



International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Web Based English to Punjabi MT System for News Headlines

Harjinder Kaur

M. Tech. CSE

Guru Kashi University, India

Dr. Vijay Laxmi

Associate Professor, CSE Department,

Guru Kashi University, India

Abstract – Machine translation (MT) processes a natural language to translate it into some other natural language for the benefits of people of various fields. In recent years, the inter regional and inter language exchange of information has increased tremendously. Due to this, the demand of translation has increased. The performance of an MT system depends on the approach used to design the system. In this paper we are presenting a web based English to Punjabi MT system that translates news headlines of an English news paper into Punjabi in a particular domain.

Index Terms- Machine Translation, Rule Based Approach Direct Approach, Hybrid Approach, EBMT, English to Punjabi.

I. INTRODUCTION

Machine translation can be defined as the study of designing the system that can translate one human language into another language. The language that is given as an input is called Source Language and the language in which we get the output is called Target language. Translation gives the same meaning of source language into equivalent target language. The Web Based English to Punjabi Translator is a machine translation system that works under news headlines from a specific domain. i.e. it can translate the headlines of an English newspaper into Punjabi in a specific domain. This system makes it easy to understand the news correctly as most of the words written in news headlines create some sort of ambiguity and the people with very basic knowledge of English may not interpret the news correctly.

Need of English to Punjabi Translation

In Punjab, most of population is not so familiar with English. As most of the information available on web or electronic information is in English, people who are unaware of English cannot make use of this electronic information without anyone's help. In order to make it possible for everyone to use web, automatic language translation is essential. Our aim is to develop the System which can translate an English headline of news paper into its Punjabi equivalent.

II. LITERATURE REVIEW

R.M.K. Sinha and A. Jain, AnglaHindi: An English to Hindi Machine-Aided Translation System. presented a machine translation system called AnglaHindi which is an English to Hindi version of the ANGLABHARTI translation methodology with a mixture of some example-based translation methodology. AnglaHindi system has been webenabled and is available at URL: <http://anglahindi.iitk.ac.in> for free translation. The system generates approximately 90% acceptable translation in case of simple, compound and complex sentences upto a length of 20 words. [3]

Latha R. Nair and David Peter S., Machine Translation systems for Indian languages, discusses various approaches to machine translation and various machine translation systems that have been developed for Indian languages. An English to Hindi MT system called Mantra was developed by CDAC, Bangalore in 1999. It translates domain specific documents in the field of personal administration.[1]. This paper also discusses various other translation systems in india that are converting English to Hindi using different approaches of MT. It is concluded that translation based systems are more flexible. Direct translation is appropriate for structurally similar language [1]. Interligua approach is used in case of multilingual translations with the help of UNL.

Cheragui A.M, Theoretical Overview of Machine Translation discusses the basics and history of machine translation from 1948 to 2010. This paper also presents various approaches of translationbased on linguistic and computational architectures. Various types of machine translations their evaluation strategies are discussed.

III. WEB BASED ENGLISH TO PUNJABI MACHINE TRANSLATION

In this system, we have used a hybrid approach for translating English sentences into Punjabi. This approach is a combination of direct, example-based(EBMT) and Rule based approach.

Using **Direct Approach** we directly map the source language to the target language. This approach is highly dependent on both the source and target languages. Only a little syntactic and semantic analysis is required in this approach.

Using **Example-Based Machine Translation (EBMT)** we consider the examples of existing translation and reuse them for new translation. Following steps are performed under EBMT.

- Examples matching with the input are found out.
- Alignment is done in order to find out the parts of translation that can be reused.
- Finally recombination is done to make sure that the reusable parts identified during alignment are put together in a legitimate way.

Using **Rule Based Approach**, we parse the source text and produce an intermediate representation. Some rules are made depending upon the pattern of the sentences that can be given as an input. If none of the above approaches are able to translate the English news to Punjabi, then these rules are applied to the input sentence one by one until we find an appropriate rule for the sentence. Then, the sentence gets translated according to that matching rule.

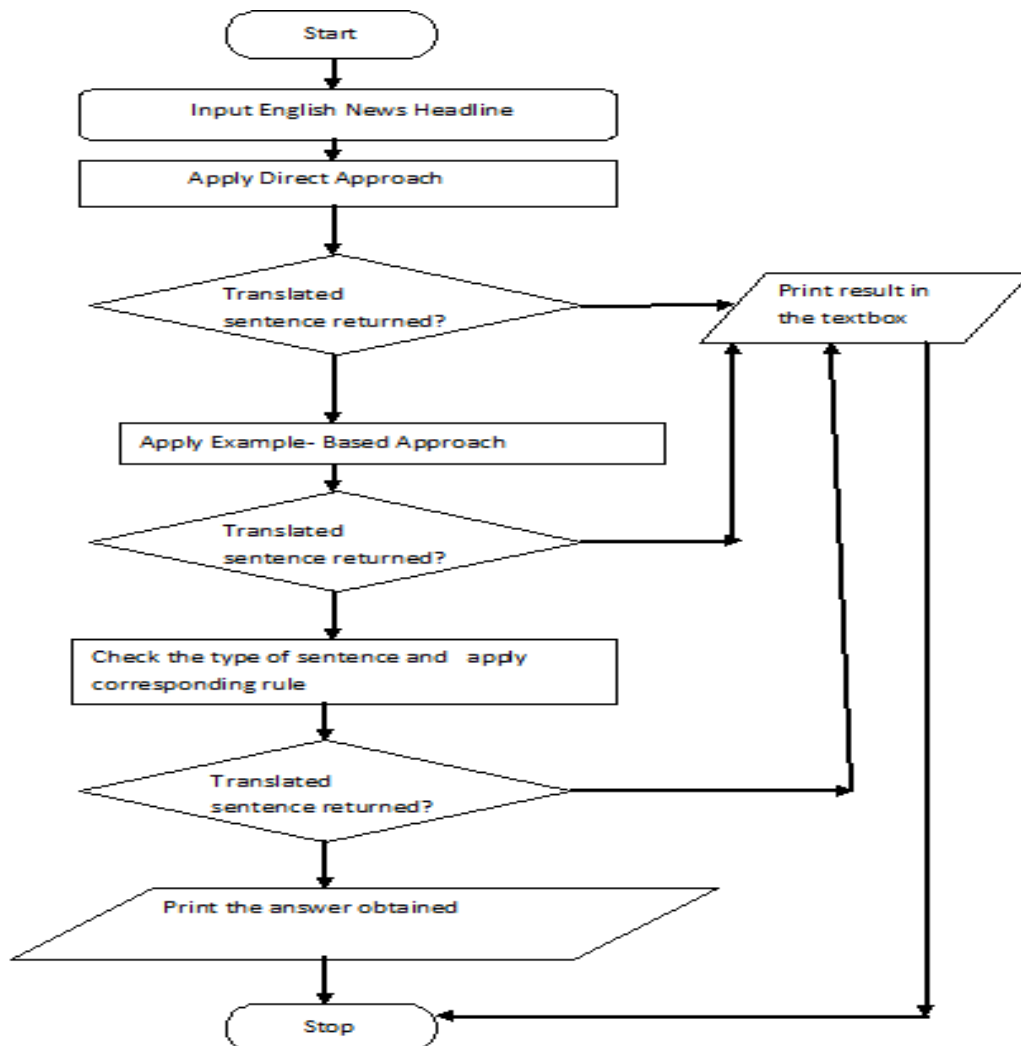
Eg. If a sentence is of the form [Noun] in [Noun], then in the translated sentence, the punjabi meaning of second noun will be at first position followed by ਵਿਚ (punjabi translation of “in”) and punjabi meaning of first noun.

Eg. The translation of English sentence “**man in delhi**” can be shown with the help of following diagram.

English : man in delhi
Punjabi : ਦਿਲੀ ਵਿਚ ਮਾਦਮੀ

IV. METHODOLOGY.

The algorithm used for performing English to Punjabi translation can be explained with the help of diagram shown below.



IV. RESULTS

The system has been tested for about 300 news headlines. Following are some of the outputs.

English	Punjabi (Our System)	Rule Applied
12 booked	ਬਾਰਾਂ ਫੜੇ	[Noun][Verb]
shakuntala devi dead	ਸ਼ਕੰਤਲਾ ਦੇਵੀ ਦਾ ਦੇਹਾਂਤ	Example based approach
Bus catches fire	ਬੱਸ ਨੇ ਅੱਗ ਫੜੀ	[Noun][Verb][Noun]
2 girls abducted in delhi	ਦਿਲੀ ਵਿਚ 2ਕੁੜਿਆਂ ਅਗਵਾਹ	[Noun][Verb]in[Noun]
Musharraf in Pakistan	ਪਾਕਿਸਤਾਨ ਵਿਚ ਮੁਸ਼ਰਫ	[Noun]in[Noun]
Shoe hurled at judge	ਜੱਜ ਤੇ ਬੂਟ ਸੁਟਿਆ ਗਿਆ	[Noun][Verb]at[Noun]
Man killed in accident ; girl abducted	ਦੁਰਘਟਨਾ ਵਿਚ ਆਦਮੀ ਮਾਰਿਆ/ਮਰਿਆ ; ਕੁੜੀ ਅਗਵਾਹ	[News] ; [News]

V. Conclusion

The system was successfully tested on 300 news headlines under the domain. The accuracy of the system has been found as 81.67 percent. The system is currently working under news headlines domain only. It can be further extended to implement a general English to Punjabi Machine Translation System.

References

1. Latha R. Nair and David Peter S., Machine Translation systems for Indian Languages, IJCA, Feb 2012.
2. Goyal V, Lehal G.S., "Advances in Machine translation System Languages in India", 2009.
3. R.M.K. Sinha and Ajay Jain, AnglaHindi: An English to Hindi Machine Translation System, MT Summit IX, New Orleans, USA, Sept.23-27, 2003.
4. Rao D.D., "Machine Translation - A Gentle Introduction", Resonance, July 1998.
5. Hutchins J.W., "Introduction to Machine Translation", Academic Press, 1992.
6. Ambati V., Rohini U., "A Hybrid Approach to Example based Machine Translation for Indian Languages", Proceedings ICON 2007.
7. Cheragui M.A., "Theoretical Overview of Machine Translation", Proceedings ICWIT 2012.

ABOUT THE AUTHORS



Harjinder kaur received her B.E. Computer Engineering Degree from AISSMS WCOE, University of Pune in 2007. Currently she is pursuing M. Tech CSE from Guru Kashi University, Talwandi Sabo , Punjab. She is doing her thesis work on Machine Translation of English to Punjabi. Her areas of interest include Machine translation, OCR.



Dr Vijay Laxmi received her B-Tech degree in computer science & Engineering from SLIET Longowal in 2003, and Ph.D degree in Computer Science & Engineering in year 2012. Her research area is Grid Computing & OCR. She has published 25 research papers in various National/International Conferences and Journals. At present, she is engaged in Guru Kashi University, Talwandi Sabo, and Punjab as Dy. Dear Research and an Associate Professor in Computer Science Engineering & Information Technology department.