RESEARCH HUB – International Multidisciplinary Research Journal (RHIMRJ)
Research Paper

Exploring Challenges and Strategies in English to Punjabi Translation: A Comparative Analysis

Vinay Garg

Department of Computer Science, Multani Mal Modi College, Patiala

Abstract: ***Machine translation has emerged as a new and essential study subject in the realm of machine translation. The primary goal of transliteration is to retain the phonetic structure of words. Proper transliteration of name entities has a substantial impact on machine translation quality. In this study, we use a rule-based method to machine transliteration for the English-Punjabi language combination. The technique of extracting or separating the syllable from the words is known as syllabification. We are computing the probability for name entities (proper names and location) in this case. Separate probabilities are generated utilizing relative frequency using a statistical machine translation for those words that do not fall within the category of name entities.***

Keywords: Machine Translation; Machine Transliteration; Name entity recognition; English to Punjabi; Text conversion.

1. INTRODUCTION

To bridge the gap in communication and understanding between people of different cultural and linguistic origins, the act of translation is crucial. This allows people of different linguistic backgrounds to communicate, share knowledge, and gain access to information. English to Punjabi is one such pair, and it refers to the process of translating any written or spoken English into Punjabi, a language spoken primarily in the Punjab region of South Asia. An example of a translation pair would be this one right here. Education, literature, business, and communication all rely heavily on the ability to translate from English to Punjabi. Punjabi speakers in India, Pakistan, and the diaspora around the world greatly benefit from accurate and effective translations. Community members are better able to understand and participate with English-language resources thanks to these translations [1].Translating from English to Punjabi is a challenging task that requires native speakers of both languages and a firm grasp of the linguistic and cultural nuances at play in the target text. They'll need to be able to work around differences in terminology, sentence structure, idiomatic expressions, and cultural allusions. The challenge of translating from English to Punjabi was communicating the meaning of the source text in a way that was idiomatic and culturally appropriate for the target language. Several methods exist for achieving this goal. Careful consideration must be given to word choice and the use of idioms in order to achieve a translation that is true to the original text in terms of both accuracy and tone. Recent advances in technology, especially in the fields of machine translation and natural language processing, have had a significant impact on the field of English to Punjabi translation [2]. In spite of the fact that automated translations produced by machine translation systems, such as neural machine translation models, have demonstrated some degree of promise, it is possible that such translations will still require human review for content that is especially difficult or highly dependent on its context. Translation from English to Punjabi is becoming increasingly important in today's globalised world. The primary role that technology plays in today's society is to make tasks performed by humans easier and lighter burdens to bear. People who are blind or have impaired vision are now able to independently use computers because to technological advancements that recognise text and speech. When combined with translation software, these technologies can be useful for communicating with individuals whose command of the English language is limited [3].Machine transliteration is critical for accessing information from sources published in multiple languages. Name entity is a common expression used in the field of natural language processing. In the field of natural language processing, name entity recognition has many potential applications, such as information extraction (more precisely, the extraction of structured text from unstructured text), data classification, and question answering. Transcribing a word from its foreign language into its native language is called backward transliteration, whereas transcribing a word from its foreign language into its native language is called forward transliteration. Forward transliteration is addressed in this article. This is also known as the process of transcribing from English to Punjabi. For example, the English word ‘silver' is transliterated into Punjabi as ‘ਸਿਲਵਰ’rather than ‘ਚਾਂਦੀ’. Students who speak Punjabi will have easier access to scholarly articles, books, and other forms of educational material as a result of this. For business reasons, it makes dialogue easier to understand and enables greater market penetration within areas

where Punjabi is spoken. It also plays a significant role in encouraging cultural awareness and the legacy of Punjabi literature, both of which are extremely essential [4].

2. PHASES OF NATURAL LANGUAGE PROCESSING

A) Lexical analysis:

At this stage, the text is divided into sentences, then phrases, and finally words. The purpose of an analysis is to identify and characterise the constituent parts of a word. It makes use of the following techniques:

- Stop word removal (removing 'and,' 'of,' 'the,' and so on from text)
- Tokenization (breaking the text into sentences or words)
- Tokenize for words
- Tokenize for sentences
- Tokenize for tweets
- Stemming (removing the letters 'ing', 'es', and's' from the end of words)
- Lemmatization is a kind of lemmatization (converting the words to their base forms)

B) Syntactic Analysis:

The study of the structure of a language in accordance with the guidelines of a formal grammar is referred to as syntactic analysis, syntax analysis, or parsing. Grammatical principles apply not to individual words but rather to categories and groupings of words rather than to individual words themselves. The meaning of a text can be better understood by first performing a structural analysis of the text.

"Truck is eating Oranges," for example, is illogical.

As a result, it is necessary to examine the intent of the words in a phrase. Among the approaches employed during this period are:

Parts of Speech (POS) tagging in Dependency Parsing

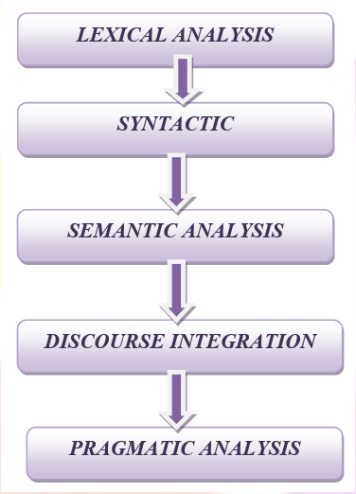

Figure 1: Phases of Natural Language Processing

C) Semantic analysis:

Semantic analysis eliminates or ignores statements that are illogical and extracts only meaningful information from the text. for instance, the sentence "Truck is eating Oranges" will be removed from the information summary,

D) Discourse integration:

Discourse integration helps with the study of the complete text because its scope is not limited to a single word or sentence. Example: "John got ready at nine in the morning. He afterwards hopped on a train headed for California.

E) Pragmatic analysis:

Robots must comprehend not just the text that is being delivered, but also the actual environment, at this challenging period. There are a number of circumstances when a phrase's intended meaning could be interpreted incorrectly by a computer if it lacks real-world experience. Example: "I apologise for the delay; the meeting has ended." (Sarcasm is present in this statement.)

3. PERFORMANCE CRITERIA

The evaluation metric that we use depends on the specific kind of NLP work that we are doing at the moment. Furthermore, the evaluation metric that we choose is influenced by the stage the project is now in. For instance, we might use a different evaluation

ISSN: 2349-7637 (Online)

measure throughout the model development and deployment phases compared to when the model is really being used in production, hence we might classify evaluation metrics in two distinct groups [5].

Accuracy(%)= (correction transliteration/total name entities)*100 **[1]**

Precision: When the reliability of the model's predictions is at issue, we talk about precision. The accuracy score would reveal what percentage of cases the classifier correctly classified as positive correspond to the positive labels.

$$Precision = \frac{TP}{TP + FP}$$ **[2]**

Recall

The model's recall assesses how it can recall the positive class

$$Recall = \frac{TP}{TP + FN}$$ **[3]**

F-MEASURES

Precision and Recall are inversely related complimentary measurements. If we are interested in both, we would utilise the F1 score to integrate accuracy and recall into a single statistic.

$$F1\ score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$ **[4]**

4. LITERATURE SURVEY

In many parts of the world, several approaches to translating languages and converting text to speech have been created. These approaches include voice recognition software. The development of a transliteration system that is capable of accurately converting words from the source language into the target language, regardless of whether or not the words in question are name entities, will serve as our major purpose here. In order to carry out our experiment, we employed a dataset consisting of proper names and geographical locations in both English and Punjabi. A great number of investigations have been carried out on the topic of machine transliteration. Kamal Deep and Goyal [2] the difficult challenge of transliterating Punjabi into English was ultimately conquered by employing a rules-based approach. The proposed transliteration strategy models the difficulty of transliteration using grapheme-based methods. Common Punjabi names were transliterated into English with a 93.22 percent success rate using this method. Sharma et al. [3] showed off Statistical Machine Translation's ability to transliterate English to Hindi using a variety of notations. By using Phrase-based statistical machine translation to train the English-Hindi corpus (of Indian names), the authors of this study showed that WX-notation transliteration is superior to UTF-notation. Kamal Deep et al. [4] a Mixed Method was designed for transliterating Punjabi into English. Utilizing a statistical method that is rule-based, this study attempted to transliterate Punjabi words into English. This was a challenging undertaking. The authors began with a letter-to-letter mapping and then attempted to apply statistical approaches in an effort to determine whether or not there were any improvements. Kaur and Josan [5] built a prototype to address the challenge of statistical methods for transcribing English into Punjabi. Statistical machine translation software MOSES is used for this English to Punjabi transliteration project. Using transliteration rules, they calculated an overall accuracy rate of 63.31 percent and a BLEU of 0.4502 for the system. Their results were then presented after this. Padda et al. [6] Describes an approach to transforming text written in Punjabi into IPA. In this investigation, they offered the methods that may be used to translate Punjabi text into IPA. In this study, they proved how a letter in Punjabi can be mapped directly to its corresponding symbol in the International Phonetic Alphabet (IPA), which represents sounds spoken in a language. Josan et al. [7] a novel approach to enhancing Punjabi-to-Hindi transliteration was proposed by combining a character-to-character mapping methodology with rule-based and Soundex-based enhancements. This resulted in the method's overall improved accuracy. The results of the experiments indicate that using this method results in a significant improvement in both the word accuracy and the average Levenshtein distance. Dhore et al. [8] dealt with the issue of machine transliteration, which requires a named entity in Hindi that is written in the Devanagari script to be transliterated into English using CRF as a statistical probability tool and n-gram as a features set. The problem was solved by using these two tools. They provided machine translation of named entities for the Hindi-English language pair by applying CRF as a statistical probability tool and n-gram as a feature set in this technique. This technique is known as "translation by machine." They succeeded in accomplishing an accuracy rate of 85.79 percent. Musa et al. [9] created an algorithm for the syllabification of the Malay language based on the matching of syllable rules. A innovative approach to syllabification for the Malay language was proposed and implemented by the writers of this study. Converting English to Arabic was made possible by Ameera Al-Rehili, Dalal Al-Juhani,

ISSN: 2349-7637 (Online)

Maha Al-Maimani, and Munir Ahmed's collaborative effort. It can translate from English to Arabic as well as the other way around [1]. Lakshmi Sahu is responsible for the conception and development of a method that can convert Hindi into Telugu and vice versa. This technology has the capability of pronouncing text in both Hindi and Telugu using male and female voices [10D. Sasirekha and E. Chandra's "Text to Speech: A Simple Tutorial" explains what is required to build a system that can convert text to speech. The report talks about the numerous TTS systems and how this industry is always evolving [11].Goyal V. Kamal Deep (2011) The accuracy of the rule-based Punjabi to English Transliteration System, developed by Kamal Deep and Dr. Vishal Goyal, was 93.23%. The transliteration system uses a grapheme-based technique to depict the transliteration problem. This approach addresses the issue of forward transliteration of human names from Punjabi to English by employing a set of character mapping rules. For Punjabi words, this method works well, but not for words from other languages.

5. COMPARITIVE STUDY OF PUNJABI AND ENGLISH

The Punjab region of India and Pakistan is home to millions of people who speak Punjabi, an ancient language. The Gurmukhi script is derived from the older Devanagri script. The term "Punjabi" is used to describe natives of Punjab or those who speak Punjabi as their native tongue.

Punjabi is a highly influenced language due to the influence of other languages such as Hindi, Urdu, Persian, and English. Because of its origins in Sanskrit, it is classified as an Indo-European language, along with several others spoken in North India. Multiple sources distinguish between western Punjabi and eastern Punjabi as distinct varieties of the Punjabi language. There are numerous scripts that can be used to write the Punjabi language. These writing systems vary by region, dialect, and religious affiliation of the speaker. Sikhs and followers of other faiths in the Indian state of Punjab often read their scriptures in Gurumukhi.

(i) Punjabi Grammar

Word order, case marking, verb conjugation, and other morphological and syntactical traits are just few of the aspects of the Punjabi language that are explored in the field of grammar. In Punjabi, SOV is the preferred word order (Subject Object-Verb). It uses postpositions rather than prepositions. Punjabi is a language with two genders, two sets of numerals, and five forms of case. The many kinds of cases are the nominative, genitive, accusative, ablative, and locative/instrumental [21]

(ii) Punjabi Word Classes and their Distinctions from English

This section covers the various word categories, inflectional patterns, and elements of a Punjabi sentence that exist in the language. In Punjabi, there are both inflected and uninflected versions of each word. A suffix that is known as an inflectional ending is frequently utilised when it is necessary to communicate grammatical relations such as number, person, tense, etc.
In English, reflexive pronouns can take either a demonstrative or personal meaning, however in Punjabi, this is not the case. For Example P: ਮੈਂਆਪਣਾਕੰਮਕਰਰਿਹਾਸੀ T: maiṃāpaṇākammkarrihāsīE: I was doing my work

(iii) Sentence

The clauses usually function as the skeleton of the sentences. In its most basic form, a sentence can only have one independent clause. Punjabi sentences follow an order known as SOV (Subject Object Verb), while English sentences follow an order known as SVO (Subject Verb Object). The subject comes first in a Punjabi phrase, followed by the object, and then the verb. Nominal phrases represent the clause's or sentence's subject and object, while verb phrases indicate the clause's or sentence's action.

(iv) Differences between English and Punjabi

The arrangement of the words in English and Punjabi is the main structural variation between the two languages. English is a language with a Subject-Verb-Verb structure, whereas Punjabi has a Subject-Verb-Object structure. For instance,

(v) The word order in English and Punjabi differs.

Whereas English has a rigid word order, Punjabi and the great majority of other Indian languages have flexible word orders. To put it another way, the arrangement of words in an English phrase may shed light on the connections between those words. The position immediately before the verb denotes the subject, and the position immediately after the verb denotes the object. However, unlike English, Punjabi does not rigidly adhere to word order.

6. APPLICATIONS

Official translation is a word that gives distinct languages and its attributes are based on languages. Some of the main applications are as follow:

1. Accuracy is the most important factor in a good translation. An accurate translation conveys the same meaning as the original text or makes every effort to come as close as feasible to the decided meaning. A translation may be flawed in

one of three ways: it may include information that was not intended, it may leave out information that the original author intended, or it may fundamentally change the original's meaning.

2. Clarity. An effective translation will make the message as clear and unambiguous as feasible. People won't read the translation and struggle to understand it.
3. Naturalness. A successful translation will read as though it were written originally in the target language. There will be no indication that a translator was used. The target language readers will assume it was originally written in at the time of publication. The cadence of any given language is based on its own unique combination of innate rhythms. A forced translation will read strangely and be difficult to comprehend. When people are not entertained by anything that has been translated artificially, they quit reading it or listening to it. For a translation to be useful and understood by its target audience, it must be as close to the original as humanly possible.
4. Acceptability. Historically, meaning-based translation emphasised solely the three attributes of a good translation: accuracy, clarity, and naturalness.

7. CONCLUSION

Finally, the importance of English to Punjabi translation cannot be overstated in terms of its impact on the success of communication, cultural interchange, and the facilitation of access to information and services for the Punjabi-speaking community. The constant development and improvement of the translation process, enabled by technological advancements and the abilities of skilled translators, makes cross-linguistic communication easier and more efficient in today's increasingly globalised society. In the near future, it is possible to implement a number of system enhancements. One possibility is implementing speech-to-text (STT) capabilities, which would eliminate the need for users to type in text and allow them to do it with their voice.

REFERENCES

1. Ameera Al-Rehili, Dalal Al-Juhani, Maha Al-Maimani and Munir Ahmed, “A NovelApproch to convert speech to Text and Vice-Versa and Translate from English to Arabic Language”, IJSAIT, Volume 1, No.2, May-June 2012. ISSN:2278-3083
2. Kamal Deep and Vishal Goyal, (2011) ”Development of a Punjabi to English transliteration system”. *InInternational Journal of Computer Science and Communication Vol. 2, No. 2, pp. 521-526.*
3. Shubhangi Sharma, Neha Bora and Mitali Halder, (2012) “English-Hindi Transliteration usingStatistical Machine Translation in different Notation” *International Conference on Computing andControl Engineering (ICCCE 2012).*
4. Kamal Deep, Dr.Vishal Goyal, (2011) “Hybrid Approach for Punjabi to English TransliterationSystem” *International Journal of Computer Applications (0975 – 8887) Volume 28– No.1.*
5. Jasleen kaur Gurpreet Singh josan , (2011) “Statistical Approach to Transliteration from English toPunjabi”, *In Proceeding of International Journal on Computer Science and Engineering (IJCSE), Vol. 3Issue 4, p1518.*
6. Er. Sheilly Padda, Rupinderdeep Kaur, Er. Nidhi, (2012) “Punjabi Phonetic: Punjabi Text to IPAConversion” *International Journal of Emerging Technology and Advanced Engineering Website:www.ijetae.com ISSN 2250-2459, Volume 2, Issue 10.*
7. Gurpreet Singh Josan, Gurpreet Singh Lehal, (2010) “A Punjabi to Hindi Machine TransliterationSystem*” Computational Linguistics and Chinese Language Processing Vol. 15, No. 2, pp. 77-102.*
8. Manikrao L Dhore, Shantanu K Dixit, Tushar D Sonwalkar, (2012) “Hindi to English MachineTransliteration of Named Entities using Conditional Random Fields.” International Journal of ComputerApplications;6/15/2012, Vol. 48, p31.
9. Musa, Hafiz, RabithA.kadir, Azreen Azman, M.taufikAbadullah, (2011) "Syllabification algorithmbased on syllable rules matching for Malay language." *Proceedings of the 10th WSEAS internationalconference on Applied computer and applied computational science*. World Scientific and EngineeringAcademy and Society (WSEAS).
10. Lakshmi Sahu, “Hindi & Telugu Text-to-Speech Synthesis (TTS) and inter-language text Conversion”,International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012. ISSN:2250-3153
11. D. Sasirekha, E. Chandra, “Text to Speech: A Simple Tutorial”, IJSCE, Volume 2, Issue 1, March 2012. ISSN:2231-2307
12. K. Deep, Dr. V. Goyal Hybrid Approach for Punjabi to English Transliteration System, International Journal of Computer Applications (0975 8887) Volume 28 No.1, August 2011.
13. 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.
14. K.K. Batra , G.S. Lehal , Automatic Translation System from Punjabi to English for Simple Sentences in Legal Domain

15. Vauquois, B. 1976. Automatic translation - a survey of different approaches. Statistical Methods in Linguistics (Stockholm). pp. 127-135.
16. S. Marinov. 2000. Structural Similarities in MT: A Bulgarian-Polish Case. Internet Source: http://www.gslt.hum.gu.se/~svet/courses/mt/termp.pdf. Accessed on Jan 21, 2011.
17. M. Corbí-Bellot, Mikel L. Forcada, Sergio Ortiz-Rojas, Juan Antonio PérezOrtiz, Gema Ramírez-Sánchez, Felipe Sánchez-Martínez, Iñaki Alegria, Aingeru Mayor, Kepa Sarasola. 2005. An open-source shallow-transfer Machine Translation engine for the Romance languages of Spain. In Proceedings of the Tenth Conference of the European Association for Machine Translation. May 30- 31. Budapest.Hungary. pp 79-86
18. J.González, A.L.Lagarda, J.R.Navarro, L.Eliodoro, A.Giménez, F.Casacuberta, J.M.de Val, &F.Fabregat. 2006. SisHiTra: a Spanish-to-Catalan hybrid Machine Translation system. LREC-2006: Fifth International Conference on Language Resources and Evaluation. 5th SALTMIL Workshop on Minority Languages: "Strategies for developing Machine Translation for minority languages", Genoa, Italy, 23 May 2006. pp. 69-73.
19. Shiu-Chang Loh, Luan Kong, & Hing-Sum Hung. 1978. Machine Translation of Chinese mathematical articles. ALLC Bulletin, Vol.6, 1978. pp. 111-120.
20. Ananthakrishnan R, Kavitha M, Jayprasad J Hegde, Chandra Shekhar, Ritesh Shah, Sawani Bade, Sasikumar M. 2006. MaTra: A Practical Approach to Fully- Automatic Indicative English-Hindi Machine Translation. In the proceedings of MSPIL-06.
21. Gill, Harjeet Singh and Gleason Jr, Henry A. 1969. A Reference Grammar of Panjabi. Patiala: Department of Linguistics, Punjabi University

  ISSN: 2349-7637 (Online)