

Urdu to English Machine Translation using Bilingual Evaluation Understudy

Asad Abdul Malik
Kohat University of Science and Technology,
Kohat, Pakistan

Asad Habib
Kohat University of Science and Technology,
Kohat, Pakistan

ABSTRACT

Machine Translation (MT) is exigent because it involves several thorny subtasks such as intrinsic language ambiguities, linguistic complexities and diversities between source and target language. Usually MT depends upon rules that provide linguistic information. At present, the corpus based MT approaches are used that include techniques like Example Based MT (EBMT) and Statistical MT (SMT). In addition to others, both of these corpus based techniques have different frameworks in the contemporary data-driven paradigm. SMT systems generate outputs using probabilities, whereas EBMT systems translate input text by matching examples from large amount of training data.

Urdu MT is in its infancy with very limited availability of required data and computational resources. In this paper, we analyzed and evaluated the main MT techniques using qualitative as well as quantitative approaches. Strengths and weaknesses of each technique have been brought to light through special focus and discussion on examples from Urdu language MT literature.

We evaluated the automated machine translated outputs using Bilingual Evaluation Understudy (BLEU). The EBMT approach produced the highest accuracy of 84.21% whereas the accuracy of the online SMT system is 62.68%. We found that BLEU scores of machine translated long Urdu sentences are low in comparison with long sentences. Similarly source text containing low frequency words affect the quality of Urdu machine translation negatively. Experiments and findings section of this paper explicate our reported results in detail. The paper concludes with proposal of future directions for research in Urdu machine translation.

Keywords

Machine Translation Comparison, Rule Based Machine Translation, Statistical Machine Translation, Example Based Machine Translation, Bilingual Evaluation Understudy, Urdu to English Machine Translation.

1. INTRODUCTION

Translation is as old as written literature. It is the art of representing text or speech of one natural language (source language) into another (target language). Currently the importance of translation is constantly increasing in different fields including education, business and medical etc. [3]. The technological advancements coupled with the application fields of translation have led us to the stage of automated translation that is commonly known as Machine Translation (MT). It is automated translation from source language to target language using computer or any other system. In order to achieve good quality in MT, a number of techniques can be employed e.g. the computer(s) may contain 1) annotated or un-annotated lexicons 2) programs required to make logical

selection based on semantics and 3) algorithms to supply missing words or rearrange word order as needed by the target language [2].

Urdu ranks 19th among 7,105 languages spoken in the world [1]. This is mostly spoken in the South Asia and in some parts of the Western world. Urdu is the national language of Pakistan. It is used as the medium of instruction in public sector schools. A considerable number of daily, weekly and monthly newspapers are published in Urdu¹. Electronic media broadcasts news, sports and entertainment etc. in Urdu². It is also used for official purposes in junior and middle level administration [19]. Urdu is not only spoken in Pakistan but also in India, Bangladesh, Afghanistan and Nepal. It has become the culture language and lingua franca of the South Asian Diaspora outside the Indo-Pak subcontinent, mainly in the Middle East, Europe, Canada and the United States [4].

In spite of large number of speakers around the world, there are very few computational natural language tools available for Urdu. We could not find any public domain machine translation tool developed specifically for Urdu. However some literature of basic MT techniques has been discovered [5-9]. In the current work a detailed survey is presented on the contemporary research in Urdu Machine translation (UMT). The weaknesses and strengths of each technique are identified. Output of each MT system is evaluated using Bilingual Evaluation Understudy (BLEU) and the guidelines are proposed for future directions in UMT research.

The remainder of paper is organized as follows. Section 2 is about the literature review where we discuss the related work and overview of three main techniques of MT. These include Rule Based MT, Statistical MT and Example Based MT. Section 3 explains the methodologies of each of these techniques whereas their respective comparison is outlined in Section 4. Section 5 explicates the BLEU technique and its calculation method. Experiments, findings and evaluation are mentioned in section 6. A brief discussion about the current work is outlined in section 7. The paper is concluded and future directions of this research are mentioned in Section 8.

2. LITERATURE REVIEW

Machine Translation is mainly divided into three techniques as shown in Figure 1. Naila et al presented a Rule Based Machine Translation (RBMT) technique for translation from English to Urdu [5]. It is primarily based on the transfer approach. In transfer approach sets of linguistic rules are used that are defined as correspondence between the structure of source and target language. The above mentioned RBMT system handles case phrases and verb postpositions using

¹ <http://jang.com.pk>, <http://www.express.com.pk/epaper/>

² <http://urdu.geo.tv>, <http://dunyanews.tv/>, <http://www.samaa.tv/>

Paninian grammar. Another common technique for MT is Statistical Machine Translation (SMT). It uses probabilities for translating text from one language to another. Bushra et al used SMT technique to investigate issues in machine translation between languages with significant word order differences [6]. The third common MT technique is the Example Based Machine Translation (EBMT) that translates input text by matching examples from large amount of training data. Maryam and Zafar presented Example based approach that translates text form English to Urdu. It supports idioms, homographs, and some other linguistic features [7]. Parallel corpus for statistical machine translation for English text into Urdu was presented by Aasim et al [8]. “Word Order Issues in English to Urdu Machine Translation” was presented by B. Jawaid and D. Zeman [9]. M. Zhang and H. Li investigated issues related to phrase reordering [20]. In addition to the aforementioned techniques, SMT systems such as Google and Bing are available online [10-11]. However they provide limited accuracy due to the intrinsic issues in Urdu Machine Translation.

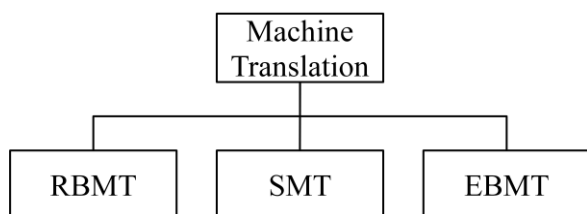


Fig 1: Paradigms of Machine Translation

Brief description on each of these paradigms of MT is given in the following section.

2.1 Rule Based Machine Translation (RBMT)

It is one of the main techniques in MT research. RBMT requires extensive linguistic knowledge for producing proper policies required for translation. This system depends upon different linguistic levels of policies for translation from source language to target language. Translation depends on formalized linguistic knowledge base that is represented in lexicon along with grammars [5].

A rule-based technique is characterized by numerous characteristics. It requires strict sense of fine tuned rules. Grammatical mistakes are strictly forbidden which means that the input data must be grammatically correct. Many rules depend upon contemporary linguistic theories. On the contrary, if the required knowledge is not available in any literature, then ad-hoc heuristic policies are used. The main benefit of rule based systems is that they allow integrating present linguistic knowledge into the translation system directly [5].

Most of the present age MT software in market are rule-based [13]. Typically, a RBMT system contains; 1) input sentence analyzer (morphological, syntactic and semantic analysis) and 2) procedure for producing sentences as outcome of a number of structural transfers based on inner structure or a few Interlingua.

2.2 Statistical Machine Translation (SMT)

The SMT systems exploit probabilities and other stochastic techniques for analysis and translation of text from source to

target language. This term occasionally refers to the use of probability-based methods in parts of MT tasks, as in word sense disambiguation or structural disambiguation. In addition the SMT also refers to pure stochastic-based system that uses probabilistic models for determining the accurate output translation of the input text [13].

In this technique, typically two statistical models are built i.e. translation model and the language model. A translation model gives probability of a target sentence given the source sentence $P(T/S)$ that is calculated by using word level aligned bilingual corpus. A language model determines the probability $P(S)$ of the string of target language actually occurring in that language. By using the language model and conditional probabilities of translation model, $P(S/T)$ is calculated using the following formula:

$$P\left(\frac{S}{T}\right) = \frac{P(S)P\left(\frac{T}{S}\right)}{P(T)}$$

Brown et al discussed this modeling process in greater detail [15]. This technique does not require explicit encoding of the core linguistic information. On the other hand, it heavily depends upon the availability of fine and large amount of bilingual data [15].

2.3 Example Based Machine Translation (EBMT)

Somers called EBMT a hybrid translation technique of RBMT and SMT systems [16]. Similar to SMT, it is depended on corpus of available translations, which are reused as a base for translation. That is why it is similar to, and sometimes confused with, the translator's aid vastly recognized as Translation Memory (TM).

EBMT and TM both involve comparing input text with the database of real examples, and then find out the nearest match. The difference between TM and EBMT is that in TM the translator decides 'what should be done with proposed match' whereas in EBMT automatic process is carried out to identify the equivalent translation fragments. Re-combination of these fragments gives us the desired target text [2][5].

The EBMT splits the process of translation into three phases. 1) Matching of fragments against the available database of real examples. This part is common between EBMT and TM 2) Alignment for identifying corresponding translation fragments and 3) Re-combination that generates the target text in surface form of the target language. A major requirement of EBMT is a database of parallel translations that are searched for source language sentences and phrases. If exact match is not found then nearest matching is performed [13]. The matched phrases are then modified and combined together to generate a transfer translation of new sentence. The closeness of a match will be determined by semantic differences between two content words as calculated by some metric based on thesaurus or ontology. Accuracy and quality of translation highly depends on the length of input sentences and the coverage of different linguistic aspects in the parallel database [13].

EBMT saves the translated examples in different manners. In simple case, examples are saved as pairs of strings with no extra information. Sometime a technique called indexing is used for this purpose. This technique is basically borrowed from the field of Information Retrieval (IR) [14].

3. METHODOLOGY

In this section, methodologies of the three Machine Translation techniques are discussed.

3.1 Rule Based Machine Translation (RBMT)

There are three stages in RBMT i.e. analysis, transfer and synthesis as shown in Figure 2.

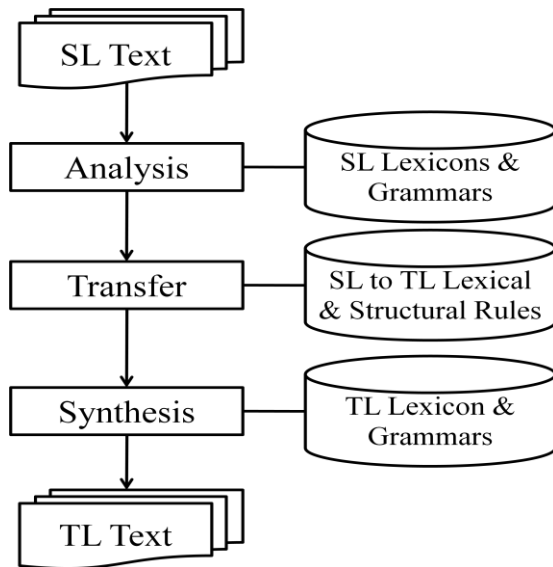


Fig 2. RBMT Model

3.1.1 Analysis

In this stage, the source language input text is analyzed by grammar and lexicon of source language. The sentence is then divided into noun phrase (NP) and verb phrase (VP). These two parts can further be divided depending upon the structure of sentence. And thus source language parse tree is created. Considering the source language is English, the parse tree of the input sentence "I called you several times" is generated as shown in figure 3.

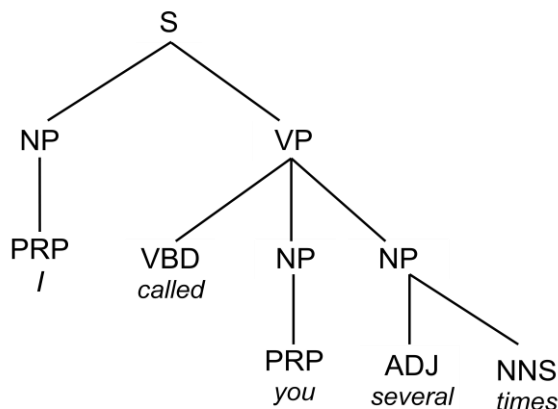


Fig 3. English Parse Tree (SVO)

3.1.2 Transfer

The parse tree of source language is 'transferred' to parse tree of target language according to the structural rules and lexicon of target language as shown in figure 4.

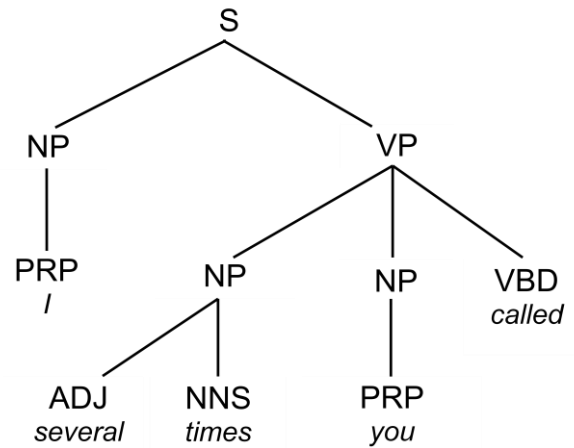


Fig 4. Parse Tree (Transferred in SOV)

The syntactical structure of English sentence is Subject-Verb-Object (SVO) and that of Urdu is Subject-Object-Verb (SOV). This means that the 'transfer' of parse tree from English to parse tree of Urdu will require recursive swapping. Subsequently re-ordering might be required to convert SVO sentence into SOV sentence. However this step would not be required at every step. For example NP in both source (English) and target (Urdu) languages follows the same rule and swapping would not be required in case of NP.

English $NP \rightarrow NP + PP$
Urdu $NP \rightarrow PP + NP$

However, if NP is having NP and PP, then we need to transform it because PP appears before NP in Urdu.

Urdu $AP + V$

If adverb phrase (AP) appears before verb then swapping is not needed. AP in English can appear in different order depending on the type of AP. Syntactically, Urdu is a partially free word order language. However, changing order of words can change the stress or focus of sentence semantically. In general though, Urdu prefers AP before verb for agreement between NP and VP of sentence S.

VP in Urdu is inflected according to gender, number and person (GNP) attributes of the head noun while form of NP depends upon tense aspect and modality of the verb phase (VP). Similarly, Urdu adjectives are modified by GNP attributes of the head noun. All the aforementioned and some additional rules must be taken into consideration in order to 'transfer' the parse tree of SVO language into SOV language.

3.1.3 Synthesis

In this stage, the target language lexicon and syntactical grammar is used to convert the parse tree (generated in previous step) to the target language surface form. Two independent monolingual dictionaries are required to generate appropriate surface form of target language.

As shown in figure 5 the source text "I called you several times" is translated into "میں نے کئی مرتبہ آپ کو بلایا" using RBMT.

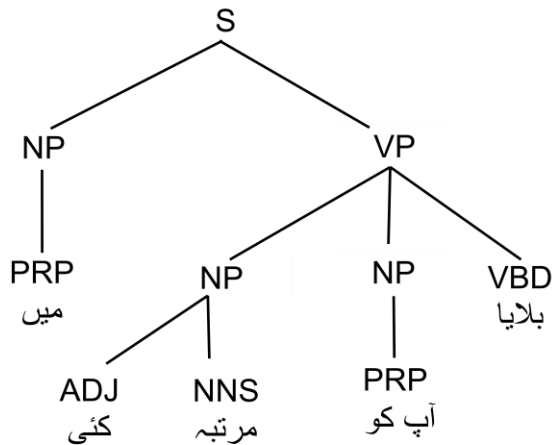


Fig 5. Urdu Parse Tree

3.2 Statistical Machine Translation (SMT)

In SMT techniques, translation from source to target language is done through the use of probability distributions and stochastic models. A typical SMT systems comprises of three stages namely translation model, language model and the decoder Algorithm as shown in Figure 6.

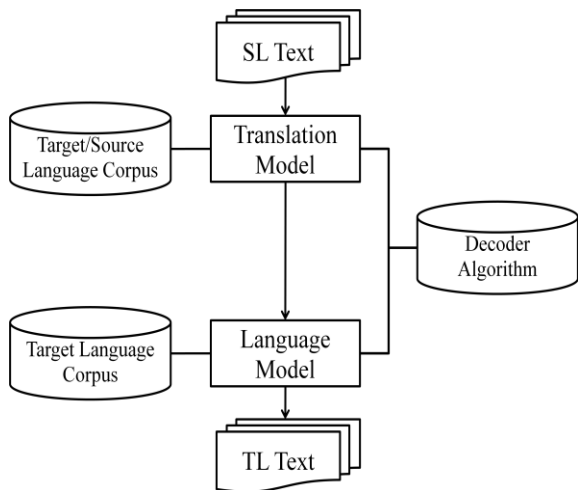


Fig 6. SMT Model

3.2.1 Translation Model

The translation model attempts to match the strings (words or phrases) of the source language to strings of the target language. The model attempts to identify and extract pairs of strings and assigns a probability value to each pair. This value is a conditional probability $P(T/S)$, and in this case is the probability of one string in the source language given the occurrence of another string in the target language. The values that are assigned to the pairs are determined based on the pre-assigned human translations of source language to target language and are stored in a parallel corpus.

This model applies machine learning techniques that help significantly to improve the system persistently over a long period of time. Frequency of translated text by human translators is stored and used for improving the quality of future translations.

3.2.2 Language Model

The language model determines the probability $P(S)$ of the string of target language actually occurring in that language. There are a number of ways to determine this value. Unlike the translation model, parallel corpus is not needed and text in only one language is required.

3.2.3 Decoder Algorithm

After calculating the product of the translation model and the language model, the decoder algorithm selects and outputs the string of target language with the highest probability.

3.3 Example Based Machine Translation (EBMT)

The methodology of a typical EBMT system for English to Urdu MT can be divided into four phases i.e. sentence fragmentation, search in corpus, N-ary Product based Retrieval and ordering of Translated Text as illustrated in Figure 7.

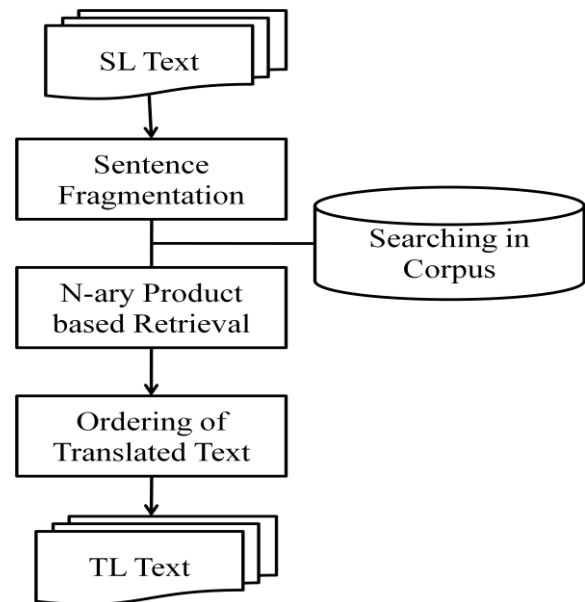


Fig 7. EBMT Model

3.3.1 Sentence Fragmentation

Division of input sentences into phrases is vital to improve the scope of input sentences that can be handled by a translator. Same result can be achieved alternatively by keeping sentences in corpus and by gaining a broad coverage by fragmentation and combination to get new sentences using the genetic algorithm at run time. The problem of fragmenting a sentence into simpler sentences and phrases is handled using idioms, connecting words and the cutter points.

3.3.2 Searching in Corpus

In this phase the bilingual corpus is searched to determine whether the input phrase is obtainable or not. If exact match is not available, then it tries to locate the closest match. Closeness is measured via threshold at two levels i.e. for exact match and for a close match. This is achieved in two ways using Levenshtein Algorithm and Semantic Distance Algorithm.

3.3.3 N-ary Product Based Retrieval

This phase consists of steps used to retrieve the translation of input text. For an input sentence there is a chance of getting more than one translation. The possibilities are computed and n-ary product is used to list all possible sentences.

3.3.4 Ordering of Translated Phrases

When a single input sentence is divided into pieces and translated into output language phrases, then ordering of the translated phrases according to syntactical structure of target language is required before generating the final output. Such a process of ordering is carried out in this phase.

4. COMPARISON

Each of the aforementioned technique has its own intrinsic computational complexity as well as requirement of data, software and human resources. The following section gives a thorough comparison of these techniques and their respective advantages and shortcomings are clearly outlined.

4.1 Rule Based Machine Translation (RBMT)

RBMT is a type of large scale rule based system. Therefore computational cost and resource requirements are high in order to create a RBMT system. It is formulated on the basis of morphological, syntactic and semantic analysis of both the source and target language. Given large scale and fine grained linguistic rules, RBMT systems have ability of generating translations with reasonable quality. Nonetheless, constructing such a system is extremely time consuming and labor intensive because such linguistic resources needs to be hand crafted. This issue is commonly referred to as knowledge acquisition problem. A RBMT system works on exact match calculation. Due to this it is unable to translate when it does not contain any knowledge about the input. Furthermore, it is very difficult to correct the input or to add more rules in the existing system in order to obtain the correct output. The strengths and weaknesses of RBMT systems are tabulated in table 1.

Table No.1 Strengths and weaknesses of RBMT

| Rule Based Machine Translation | |
|---|--|
| Strengths | Weaknesses |
| <ul style="list-style-type: none"> • Effective for core phenomena • Based on linguistic theories • Easy to build an initial system | <ul style="list-style-type: none"> • Rules are formulated by experts • Sometimes the experts do not agree hence the system remain unreliable • Difficult to maintain and extend • Ineffective for marginal phenomena |

4.2 Statistical Machine Translation (SMT)

A SMT system employs stochastic models where translation knowledge is evolved and learned automatically from the example data. Due to this reason the development and testing of SMT systems is not much time consuming as compared to the RBMT systems. It is an efficient method when a large corpus is available and where limited linguistic knowledge is available. It is a good choice when both the source and target languages are not rich in terms of their morphological and syntactic structures. The efficiency and quality of translation generated by SMT system depends upon the quality of bilingual corpus available. Bilingual dictionary is not a requirement of a SMT system. For unobserved domain, the quality of translation and performance of SMT system is poor because there is no availability of highly trained database. The

qualitative strengths and weaknesses of SMT systems are tabulated in table 2.

Table No.2 Strengths and weaknesses of SMT

| Statistical Machine Translation | |
|---|--|
| Strengths | Weaknesses |
| <ul style="list-style-type: none"> • Numerical knowledge • Extracts knowledge from corpus • Reduces the human cost • Model is mathematically grounded | <ul style="list-style-type: none"> • Less linguistic background • Overall lookup cost is high • Hard to capture long distance phenomena • Authenticity of results can be questionable. • Not suitable for free word order languages |

4.3 Example Based Machine Translation (EBMT)

An EBMT system needs bilingual dictionary. It directly provides translation by adapting examples with no calculations of extensive chain of rules. In EBMT technique, the computational cost of translation is less than the computation cost of RBMT. Updation in syntactic or semantic rules is not required. This makes EBMT system easy to improve simply by entering suitable examples into database. Due to availability of large amount of text and its respective translation, an EBMT system is easy to build. It works on best match reasoning. The translation becomes difficult if there is no corresponding example in the corpus. The corpus containing overlapping sentences is good for extracting multiple translated phrases for a matched source language phrase. It basically translates in a fail-safe way. Reliability factor is given to translation result according to the distance between input text and similar examples found in the database. EBMT can also notify us when its translation is improper.

Table No.3 Strengths and weaknesses of EBMT

| Example Based Machine Translation | |
|---|--|
| Strengths | Weaknesses |
| <ul style="list-style-type: none"> • Extracts knowledge from corpus • Based on translation patterns in corpus • Reduces the human cost | <ul style="list-style-type: none"> • Similarity measure is sensitive to system • Lookup cost can be high • Knowledge acquisition is problematic • Trade off is required between corpus size and performance. |

5. The Bilingual Evaluation Understudy (BLEU)

The evaluation of machine translated text by human experts is extensive and intelligent but on the other hand it is expensive in terms of money, time and labor. Human evaluation of machine translated text can take days or even months. Due to these problems an automatic evaluation of machine translated text was proposed by Kishore et al which is known as Bilingual Evaluation Understudy (BLEU) [17]. As compared to human evaluation, this system is economical, fast and language independent.

BLEU is an algorithm used for evaluation of the quality of machine translated text. The main idea behind BLEU is "the closer a machine translation is to a professional human translation, the better it is" [17]. Output value is typically

reported between 0 and 1 that can be easily converted into percentages if required. A higher BLEU score shows better quality of machine translation.

BLEU uses a modified form of precision to compare output text against multiple reference sentences. The reference sentences are human translated text. Larger number of reference sentences will result in higher BLEU scores. The BLEU precision value can be calculated using the following formula.

$$P = m_{max} / W_t$$

Where P is precision value of BLEU, m_{max} is the number of words found similar in reference sentence and output sentence and W_t is total number of words in output sentence.

We explain the BLEU metric calculation through an example of a sentence translated from Urdu to English using an automated translation system. For calculating BLEU value, we compare the machine translated text with the reference sentences that are translated by language experts.

Example:

In the following, we consider two reference sentences of the same semantic class. These sentences are translated by human experts.

Reference Sentence 1:

The Pakistani weapons are to be handed over to the army within two weeks.

Reference Sentence 2:

The Pakistani weapons will be surrendered to the army in two weeks.

Output Sentence:

In two weeks Pakistan's weapons will give army.

The output sentence shares “in” with Reference 2, “two weeks” with Reference 1, “weapons” with Reference 1, “will” with Reference 2, and “army” with Reference 1. So the total similar words (m_{max}) between output sentence and both reference sentences are 6, and the number of words in output sentence (W_t) is 8. Thus using the BLEU formula gives us the precision value of **0.75** for this output sentence.

6. EXPERIMENTS AND FINDINGS

The following seven carefully selected English sentences from already published literature were used as the input text for translation into Urdu language.

1. *The conspiracy was brought to light by policeman.*
2. *He has come of age today.*
3. *He gets an apple.*
4. *He gets an idea.*
5. *He works in a bank.*
6. *He is waiting near the bank of a river.*
7. *They are playing in the garden.*

These sentences were translated by seven human experts based on which the BLEU score of each sentence was computed using the standard procedure and formula mentioned in the previous section. This BLEU metric is then used for evaluation of the machine translated text. A graph containing BLEU values for all example sentences is shown

in figure 8 whereas cumulative average BLEU values are illustrated in the graph shown in figure 9.

Our study reports two types of findings i.e. the qualitative as well as the quantitative findings. The qualitative findings are tabulated in table 1, 2 and 3, whereas the quantitative findings are mentioned in table 4-10. Each of the tables contains the input text, the translated output text generated by RBMT, EBMT, Google and Bing translators along with their respective BLEU metric. Cumulative average BLEU scores of all these techniques are tabulated in table 11.

Table No.4 Example Sentence 1

| Input | <i>The conspiracy was brought to light by policeman.</i> | BLEU |
|--------|--|------|
| RBMT | سازش پولیس کے افسر کے پاس روشنی کی طرف لابی گئی | 0.64 |
| EBMT | سازش پولیس افسر سے منظر عام پر آئی | 0.63 |
| Google | سازش پولیس اہلکار کی طرف سے روشنی میں لایا گیا تھا | 0.82 |
| Bing | سازش کی روشنی ایک پولیس اہلکار کی طرف سے لائی گئی | 0.54 |

For the input text in table 4 the BLEU score of translation produced by the Google translator (a SMT system) is higher than the rest of systems. Next higher value is of RBMT, and then EBMT systems. The least value is produced by the Microsoft's Bing translator (another SMT system).

Table No.5 Example Sentence 2

| Input | <i>He has come of age today.</i> | BLEU |
|--------|-------------------------------------|------|
| RBMT | وہ آج عمر کا آیا ہے | 1.0 |
| EBMT | وہ آج بالغ ہوا ہے | 0.6 |
| Google | انہوں نے کہا کہ آج کی عمر کے آیا ہے | 0.6 |
| Bing | وہ آج عمر کے آیا ہے | 1.0 |

For the input text in table 5, both the RBMT and Bing translator generate the higher BLEU values than EBMT and Google translator. The value generated by EBMT and Google translator is also same.

Table No.6 Example Sentence 3

| Input | <i>He gets an apple.</i> | BLEU |
|--------|------------------------------------|------|
| RBMT | اسے سیب ملتا ہے | 1.0 |
| EBMT | اسے ملا ایک سیب | 0.75 |
| Google | انہوں نے کہا کہ ایک سیب ہو جاتا ہے | 0.44 |
| Bing | وہ ایک سیب ہو جاتا ہے۔ | 0.5 |

Input text in table 6 gives higher BLEU value than the EBMT and the Bing SMT system. The Google SMT gives the smallest value.

Table No.7 Example Sentence 4

| Input | <i>He gets an idea.</i> | BLEU |
|--------|--------------------------------|------|
| RBMT | اسے خیال ملتا ہے | 0.75 |
| EBMT | اسے سوچھا ایک خیال | 1.0 |
| Google | انہوں نے کہا کے خیال بوجاتا ہے | 0.38 |
| Bing | وہ ایک خیال بو جاتا ہے | 0.5 |

For input text in table 7, the BLEU value of EBMT is greater than RBMT, Google and Bing SMT systems. However, Google translator produces the smallest BLEU score.

Table No.8 Example Sentence 5

| Input | <i>He works in a bank.</i> | BLEU |
|--------|--|------|
| RBMT | وہ بینک میں کام کرتا ہے | 1.0 |
| EBMT | وہ کام کرتا ہے بینک میں | 1.0 |
| Google | انہوں نے کہا کے ایک بینک میں کام کرتا ہے | 0.5 |
| Bing | انہوں نے ایک بینک میں کام کرتا ہے | 0.5 |

The translation and BLEU scores of translated text in table 8 shows that the value of RBMT and EBMT are the same and double than both the SMT systems i.e. the Google and Bing translators. The value of Google and Bing are also the same.

Table No.9 Example Sentence 6

| Input | <i>He is waiting near the bank of a river.</i> | BLEU |
|--------|--|------|
| RBMT | دریا کے بینک کے قریب ہے از ویشنگ | 0.5 |
| EBMT | وہ دریا کے بینک کے قریب انتظار کر رہا ہے | 0.9 |
| Google | انہوں نے کہا کے دریا کے کنارے کے قریب انتظار کر رہا ہے | 0.69 |
| Bing | وہ ایک دریا کے کنارے کے قریب کا انتظار کر رہا ہے | 0.83 |

For the input text in table 9, the BLEU value of EBMT is greater than the rest of systems. Bing generates the next higher value followed by the Google translate. RBMT generates the least value.

Table No.10 Example Sentence 7

| Input | <i>They are playing in the garden.</i> | BLEU |
|--------|--|------|
| RBMT | وہ باغ میں کھیل رہے ہیں | 1.0 |
| EBMT | وہ باغ میں کھیل رہے ہیں | 1.0 |
| Google | وہ باغ میں کھیل رہے ہیں | 1.0 |
| Bing | وہ باغ میں کھیل رہے ہیں | 1.0 |

For the input text in table 10, all the four systems generate the same quality translation and the BLEU scores of each of these systems are the same.

The cumulative average of BLEU metric values for all the four translation systems taken into consideration in our experiments are tabulated in table 11. These four systems include the RBMT, EBMT and the two contemporary online SMT systems namely Google translate and the Microsoft Bing translator. The table clarifies that the EBMT system generated the highest quality translation. Next higher quality translation was done by the EBMT system. The BLEU scores of SMT systems suggest that there is significant need for improvement in the SMT systems for Urdu language. The values can be easily converted into percentage values, if required.

Table 11 Comparison of cumulative BLEU metric averages of RBMT, EBMT and SMT systems

| | RBMT | EBMT | SMT | |
|------------|-------|--------|--------|-------|
| | | | Google | Bing |
| BLEU Value | 0.800 | 0.8421 | 0.6268 | 0.709 |

Our example sentences are of variable length and semantic complexity. The BLEU values of each sentence based on RBMT, EBMT and the two online SMT systems (Google translate and Bing Translator) are illustrated using a graph as shown in figure 8. The horizontal axis shows the sentence sequence number whereas the vertical axis shows the corresponding BLEU scores of each sentence. From the graph in figure 8, we can deduce that long Urdu sentences result in low BLEU scores. Habib et al have analyzed Urdu word frequencies in detail [12][18]. Considering the word frequencies, we reached a conclusion that sentencing containing low frequency words also result in low BLUE scores. On the contrary, the BLEU value is high for simple, short sentences containing words that are much frequently used in Urdu language.

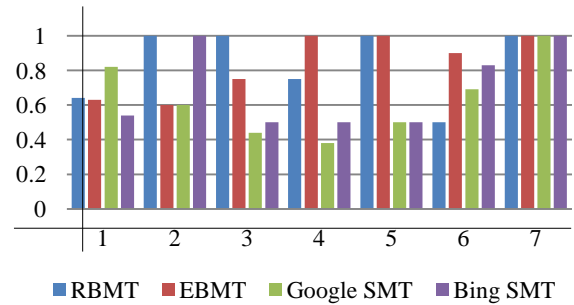


Fig 8. Blue score of each example sentence

The average BLEU value in table 11 shows that EBMT performs better than the rest of the three MT systems discussed in this paper. RBMT was found to be better than both the SMT systems. Out of the two SMT (Google translate and the Microsoft Bing), the Bing translator gave better results than the Google translator. This is illustrated in the graph shown in figure 9.

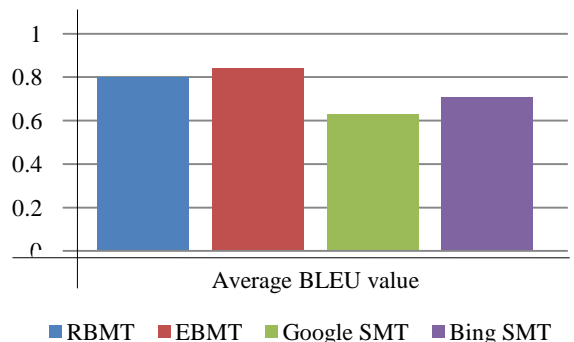


Fig 9. Average BLEU value of all systems

7. DISCUSSION

After detailed literature study and evaluation of the above mentioned three MT systems, we can conclude that for languages with similar lexical and syntactic structure e.g. Urdu and Hindi, the Rule based MT technique gives better results. The SMT systems perform better if necessary resources such as annotated corpora etc. are available. At present, most of the systems translate text from source to target language on the basis of single sentence whereas in real life text for translation is much larger than one sentence. Nonetheless, the continuous process of repetitive translation and improvements by human annotators and translators contribute significantly to any MT system.

8. CONCLUSION AND FUTURE DIRECTIONS

In this paper we explained three main techniques of machine translation; Rule Based Machine Translation, Statistical Machine Translation and Example Based Machine Translation. We explained the methodology of each of these systems and found their comparison based on their respective outputs using BLEU. Our current work is preliminary in nature. However it reports significant results based on qualitative and quantitative analysis.

In order to contribute a significant role to Urdu machine translation research, at present we are in the process of building the required corpora. We intend to use our corpora to conduct larger scale automated experiments and report quantitative results that are comparable to human translators. Based on our qualitative and quantitative results, we aim at proposing a new model that minimizes flaws in the existing Urdu MT systems. Ideally, we would like to implement our proposed system with fewer requirements of computational and human resources.

9. REFERENCES

- [1] M. P. Lewis, G. F. Simons, C. D. Fennig, "Ethnologue: Languages of the World", Summer Institute of Linguistics International, 2013
- [2] J. Hutchins, "Latest Developments in Machine Translation Technology: Beginning a New Era", MT Summit IV, 1993, 11-34
- [3] A. H. Homiedan, "Machine Translation", Journal of King Saud University, Language & Translation, Vol. 10, 1998, 1-21
- [4] D. Attash, "Urdu Informatics", National Language Authority Press, Vol. 1, 2008, 102-112
- [5] N. Ata, B. Jawaid, A. Kamran, "Rule Based English to Urdu Machine Translation", Conference on Language and Technology, 2007
- [6] B. Jawaid, "Statistical Machine Translation between Languages with Significant Word Order Difference", University of Malta & Charles University in Prague, 2010
- [7] M. Zafar, A. Masood, "Interactive English to Urdu Machine Translation using Example-Based Approach", International Journal of Computational Science and Engineering, Vol. 1(3), 2009, 276-283
- [8] A. Ali, S. Siddiq, M. K. Malik, "Development of Parallel Corpus and English to Urdu Statistical Machine Translation", International Journal of Engineering & Technology, International Journal of Engineering & Sciences, Vol. 10(5), 2010, 30-33
- [9] B. Jawaid, D. Zeman, "Word-Order Issues in English-to-Urdu Statistical Machine Translation", The Prague Bulletin of Mathematical Linguistics, 2011, 87-106
- [10] Online machine translation system, The Bing Translator by Microsoft Inc. <http://www.bing.com/translator>
- [11] Online machine translation system, The Google Translate by Google Inc. <http://translate.google.com>
- [12] A. Habib, M. Iwatate, M. Asahara, Y. Matsumoto, "Keypad for large letter-set languages and small touch-screen devices (case study: Urdu)", International Journal of Computer Science Issues, Vol. 9(3), 2012, 1694-0814
- [13] S. D. Samantaray, "Example Based Machine Translation approach for Indian Language", International Center for Chemical and Biological Sciences, 2004, 1-10
- [14] P. F. Brown, J. Cocke, S. A. D. Pietra, V. J. D. Pietra, F. Jelinek, J. D. Lafferty, R. L. Mercer, P. S. Roossin, "A Statistical Approach to Machine Translation", Computational Linguistics, Vol. 16(2), 1990, 79-85
- [15] N. Karamat "Verb Transfer for English to Urdu Machine Translation (Using Lexical Functional Grammar (LFG))", National University of Computer & Emerging Sciences, Lahore, Pakistan, 2006
- [16] H. Somers, "Machine translation and Welsh: The way forward". A Report for the Welsh Language Board, Centre for Computational Linguistics, University of Manchester Institute of Science and Technology, 2004
- [17] K. Papineni, S. Roukos, T. Ward, W. J. Zhu, "BLEU: A Method for Automatic Evaluation of Machine Translation", Association for Computational Linguistics, 2002, 311-318
- [18] A. Habib, M. Iwatate, M. Asahara, Y. Matsumoto, W. Khalil, "Optimized and hygienic touch screen keyboard for large letter set languages", International Conference on Ubiquitous Information Management and Communication, Association for Computing Machinery, 2013
- [19] A. A. Malik, A. Habib, "Qualitative Analysis of Contemporary Urdu Machine Translation Systems", Logic Programming and Nonmonotonic Reasoning, Natural Language Processing and Automated Reasoning 2013, 27-36.
- [20] M. Zhang and H. Li, "Tree kernel-based SVM with structured syntactic knowledge for BTG-based phrase reordering", Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2009, 698-707.