Machine Translation Techniques and their Comparative Study

Neeha Ashraf
Research Scholar computer sciences
University of Kashmir
Jammu and Kashmir, India

Manzoor Ahmad
Research Supervisor computer sciences
University of Kashmir
Jammu and Kashmir, India

ABSTRACT

Human beings exhibit a striking quality of communicating with each other. Communication by means of a system of communication based upon words and the combination of words into sentences, referred as linguistic communication. None of non-human species have such a system of communication in place that's comparable to human language. What makes languages of human varied and different, are features of duality and arbitrary. In annuals of Anthropology, language is considered as a primary tool for studying the culture of a civilization, what we speak influences what we think, what we feel and what we believe. Culture is transmitted through language. Humans learn their culture through language. Its inquisitive nature of human and passion to travel across the world, warrants different cultures interact with each other, the means to achieve this is through human language, often interacting cultures communicate through different languages. As such, it's essential that humans translate and interpret languages of different cultures for understand their rituals, business and allied activities. With advancements in technology, computer systems have facilitated the translations of languages and achieved results in minimal amount of time, though these systems do not produce exact translated verse but enough and relevant information that could be used by the information professionals to understand the nature of information contained in the document, tools like Babelfish and Google Translator are examples of such systems. Numerous techniques have been developed to automate the translation process and these are termed under Machine Translation, which can be defined as a task of automatically converting one natural language into another, preserving the meaning of the input text, and producing fluent text in the output language. These automated translation systems use state of art technology with wide-ranging dictionaries and a collection of linguistic rules that translate one language into another without relying on human translators. The motivation of this research is to have a comparative study of machine translation techniques used for multilingual translation vis-à-vis efficiency, ease of use, space-time complexity and creation of experimental framework for comparing machine translation techniques using open-source translation tools.

Keywords

Machine translation; translation techniques; comparative study; RMT; SMT;

1. INTRODUCTION

Human beings exhibit a striking quality of communicating with each other. Communication by means of a system of communication based upon words and the combination of words into sentences, referred as linguistic communication. Language of human beings is diverse and multifarious. Diverse due fact that more than 5000 dialects exist on earth which mirrors the phonetic differing qualities. What makes languages of human unique, are features of duality and arbitrary. Duality arises from using arbitrary vocal symbols used for communication, which can further be attributed to

coexistence of the system of sounds and the system of meanings. Predicting which features exist in any language is difficult, rather impossible, making language arbitrary, on similar lines, language is symbolic, as words uttered are related with objects, ideas, and action. In annuals of Anthropology, language is considered as a primary tool for studying the culture of a civilization, they maintain what we speak influences what we think, what we feel and what we believe. Culture is transmitted through language. Humans learn their culture through language. Keeping in mind the end goal to overcome any and all hardships of dialect assorted qualities numerous language tools are broadly utilized by bookkeepers and data experts. These tools give essence of the data that could be utilized by the client to comprehend the sort of information contained in the archive.

Language - as characterized above - is a solely human property. Among the attributes that make a moderately clear refinement correspondence significant, two are especially critical: twofold explanation and grammar. Dialects comprise of a huge number of signs, which are mixes of structure and importance. Dialects have countless signs, and the term twofold explanation alludes to the way that the formal sides of these sign are assembled from a generally little collection - ordinarily somewhere around 10 and 100 - of trivial sounds. The smart creation that empowered individuals to talk envision, is language structure. Sentence structure is utilized to assemble signs communicating generally straightforward implications into sign mixes communicating more unpredictable implications.

Information communication and technology (ICT) have facilitated development of efficient procedures of machine interpretation. Research endeavors have been on to learn the likelihood of programmed interpretation of one dialect (source content) to another dialect (target content). A few free and additionally exclusive instruments are presently accessible which bolster interpretation of content in one or more dialects. Over internet, Yahoo and AltaVista offers online translation through Babelfish. Librarians use Bing translator of Microsoft and Google Translator from Google. Firefox uses Greasemonkey tool to translate the text in other languages. Google chrome browser offers translation if URL is in language other than default language (mostly English).

2. MACHINE TRANSLATION

Machine translation (MT) is a procedure whereby a computer procedure/program assesses a source content and, on a basic level, delivers an objective content, target text without human mediation. In like manner speech, it's decoding the meaning of the source text and re-encoding the meaning in the target language. As a general rule, on the other hand, machine translation ordinarily does include human mediation, as pre-altering and post-altering. With legitimate phrasing work, with readiness of the source content for machine translation (pre-altering), and with modifying of the machine translation by a human interpreter (post-altering), business machine-translation devices can create helpful results, particularly if the machine translation framework is coordinated with a

translation-memory or globalization-administration framework [1]. Prime element of machine translation is increment in efficiency by quicker translation of source to target content.

There have been real activities from different exploration associations and government offices to create instruments for programmed interpretation of content otherwise called Machine Translation. Machine translation is one of the exploration areas under "Computational Linguistics" [2]. Different systems have been contrived to robotize the translation process. Then again, the goal has been "to restore the significance of unique content in the deciphered verse". All in all, the translation procedure has two levels

2.1 Level 1- Metaphrase

Meta-phrase is "word-to-word" translation. It relates to "formal equivalence", i.e. the translated version will have "literal" translation for each word in the text. However, the translated text might not essentially convey the meaning of the original text. That means sometimes the semantics may differ from the original text.

2.2 Level 2 - Paraphrase

In Para-phrase, translated text contains the main point of the original text but may not necessarily contain the word-to-word translation. It relates to "dynamic equivalence" (i.e. the translated text is the main point of original text).

3. METHODS OF MACHINE TRANSLATION

3.1 Dictionary Based

This translation method takes into account sections of a language dictionary, translated verse is produced using the word's matching or equivalent words in dictionary. It implies that the words will be interpreted as a dictionary does - word by word, for the most part without much relationship of meaning between them. Dictionary lookups may be done with or without (related to shape and structure) analysis or lemmatization. While this way to deal with machine translation is most likely the slightest fancy, dictionary-based machine translation is good for the translation of long lists of phrases on the sub-sentential (i.e., not a full sentence) level, e.g. (items that are stored and available now) or simple (big lists of items) of products and services.

It can likewise be utilized to accelerate manual translation, if the individual doing it is familiar with both dialects and accordingly fit for rectifying language structure and linguistic use. The original of machine translation which is from late 1940s to mid-1960s was completely in view of machine-decipherable or electronic word references. This technique is still useful in translation of phrases however not sentences to some degree. The vast majority of the translation approaches that were created later on utilize bilingual lexicons with syntactic principle [3].

3.2 Rule Based

This machine translation framework is taking into account etymological data about source and target dialects fundamentally recovered from (unilingual, bilingual or multilingual) lexicons and sentence structures covering the primary semantic, morphological, and syntactic regularities of every dialect individually. Having data sentences (in some source dialect), a RBMT framework produces them to yield sentences (in some objective dialect) on the premise of morphological, syntactic, and semantic examination of both the source and the objective dialects included in a solid

translation assignment. The principal RBMT frameworks were created in the mid-1970s.

Rule-based machine translation manages the morphological, syntactic and semantic data about the source and target dialect. This data is utilized to manufacture linguistic tenets. The principle methodology of RBMT frameworks is in light of connecting the structure of the given information sentence with the structure of the requested yield sentence, fundamentally saving their novel importance. The accompanying sample can outline the general edge of RBMT: "A boy eats an apple."

Source Language = English; Demanded Target Language = Kashmiri/Kashur. Minimally, to get a Kashmiri translation of this English sentence one needs:

- A dictionary that will outline English word to a fitting Kashmiri word.

- Stay away from Rules speaking to consistent English sentence structure.

- Principles representing regular Kashmiri sentence structure

- Lastly, we need standards as per which one can relate these two structures together.

As needs be we can express the accompanying phases of interpretation:

1. **Getting basic part-of-speech information of each source word**
A = indef.article; boy = noun; eats = verb; an = indef.article; apple = noun

2. **Getting syntactic information about the verb "to eat":**
NP-eat-NP; here: eat – Present Simple, 3rd Person Singular, Active Voice

3. **Parsing the source sentence**
(NP Akh choonth) = the object of eat

Usually, halfway parsing is adequate to get to the syntactic structure of the source sentence

4. **Translate English words into Kashmiri**
A (category = indef.article) => Akh (category = indef.article)

boy (category = noun) => Lådkhi (category = noun)

eat (category = verb) => khyavan (category = verb)

an (category = indef. article) =>Akh(category = indef.article)

apple (category = noun) => choonth (category = noun)

5. **Mapping dictionary entries into appropriate inflected forms (final generation):**
A boy eats an apple. => Akh Lådkhi khyavan Akh choonth.

Also millions of bilingual dictionaries are used for source as well as target languages. Rule-based machine translation is extensible and maintainable [4] and is able to deal with the needs of wide variety of linguistic phenomena. However, the system may be limited by some factors such as:

- Exceptions in grammar add difficulty to the system.

- The research process requires high investment.

Example: IIT Kanpur developed Anglabharati which is a rule based machine translation system from English to Hindi and other Indian Languages. For AnglaBharti Mission,

CDAC(GIST group) Pune will be working on Urdu, Sindhi & Kashmiri and will develop AnglaUrdu, AnglaSindhi & AnglaKashmiri.

Objective: The objective of Rule-based machine translation is to convert source language structures to target language structures.

5.1 Knowledge Based Machine Translation

This kind of system is concerted around "Concept" lexicon representing a domain.

E.g. KANT is an example of Knowledge Based Machine Translation System for multilingual translation. It is developed using large scale knowledge base and controlled language system.

5.2 Corpus Based Machine Translation

Since 1989, Corpus based methodology for machine translation has developed as one of the broadly investigated region in machine translation. This technique has ruled over different methodologies, in view of high level of exactness accomplished amid the translation of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For three authors, you may have to improvise.

5.3 Example Based Machine Translation

Example based translation also k/a memory based translation is based with respect to reviewing or discovering closely resembling cases of the dialect pair. This idea of "Translation by Analogy" was initially proposed by Makoto Nogao in 1981[5]. The framework is given a situation of sentences in the source dialect from which one is translating and comparing translations of every sentence in the objective dialect with point to point mapping. These samples are utilized to decipher comparable kind of sentences of source dialect to the objective dialect. The fundamental reason is that, if a formerly translated sentence happens once more, the same translation is liable to be right once more. [6].

Frederking stated some advantages of example based system over statistical machine translation system and these are:

- This can work with small set of data even with one sentence pair.

- Trains translation program and decodes more quickly.

- Less principles at least in theory

5.4 Context Based Machine Translation (CBMT)

The context based machine translation is being produced as a corpus-based technique that obliges neither principles nor parallel corpora. Rather, CBMT connection based framework needed

- An extensive monolingual target text corpus.

- A full-form bilingual dictionary

- To further improve translation quality it requires a smaller monolingual source-text corpus to run its algorithms [7].

CONTRAST [8] and REFTEX [9] are examples of Context Based Machine Translation System.

6. DIFFERENT APPROACHES TO RULE-BASED MACHINE TRANSLATION
6.1 Direct Approach

Words of source dialect are translated without going through a middle-person representation. Anusaarka is a sample of machine translation framework in light of direct approach. It has been created at Indian Institute of Information Technology, Hyderabad and covers all major Indian dialects.

6.2 Transfer Based Approach

This methodology fits in with the second era of machine translation (mid 60s to 1980s). In this methodology source dialect is changed into a dynamic representation which is less dialect particular. An equal representation with same level of reflection is then produced for the objective dialect utilizing bilingual dictionaries and grammar rules. This frameworks have three noteworthy parts

Components of Transfer Based Approach:

1. Analysis: The source text is analyzed based on linguistic information such as morphology, part-of-speech, syntax, semantics etc. Algorithms as well as heuristics are applied to parse the source language. This derives:

2. The syntactic structure of the text to be translated for language pair of the same family e.g. Tamil and Telugu are siblings of same family (i.e. Dravidian Languages).

   OR

   The semantic structure (for language pair of different families, Hindi from Devanagari family and Telugu from Dravidian family) of the text to be translated.

3. Transfer: The syntactic/semantic structure of source language is then moved into the syntactic/semantic structure of the target language.

4. Synthesis: Synthesis module replaces the constituents in the source dialect to the target language reciprocals

Transfer-based methodology has reliance in light of the language pair included. Therefore in Eurotra Project [10] two autonomous monolingual word references were recommended. Likewise, there are diverse representations for distinctive dialects.

Example of Transfer-Based approach:

- PaTrans(Translation for Patents) is in light of transfer based approach and is one of the results of Eurotra Research.

- Mantra, a translation model for Indian dialects taking into account transfer based approach. It is a task subsidized by Government of India and the parser utilized for dialect preparing is known as Vyakarta.

4.1 Interlingua Approach

This is considered to have a place with third era of machine translation. It is an inborn piece of a branch called Inter-linguistics. Interlingua means to make etymological homogeneity over the globe. Interlingua is mix of two Latin words Inter and Lingua which means in between/intermediary and language respectively.

In Interlingua, source dialect is changed into an assistant/mediator dialect representation which is autonomous

of any of the dialects included in the translation, the deciphered verse for the objective dialect is then determined through this helper representation. Henceforth just two modules i.e. analysis and synthesis are needed in this sort of framework. This methodology is autonomous of the dialect pair for translation, hence it is pertinent in multilingual machine translation. This methodology underscores on single representation for distinctive dialects

Examples:

- The parameterization model proposed by Ali [11] is one of the upgrades over between lingua model with standout investigation parts (multi-lingual parser) and one amalgamation segment which work multi-lineally.

- UNITRAN [12] framework is one execution of this model. It utilizes parameterization as a part of both the syntactic and lexical qualifications. Indian Institute of Technology, Powai is dealing with creating interpretation framework for Indian dialects taking into account Interlingua

5. DIFFERENT APPROACHES TO CORPUS-BASED MACHINE TRANSLATION

5.1 Statistical Machine Translation Approach

In 1949, Warren Weaver presented the thought of statistical machine translation. In this methodology, statistical methods are employed to create translated form utilizing bilingual corpora. Statistical machine translation uses factual translation models whose parameters stem from the examination of monolingual and bilingual corpora. Building statistical translation models is a fast process, however the innovation depends intensely on existing multilingual corpora. At least 2 million words for a particular space and considerably more for general dialect are needed. Hypothetically it is conceivable to achieve the quality edge however most organizations don't have such a lot of existing multilingual corpora to construct the important translation models. Also, statistical machine translation is CPU concentrated and requires a broad equipment arrangement to run translation models for normal execution levels.

Examples

- n-gram based SMT[13]

- Occurrence based SMT[14]

Macherey [15] has experimented statistical methods for spoken language understanding for SMT.

Google, Microsoft and SDL Language Weaver have also created SMT systems, some publicly accessible

5.2 Statistical Word Based Machine Translation Model Approach

The essential unit of this methodology is word. Algorithms identified with arrangement of words are obliged to accomplish most extreme exactness in sentence translation. Compound words, expressions, and homonyms make many-sided quality for basic word based translation.

5.3 Statistical Phrase Based Model Approach [16, 17]

The crucial unit of this model is a phrase or grouping of words. A grouping of words in the source and the objective dialect is created. Decoding is done taking into account the vector of elements with coordinating qualities for the dialect succession pair.

5.4 Statistical Syntax Based Model Approach

The fundamental unit of this model is the translation rule. The translation rule consists of:-

- Sequence of words and variables in the source language.

- A syntax tree in the target language having words or variables at leaves.

- Vector of feature values which describe the language pair's likelihood [18, 19].

Liu and Gildea[20] in one of their studies have explored the semantic roles to improve syntax based machine translation.

6. COMPARATIVE
6.1 Rule-Based MT and Statistical MT

Rule-based MT gives great out-of-domain quality and is by nature unsurprising. Dictionary based customization assurances enhanced quality and consistence with corporate phrasing. Be that as it may, translation results may do not have the familiarity per users anticipate. As far as speculation, the customization cycle expected to achieve the quality edge can be long and unreasonable. The execution is high even on standard hardware equipment.

Statistical MT gives great quality when expensive and qualified corpora are accessible. The translation is familiar, significance it peruses well and accordingly meets client desires. Be that as it may, the translation is neither unsurprising nor predictable. Preparing from great corpora is computerized and less expensive. In any case, preparing on broad dialect corpora, importance content other than the predetermined space, is poor. Moreover, statistical MT requires huge equipment to assemble and oversee vast translation models.

Table 1. Comparison RMT and SMT

Rule-Based MT	Statistical MT
Consistent and predictable quality.	Unpredictable translation quality.
Out-of –domain translation quality.	Poor out-of-domain quality.
Knows grammatical rules.	Does not know grammar.
High performance and robustness.	High CPU and disk space requirements.
Consistency between versions.	Inconsistency between versions.
Lack of fluency.	Good fluency.
Hard to handle exceptions to rules.	Good for catching exceptions to rules.

High development and customization costs.	Rapid and cost-effective development costs provided the required corpus exists.

Experimental framework for comparing RBMT and SMT

English-Hindi MT frameworks accessible on the web include RBMT system: Anusaarka (http://anusaaraka.iiit.ac.in) and SMT system: Google (http://translate.google.com).This segment characterizes a test structure in which the frameworks presented above can be compared. The thought is to report the principle contrasts in execution terms between the best in class rule-based and statistical systems for English to Hindi translation. In the accompanying segments, the outcomes are accounted for through human assessment. Test set is defined in order to perform the evaluation which comprises of a compilation of excerpts of William Shakespeare in the English source test corpus consists of 8 sentences extracted from poem "The Seven Ages of Man ("All the world's a stage") ".

Table 2. Corpus statistics for English-Hindi test set

	Source Text in English
Sentences	**8**
Words	**211**
Paragraphs	**28**
Characters	**1187**

The examination between diverse translation system outputs was performed by 8 different human evaluators. Every one of the evaluators were bilingual in English and Hindi, accordingly, no reference of interpretation was indicated to them, with a specific end goal to maintain a strategic distance from any inclination in their assessment. Translation using Anusaarka was labelled as Technique-1 and Google was labelled as Technique-2, as shown in Figure 1.

Every evaluator was asked to compare two systems: Anusaarka and Google. Figure 1 demonstrates a case of the screenshot that is indicated to the annotator. Every evaluator looked at 8 extracted translation pairs, and surveyed for every situation whether one framework delivered a superior translation, or whether two or more were proportional. Every judge assessed an alternate arrangement of (conceivable covering) sentences. Keeping in mind the end goal to maintain a strategic distance from any predisposition in the assessment, the particular position in the showcase of the sentences comparing to every framework was likewise arbitrary.
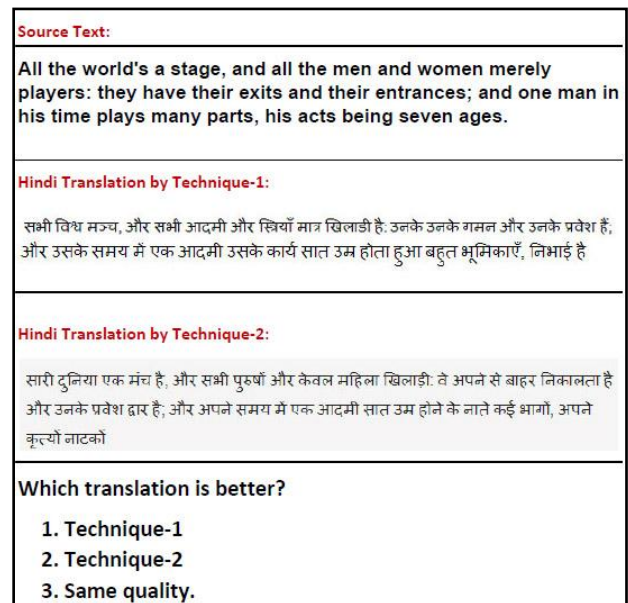

Source Text:

All the world's a stage, and all the men and women merely players: they have their exits and their entrances; and one man in his time plays many parts, his acts being seven ages.

Hindi Translation by Technique-1:

सभी विश्व मञ्च, और सभी आदमी और स्त्रियाँ मात्र खिलाडी है: उनके उनके गमन और उनके प्रवेश हैं, और उसके समय में एक आदमी उसके कार्य सात उम्र होता हुआ बहुत भूमिकाएँ, निभाई है

Hindi Translation by Technique-2:

सारी दुनिया एक मंच है, और सभी पुरुषों और केवल महिला खिलाड़ी: वे अपने से बाहर निकालता है और उनके प्रवेश द्वार है; और अपने समय में एक आदमी सात उम्र होने के नाते कई भागों, अपने कृत्यों नाटको

Which translation is better?
1. Technique-1
2. Technique-2
3. Same quality.

Fig 1. Screenshot of the human evaluation when comparing the two systems.

The correlation among the two systems was not considered when two systems or three frameworks were superior to the other(s) one(s). It tallies entirely when one framework is superior to the others. Making the 8 sentence judgments on the two frameworks typically takes 20 minutes.

We collected a total of 64 judgments in the comparison of the 2 systems, results present a statistical significance equivalence similar to the ones in that international evaluation of English-Spanish WMT 2009 task

Table 3. Human judgments regarding comparison of two systems. Each column indicates the number of times (in percentage) in which one system was chosen as better than the other

English to Hindi Translation	
Anusaaraka	**Google**
74	**26**

Total of 64 judgments were gathered in the correlation of the 2 systems, results show that Anusaaraka employing RMT approach did fairly well against Google with SMT. With regard to syntax and semantics also, Anusaaraka with RMT approach aptly recreated Hindi sentences keeping intact context, grammar and meaning of source English text.

6.2 Rule-Based and Corpus-Based machine translation

Rule-based machine translation (RBMT) is produced on the premise of morphological, syntactic, and semantic examination of both the source and the objective dialects. Corpus-based machine interpretation (CBMT) is created on the examination of bilingual content corpora. The former has a place with the space of realism and the latter experimentation. Given extensive scale and fine-grained phonetic guidelines, RBMT frameworks are fit for delivering translation with sensible quality, however building the framework is exceptionally drawn out and work escalated in light of the fact that such etymological assets should be hand-made, habitually alluded to as learning securing issue.

Besides, it is of extraordinary trouble to redress the data or add new principles to the framework to produce an interpretation. By complexity, be that as it may, adding more samples to a CBMT framework can enhance the framework since it is in view of the information, however the collection and administration of the enormous bilingual data corpus can likewise be exorbitant.

6.3 Direct, Transfer and Interlingua machine translation

Rule- The direct, transfer-based machine translation and Interlingual machine translation routines for machine translation all fit in with RBMT yet vary in the profundity of examination of the source dialect and the degree to which they endeavor to achieve a dialect autonomous representation of significance or expectation between the source and target dialects . Their dissimilarities can be clearly seen through the Vauquois Triangle [21], which outlines these levels of investigation. Beginning with the shallowest level at the base, direct exchange is made at the word level. Contingent upon discovering direct correspondences between source dialect and target dialect lexical units, direct exchange is a word-by-word translation approach with some basic syntactic alterations. An immediate exchange framework is intended for a particular source and target dialect pair and the translation unit of which is normally a word. And after that translation happens on representations of the source sentence structure and meaning respectively through syntactic and semantic exchange approaches.

A Transfer-based machine translation framework includes three stages. The principal stage makes examination of the source content and proselytes it into dynamic representations. The second stage changes over those into equal target dialect situated representations. The third produces the last target content. The representation is particular for every dialect pair. The exchange method can be seen as "a reasonable bargain between the effective utilization of assets of interlingua frameworks, and the simplicity of execution of direct frameworks".

At long last, at the interlingual level, the thought of exchange is supplanted by the Interlingua. The interlingual MT works more than two stages: dissecting the source dialect content into a conceptual all inclusive dialect autonomous representation of importance, i.e. the Interlingua, which is the period of examination; producing this significance utilizing the lexical units and the syntactic developments of the objective dialect, which is the period of combination. Hypothetically, the higher the triangle, the less cost the investigation and union. For instance, to make an interpretation of one source dialect to N target dialects, (1+N) stages are required utilizing an Interlingua contrasted with N ventures of exchange. Yet, to decipher every one of the dialects, just 2N stages are required by the Interlingua methodology contrasted with N² by the Transfer based methodology, which is a noteworthy diminishment. In spite of the fact that no exchange segment must be made for every dialect combine by receiving the methodology of IMT, the meaning of an Interlingua is of extraordinary trouble and even perhaps unthinkable for a more extensive space.

6.4 Statistical and Example-Based machine translation

Statistical machine translation (SMT) is produced on the premise of factual models whose parameters are taken from the examination of bilingual content corpora. The beginning model of SMT, in view of Bayes Theorem [22], proposed by

Brown et al. takes the perspective that each sentence in one dialect is a conceivable translation of any sentence in the other and the most fitting is the translation that is allotted the most noteworthy likelihood by the framework. Example based machine interpretation (EBMT) is portrayed by its utilization of bilingual corpus with parallel messages as its fundamental learning, in which translation by relationship is the principle thought. There are four assignments in EBMT: case procurement, case base and administration, case application and union.

Both fitting in with CBMT, now and again alluded to as information driven MT, EBMT and SMT have something in like manner which recognize them from RBMT. In the first place, they both utilize a bi-message as the basic information source. Second, they are both observational with the standard of machine adapting rather than discerning with the rule of language specialists composing tenets. Third, they both can be enhanced by getting more information. Fourth, new dialect sets can be created just by discovering suitable parallel corpus information, if conceivable. Aside from these likenesses, there are likewise a few dissimilarities. SMT basically utilizes factual information, for example, parameters and probabilities got from the bi-content, in which preprocessing the information is vital and regardless of the possibility that the input is in the preparation information, the same translation is not ensured to happen. By complexity, EBMT utilizes the bi-message as its essential information source, in which preprocessing the information is discretionary and if the data is in the case situated, the same translation is to happen.

7. CONCLUSION

By definition, corpus-based methodologies are principally dependent in light of a bilingual corpus, with which there may be versatility issues. On the off chance that the corpus is too little or the topic excessively hypothetical, then close matches will be hard to retrieve. Alternately, on the off chance that it is too extensive, the redundancy of samples may have antagonistic impacts on execution. Some consider corpus-based ways to deal with be most appropriate to sublanguages, in spite of the fact that this may be a coincidental suspicion brought on by the predefined area sorts of corpora utilized.

SMT is as of now seen as the most prevailing MT approach (Hutchins succinct), whilst EBMT has attempted to find its personality. There are such a large number of varieties of the methodology, for example, the consideration of standard based routines (e.g. parsing) and factual investigation, that there is "no reasonable agreement on what EBMT is or isn't". The outcome is the absorption of EBMT approaches into different standards, prompting half breed EBMT-RBMT or EBMT-SMT frameworks. It is presently acknowledged that a solitary methodology won't augment yield quality, and cross breeds are seen as the best alternative. The best of all methodologies can be joined to augment results, for instance utilizing measurable techniques for exchange and era with a syntactic and morphological base for examination. Perhaps neither one of the approaches has built up another ideal model, however they have conveyed crisp thoughts to the MT table and upgraded existing frameworks.

It would be ideal to get the best of each world: RBMT's grammatical correctness, SMT's lexical selection and SMT tolerance to unexpected structures. As such machine translation (MT) system based on combination of approaches is an interesting challenge and will go long way in resolving issues in existing MT techniques. The main objective of this hybrid approach is to take advantage of the strengths of both linguistic rules and statistical techniques. Hybrid machine

translation (MT) is motivated by the fact that hybridization techniques have become popular as they attempt to combine the best characteristics of highly advanced pure rule or corpus-based MT approaches. In future our endeavor will be to evaluate hybrid approach to machine translation.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Translations.com, GMS Technology Solutions, Globalization Management Systems. Available at: http://www.translations.com/quality/solutions/gms-Technology-solutions.html (Accessed on 19th June, 2015)

[2] Uszkoreit H, What is computational linguistics? (2000) Available at: http://www.coli.uni-saarland.de/-hansu/What _is_cl.html (Accessed on 19th June, 2015)

[3] Yang V S-C, Electronic dictionaries in machine translation Encyclopedia of Library and Information Science,48 (1991) 74-92..

[4] Kaji H,An efficient execution method of rule-based machine translation (1998).Available at: http://www.aclweb.org/anthology/C/C88/C88-2167.pdf (Accessed on 19th June, 2015)

[5] Hutchins J, Example based machine translation- a review and commentary .In Recent advances in example-based machine translation .(Ed. Micheal Carl and AndyWay) (2003).Available at: http://www.hutchinsweb.me.uk/MTJ-2005.pdf (Accessed on 21- June, 2015)

[6] Carbonell J G and Brown R D,Example based machine translation (2004) .Available at: http://www.cs.cmu.edu/~ralf/ebmt/ebmt.html (Accessed on 21- June, 2015)

[7] Carbonell J G , Klein S , Miller D, Steinbaum M,Grassiany T and Frey Context-based machine translation Proceedings of the 7th Conference of the Association for Machine Translation in the Americas(2006) pages 19-28

[8] Isahara H and Uchida Y,Analysis,generation and Semantic representation in CONTRAST – a context-based machine translation system Systems and Computers in Japan, 26 (14) (2007) pp. 37-53.

[9] Kjaersgaard P S , REFTEX – A context-based translation aid Proceedings of the 3rd Conference of the European Chapter of the Association for Computational Linguistics, (1987) pp. 109-112.Available at: http://acl.ldc.upenn.edu/E/E87/E87-1020.pdf (Accessed on 04- July, 2015)

[10] Krauwer S, The Eurotra Project (2008).Available at: http://www.sk.let.uu.nl/stt/eurotra.html (Accessed on 05- July, 2015)

[11] Ali N,Machine translation:a contrastive linguistic perspective.Available at: http://www.unesco.org/comnat/france/ali.htm

[12] Dorr B J, A parameterized approach to integrating aspect with lexical-semantics for machine translation "Proceedings of the 30th annual meeting on Association For Computational Linguistics, (1992) pp. 257-264.Avialable at: http://www.aaai.org/Papers/AAAI/1987/AAAI87-095.pdf (Accessed on 05- July, 2015)

[13] Costa-jussa M R,Crego J M, Lambert P, Khalilov M,Fonollosa J A R ,Marino J B and Banchs R E, Ngram Based statistical machine translation enhanced with multiple weighted reordering hypothesis Proceedings of the Second Workshop on Statistical Machine Translation . Prague, June 2007, pp. 167-170. (Accessed on 05- July, 2015)

[14] Ambati V and Lavie A, Occurrence based statistics in Machine translation.Available at: http://www.cs.cmu.edu/~vamshi/publications/obs.pdf (Accessed on 06- July, 2015)

[15] Macherey K,Bender O and Ney H,Applications of statistical machine translation approaches to spoken languages understanding,IEEE Transactions On Audio Speech and Language Processing, 17(4) (2009) pp.803-818

[16] Koehn P, Och F J and Marcu D, Statistical phrase based Machine translation (2003).Available at: http://www.aclweb.org/anthology/N/N03/N03-1017.pdf (Accessed on 07- July, 2015)

[17] Zens R, Och F J and Ney H, Phrase based machine Translation .Lecture Notes in Computer Science;Springer (2002) pp. 35-56

[18] DeNeefe S, knight K, Wang W and Marcu,D, What can syntax-based MT learn from phrase –based MT.Available at: http://www.isi.edu/natural-language/mt/ats-vs-ghkm.pdf. (Accessed on 07- July, 2015)

[19] Yamada K and Knight K, A syntax-based statistical translation model.Available at: http://www.aclweb.org/anthology/P/P01/P01-1067.pdf.(Accessed on 07- July, 2015)

[20] Liu D and GIldea D, Can semantic roles improve syntax based machine translation? (2008) .Available at: http://www.cs.rochester.edu/~dliu/pub/liu-gildea-nowhere08.pdf (Accessed on 09- July, 2015)

[21] B Dorr,E Hovy, L Levin ,Machine Translation,Interlingual Methods,Elsevier (2008).

[22] Available at: http://www.cut-the-knot.org/Probability/BayesTheorem.shtml (Accessed on 09- July, 2015)

[23] A dictionary of the Kashmiri language - Internet Archive https://archive.org/stream/avocavularykash00elmsgoog/avocavularykash00elmsgoog_djvu.txt (Accessed on 11- July, 2015).

[24] Learn Kashmiri, http://www.languageshome.in/English-Kashmiri.htm . (Accessed on 11- July, 2015).