Abstract*:*

With the increasing volume of data and information as well as expanding international interactions, communication is one of the important aspects of today's life. A major problem That there is in here is the impossibility of communicating with the data in another language. Hence one of life's most important issues is to provide an automated solution for the translation from one language to another. Machine translation is one way that is presented to solve this problem and because of its importance, it has a lot of attention in recent years. Statistical machine translation is one of the best methods to translate from one language to another language. For languages that are very similar to each other, the output of this translator is very good. But for some pairs of languages such as English and Persian, structural differences between the two languages and also the lack of a large bilingual corpus is made that this technique does not produce the desired language translation of English to Farsi translations.

This thesis is trying to overcome the problems of this method to translate English into Persian partly, with the help of linguistic information. To do this, first try to reduce the structural difference between sentences in English and Persian. This can lead to a better translation model. For this purpose a series of rules extracted and on the English sentences were applied. These changes lead to 17% improvement in BLEU score and 21% improvement in NIST score. Then enrichment within the body is done using some linguistic information, including part of speech tags and the word stems. With this information a factor-based translation system was created. Output system 17 percent improvement in the BLEU score and 25% improvement in NIST score.

Keywords:

Machine translation,Statistical machine translation, Language model, Translation model, Syntactic reordering, Factor based model