

# Addressing Challenges in Multilingual Machine Translation

Prof. Rekha Sugandhi, Sayali Charhate, Anurag Dani, Amol Kawade

**Abstract-** The machine translation process may be unidirectional or bidirectional between a pair of languages. Or it can be multilingual too. A number of software's are developed till date and different advancements are taking place in this field to overcome the language barriers and create borderless marketplace. Still there are many challenges involved in this field of AI which are yet to be overcome. The translation quality of MT systems may be improved by developing better methods as well as by imposing certain restrictions on the input. All sort of challenges peak in case of Multilingual Machine Translation as compared to bilingual one. Paper focuses on long term challenges like High-Quality MT for many more language pairs, training with limited data resources, robustness across domains, genres and language styles, Achieving human-level translation quality and fluency.

**Index Terms—** Machine Learning, Machine Translation (MT), Natural Language Processing (NLP), Word sense disambiguation (WSD), Interlingua, Source Language (SL), Target Language (TL), Artificial Intelligence.

## 1 INTRODUCTION

Machine Translation (MT) is a sub-field of Artificial Intelligence (AI) which involves automated translation of text from one natural language to another with the help of computer. At basic level, Machine translator performs simple substitution of words in one natural language for words in another, but that alone usually cannot produce a good translation of a text, because recognition of whole phrases and their closest counterparts in the target language is needed. Every natural language has got its own grammatical structure and certain set of rules. Thus, during task of translation, two things should be taken care of, pertain the meaning as of source language (SL) and output must satisfy lexical rules of target language (TL).

There are many challenges involved in this field of AI which are yet to be overcome. Multilingual MT mainly suffers from 4 types of ambiguity [13]; Lexical ambiguity, Referential ambiguity, scope ambiguity, structural ambiguity. Lexical and structural ambiguities affect quality of translator the most. Lexical ambiguity arises due to multiple meanings of same word, while structural ambiguity arises due to multiple interpretations of the same sentence. During translation, corpus is processed for a

number of times to perform different operations rather than directly translating it to the TL. In this paper we present overview of challenges involved in these processes and long term challenges in this field.

Section 2 focuses on challenges in preprocessing of text i.e. different types of analysis. Section 3 involves challenges during training of machine which affect overall performance and accuracy. Section 4 is about challenges in dealing with multiple languages from different origin and having different characteristics; how these aspects were considered till date. Section 5 comprises of long term challenges followed by some suggestions in section 6. Section 7 contains a proposed model.

## 2 CHALLENGES IN ANALYSIS

Analysis of input text from SL consists of morphological analysis, syntactic analysis and semantic analysis. Each kind of analysis poses some challenges in MT.

### 2.1 In Morphological Analysis

There are no generalized grammatical rules in any language [1] which we can use to reduce the size of lexicon. E.g. suffix 'er' is used to indicate a person performing an action like the one who 'Farms' is a Farmer. But this is not always the case e.g. one who 'Cooks' is known as 'Cook', one who 'Drafts' is called as 'Draftsman' and so on. Similarly different forms of a verb are obtained like 'book', 'book-s', 'book-ing', 'book-ed'. But there are many exceptions too like 'do', 'does', 'doing', 'did', 'done' or 'sing', 'sings', 'singing', 'sang', 'sung'. So if we reduce size of lexicon using this approach, it will certainly display the minimal units of grammatical analysis in a vast amount of language data. This technique may also lead to an imperfect attempt to describe something which is too complex [17].

Moreover derivational morphology will tend to change category of the word (POS) unlike the inflectional morphology [1]. As each language varies greatly, even if we find out a rule which is applicable in most of the cases then it will be based on the experience of only one language. Consequently, it will get counter exemplified from other

- *Rekha S. Sugandhi*, working as Assistant Professor at the Department of Computer Engineering, MIT College of Engineering Pune is currently pursuing her Ph.D. in Computer Engineering at the SGB, Amravati University. She has completed her M.Tech in Computer Engineering and B.E. in Computer Engineering from the University of Pune. Her research area and areas of interest include Machine Learning, Natural Language Processing and Theory of Computation. PH-+91-02030273130. E-mail:rekha.sugandhi@gmail.com
- *Sayali Charhate*, is currently pursuing her bachelor degree in Computer Engineering from the University of Pune
- *Anurag Dani*, is currently pursuing his bachelor degree in Computer Engineering from the University of Pune
- *Amol Kawade* is currently pursuing his bachelor degree in Computer Engineering from the University of Pune

languages. It becomes difficult to handle this issue in case of multilingual MT. Different languages have different way of expressing the same concept. When it comes to multilingual machine translation, it will have to frame different strategies for different TL.

A morphological analyzer has to be able to separate the clitic from its attached morpheme. On the other hand when it comes to agglutinative languages like Turkish, morpheme concept is invaluable [17], as morphemes are combined without fusion or morphophonemic changes.

If we think of listing all inflectional and derivational forms regardless of their type as well as learning program to learn appropriate rules, this may result into redundancy involved in typing out such a huge data and most machine algorithms are designed so as to deal with single token entities.

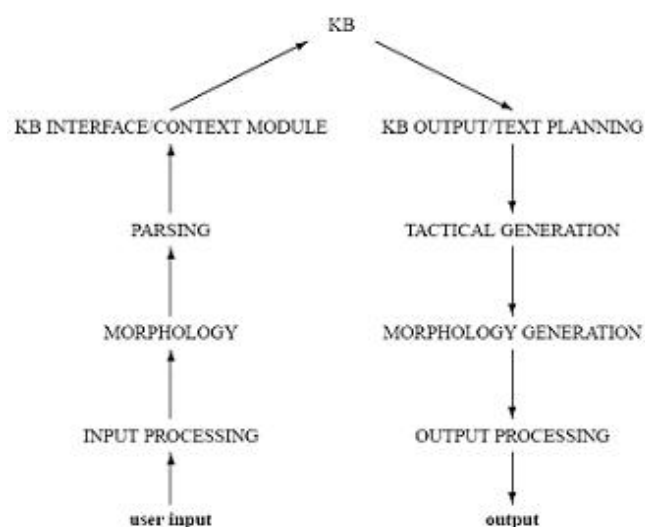


Fig 1 Steps involved in machine translation [11]

There are various methods that could be used to abate this problem, such as passing as many alternatives as possible from the morphological analyzer to the syntactic parser and hoping that the latter can resolve the ambiguities. Another method might be to attempt to operate morphological and syntactic analysis in parallel. But these can be imperfect ways of describing language and we should find a better descriptive model.

### 2.1.1 Stemming challenges

Challenges arise due to same problem of no generalization of grammatical rules. They work only for few commonly used suffixes [2] e.g. remove 'ed', 'ing', 'ness', 'tion', 'ly', 's' etc.

- Actively → active (stem) + ly (suffix)
- Wolves → wolve (stem) + s (suffix)
- (2nd example does not work well.)

As all parts of speech have no such well formulated set of rules it's not always straight forward to break word to its

stem and affixes which may lead to two types or errors after stemming-

1. Understemming error- word with same stem are converted to different stems
2. Overstemming error-word with different stem converted to same stem.

Multilingual Stemming is similar to regular stemming but uses morphological rules of several languages at the same time instead of rules relative to only one language [10]. When rules of more than one language apply to a word, extra procedures are used to determine the right rule. Although these procedures cover the majority of possible conflicts, there are cases that cannot be solved. As a result, multilingual stemming is slightly less precise than monolingual stemming for a given language.

So stemming is not suitable when-

1. Precision is major concern
2. index is too large

#### 2.1.1.1 Lemmatization

Lemmatization can be a solution for this problem. It involves detection of parts of speech (POS) followed by stemming rules as per POS [1]. It operates with the help of the dictionary. We can make use of small look up tables to store few frequent exceptions. If the word is not there in this list, then only suffix strapping or lemmatization is done. Lemmatization is more accurate than stemming requires additional resources unlike stemming. Moreover it might not be available for all the languages or it might be costly.

#### 2.1.1.2 Challenges in information retrieval

Problems in stemming impose challenges in IR. Single query used to retrieve the information may return multiple related documents along with the desired document. In such a case, we can narrow it down by adding more words to the query. But still there will be some exceptions. It would be expensive to handle exceptions in terms of memory and speed considering their relative rareness. Exact match queries can also occasionally retrieve non expected results. For instance, the exact match string "sales reports" matches "sal- report-". Moreover it is not preferred to stem short words to favor performance. So, morphological variations of short words should be implicitly included in a query.

Problem becomes severe in case of multilingual machine translation because translating different languages is not necessarily transitive. Proper handling should be done in multilingual MT.

E.g. For example, using Google, "River Bank" in English translates to "Rive" in French [15], which translates to "Ufer" in German, which translates to "Shore" in English.

## 2.2. In Syntactic Analysis

It includes challenges in tagging and parsing.

### 2.2.1 Tagging

When data is annotated using part of speech (POS), it is used for parsing and chunking. It can also be used for word alignment. But this type of tagging is not suitable for small training data. Another approach i.e. sense tagging makes use of several predetermined tags which denote the sense of the word in current context. So, this helps in task of disambiguation which is one of the important steps in machine translation. The problem in this technique is that there are no standard tags used by all people or tagging system. This tagging is difficult as compared to POS tagging.

### 2.2.2 Parsing challenges

It is necessary that every sentence will have only single parse tree. Many times parsing results in multiple parse trees. Meaning or translation relationship changes with each representation. Multiple trees for a single sentence will lead to ambiguities. Moreover it becomes highly complex to represent large sentences using parse trees. Parsing takes long time for large corpora. It slows down processing too because of high computational complexity (of the order of  $O(n^3)$  considering length of sentence and of the order of  $O(G^2)$  where  $G \rightarrow$  grammar size). Full parsing is not very robust too. These limitations affect the performance of Syntax -Based Statistical Machine Translation. Chunking can be an option for parsing being more efficient [4].

In case of multilingual MT, parsing technique changes as per the structure of language which of course different for languages with different origin.

#### 2.2.2.1 Chunking

Chunking divides sentences into their sub constituents, such as noun, verb, and prepositional phrases. Chunking might be a first step in full parsing or it's known as partial parsing. Chunks are non-overlapping regions of text which are non-recursive, non-exhaustive [3].

But within chunks, (syntactic) attachment ambiguities do not arise, and simple context-free parsing techniques are very effective. By having separate chunker and attacher, we can limit the use of expensive techniques for dealing with attachment ambiguities to the parts of grammar where they are really necessary—i.e., in the attacher.

Chunk parsing [4] approach can also be used in which parse trees are formed for individual chunk rather than for whole sentence. Alignment of these chunks will differ for multilingual MT as per every TL. It should not be a limitation of MT.

## 2.3 In Semantic Analysis

It mainly includes the challenges in disambiguation. First sense baseline is a real challenge for all-words WSD systems. In case of supervised learning, homonym disambiguation challenges are there. Structural ambiguities also come under semantics analysis. Different languages follow different structure of sentence e.g. English follows 'Subject Verb Object' structure, Indian languages follow 'Subject Object Verb' structure while Arabic language follows 'Verb Subject Object' kind of structure. A sentence can also be in passive voice. Phrases must be properly rearranged after the translation in order to have proper grammatical structure. So in multilingual MT it becomes difficult to handle different structure of different TL. Two steps can be implemented in semantic analysis, context independent interpretation and context interpretation [13].

## 3 CHALLENGES IN TRAINING

Every system should undergo proper training phase which is done using training data i.e. some part of corpus. More the training data more is the system efficiency. This training data is tagged. Corpus contains at least few billions of words. If we consider hand labeling of data, then it's a very tedious job to annotate it. As human interference is there its more error prone. If we consider automated taggers, then there are challenges due to ambiguity of POS of word in current context. In such a case automated tagger need help of morphological analyzer.

### 3.1 Data bottlenecks

As performance of MT depends on training phase, it may suffer from the problem of data sparseness. In case of small training set, there is unreliability of data prediction. This requires wide coverage system. It must cover full lexicon of languages of interest. Problem arises even when training data is from one domain and system is applied to corpora in some other area.

Now a day, parallel corpus are becoming increasingly available for different language pairs, the magnitude of such corpora that is likely to be available for most language pairs in the foreseeable future is limited. The levels of translation performance that can be achieved using today's MT models with such limited amounts of data are rather unsatisfying. Significant progress is therefore required in developing new types of translation models that better generalize from limited amounts of available training data, in order to enable the development of MT systems for a far broader range of language pairs.

### 3.2 Knowledge bottleneck

All WSD systems heavily rely on knowledge bases. Lexicon has to be morphologically rich in order to guarantee proper translation. The problem comes in creating and maintaining the knowledge base. Insufficient knowledge base will obviously affect throughput of system. Knowledge base contains billions of words. So, whenever

system searches for a specific word in lexicon, it must retrieve the target meanings in negligible time and then perform the task of disambiguation if needed. It's a costly affair to find out single word each time from billions of words. So it is better to go for domain model in which translator will deal with corpus from a specific domain only. This will result into smaller granularity of knowledge base, but we will have to compromise with precision. Such type of MT is not capable of operating on a morphologically rich language.

Knowledge acquisition bottleneck is more serious kind of problem. Manual creation of knowledge is time consuming and expensive effort. Manual identification and entry of relevant information into knowledge base with change in disambiguation scenario is practically impossible. When we consider automated knowledge acquisition, it is must for every MT so as to improve its morphological knowledge. The basic concept behind this is that what the machine has used recently, which was not there in its knowledge, will be needed in future also. And when we come to a multilingual MT, this problem becomes more severe because quality of translation depends on existing multilingual corpora. A minimum of 2 million words for a specific domain and even more for general language are required. Theoretically it is possible to reach the quality threshold but many times such large amounts of existing multilingual corpora are not there.

## 4 LANGUAGE-CONSTRAINTS

### 4.1 Due to the concepts specific to language

Cultural differences also impose challenges during translation. Some words normally related to culture have no equivalent words in other languages. E.g. Indian word 'sari' (traditional dress of Indian women) has no equivalent word in other languages. Same is the case of 'kimono' in Japanese.

In Hindi, person is addressed according to his age. Elder person is addressed using 'aap' while younger one with 'tum'. In most of the Indian languages, same scenario is there. But in language like English there is no such differentiation. 'You' is the only word used for addressing. However in German, unknown person is addressed using 'du', but a known person is addressed using 'sie'.

In many Indian languages, morphological forms of verbs are depends upon gender of the subject or the object usually observed at the end of the sentence which is the not case in English.

In Spanish, appendages of hand or foot are not differentiated, 'dedo' is the only word used to refer those. But in English tow different words are used namely 'toe' and 'finger'.

### 4.2 Due to change in Linguistic theory

Developments in MT as well as in linguistic theory must go hand in hand considering the performance of system. But practically this is very difficult to achieve. Modification of knowledge base is not simple. Especially if we consider the colloquialism, it's not feasible to modify lexicon time to time neither it's simple to go for online knowledge base as far as MT is concerned. This results in a wide communication gap between theoretical linguistics and practical MT research. Multilingual MT, being concerned with multiple languages, suffers from this problem the most.

Care should be taken for acronyms and official words from languages of interest especially in multilingual MT. Acronyms do not translate well because of different letters used in different languages in different order. To avoid this, the acronyms should become accepted words in multiple languages, they are difficult to translate. Translations of official names as opposed to unofficial and shortened names also must be accounted (e.g., United States, vs. United States of America).

### 4.3 Focused only on English

USA is the heart of development of computer technology. It is one of the most thoroughly monolingual societies in the world. For this reason, problems arising from the use of the computer with languages other than English did not at first receive much attention, and even today only few languages are there in focus. Most linguistic theory of MT is based upon phenomenon observed in English [14]. This results in less impact of linguistic theory on MT. One of the reasons is that MT research is sometimes regarded as an 'engineering' task, a search for computational methods that work with the facts of language. On the other hand, the aims of many theoretical linguists are more concerned with investigations of human aptitude. Statistical MT in particular, has focused on a small number of language pairs for which vast amounts of sentence-aligned parallel text have become available or was explicitly constructed. However, recent theories such as Lexical Functional Grammar or Generalized Phrase-Structure Grammar and their various derivatives have set out explicitly to cover as broad a range as possible, not only within one specific language, but also for different types of languages.

When we compare different languages with English, they vary in writing system, grammatical structure, way of expressing similar meanings and intentions. Many languages of the world use either a completely different writing system from English, or else a slightly different variant of the English writing system, involving in particular special characters. All these differences are problems for which computational solutions must be found. Languages which are not used widely will need deep study and special attention. Languages like Arabic and Hebrew are written from right to left. This needs change in processing technique of different input output

devices. Some languages do not even use alphabetical writing systems and character sets of about 3,000 or 6,000 characters [5]. Japanese in particular have achieved much advancement overcoming this problem.

#### 4.4 Morphological complexity

Every language has different level of morphological complexity and as far as English is concerned, the morphology is quite simple. As a result of this, recent MT research work has placed only limited focus on issues of effectively handling languages with complex morphology especially as a target language. Furthermore, all methods of MT tend to retain SL structural features; however, theoretically MT methods should be more TL oriented. But in multilingual MT, translator will have to have different MT methods for different TL.

To address this issue, researchers may have to work upon new translation models that can effectively deal with complex morphology. Suitable language model or n-gram model should be developed [14]. The issue of data sparseness should be taken care of in this case. This can be done by developing translation models that generalize far better from their given training data which is a grand challenge.

As the field develops advanced new types of models, new adaptation techniques, most suitable for these new models will need to be developed and explored.

## 5 LONG TERM CHALLENGES

The aim behind this field of AI is to achieve human level translation with or without human intervention which will need minimum efforts and minimum time to develop and to process. To achieve this there are some grand challenges in this field.

### 5.1 Domain Model

In many scenarios, ample amounts of data for training MT systems are available in specific domains and text styles. In that case, machine should consider most common general patterns of mappings between basic syntactic structures in two languages that are likely to hold across genres and text styles. Such patterns can be acquired from training data in one domain, and yet be effective when translating text from a different domain. Word-level translation pairs for the new domain still must be acquired somehow, but this can conceivably be done using far smaller amounts of new domain-specific training data.

When very limited amounts of training data is available in targeted genres and domains, a major challenge that how to adapt or extend MT systems to such new domains and genres to utilize the most out of the limited data that is available. Very little research has been done to date on methods that can identify the differences between genres and domains and use this information for targeted learning of new models [14]. Furthermore, in some scenarios, it may

be possible to actively create small amounts of targeted new training data that are most useful for improving MT system performance.

### 5.2 Human level translation quality

To achieve the MT performance close to human performance level, translation models must be able to capture advanced syntax and semantic representations and their correspond across languages. The key technical challenge is identifying representation formalisms that are rich and powerful enough on the one hand, yet are simple enough to support the development of algorithms for automatic acquisition of the models from training data, appropriately annotated. The main learning task in this case is to discover the correspondences between the structures in the two languages and to model these correspondences statistically [14]. A more challenging scenario is to learn such models using parallel data where syntax and semantic structures are available for only one of the two sides. The learning task in this case is to project the structures from one language to their corresponding structures in the other language, using word-to-word correspondences. The bottleneck is obtaining the needed training resources. Only small and very limited annotated corpora of this kind currently exist. The development of such annotated corpora is a critical enabling step, without which this research direction cannot hope to even get started. In its easiest to implement multilingual form, multilingual users contribute in multilingual settings. Unfortunately, the number of multilingual users is limited and the number of languages that any one individual communicates in is limited. As a result, in some cases users can employ their own language and that language will be translated.

### 5.3 Inter-sentential context

During translation, the focus is on an individual sentence. Thus for the correct translation of pronominal referents inter-sentential context has to be considered to resolve specific types of ambiguity. Result should be a coherent multi sentence discourse structure in the target language.

### 5.4 In Interlingua approach

Interlingua is standard of representing a natural language. It has got its own rules and structure independent of any language [18][19]. Source text is represented using Interlingua. To represent source language sentences all the way into a language-independent representation requires a complex series of NLP components. It is very complex and requires experts. Generating target sentences from Interlingua is similarly challenging. Furthermore, researchers have broadly recognized the extreme challenge of devising a true Interlingua representation that is simultaneously adequate for all languages it is intended for, rich enough to represent all intended meaning, and simple

enough for humans to agree upon and for NLP algorithms to analyze and generate from.

If such a system is developed then we can have multiple source languages for a given system.

### 6 SUGGESTIONS

The previous section of the paper discusses various grave challenges that come in the way of accurate and efficient translated target language outcome for any source input. We therefore suggest a generalized intermediate language representation that will easily map SL and TL constructs with the help of auxiliary tags that will take care of extra information that may or may not be considered (depending on the target language constructs) in the translation. Though this would involve storing additional tags in the intermediate language document, it stills proves feasible considering the various choices of target languages that we can output.

Secondly, we can also maintain tuning parameters that can decide the granularity to be considered during the translation. For highly granular systems, the word sense disambiguation can be precisely or accurately done, while low granularity, some disambiguation steps may be speeded up by compromising between precision and translation time.

### 7 PROPOSED MODEL

Based on the overall study of challenges and suggestions from previous section, we are proposing a new model for machine translation. Following figure shows the block diagram of the model.

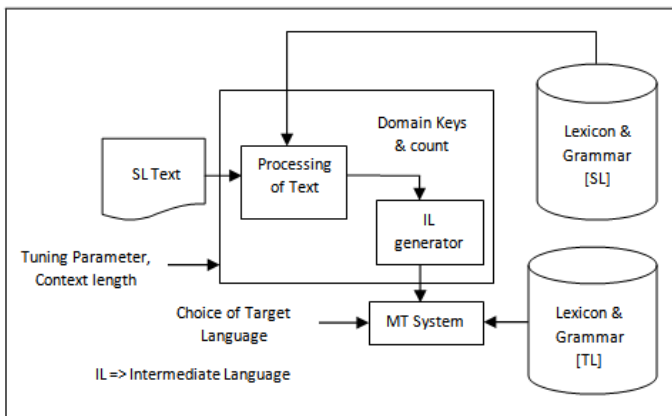


Fig2 Improved Model for Multilingual Machine Translation

Considering structural complexity of Interlingua, it is better to go for a simplistic approach like an intermediate language. It is better to consider the structural constructs for basic languages like *Sanskrit*, for the structure definition of the IL and the mapping of SL with IL and then with the TL. The reason for consideration of basic and simpler languages for the IL structure definition is because

inconsistencies in sentence formation are least in such languages. A major problem with derived languages such as *English* is that many sentences may be correctly formed in more than one ways. Hence automated syntax checks in such cases become very difficult. Also some languages are derived further into dialects for which the grammar may not be well-defined or consistent. The solution we propose to this problem is to convert the source language into its base language and generate some auxiliary tags for extra information by applying context checks on the input text. Hence, size of the context window can be varied depending on the tuning required for accuracy of meaning extraction.

Context check is very important in multi-lingual translation for various reasons explained in the previous section. One example for such case would be where in *English* a simple sentence like "Where are you going?" cannot be properly converted into *Hindi* equivalent unless the gender of the person (to whom the question is addressed) is known. This knowledge can only be extracted by mining the text surrounding the sentence to be translated.

Languages, their evolvement into other languages and their derivatives can be represented in a tree structure. Thus a language that exists at the root of this tree is the ancestral language. It is easier to map one ancestral language to another ancient language. Even we can easily map a language to its ancestral language and vice versa. By making use of this concept we can perform translation.

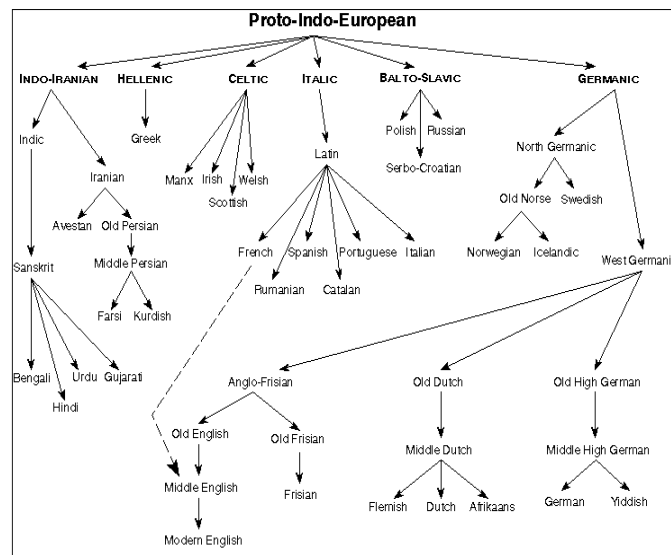


Fig 3 Language Tree [21]

In first phase, text from SL is processed using knowledge repository comprising of lexicon and grammar rules. Different domain keys are provided to processing module. Domain count will be calculated during processing and passed on to IL generator. Processed text will be

converted to intermediate language in second phase. Along with the text to be translated, other inputs can be tuning parameter and context length or context window [20] for disambiguation. This overall process will result into intermediate language file. Now the remaining task is to convert it to the desired TL text which is done in third phase by MT system. Being multilingual MT, there will be choice of TL. MT will be provided with the lexicon of corresponding TL as well as grammar rules i.e. knowledge repository for TL.

Implementation will be easy if SL and TL are having same ancestral language. For example to translate text in Marathi to Malayalam (both of which are Indian Languages), Sanskrit seems to be the obvious choice for the Intermediate language representation. Furthermore, we can even use two intermediate languages. One will be holding characteristics of SL and another that of TL.

## 8 CONCLUSION

Although the theory of machine translation is quite old but it still poses long term challenges. There is good scope of substantial improvement in translation quality, robustness across domains, genres and language styles. Research till date is mainly focused on English. Except very few languages like Japanese, German and Chinese, other languages had been neglected. To increase the scope of application on MT, Research on other languages should be equally emphasized. It is needed to develop high quality MT for other languages too. Multilingual MT will be effective one rather than having separate MT for each pair of language. MT quality should be so high so that it should match with human-level translator quality and fluency which is the ultimate goal of MT. The MT Models should be Trainable with Limited Data Resources so as to train MT in short span of time. It will reduce cost of training. Low cost techniques should be explored without compromising the quality of output. With the reduction of cost of Machine Translation, its area of application will be augmented.

## REFERENCES

- [1] Philipp Koehn, "Advanced Natural Language Processing," Lecture 2, Morphology, 24 September 2010.
- [2] D. Grimshaw, "Artificial Intelligence Topics With Agents," Fall 2001.
- [3] W. Raynor, "International Dictionary Of Artificial Intelligence," (Amacom, 1999). (Book)
- [4] Shin-ichiro KAMEI and Kazunori MURAKI, "Interlingua Developed And Utilized In Real Multilingual MT Product Systems," 1997.
- [5] W. John Hutchins, Harold L. Somers , " An Introduction To Machine Translation," The Library,University of East Anglia, Norwich, UK, 1992.
- [6] Cyril Goutte, Nicola Cancedda, Marc Dymetman, " Learning

Machine Translation," George Foster, Massachusetts Institute of Technology.

- [7] Yaser Al-Onaizan, Ulrich Germann, Ulf Hermjakob, Kevin Knight, Phillip Koehn, Daniel Marcu, Kenji Yamada, " Translation with Scarce Resources," American Association For Artificial Intelligence, 2000.
- [8] Fuji Ren, Hongchi Shi , "A General Ontology Based Multi-Lingual Multi-Function Multi-Media Intelligent System" .
- [9] Philipp koehn, <http://homepages.inf.ed.ac.uk/pkoehn> (URL)
- [10] Coveo knowledge base, <http://www.coveo.com/en/support/articles/Information%20-%20CES4-060330-3%20-%20Understanding%20Stemming.htm> (URL)
- [11] [http://www.rustyspigot.com/Computer\\_Science/Natural%20Language%20Processing2.html](http://www.rustyspigot.com/Computer_Science/Natural%20Language%20Processing2.html) (URL)
- [12] LIM Lian Tze, " Multilingual Lexicons for Machine Translation"
- [13] Shaidah Jusoh , Hejab M. Alfawareh , "Automated Translation Machines: Challenges and a Proposed Solution," Second International Conference on Intelligent Systems, Modelling and Simulation, 2011.
- [14] Alon Lavie, David Yarowsky, Kevin Knight, Chris Callison-Burch, Nizar Habash, Teruko Mitamura, "MINDS Workshops Machine Translation Working Group Final Report".
- [15] Daniel E. O'Leary, " Multilingual Knowledge Management " .
- [16] Hozumi Tanaka, "Progress in Machine Translation", ed. Sergei Nirenburg, "Multi-Lingual Machine Translation" Systems in the Future", Amsterdam: IOS Press, 1993.
- [17] [http://www.cs.bham.ac.uk/~pjh/sem1a5/pt2/pt2\\_intro\\_morphology.html](http://www.cs.bham.ac.uk/~pjh/sem1a5/pt2/pt2_intro_morphology.html) (URL)
- [18] Teruko Mitamura, Eric Nyberg, Jaime Carbonell , "Kant knowledge based , accurate natural language translation," Center for Machine Translation
- [19] Sergei Nirenburg, "New developments in knowledge-based machine translation," Carnegie Mellon University.
- [20] ROBERTO NAVIGLI, "Word Sense Disambiguation: A Survey," ACM Computing Surveys, Vol. 41, No. 2, Article 10, February 2009.
- [21] <http://andromeda.rutgers.edu/~jlynch/language.html> (URL)