging problem in Machine Translation. Apart from it, some other similar and related problems also persist in the Natural Language (NL) text that affects the translation accuracy of the text. Most of the Natural Languages are affected by these issues.

A large no of researches has been carried in the area of MT for various language pairs. A number of MT tools have also been developed in India and across the world. These tools are either open domain or are based on the specific area of applications. Many popular MT tools though translate all types of sentences to the varying degree of accuracy; they have not been specifically designed for translation of questions from one Natural language to another. Specially in India, when many exam questions are needed to be translated in various Indian Languages, the general domain MT tools often fail to produce the desired accuracy while translating questions from English to Indian Languages.

There has not been much effort or study on knowing the issues in automatic translation of questions specially from English to Hindi, though WSD and other important issues have been widely discussed in the literature on various sentences. Therefore this research aims to specially target various questions sentences to analyze how they behave with the existing MT tools which are otherwise very popular.

The major focus of this research work is to analyze the impact of ambiguity in Question Sentences. Limited work has been carried (to the best of our knowledge) in this narrow domain of MT, to study, analyses and to remove ambiguity from this Question Paper Translation to improve the performance of English to the Hindi language.

The overall objective of the proposed research, therefore, is to study and analyses the underlying problems in the translation of exam questions from English to Hindi through a detailed experimental analysis of real exam questions. The specific objectives of the present research work are the study and analyses various WSD approaches and specific techniques used by researchers, their impact and suitability. To achieve the objectives, the thesis makes the comprehensive analysis of types of ambiguity and other related problems that affect the automated translation of question papers. Further, detailed experimental analysis according to the size of questions using the supporting MT tool has also been carried to understand the impact on translation in this aspect of questions. Accordingly based on the impacts and issues identified through different analysis an effective algorithm may be proposed to deal with such issues to incorporate into a personalized MT system. In order to perform the detailed experimental analysis, a set of questions in English has been used which were asked in various examinations in India and utilized five major Machine Translation tools which are popular and efficient.

The work carried in this thesis clearly reveals that there are a number of issues when it comes to question paper translation which should be effectively handled by applying suitable approaches and WSD algorithm in order to have an MT system which could be used for practical purposes. The further study and the analysis of work carried in the present research may help to develop an efficient machine translation system would greatly reduce the dependency on human experts in translating questions into different Indian languages for various exams that require bilingual papers.

Our experimental analysis clearly shows that none of these are capable of appropriately handling the issue raised in our experiments. It is also been observed in many cases that translators while translating the question in Hindi have changed the overall interpretation of the questions. This is due to the fact that these translators could not have appropriately resolved various type of ambiguity that may have occurred in such questions.

In another observation, it is found that as the size of questions increases then translation accuracy decreases. Almost all translators have performed fairly well with category I type (Small Size) questions whereas they perform poorly for category III

(Large Size) questions. Interestingly, the average accuracy of each of the five translators for all three categories of questions is found to be below 50%.

In another important analysis with different types of questions (Wh-questions, Objective, Match, Fill in the blank and Keyword Specific), our analysis show that among the performances of different types of questions (Wh-questions, Objective, Match, Fill in the blank and Keyword Specific having accuracy 52.2%, 55.5%, 47.1%, 49.3%, and 50.3% respectively), the objective-questions perform better than other questions and match questions performed poorly.

The work described in this thesis leads to a number of distinct area for future investigation. Various studies on a different aspect of ambiguity and development of an MT with automated disambiguation can be done for Question Paper Translation from English to the Hindi language. An efficient MT system based on our analysis of WSD impact can help the users a lot and they can be getting rid of the question translation refinement process to some extent. Subsequent experimental analysis using real life questions in English revealed some very interesting result and facts which could be effectively used to develop a suitable WSD algorithm to be effectively used specifically for questions translation.