ISSN:0975-9646

Pramod Salunkhe et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2557-2560

# A Research Work on English to Marathi Hybrid Translation System

Pramod Salunkhe
*M.Tech Scholar*
*Dept of Computer Engineering,*
*Bharati Vidyapeeth College of Engineering*
*Pune, India*

Mrunal Bewoor
*Professor*
*Dept of Computer Engineering,*
*Bharati Vidyapeeth College of Engineering*
*Pune, India*

Dr.Suhas Patil
*Professor*
*Dept of Computer Engineering*
*Bharati Vidyapeeth College of Engineering*
*Pune, India*

***Abstract*—Research Evaluation is major work in achieving a success. Which is been evaluated in terms of comparison of methodology or technique i.e. algorithm. Machine Translation is research domain with numerous techniques and Methodologies with every technique having advantage over other in terms of evaluation parameter. Statistical approach achieves fluency parameter in translation but lacks accuracy whereas rule based system overcome accuracy to present accurate translation but requires large time for development whereas statistical requires shorter time in development. The novel approach is to bring twin techniques in combination to achieve better results in terms of evaluation parameters of system achieve better accuracy and fluency. In hybrid approach statistical methodology runs over with rule based methodology to achieve better result output of statistical system is corrected in comparison to rule based system. This manuscript the implementation work is presented with evaluation parameters to compare system in terms of research hypothesis proposed and its achieved is presented**

***Index Terms*—H-Machine Translation, Rule based translation, Statistical based translation, Natural language Processing, Marathi, Precision, Recall, Evaluation Metric**

## I. INTRODUCTION

Machine Translation is translation obtained by machine on large scale from source to target language. India is nation where large diversity is observed in culture with diversity in spoken language .five major division hold the Indian Languages with Hindi official language and English foreign adopted language. English is the far most used language all over world and require research work to translate these English documents to native languages for knowledge gain process. World has accepted the English as major communication language .Marathi is state language of Maharashtra with world of information many articles and web articles are been written in English languages mostly in regional language .rich amount of information is written by language experts on particular topic related to local spoken language .problem lies to understand this doc's and articles in different language for which computer assisted translation is faster and better solution than human assisted translation. Government needs a Translation system for assistance and communication orders.

Large research effort is been taken by major organizations like IIT Bombay ,C-DAC ,IIT Hyderabad for better fully automated machine translation system. the research evaluation of work carried out by them shows a major upliftment in software development. Diverse approaches with unique methodology have been under taken by them to solve various research issues and problem in machine translation. Although major projects have been on English to Hindi, Tamil, Bangala, Urdu...etc. a Smaller amount of work exists in English to Marathi Translation IIT Mumbai P.S.Battacharya's work is appreciate with formation of wordnet in a Marathi parallel corpus for Marathi English and Hindi and interlingua approach in translation which is also a hybrid approach in machine translation. This research article is been documented in 5 subgroups. Subdivision I gives introduction on subject, subdivision II Related work and survey on systems, Subdivision III core technique in our work, IV Implementation details, V comparative study and Evaluation parameters, concluding mark on work.

.

## II. RELATED WORK

This subdivision currently aspect at Major Indian translation research work. The restrictions we gaze at stand: linguistic couple, formalism and approach for supervision difficulties, ambiguity, difficulty & presentation domain of every scheme. The article presents all major research work with languages as Hindi Bengali, Tamil, and Marathi, Punjabi as source or target but major approach in English to Marathi as that is our aim of research.

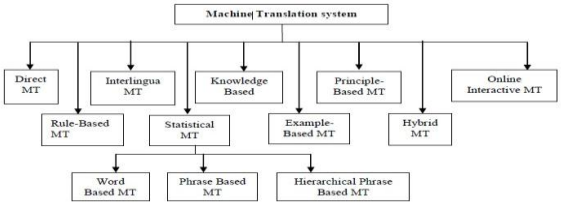


Fig 1: MT Schemes

The literature review mainly focuses on Marathi Translation Schemes [12] author has proposed UI tags for web pages translation which proposes hybrid process that builds bilingual dictionary on RBI portal and parser is built in C.[13] research scholar as built rule based scheme for E-M translation which

extend better process with grammar correction sentiment examination and spell check.[14] the writer has constructed E-M scheme with on rule scheme for Assertive type Sentences OPEN NLP TOOLS Have been integrated .[15] rule based scheme is been built with a comparative Evaluation to google Translator large work on Grammar lexicaon is been built for morphological better output. [16] the hybrid approach is been buit with intermediate language Selection for machine translation system is been developed on parallel corpa and match mapping is process. In all a hybrid approach is suited for E-M Translation and can develop Better System in terms of Acuuracy so the research work focus is on Hybrid System. The Summarization work is incorporatated from previous work of author[7][11][17].

## III. CORE METHODOLOGY

Machine assisted translation can be classified in broad way as rule method, statistical Method ,example based method, hybrid system, in every domain it can be further classified as knowledge based or corpus based system , word ,phrase based system which are statistical scheme. most system are been designed on rule based and statistical approach. hybrid approach is the best one and minimal implementation is been seen on it.In this research work statistical system is been built on corpus downloaded from IITB produced by pushpak Bhattacharya. The corpus is been cleaned and developed for agriculture Domain and consists of Inter- lingual Translations for English Marathi English. At current working only for English to Marathi only relevant corpus clusters have been taken into consideration for further the Hindi cluster can be taken. This research work Core Methodology in every module is novel and unique that helps to achieve better system performance.

**Module A:** Statistical Translation: Corpus Based matching of sentences to sentence→Corpus Matching

**Module B:** Rule based Translation: Mapping of English Generated Rules to Marathi Rules for accurate Translation→Sematic Mapping**.**

**Module C:** Hybrid Translation: get Statistical output and compare with Rule based system to replace with correct appropriate words and correct grammatical sentences →A + B Methodology

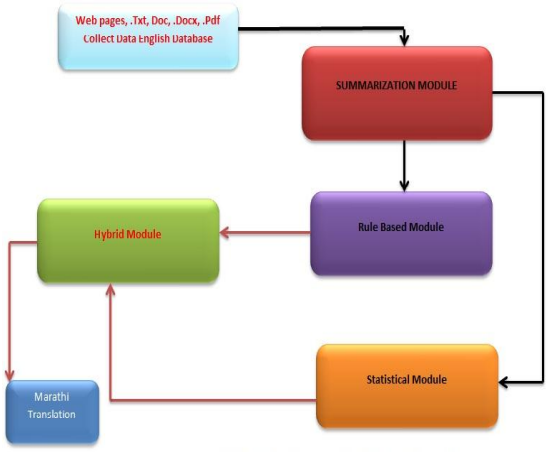


**Fig 2 : Implementation Architecture of System**

www.indiaagronet.com is been selected for extracting information from web the extracted web pages are then been summarized to extract main content from web this extracted information is passed to Translation system. The proposed translation system consists of 3 individual systems that run in parallel to same input translation and produce three Different outputs for given input .hybrid system takes input from statistical system and apply accuracy parameter i.e. .grammatical corrections on sentence to produce better output. The detailed algorithm for rule based implementation is.

**Algorith for Rule based System**

**Process 1**: Create Database of words and rule
**Process:**2 craft marathi rules for dataset based english sentences.
**Process 3**:Semantic mapping of new Translation input to English sentence Srcuture→mapping it to marathi rule
**Process:4** generate output

**Algorithm for Statistical system steps are as:**

1. Creation of Parallel Corpus .
2. for Every Sentence Si in Paragraph (P) Match Sj {E}
3. Sentence Detection for English Language
4. Match Equivalent English to Marathi sentence
5. Extract detected sentence from Marathi corpus as output.

**Algorithm for Hybrid based system steps are as:**

1. Input from statistical system
2. Apply grammatical correction rule on statistical output.
3. Replace appropriate words from rule based to statistical output
4. The result is better output.

## V. COMPARATIVE STUDY AND EVALUTION

A study and scrutiny of literature examination reveals following comparative study on Machine Translation system which in further confidence support your implementation. The detailed comparative study of our proposed system is been given in with comparative analysis to Google Translator Transmuture system and other online Translators**.**

| MT approach | Advantages | Disadvantages |
|---|---|---|
| Rule-based | 1. Easy to build an initial system<br>2. Based on linguistic theories<br>3. Effective for core phenomena<br>4. Better choice for domain specific translation<br>5. The quality of translation is good for domain specific systems | 1. Rules are formulated by experts<br>2. Difficult to maintain and extend<br>3. Ineffective for managerial phenomena<br>4. The number of rules will grow drastically in case of general translation systems |
| Knowledge-based | 1. Based on taxonomy of knowledge<br>2. Contains an inference engine<br>3. Interlingual representation | 1. Hard to build knowledge hierarchy<br>2. Hard to define granularity of knowledge<br>3. Hard to represent knowledge |
| Example-based | 1. Extracts knowledge from corpus<br>2. Based on translation patterns in corpus<br>3. Reduces the human cost | 1. Similarity measure is sensitive to system<br>2. Search cost is more<br>3. Knowledge acquisition problem still persists |
| Statistical | 1. Does not consider language grammar for translation<br>2. Extracts knowledge from corpus<br>3. Reduces the human errors<br>4. Model is mathematically grounded | 1. No linguistic background<br>2. Search cost is expensive<br>3. Hard to capture long distance phenomena<br>4. Require huge amount of parallel corpora<br>5. The translation quality will be very coarse due to lack of sufficient corpora |

Fig 3:Evaluation of MT schemes

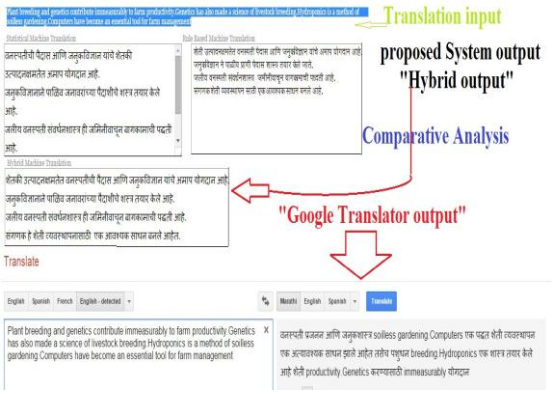


Fig 4: comparative Examination 1

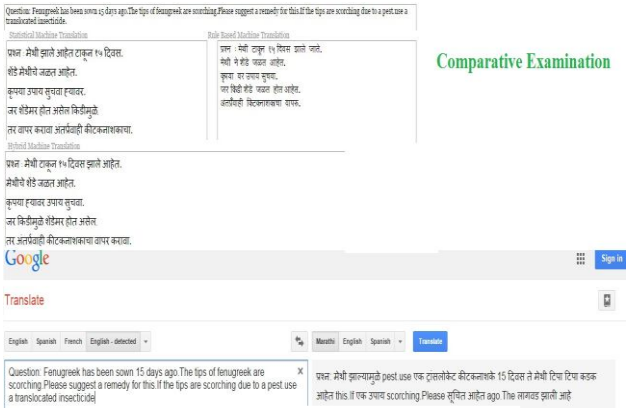


Fig 5: Comparative Examination 2

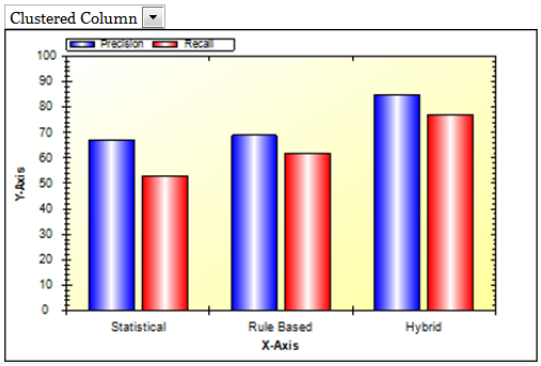


Fig 6: Graphical Evaluation

## VI. SYSTEM EVALUATION PARAMETERS

In Any system Evaluation is very important factor that certificates success of research task accomplished. Is quite a divergent task to evaluate Machine translation systems Translation output can have similar and semantic equivalent result .A system can be judged with certain parameters in constraints to other, a graphical evaluation is best that maps system inputs to outputs. A metric is setup as golden evaluation that apply in generalization to every system .in our evaluation approach we used generalized evaluation to specialized evaluation for our hybrid system as one particular metric would always present the output of system in overall effective way. Every assessment has→metric and every metric has →goals →that evaluate success and failure rate of system→ future improvement in system. Manual evaluation is dependent on adequacy and Fluency.

| Adequacy | |
|---|---|
| 5 | all meaning |
| 4 | most meaning |
| 3 | much meaning |
| 2 | little meaning |
| 1 | none |

| Fluency | |
|---|---|
| 5 | flawless English |
| 4 | good English |
| 3 | non-native English |
| 2 | disfluent English |
| 1 | incomprehensible |

Fig 6: Manual Evaluation

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

**PRECISION**

$$\frac{correct}{output\text{-}length} = \frac{3}{6} = 50\%$$

**Recall**

$$\frac{correct}{reference\text{-}length} = \frac{3}{7} = 43\%$$

## VII. DISCUSSION OF EVALUATION METRIC

Automated Evaluation is done with precision and Recall word Error rate (WER) and BLUE score, METEOR. manog this Precision and recall is found to be the best and most reliable method for dynamic evaluation of system .the above comparative analysis show that system achieves adequacy of 85% and Fluency of 90% on basis of manual evaluation and precision of 90 and 85% respectively with hybrid system .on and average a precision of 87.5% with comparison to machine translation. A new evaluation parameter is required that overcomes semantic understanding of words and evaluates the output in better way the new parameter we propose is "S-mark" a semantic or similarity score evaluation taking in consider synomous meaning of translations input to reference or golden standard evaluation stored sentences .

## VIII. REMARK AND AREA OF FURTHER SCOPE

The hybrid system achieves a better output considered to statistical and rule based .the system overcomes the lacking and cons of both to form a better result combining the two translation schemes and gives a by a good quality translation in terms of Accuracy and fluency also. A research scope in area of research is to evaluate the system in different metric as we proposed a new metric "S-mark" which would make the evaluation of system better and also evaluate the effort of research scholar in better and statistical way to prove the effectiveness of his system. The system can be made in better translation with application of Marathi wordnet or development of trans-lingual word net for all language. This proposed system can be extended to Interlingua example or statistical rule based or interlingua-hybrid translation would make translation in three languages English-Marathi-Hindi

## ACKNOWLEDGMENT

I firstly acknowledge Prof. Pushapak Bhattacharya of IITB for the corpus provided for statistical machine translation the corpus being clean and better arranged made our research work easy I also than him and his team for their research effort in machine translation. I truly acknowledge Prof.M.S.Bewoor Prof.S.H Patil. The core motivation for research better has been prof.D.M.Takhore and I thank all other in and indirectly all authors for their research work.

## REFERENCES

[1] Abhay Adapanwar, Anita Gajare "Rule Based English to Marathi Translation of Assertive Sentence ,International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013.

[2]S. B. Kulkarni , P. D. Deshmukh "Linguistic Divergence Patterns in English to Marathi Translation International Journal of Computer Applications (0975 – 8887) Volume 87 – No.4, February 2014.

[3] J. Ramanand, Akshay Ukey, Brahm Kiran Singh, Pushapak Bhattacharyya "Mapping and Structural Analysis of Multi-lingualWordnets" Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.

[4] Sitender, Seema Bawa "Survey of Indian Machine Translation Systems International Journal of Computer ScienceAnd Technology.

[5] Translation Resources, Services and Tools for Indian Languages Salil Badodekarsalil@cse.iitb.ac.inComputer Science and Engineering DepartmentIndian Institute of Technology [online pdf]

[6] Sandeep Kumar Vishwakarma "Graph Based Approach to Word Sense Disambiguation forHindi Language" International Journal of Scientific Research Engineering & Technology (IJSRET)

[7] Pramod Salunkhe, Prof.M.S.Bewoor, Dr.S.H Patil ,Prof.Dr.S.D.Joshi "Summarization and Hybrid Machine TranslationSystem for English to Marathi:(H-Machine Translation)" International conference IEEE CCEEDS 2015 Vijayawada.

[8] Sreelekha.S, Raj Dabre, Pushpak Bhattacharyya "Comparison of SMT and RBMT,The Requirement of Hybridization for Marathi – Hindi MT" [online pdf]

[9] Amruta Godase, Sharvari Govilkar"Survey Of Machine Translation Development for Indian Regional Languages" International Journal of Modern Trends in Engineeringand Research IJMTER-2014.

[10] Tagyoung ChungDaniel Gildea "Unsupervised Tokenization for Machine Translation"[online available]

[11] Text Rank: A Novel Concept for Extraction Based Dipti D.Pawar,Mrs.Prof.M.S.Bewoor,Prof.Dr.S.H.Patil International Journal of Computer Science and Information Technologies 2014,May 2014

[12] M.L.Dhore & S.X.Dixit (2011) "English to Devnagari Translation for UI Labels of Commercial web based Interactive Applications.

[13] Devika Pishartoy, Priya, Sayli Wandkar (2012) "Exteneding capabilities of English to Marathi machine Translator.

[14] Abhay A, Anita G, Paurnima T, Prajakta G (2013), "Rule based English to Marathi translation of Assertive sentence

[15] G.V.Gajre, G.Kharate, H. Kulkarni (2014), "Transmuter: An approach to Rule-based English to Marathi Machine Translation"(2014)

[16]priyanka-choudhary " A approach for interlingua example basedE-MTranslation,IJCSIT,2015.

[17] Survey on Extractive Text Summarization Approaches M.S patil,,Prof.M.S.Bewoor,,Prof.Dr.S.H.Patil,NCI2TM:2014.