International Journal of Advanced Engineering: Artificial Intelligence, Robotics and Automation
ISSN XXXX-XXXX
Vol.:1 Issue: 1
Published online on: Dec 2024
PP. 12- 21

Machine Translation Model for Hindi-English Poetry using Moses for Primary Level Education

Pratosh Bansal[1], Jayshri Bansal[2], Premanand Singh Chauhan[3], Rajesh Kumar Chakrawarti[4]
[1]Institute of Engineering &Technology, DAVV, Indore ,India
[2]Human Resource Development Centre,DAVV, Indore, India
[3]Vikrant University, Gwalior, India
[4]Sushila Devi Bansal College of Technology, Indore , India

Abstract —Primary Level Education is a place for every child where he/she learns the information pertaining to the languages, cultures and societies of the country. India is one of the best examples for possessing such kind of mass information. Most of the literatures of India are available in Hindi, Sanskrit and in native languages of various linguistic regions. The available literatures are in the form of either prose (textual) or verse (poetic) form and sometimes both. There are numerous translators (like Google Translate) available which can translate the prose (textual) form of literatures from Hindi to English correctly. However, correct translation of the verse (poetic) literature from Hindi to English is still an issue. Further, translation of Hindi poems into English must preserve the similar nature, emotions and behavior so that the children from another country may grasp the Indian flavor in a way similar to the Indian children. Through this research, efforts have been made towards the translation for Hindi poems at Primary Education level into English Poems. A translation model for Hindi-English Poetry Machine Translation has been built which shall be very helpful to develop learning material for the children at Primary Education level. This Translation Model has been built with the use of Monolingual and Parallel corpus and has been implemented in Machine Translation tool MOSES. Further, the developed model has been evaluated with BLEU (Bilingual Evaluation Understudy).

Keywords — Primary Education, Language, Hindi, English, Poetry, Machine Translation, MOSES.

I. INTRODUCTION

India is a multi-lingual, multi-cultural and multi-social country. This type of diversity is the oldest among the languages, cultures and societies available in various countries around the world. Nowadays, almost everyone, who is living at any location of the world, is impressed with the languages used or in vogue, cultures and social structure of India and even eager to learn them. Most of the literature of India is available in Hindi, Sanskrit and in native languages of various linguistic regions. The available literature is in the form of either prose (textual) or verse (poetic) form and sometimes both. Hence, before learning and understanding the languages, cultures and societies of India, the knowledge of spoken and written forms of Indian languages is essential. Specially, the knowledge of the Hindi language is must as it is the national language of India. In India, people have tried to translate most of the literatures in Hindi to provide a common platform for understating for every Indian. If anyone is interested to learn about Indian languages, cultures and societies, he/she must have the knowledge of at least Hindi language. In fact, it is not possible for everyone, to possess the knowledge of a foreign (second) language. Translation of Indian literature into English is essential as this will create an international platform for understanding the literature, culture and social environment of India for those who interested, all over the world. There are numerous translators (like Google Translate) available that can translate the prose (textual) form of literature from Hindi to English correctly. However, translation of the verse (poetic) literature from Hindi to English is still an unresolved issue. Nowadays, any information can be easily understood via poems rather than prose. If focus goes on imparting education to children at primary level, learning anything is very easy for them in the play way format by just singing and understanding the poems. That's why, if people across the countries around the world want to bring the social and cultural environment of India into their own places, learning must starts from their childhood or at least from primary level education. A lot of poetries have been written by the Indian poets for primary level educations which themselves express and provide the learning habits of languages,

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

cultures and societies of Indian cultures itself. Hence, for the others counties, the efforts for the translations of the Hindi poetries into English might play very important roles in the futuristic life of children at primary level educations by just learning those translated poetries by studying and singing [1][2].

II. LITERATURE SURVEY

Since last 03-04 decades, around the world, a lot or research have been done in the area of translations of the contents from one language to another language. Among these translations, the poetry translation is an important area, whose correct translation for one language (called source) to another language (called target) is still an issue. A number of Machine Translations systems have been developed for poetries. In India itself, prominent institutions and research organizations have worked on various research projects in the area of Machine Translation(MT) Systems and they are still working in the same direction. World's renowned IT industries, research laboratories and research organizations have also made tremendous efforts towards this. They have produced good Machine Translations. Observations can be done by analyzing the use of Google Translate, IBM Translator (Watson), Microsoft Translator (Bing), Worldlingo Translator, SYSTRAN Translator, Yandex Translate and many more. Mostly, these translators are working fine for the translations of literatures (texts) but correct translation of poetries are still issues. However efforts made by them can be accepted at utmost level and are helpful towards more and more research for development of extent level of Machine Translation of the poetries.

Diverse countries like India, a lot of historical information are available in the form of either literature (texts) or Hindi and Sanskrit Poetries. Ramcharitmanas (□□□□□□□□□□□) is one of the important and well known examples of Hindi Poetries called as Hindi-KaavyaShaastra (□□□□□-□□□□□□□□□□□□). To know and understand, the cultures, religion, languages etc. the knowledge of these kinds of poetries is very helpful. Hence, the translations of these poetries into other languages (especially into English language, because it is a global and internationally accepted language) might help in learning the cultures, religion, languages of India. Hence, before going for translation from Hindi to English Poetries, the knowledge of the Poetries, their types and structures might play very important role. A number of poets have defined the types and structures of verities of poetries. Poets have also tried to distinguish what are the major differences between the structures of Poetry written in Hindi as well as English language. A lot of information is already available for understanding the structures of Hindi and English language and their construction procedures.

To understand the cultures, religions, societies and many more of a particular country, the deep knowledge of foreign (second) i.e. the language used for that country is very essential. Even if discussion goes for transferring the same feelings and emotions, the knowledge of the Historical literatures and poetries might play an important role at here. Hence, to know the India, the knowledge on Hindi (at it is National language and also mother tongue) language is very important. But it is not an easy task to learn a foreign (second) language. Hence, translations of foreign (second) language into native (own) language might play important role. Keeping this in mind, a number of MT Systems have been developed for understanding the foreign (second) language e.g. a number of Hindi to English MT Systems are already available. Further, to understand the cultures, religions, societies of any diverse country like India with the same feelings and emotions by another country, it is better to teach the translated information to the children at primary level. Teaching of the translated poetry might be the biggest learning tool for the children at primary level to understand the cultures, religions, societies of any diverse country like India with the same feelings and emotions. Children learn anything majorly by singing anything. Learning cannot be based on the only singing of poetries; they may also be done by singing the Idioms, Couplets and other allied phenomenon as or when required by the instructor to children. If the children may grasp the information of the foreign (second) language during their primary level education, in near future, they must understand all the things related to the foreign (second) language easier during their high classes and higher level educations.

A number of tools have been developed around the world and working fine. At first, they work based on word, phrase, sentence, paragraph levels and sometimes discourse level. Secondly, they are built on the algorithms/techniques used for the translations like Rule-bases, Example-based, Pseudo-interligua-based, Statistical-based, Neural-based and most of the times Hybrids. Further, they are based on the learning approaches Knowledge-based, Corpus-based, Supervise-based, unsupervised-based and unsupervised based and sometimes hybrids. Every tool has its own merits and demerits. Among these tools, the *"MOSES"* is an important Machine Translation tool which is based on phrase/sentence-level (might work on upper level also), Statistical-based and Corpus-based approaches. It can translate most of the world-wide spoken languages (from one language to another). In today's scenario, most of the translations between various languages are happening with this.

Nowadays, most of the available MT Systems are from the Corpus based approaches. They may choose the corpus like Monolingual, Bilingual, Parallel and multilingual etc. depending upon the language and translation modeling characteristics. The MOSES is based on the Monolingual and Parallel corpus at where the Monolingual corpus is used for language modeling and the Parallel corpus is used translation modeling.

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

For Hindi to English Poetry Translation, the proper and true construction of Monolingual (English) and Parallel (Hindi-English) corpus, under the umbrella of the MOSES: Statistical Machine Translation, might provide the correct translation at utmost level. The poetries available (Hindi) language might be given to the MOSES with some level of pre-processing and for the translated poetries (English) the further post-processing may be applied for getting the proper and true translations of the poetries with same emotions and feelings.

The translated information through Machine Translation Systems can be further evaluated for the checking the quality of the translation with a number of tools. Among those Tools, BLEU (Bilingual Evaluation Understudy) is important one. Further, this tool has been considered for evaluation almost most of the translations which may done at around the world[1][2].

III. MACHINE TRANSLATION SYSTEM

A very important research and well-known area in the field of Natural Language Processing (NLP) is Machine Translation. Basically, it is an automated and a computerized system which is responsible for translating the information available in one language (called source) into another language (called target). A block diagram for a simple Machine Translation System has been shown in ***Fig. 1***.

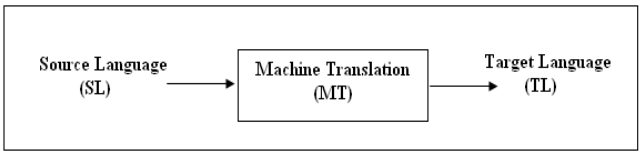

Fig.1. A simple Machine Translation (MT) system [1][2]

For the example, if Hindi and English are the Source and Target Languages for a Machine Translation System. Then, a statement written in Hindi Language *"□□□□□□ □□□□ □□□□□ □□।"* should be converted into English Language as *"Prithvi wants to sleep."*

Google Translate, IBM Translator (Watson), Microsoft Translator (Bing), Worldlingo Translator, SYSTRAN Translator, Yandex Translate and many others are very important examples of the MT Systems.

Researches in MT areas have been going on since several decades but developments of an efficient MT Systems are still a challenging task. Multicultural and Multilingual countries like India, the market are very big for the developments of such kind of MT Systems.

IV. POETRY

By definition, a poem must either singable or at least hummable. Essentially, Poem (/Poetry /Verse – □□□□□ /□□□□□ /□□□□) is a piece of writing which express feelings and thoughts with either silent or violent intensity as per the context. The words are chosen for the Poem as per its beauty and sound and then it is arranged systematically and carefully. These features separate a Poem (Verse-□□□□) from Prose (□□□□).

Although, over the years, certain rules of poetry structure has been developed depending upon their characteristics. But, presumably, all languages have their own rules of structure. Every language has its own structure of units for writing the poetries but commonalities are found among them which conclude that every poem has some kind of start, middle and end parts. For discussions here, most of the general characteristics (also known as the ingredients) of Hindi and English Languages have been discussed which are helpful in construction of beautiful poetries. These characteristics are also helpful writing poems in other languages and also for poetry translations from one language to another[3][4].

The following are the most common Characteristics of the poetries (nominations might be changes depending upon the languages):

a. Verse (□□□□□) and Free-verse {Mukta-Kavita (□□□□□- □□□□□)}
b. Syllable {Maatraa(□□□□□□)}
c. Rhyme {Tuk(□□□)}
d. Rhythm {Laya(□□)}
e. Meter
f. Form

V. HINDI AND ENGLISH POETRY

A unit of Hindi poetry consists of sthaayee or 'tek' (□□□□□□ □□ □□□), an antara (□□□□□), a samaapti (□□□□□□□). Sometimes a sthaayee is accompanied by another line sahyogee (□□□□□□) and it is also possible to have a sahyogee of samaapti.

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

The following is one of the famous poems in Hindi among children:

□□□□ □□ □□ □□□□ □□,
□□□□ □□□□ □□□□ □□ |
□□□ □□□□ □□ □□□□□,
□□□□ □□□□□□ □□ □□□□□ |

□□□□□-□□□□□□□□□□□□ {Hindi-KaavyaShaastra}, □□□□□□□□□□□ (Ramcharitmanas) written by Goswami Tulsidas, known for one of the greatest □□□□□□□□□□□□ for the all Indians (especially for Hindus). □□□□□□□□□□□ might have different kinds of poetries like □□□□□, □□□□ (couplet), □□□□□ and □□□□□ etc[3].

Similarly a unit of English poetry is called as Stanzas which is basically a collection of a series of lines which have been grouped together and separated with each other through an empty line from other stanzas. Other than stanza there are also few more nominations have been defied by the English language used for the components of the poetries.

The following is one of the famous poems in English among children is:

Twinkle, twinkle, little star,
How I wonder what you are.
Up above the world so high,
Like a diamond in the sky.

One of the important book entitled "The Oxford book of English Verse", 1250-1900, is one of the largest collection of English Poems which has been Chosen and Edited by A. T. Quiller-Couch[4].

The sample methodology of □□□□□□ (Maatraa) counts for the Hindi Poem and Syllable counts for the English Poem both have been displayed in ***Table 1*** and ***Table 2***. This is only one of the approaches but other than these there are other techniques which are helpful in understanding the counting methodology of □□□□□□ (Maatraa) and/or Syllables (these might be treated with different nominations for different languages) for poetries.

TABLE 1. HINDI POETRY AND □□□□□□ COUNT [5]

□□□□	□□	□□	□□□□	□□	Total
(1+1+2)	(1+1)	(2)	(2+2)	(2)	= 14
□□□□	□□□□	□□□□	□□		Total
(2+1+1)	(1+1+2)	(2+2)	(2)		= 14
□□□	□□□□	□□	□□□□□		Total
(2+1)	(1+2+2)	(1+1)	(2+2+2)		= 16
□□□□	□□□□□□	□□	□□□□□		Total
(2+1+1)	(1+2+2)	(1+1)	(2+2+2)		= 17

TABLE 2. ENGLISH POETRY AND SYLLABLE COUNT [6]

Twinkle	Twinkle	little	star			Total
(2)	(2)	(2)	(1)			= 7
How	I	wonder	what	you	are	Total
(1)	(1)	(2)	(1)	(1)	(1)	=7
Up	above	the	world	so	high	Total
(1)	(1)	(1)	(2)	(1)	(1)	= 7
Like	a	diamond	in	the	sky	Total
(1)	(1)	(2)	(1)	(1)	(1)	= 7

VI. TEACHING POERTY AT PRIMARY LEVEL EDUCATION

"Experts in literacy and child development have discovered that if children know eight nursery rhymes by heart by the time they"re four years old, they"re usually among the best readers by the time they"re eight."

The above single small paragraph is itself enough to prove that how much the poetry teaching is important for the children at primary level education? Still, more discussions and information are essential for the importance of the poetries that how they are helpful in the overall development of the children. Poetries can play the significant roles in lives. Poetry builds resilience in kids and adults both. Poetries might be considered and treated as central part of the teaching of languages and literatures in schools, colleges and universities. The central idea of teaching poetry is basically to provide the knowledge and expand understanding of almost all human beings. That's why, most of the times, Poetry is used as a primarily and important tool for teaching to overall and systematic

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

development of a child rather than just considering it as a medium of exploring experience only. Although, the April is known as National Poetry Month, but, poetries can be learn any time or place besides specific one [7].

The following are the important developments which can be found in a child studying at primary level education and learning the poetries.

(i) *Personal development,*
(ii) *Build reading and writing skills,*
(iii) *Build the speaking and listening skills,*
(iv) *Promotes literacy.*
(v) *Language development,*
(vi) *Physical development,*
(vii) *Memorizations and emotional developments,*
(viii) *Building Community and Social developments,*
(ix) *Cognitive and affective learning.*

Definitely, the all developments are very-very important for the children for their overall developments [8].

VII. FOREIGN(SECOND) LANGUAGE LEARNING

If anyone, who might not write or speak a foreign language (second language) just like as native (first) speakers of that language, they can be consider as a category of learners for foreign (second) languages. Around the world, English is considered as an international language, hence, information available in other languages need to be translated into this for every kinds of communications like personals, educational, commercial, or for so many reasons. Generally, it can be found that there are not so many native speakers of English who might also know an Indian language say Hindi like the most native speakers of Hindi. In this situation, people have to undertake the task of translating into English. Even, if talk about the native speakers of Hindi, there is only a only a small group of people who will have native speaking like ability in English language as well and, definitely, even most of them will have the different proficiency levels in learning of English. This is enough to show and clarifying that there is a big demand in among the students all over the world to learn English as a foreign (second) language. Further, it might be easily observed that knowledge of a learner for the foreign (second) language is irregular rather than systematic. Again, being a learner of a foreign (second) language means the language spoken by such a person is always a kind of 'inter-language'. Generally, it can also be observed that the target language spoken by the learners of that language is always "defective", or incomplete, in few ways i.e. it does not correspond to the language normally used by adult native speakers.

The term language learning is vague enough as it may ask either as "about the ability to use the language" or as "about the knowledge of the language". The case one asks the ability to use the language i.e. demand for satisfying the criteria for comprehension level as the adult native speakers of that language. On contrast, the case second is simple as it asks only the knowledge of the language which might be either simple or comprehension levels or in between.

The Translation from one language to another requires knowledge of two languages. At the same place, one could say it as either easier or more difficult in defining. It can be consider it easier in defining because the tasks of machine translation is only for a small subset of the whole (involving language) and thus it is very easier to separate and go for the study. On the other hand, it is might be difficult to defining, because it involves two languages, the problems of defining translation, the union of the problems of defining the linguistic competence of each language and many more from certain aspects [9-13].

VIII.MOSES: MACHINE TRANSLATION TOOL

Moses is a Statistical Machine Translation (SMT) tool which allows training translation models automatically for any language pair. It is an implementation of the statistical (or also so called data-driven) approach of machine translation. It is also employed by the online translation systems deployed by Google and Microsoft.

In SMT, translation systems are trained on large quantities of parallel corpus (from which the systems learn how to translate small phrases or segments) as well as even larger quantities of monolingual corpus (from which the systems learn what the target language should look like). The training process in Moses takes in the parallel data and uses co-occurrences of words and phrases to infer translation correspondences between the two languages of interest [14].

The main components in Moses are:

A. *Training Pipeline:* The training pipeline in ***Fig. 2***. is really a collection of tools (mainly written in perl, with some in C++) which take the raw Corpora (parallel and monolingual) and turn it into a machine translation model.

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

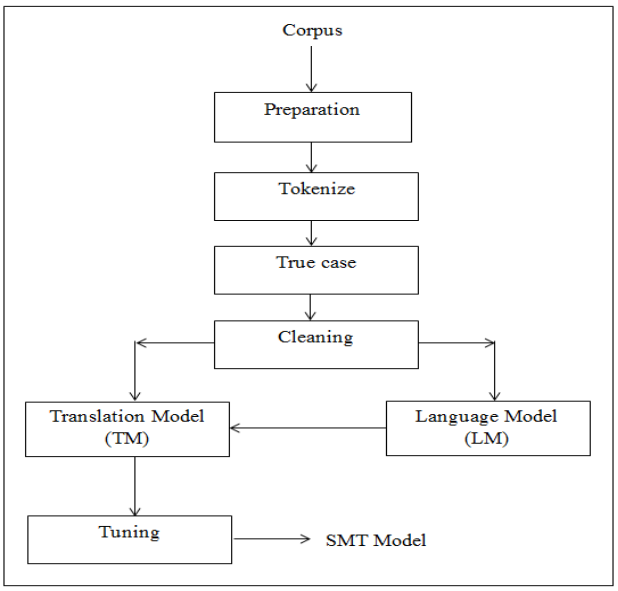

Fig. 2. Training Pipeline through Moses SMT [15].

There are various stages involved in producing a translation system from training data. The data typically needs to be *prepared* before it is used in training, tokenizing the text and converting tokens to a standard case. Heuristics are used to remove sentence pairs which look to be misaligned, and long sentences are removed. The parallel sentences are then *word-aligned*. These word alignments are used to extract phrase-phrase translations. An important part of the translation system is the *Language Model (LM)*, a statistical model built using monolingual data in the target language and used by the decoder to try to ensure the fluency of the output. The final step in the creation of the machine translation system is *tuning* where the different statistical models are weighted against each other to produce the best possible translations. Moses contains implementations of the most popular tuning algorithms.

B. Decoder: The decoder in ***Fig.3***. is a single C++ application which, given a trained machine translation model and a source sentence, will translate the source sentence into the target language.

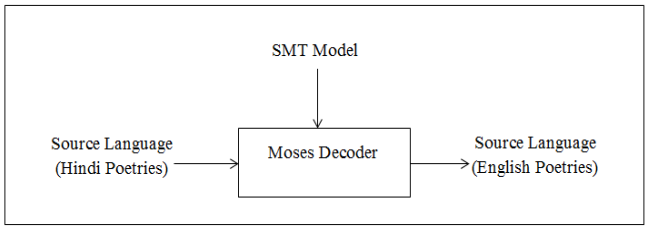

Fig. 3. Moses Decoder [15]

The job of the Moses decoder is to find the highest scoring sentence in the target language (according to the translation model) corresponding to a given source sentence. It is also possible for the decoder to output a ranked list of the translation candidates, and also to supply various types of information about how it came to its decision (for instance the phrase-phrase correspondences that it used).

Other than these two, there are also a variety of contributed tools and utilities. There are also tools to evaluate translations, alternative phrase scoring methods, an implementation of a technique for weighting phrase tables, a tool to reduce the size of the phrase table, and other contributed tools [15].

IX. CORPUS

A collection of linguistic data, either compiled as written texts or as a transcription of recorded speech. A corpus is defined here as a principled collection of naturally occurring texts which are stored on a computer to permit investigation using special software. A corpus is principled because texts are selected for inclusion according to pre-defined research purposes. Usually texts are included on external rather than internal criteria. Here are some examples of some of the different types of corpora and how they represent a particular variety:

(i) Annotated Corpus: An annotated corpus may be considered to be a repository of linguistic information, because the information which was implicit in the plain text has been made explicit through concrete annotation.

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

(ii) *Comparable Corpus:* A type of corpus used for comparison of different languages.

(iii) *Monitor Corpus:* A type of corpus which is a growing, non-finite collection of texts, of primary use in lexicography. Monitor corpus reflects language changes in a constant growth rate of corpora, leaving untouched the relative weight of its components (i.e. balance) as defined by the parameters.

(iv) *Monolingual Corpus:* A type of corpus which contains texts in a single language.

(v) *Multilingual Corpus:* A type of corpus which represents small collections of individual monolingual corpora (or sub-corpora) in the sense that they use the same or similar sampling procedures and categories for each language but contain completely different texts in those several languages (for two languages bilingual corpus).

(vi) *Parallel (Aligned) Corpus:* A type of multilingual corpus where texts in one language and their translations into other languages are aligned, sentence by sentence, preferably phrase by phrase.

(vii) *Reference Corpus:* A type of corpus that is composed on the basis of relevant parameters and should include spoken and written, formal and informal language representing various social and situational strata.

(viii) *Spoken Corpus:* A type of corpora that contain texts of spoken language.

(ix) *Speech Corpus:* A large collection of audio recordings of spoken language. Most speech corpora also have additional text files containing transcriptions of the words spoken and the time each word occurred in the recording.

(x) *Unannotated Corpus:* A type of corpora that are in raw states of plain text; opposed to annotated corpora [16][17].

X. PARALLEL AND MONOLINGUAL CORPUS

As the Moses SMT uses the Parallel Corpus and Monolingual Corpus and the present system is responsible for translations of Hindi poetries into English. Hence, a Parallel Corpus for *English* translations (EngPrlTL.en) of *Hindi* Poetries (HinPrlSL.hi) has been prepared which contains around 5,000 lines of poetries. Also a Monolingual Corpus for *English* Poetries (EngMonoLM.en) has been prepared which contains around 10,000 lines of poetries [16][17].

XI. POETRY TRANSLATION USING MOSES

The complete machine translation model with step by step process using Moses SMT tool for Hindi Poetries into English Poetries has been depicted in the ***Fig. 4***. This process can be assumed the complete explanation or implementation of the main components of Moses discussed previously i.e. Translation Pipeline and Decoders and all. The process may increase or decrease the steps depending upon the natures of the data. However, the following 09 important steps have been specified here properly which might be sufficient for the true translation of Hindi Poetries into English Poetries:

1. *Data Categorisation:* Initially, the Parallel corpus is divided into four categories: Training, Development, Testing and Evaluation Data. For example, out of 100% of parallel corpus, 90% of both Hindi and English poetries both have been chosen for Training purpose, 5% of both have been chosen for Development purpose and among the remaining 5%, Hindi Poetries have been chosen for Testing and English Poetries have been chosen for Evaluation purpose.
2. *Tokenisation:* All categories of parallel corpus as well as 100% of Monolingual corpus should have to be tokenized. This means that spaces have to be inserted between (e.g.) words and punctuation.
3. Truecasing: Further, All categories of parallel corpus as well as 100% of Monolingual corpus sent for truecasing in which the initial words in each sentence are converted to their most probable casing. This helps reduce data sparsity.
4. *Cleaning:* After the Truecasing, the Training Data is sent for cleaning in which Long sentences and empty sentences are removed as they can cause problems with the training pipeline, and obviously mis-aligned sentences are removed.
5. *Language Model (LM):* The language model is used to ensure fluent output, so it is built with the target language (i.e Pure English Poetries Monolingual corpus in this case). The IRSTLM is used to build an appropriate 3-gram language model which also has a binary format that Moses supports.
6. *Translation Model (TM):* Finally the main event is training the translation model (built on the Training Data i.e. 90% of Parallel Corpus). To do this, word-alignment (using GIZA++), phrase extraction and scoring, create lexicalised reordering tables and create Moses configuration file.

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

7. *Tuning:* This is the slowest part of the process which might want to line up something to read whilst it's progressing. Tuning requires a small amount of parallel data (here is called development data which is 5% of Parallel Corpus), separate from the training data.
8. *Testing:* In this step, the decoder can be tested by translating the Testing Data i.e. 5% of Hindi Poetries of the Parallel Corpus as Input into the English Poetries as Output.
9. *Evaluation:* Finally, the for evaluating the English Poetries produced as Output during Testing Phases, It is passed over BLUE script along with the Evaluation Data i.e. 5% of English Poetries of the Parallel Corpus. It generates the BLUE Score which has been discussed in details in the proceeding section [14][15].

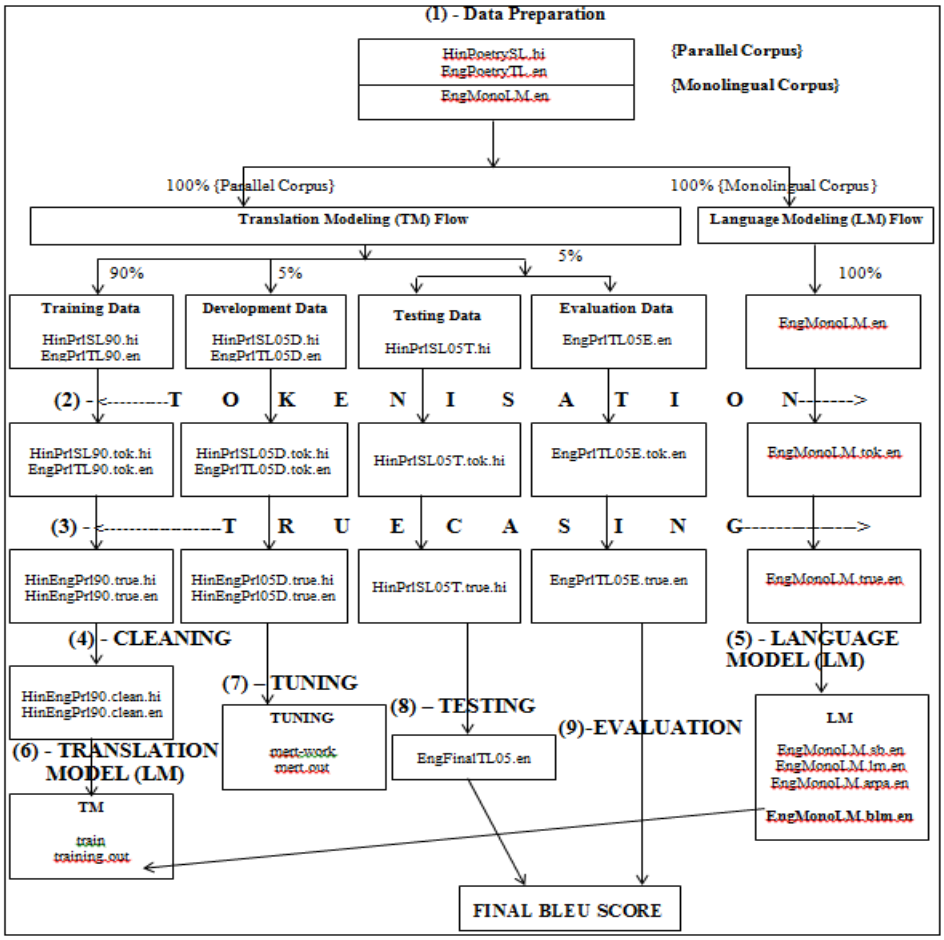

Fig. 4. Machine Translation Model for Hindi-English Poetry using Moses for Primary Level Education [14][15].

XII. BLEU: MACHINE TRANSLATION MEASUREMENT

The Bilingual Evaluation Understudy (BLEU) score had been developed for evaluations of the outcomes made by the developed automatic machine translation systems. It is a metric for evaluating the translated (output) statement(s) with the reference statement(s). The BLEU Score is measured in the range of 0.0 to 1.0. If there is an imperfect match between the reference statement(s) to the translated (output) statement(s), the BLEU results score is 0.0 and for a perfect match the BLEU results score is 1.0. BLEU score has significant benefits in terms of measurements like [18][19]:

- *Fast and cost-effective in translation.*
- *Easily understandable.*
- *Language is no bar.*
- *Behaviors are just like as human evaluations.*
- *Widely used.*

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

XIII. Result and Discussion

The prepared Machine Translation (MT) model, for translation of Hindi Poetries into English language from the children at primary level education, can be assumed working fine at utmost level. It is translating the Hindi poetries into the English significantly. With Parallel Corpus (around 5,000 lines of *English* translations aligned with *Hindi* Poetries) and Monolingual Corpus (around 10,000 lines of English Poetries), the Machine translation model has given BLEU Score of 0.0365 for a small set of Hindi Poetries (around 250 lines). If the BLUE Score is analyzed then it looks like that the quality of Translation is not up to a level. But it's not true because if the translated poetries are looked up then the translation level is good but as the size of the Parallel Corpus and the Monolingual Corpus both is very less that why the BLEU Score is less. If there will be large sized the Parallel Corpus and the Monolingual Corpus both then the BLEU scores will be slightly increase. Further if large sized the Parallel Corpus and the Monolingual Corpus will be there then the Large sized (also called as the large shallow depth) Hindi also might be translated into the English Poetries. Again, with this extension, definitely, the quality of the translation will be enhanced and precision will be improved also.

Further, if the translated poetries into English are analyzed then their nature, emotion and behavior looks like similar to the Hindi Poetries. Hence, these might also helpful and lead to the cultural transformation of India at around the world.

XIV. Conclusion and Future Scope

Although the prepared Machine translation model, for translation of Poetries from Hindi to English language from the children at primary level education, is working fine with the used parallel corpus (around 5,000 lines of *English* translations aligned with *Hindi* Poetries) and the monolingual corpus (around 10,000 lines of English Poetries), however, firstly, it is not producing a good BLEU score. Secondly, it will not work properly for the large shallow depth Hindi poetries. Thirdly, quality of translation not might be considered with utmost level. At last, its precision is also not so good.

Hence, till now it could be assumed that the Machine translation model, for translation of Poetries from Hindi to English language from the children at primary level education is prepared and developed properly and working fine at utmost or so called acceptable level. But, on contrast, in considering the four (04) important concerns i.e. BLEU Score, Large Shallow depth, Translation Quality and Precision, the size of the parallel corpus and monolingual corpus both must increase significantly. For example, if increase of 10% in BLUE score is expected then the corpus must increase with 10% i.e. Parallel corpus and monolingual corpus must have around 50,000 and 1,00,000 lines of poetries respectively. Furthermore, little bit additional study is essential for true preparations of both corpuses i.e. parallel corpus as well as monolingual corpus.

References

[1] R.K. Chakrawarti and P. Bansal, "Hindi-To-English Machine Translation System for Primary Education", *Doctoral Conference (DocCon-2016), Janardan Rai Nagar Rajasthan Vidyapeeth University, In collaboration with CSI and ACM Udaipur Chapter*, Udaipur, Rajasthan, March 2016.

[2] R. K. Chakrawarti and P. Bansal, "Approaches for Improving Hindi to English Machine Translation System", Indian Journal of Science and Technology, vol. 10 no. 16, April 2017, ISSN 0974-6846. DOI: 10.17485/ijst/2017/v10i16/111895

[3] Hindi Poetry Structure [Online] available at: https://www.kaavyaalaya.org/hindipoetrystructure2.

[4] English Poetry Structure [Online] available at: https://www.howmanysyllables.com/howtocountsyllables

[5] 'GeetGatiroop', Syllable counter for Hindi Poetry [Online] avaialble at: http://manaskriti.com/geet-gatiroop/

[6] 'SyllableCounter', Syllable counter for Hindi Poetry [Online] avaialble at: https://syllablecounter.net/

[7] Karen Simecek & Kate Rumbold (2016), The Uses of Poetry, Changing English, 23:4, 309-313, DOI: 10.1080/1358684X.2016.1230300

[8] Importance of teaching poetries at primary level education: [Online] avaialble at:
https://proudtobeprimary.com/reasons-teach-poetry-classroom/
http://musicearlychildhoodpresenter.com/blog/why-poetry-is-great-for-children/
http://timbuktu.me/blog/why-do-children-love-poems/
https://www.edutopia.org/blog/five-reasons-poetry-needed-schools-elena-aguilar

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*

[9] Shei, Chi-Chiang. "Combining Translation into the Second Language and Second Language Learning: An Integrated Computational Approach", *Doctor of Philosophy*, University of Edinburgh 2002.

[10] R. K. Chakrawarti, H. Mishra and P. Bansal, "Review of Machine Translation Techniques for Idea of Hindi to English Idiom Translation", International Journal of Computational Intelligence Research, vol.13, no. 5, May 2017, ISSN 0973-1873.

[11] H. Mishra, R. K. Chakrawarti and P. Bansal, "A new approach for Hindi to English idiom translation", 3rd International Young Scientist Congress (IYSC-2017), Ganpat University, Mehsana Gujrat, 08-09 May 2017.

[12] H. Mishra, R. K. Chakrawarti and P. Bansal, "Implementation of Hindi to English Idiom Translation System", In: Kamal R., Henshaw M., Nair P. (eds) *International Conference on Advanced Computing Networking and Informatics. Advances in Intelligent Systems and Computing, vol. 870,* Springer, Singapore, November 2018.

https://doi.org/10.1007/978-981-13-2673-8_39

[13] A. Yadav, R. K. Chakrawarti and P. Bansal, "Couplets Translation from English to Hindi Language*", International Conference on Social Networking and Computational Intelligence (SCI-2K19),* Springer, Rajiv Gandhi Proudyogiki Vishwavidyalaya and Sagar Institute of Research and Technology, Bhopal (MP), India, Mar 2019. "Lecture Notes in Networks and Systems", ISSN: 2367-3370.

[14] Koehn, Philipp. "MOSES: Statistical Machine Translation System, Machine Translation System User Manual and Code Guide." (2011).

[15] MOSES: Statistical Machine Translation System [Online] available at: http://www.statmt.org/moses/index.php?n=Main.HomePage

[16] TDIL-Technology Development for Indian Languages: available at: https://tdil-dc.in/index.php?option=com_vertical&parentid=58&lang=en

[17] David Evans, Irina Dahlmann and Juliet Herring, "Corpus building and investigation for the Humanities", University of Nottingham and University of Birmingham, available at: https://www.birmingham.ac.uk/research/activity/corpus/publications/introduction-corpus-investigative-techniques.aspx

[18] D. Genzel, J. Uszkoreit and F. Och, *"Poetic" Statistical Machine Translation: Rhyme and Meter, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, MIT, Massachusetts, USA, 9-11 October 2010*. 2010 Association for Computational Linguistics, pp. 158–166.

[19] K. Papineni, S. Roukos, T. Ward and W.J. Zhu, "BLEU: a method for automatic evaluation of machine translation", *In Proceedings of the 40th annual meeting on association for computational linguistics*, Association for Computational Linguistics. pp. 311-318, July 2002.

Corresponding Author's E-mail ID: *rajesh_kr_chakra@yahoo.com*