Research Article Impact Factor: 4.226 ISSN: 2319-507X
K. S. Gilda, IJPRET, 2017; Volume 6 (2): 275-283 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

SPECIAL ISSUE FOR
INTERNATIONAL LEVEL CONFERENCE
"ADVANCES IN SCIENCE, TECHNOLOGY & MANAGEMENT"
(IC-ASTM)

MACHINE TRANSLATION: APPROACHES AND EVALUATION

K. S. GILDA, S. S. DIXIT

Assistant Professor, Department Of Computer Science & Engineering, C.O.E.&T., Akola, India.

Accepted Date: 05/09/2017; Published Date: 10/10/2017

Abstract: Machine translation is gaining attention in recent years, with the wide usage of search engines. MT is sub field of computational linguistics that translates one natural language to another. The concept of MT is based on Natural Language Processing. However the widespread of MT systems depend on quality of MT output. This paper takes overview of different approaches to MT, challenges in MT, need for MT evaluation and several such issues.

Keywords: Machine Translation, Natural Language Understanding, MT evaluation, Interlingua.

Corresponding Author: K. S. GILDA

Co Author: - S. S. DIXIT

Access Online On:

www.ijpret.com

How to Cite This Article:

K. S. Gilda, IJPRET, 2017; Volume 6 (2): 275-283

PAPER-QR CODE

INTRODUCTION

Machine Translation (MT) is one of the emerging areas of Artificial Intelligence. MT is subfield of computational linguistics that translates one natural language into another. It involves translation of either text or speech. This paper focuses on text translation only. MT includes understanding the source text and generation of target text. Therein lies the challenge in machine translation. MT is not just word to word translation but meaning of source text should be reflected in target text as if translation is done by human, at the same time target text should be in well-formed according to target language. This is beyond the current technologies. Definitely, MT saves the time and efforts of human translator and very helpful in the cases where no human translator is available.

Success of MT needs issue of *Natural Language Understanding (NLU)* to be solved first. Natural language refers to the language spoken by people such as Hindi, Marathi, Sanskrit, English, French etc. as opposed to artificial languages such as C, C++, java etc. NLU is to build intelligent computers that can interact with human being like a human being!! NLU motivated range of computational techniques for analyzing natural text or speech for the purpose of achieving human like processing for range of applications.

Another issue is - no standard is available for testing the results of MT. Human *evaluation of MT* output is an expensive process and not preferred when evaluations must be performed quickly and frequently in order to measure progress. There are few automatic evaluation methods which can be employed as an alternative to human evaluation methods.

Rest of the paper is organized as follows. Section II describes MT process in brief. Approaches to MT are elaborated in section III. Challenges in MT are discussed in section IV. Section V describes MT evaluation methods.

II. MACHINE TRANSLATION (MT)

MT investigates the use of software to translate text from one natural language to another. The translation is a two step process- 1.Understanding source text 2.Generation of target text. MT requires in depth knowledge of all features of source and target language such as grammar, semantics, syntax, idioms etc. There are various approaches for the process of MT. These approaches employ roughly one or more phases of *NLP[1]*. General description of these phases is as follows.

- *Lexical/Morphological analysis*: Individual words are analyzed in their components (such as nouns, verbs, adverbs etc.) and nonword tokens (such as punctuations) are separated from the sentence.
- *Syntactic analysis*: Linear sequences of words are transformed into structures that show how the words relate to each other according to source language grammar. For example, the sentence “The boy the go the school” is rejected by syntactic analyzer.
- *Semantic analysis*: The structures created by syntactic analyzer are assigned meanings. Here is the mapping between syntactic structures and task domain. For example, the sentence “65 years old baby win the game.” is rejected as semantically anomalous.
- *Contextual analysis*: The sentence is related with current context in the text. The sentence may depend on previous sentence. For example, in the sentence, “He goes there.” Depends on the context to identify who “He” is.

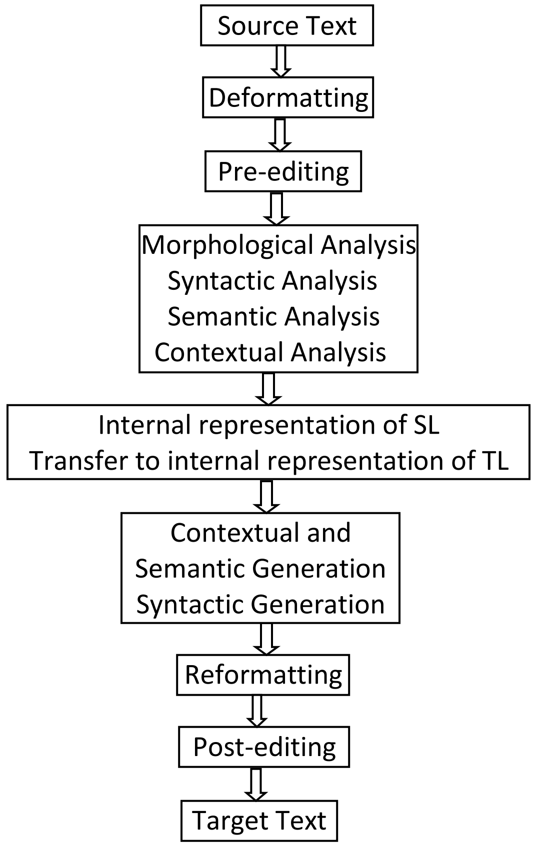

Fig.1: Typical MT process

Fig.1 [6] shows the typical sequence of machine translation process. Preprocessing the input include de-formatting and pre-editing source text. This is followed by three modules-source text understanding (as discussed above), internal representation of text and target text generation. Then, post processing of target text is done which include reformatting and post-editing.

II. APPROACHES TO MACHINE TRANSLATION

Approaches to MT can be classified into five major categories: Direct, Rule-based, Corpus-based, Knowledge-based and Hybrid MT. Fig.2[3] show the classification of approaches to Machine Translation.

Machine Translation

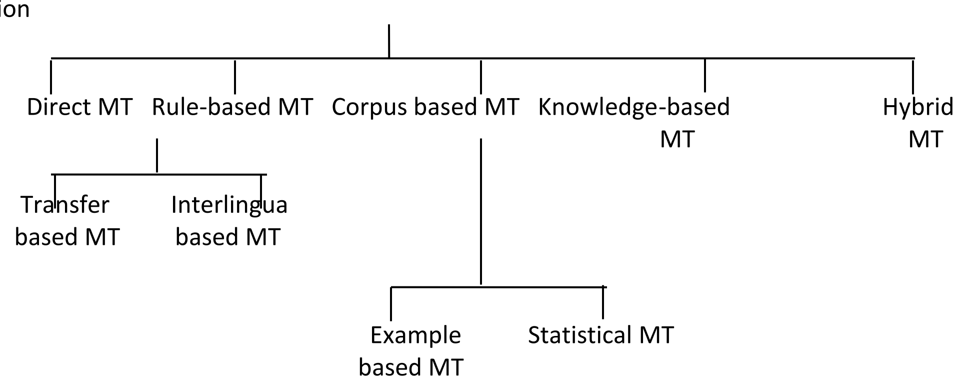

Fig.2: Approaches to MT

1. Direct Machine Translation (DTM) [3]

Direct MT system provides direct translation. No intermediate representation will be involved in the approach. It carries out word by word translation with the help of a bilingual dictionary, followed by some syntactic rearrangement. It involves little analysis of source text, no parsing and mainly relies on the quality of bilingual dictionary. Consider the example 'Rita goes to the school'. Let's see how it is treated by DTM for translation into Hindi.

Input (English sentence)	- Rita goes to the school.
Word by word translation	- jhrk Tkkrh esa ikB'kkyk
Syntactic rearrangement	- jhrk ikB'kkyk esa Tkkrh A

The limitation for this approach is it does not consider the structure and relationship between words. Another limitation is that no attempts to disambiguate the sense as majority of words in our natural language are ambiguous. Also, the system which is developed for a particular language pair will not be suitable for another language pair. So, the system is not adaptable.

2. Rule-based Machine Translation (RBTM)

In rule based approach the system relies on hand made linguistic rules for performing the MT process. There are two types of rule-based MT approaches-Transfer-based and Interlingua based Machine Translation.

2.1 Transfer-based Machine Translation [8, 3]

In this approach the source text is analyzed according to source language to produce syntactic structure representation. Then parsing of syntactic structure may be done to produce tree structure according to source language grammar. Then, source language syntactic structure is transferred to target language syntactic structure. After that the system uses bilingual dictionary for word by word translation. The disambiguities occurring during translation are resolved at the end according to target language grammar. Let us workout the system with our previous example.

Input	- Rita goes to the school.
Analysis output	- (S (NP (NNP Rita)) (VP (VB goes) (PP (TO to) (NP (DT the) (NN school)))))
Syntactical transfer	- (S (NP (NNP Rita)) (VP (PP (NP (DT the) (NN school)) (TO to)) (VB goes)))
Hindi lexicalization	- (S (NP (NNP jhrk)) (VP (PP (NP (NN ikB'kkyk)) (IN में)) (VBD Tkkrh)
Hindi Sentence	- jhrk ikB'kkyk esa Tkkrh A

The advantages of the system are modular structure and capability to disambiguate the word sense. Disadvantage of the system is its adaptability or extensibility for a group of language pairs. If we are trying to develop a system for English to Hindi and Sanskrit to Hindi we need two different source language analyzers.

2.2 Interlingua-based Machine Translation [14]

In Interlingua based approach, the source language will be converted in to a language independent meaning representation called 'interlingua'. From this interlingual representation, the target language text can be generated. In short the translation in this approach is a two-stage process, i.e. analysis and synthesis.

Rita goes to the school.
(*goes
(tense present)
(mood declarative)
(punctuation period)
(subject (*Rita
(number singular)
(Location (*school
(reference definite)
(number singular)))

Interlingua is a recursive list-based structural representation of the information content of individual source sentences. An interlingua frame consists of a head concept, feature-value pairs, and semantic slots which in turn contain nested interlingua frames. Many efforts are required in defining a universal and abstract interligual representation. Above is the interligua representation of the example.

Advantage of Interlingua-based MT is that source language can be converted to multiple target languages. For example, same Interlingua can be used for English to Hindi and English to Marathi translation.

3. Corpus Based Machine Translation [3]

Corpus is a large collection of text or speech in a language. A dictionary is a description of the vocabulary of a language arranged alphabetically whereas; corpus is large and structured set of texts. A corpus may contain texts in a single language (monolingual corpus) or text data in multiple languages (multilingual corpus). Corpus based MT needs less effort from the side of language experts. But, they require large amount of sentence aligned parallel corpus. Multilingual corpora that have been specially formatted for side-by-side comparison are called aligned parallel corpora. The corpus based approach can be further classified into two types statistical and example based Machine Translation.

3.1 Statistical Machine Translation (SMT) [5, 3]

The SMT is inspired by the noisy channel and mainly used in Automatic Speech Recognition (ASR). SMT is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora.

P(e) is the chance that e happens. For example, if e is the English string “I like snakes,” then P(e) is the chance that a certain person at a certain time will say “I like snakes” as opposed to saying something else. Conditional probability P(h | e) is the chance that upon seeing Hindi sentence h, a translator will produce English sentence e. In SMT, given a Hindi sentence h, we seek the English sentence e that maximizes P(h | e) (The most likely translation).

The SMT system models a target language sentence T, given a source language sentence S, as the product of translation probability P(S,T) and target language probability P(T). We can assume that when a native speaker of Hindi produces an English sentence he will be having a Hindi sentence in mind and will be translating it in to English mentally. The goal of SMT is to find the sentence h that the native speaker in his mind when he produces e.

3.2 Example-based Machine Translation (EBMT) [3]

EBMT system uses past translation examples to generate translation for a given source language text. An EBMT system maintains an example-base consisting of translation examples between source and target languages. The system has two main modules- retrieval and adaption. When a sentence is given to the system, the system retrieves a similar sentence from the example-base and its translation. Then it adapts the example to generate the target language sentence of the input sentence. The adaption may involve addition, deletion, insertion of morphological words, constituent words or suffixes. The EBMT system is based on the idea that similar sentence will have similar translation.

Let’s elaborate the concept with the help of an example. Consider English- Hindi translation as shown below. Using this retrieved pair the system will replace Ravi with Santhosh and उपन्यास with पत्र in the translation.

Input -	Santhosh is writing a letter.	
Example base -	Vikram wrote a poem.	(1)
	Anand is writing.	(2)
	Ravi is writing an essay.	(3)
	Mukesh writes a Malayalam poem.	(4)
Selection by the retriever		
	Ravi is writing an essay	
	रवि एक उपन्यास लिख रहा है ।	

4. Knowledge-based Machine Translation (KBMT) [1, 3, 12]

Semantic based language analysis has been introduced by Artificial Intelligence (AI) researchers. This approach requires a large amount of ontological and lexical knowledge. The KBMT approach includes semantic parsing, lexical decomposition in to semantic networks and resolution of ambiguities and uncertainties by reference of knowledge-base. Here is the example of ontology for KBMT.

```
person ::= ('person'
            ('isa' creature)
            ('agent-of' (Eat, Drink, Move, Attck, Love ....))
            ('consists-of'(Hand Foot, ....)))
      computer-user ::= ('computer-user'
            ('isa' person)
            ('agent-of (+(Operate)))
            ('subworld' computer-world))
```

5. Hybrid Machine Translation (HMT) [4]

Hybrid machine translation is a method of machine translation that is characterized by the use of multiple machine translation approaches within a single machine translation system. The HMT systems differ in a number of ways; some of them are discussed here:

- *Rules post-processed by statistics*: Translations are performed using a rule-based engine. Statistics are then used in an attempt to adjust/correct the output from the rules engine.
- *Statistics guided by rules*: Rules are used to pre-process data in an attempt to better guide the statistical engine. Rules are also used to post-process the statistical output to perform functions such as normalization. This approach has a lot more power, flexibility and control when translating.

IV. CHALLENGES IN MACHINE TRANSLATION [6]

In the easiest form of MT, one just puts the original text in and the machine gives the translation. MT is just one click away and cheap alternative to human translation. Though the MT is easy to use, fast and cheap, there are many challenges one has to face when attempt to do machine translation.

- Not all the words in one language have equivalent words in another language. In some cases a word in one language is to be expressed by group of words in another.
- Two given languages may have completely different structures. For example English has SVO structure while Tamil has SOV structure.
- Sometimes there is a lack of one-to-one correspondence of parts of speech between two languages. For example, color terms of Tamil are nouns whereas in English they are adjectives.
- Words can have more than one meaning and sometimes group of words or whole sentence may have more than one meaning in a language. This problem is called ambiguity.
- Translation requires not only vocabulary and grammar but also knowledge gathered from past experience.
- The programmer should understand the rules under which complex human language operates and how the mechanism of this operation can be simulated by automatic means.
- The simulation of human language behavior by automatic means is almost impossible to achieve as the language is open and dynamic system in constant change. More importantly the system is not yet completely understood.
- It is too difficult for the software programs to predict meaning.

V. MACHINE TRANSLATION EVALUATION (MTE)

MT has gained attention in translation community during recent years and has become crucial in search engines. However the widespread of MT systems largely depends on the quality of machine translation output [10]. Therein lies the need for machine translation evaluation (MTE). Though MTE has a long history, advancements in technology has drastically changed the MT approaches, so also the MTE methods. This section takes an overview of MTE methods in brief. Different approaches have been introduced to address the issues of evaluating translations from one natural language to another. MT evaluation approaches can be classified as follows:

MTE Methods

Human Evaluation	Diagnostic Evaluation	Automatic Evaluation
Accuracy	Morphology	BLEU
Comprehensibility	Syntax	METEOR
Fluency	Semantics	NIST
Adequacy	Pragmatics	WER
	Lexicology	LEPOR

Fig.3: MT evaluation methods

5.1 Human Evaluation [10,19]

Human evaluation methods involve various metrics like Accuracy, Comprehensibility, Fluency, Adequacy, Fidelity etc. Although human evaluation has lots of advantages and can evaluate many aspects of translation, it is not preferred, as it is expensive, time consuming and cannot be reused. Moreover human evaluation is subjective.

5.2 Diagnostic Evaluation [19]

Along with the subjective (human) evaluation, in-depth evaluation of MT system to handle various linguistic processes is also important. The aim of the linguistic / diagnostic evaluation is to test whether the translation system can handle morphology, syntax & lexis when translating from source language to target language.

5.3 Automatic Evaluation [20,22,23]

Although manual evaluation has a lot of advantages and can evaluate many aspects of successful translations, it is discouraged because human resources are more expensive specially language experts. Automatic evaluation is preferable as it is quick, inexpensive and mostly language independent. It is useful wherein frequent evaluations are required. In automatic evaluation, translations are compared with reference sentences produced by human. Effective automatic evaluation metric has to satisfy some requirements. A good metric should be

- as sensitive as possible to differences in MT quality between different systems, and between different versions of the same system.
- consistent (same MT system on similar texts should produce similar scores)
- reliable (MT systems that score similarly can be trusted to perform similarly)
- general (applicable to different MT tasks in a wide range of domains and scenarios).

Some of the common Automatic Evaluation Metrics are discussed below.

5.3.1 Word error rate (WER) [23] is a metric based on the Levenshtein distance, where the Levenshtein distance works at the character level, WER works at the word level. It was originally used for measuring the performance of speech recognition systems, but is also used in the evaluation of machine translation. The metric is based on the calculation of the number of words that differ between a piece of machine translated text and a reference translation. A related metric is the Position-independent word error rate (PER), this allows for re-ordering of words and sequences of words between a translated text and a reference translation.

5.3.2 Bilingual Evaluation Understudy (BLEU) [21]: BLEU was one of the first and the most popular metric for automatic machine translation evaluation metric, to report high correlation with human judgments of quality. The main idea of BLEU is to consider matches of larger n-grams between the input and reference translations. The metric calculates scores for individual segments, generally sentences—then averages these scores over the whole corpus for a final score. It has been shown to correlate highly with human judgments of quality at the corpus level.It also handles the role of word order.

BLEU is a precision-oriented metric in that it measures how much of the system output is correct, rather than measuring whether the references are fully reproduced in the system output [20]. BLEU uses a modified form of precision to compare a candidate translation against multiple reference translations. The metric modifies simple precision since machine translation systems generate more words than appear in a reference text.

Despite several shortcomings, BLEU remains the most commonly used automatic metric both for the optimization of system parameters and for final evaluation of the quality of an MT system. The use of the BLEU metric has driven development in the MT research community, and it is now the automatic evaluation metric against which all new metrics are compared.

5.3.3 METEOR: metric is designed to address some of the deficiencies inherent in the BLEU metric. The metric was designed after research by Lavie (2004) and Banerjee(2005) into the significance of recall in evaluation metrics. METEOR (Metric for Evaluation of Translation with Explicit ORdering)[20] is a recall-oriented metric. The metric is based on the weighted harmonic mean of unigram precision and unigram recall. METEOR uses several stages of word matching between the system output and the reference translations in order to align the two strings.

METEOR also includes some other features not found in other metrics, such as synonymy matching, where instead of matching only on the exact word form; the metric also matches on synonyms. For example, the word "good" in the reference rendering as "well" in the translation counts as a match. The metric also includes a stemmer, which lemmatises words and matches on the lemmatised forms. The implementation of the metric is modular as the algorithms that match words are implemented as modules, and new modules that implement different matching strategies may easily be added.

5.3.4 NIST: metric is based on the BLEU metric, but with some alterations. Its name comes from the US National Institute of Standards and Technology. Where BLEU simply calculates n-gram precision adding equal weight to each one, NIST also calculates how informative a particular n-gram is. NIST uses arithmetic mean of n-gram counts whereas BLEU uses geometric mean [20]. When a correct n-gram is found, if it is rarer then more weight is given. For example, if the bigram "on the" correctly matches, it receives lower weight than the correct matching of bigram "interesting calculations," as this is less likely to occur. NIST also differs from BLEU in its calculation of the brevity penalty, insofar as small variations in translation length do not impact the overall score as much.

5.3.5 LEPOR[23]: A new MT evaluation metric LEPOR was proposed as the combination of many evaluation factors including existing ones (precision, recall) and modified ones (sentence-length penalty and n-gram based word order penalty). The experiments were tested on eight language pairs including English-to-other (Spanish, French, German and Czech) and the inverse, and showed that LEPOR yielded higher system-level correlation with human judgments than several existing metrics such as BLEU, Meteor-1.3 etc. It is a metric considering Enhanced **Le**ngth penalty, **P**recision, n-gram **Po**sition difference penalty and **R**ecall. Furthermore, they design a set of parameters to tune the weights of the sub-factors according to different language pairs. It is an open source metric.

There are several other metrics available, but least frequently used. Though these metrics yield a considerable degree of output, still they suffer some weaknesses.

5.4 Weaknesses of Machine Translation evaluation [23]

- Good performance on certain language pairs
- Rely on many linguistic features for good performance
- Employ incomprehensive factors.

VI. CONCLUSION

Machine translation has been an active research subfield of artificial intelligence. It is a difficult task as natural languages are highly complex. Thus we find several different approaches for machine translations and still are lacking in various aspects. The widespread of machine translation depends on the MT output quality. Hence MT evaluation is an essential task. Both human evaluation (subjective) and automatic evaluation (quantitative) have some advantages and disadvantages. Human evaluation is expensive, slow and not reusable and it has inherent problem of inter-annotator agreement. On the other hand automatic evaluation is quick, inexpensive and mostly language independent. For a reliable outcome human evaluation along with error analysis is a preferred combination.

References

[1] Kevin Knight , Elaine Rich and B. Nair, “Artificial Intelligence”, 3E, MGH Education, 2009

[2] Akshar Bharati, Vineet Chaitanya and Rajeev Sangal, "Natural Language Processing: Paninian perspective", Prentice Hall of India, 1994
[3] https://www.slideshare.net/jaganadhg/a-tutorial-on-machine-translation
[4] https://en.wikipedia.org/wiki/Machine_translation
[5] https://www.isi.edu/natural-language/mt/wkbk.rtf
[6] http://language.worldofcomputing.net/machine-translation/machine-translation-process.html
[7] https://blog.algorithmia.com/introduction-natural-language-processing-nlp/
[8] G. V. Garje, G. K. Kharate and H. Kulkarni, " Transmuter: An Approach to Rule-based English to Marathi Machine Translation", International Journal of Computer Application, Volume 98 – No.21, July 2014
[9] Simon Corston-Oliver, Michael Gamon and Chris Brockett, "A machine learning approach to the automatic evaluation of machine translation", Microsoft Research
[10] Ibrahim Ahmed Ibrahim Saleh Sabek, "Intelligent Hybrid Man-Machine Translation Evaluation",Thesis, Alexandria University, June 2014
[12] Sergei Nirenburg, "Knowledge based machine translation", Kluwer academic publishers, 1989
[13] Anand Ballabh and Dr. Umesh Chandra Jaiswal, "A Study Of Machine Translation Methods and their challenges", International Journal of Advance Research In Science And Engineering, Vol. No.4, Special Issue (01), April 2015
[14] DeryleW. Lonsdale, Alexander M. Franz, and John R. R. Leavitt, "Large-Scale Machine Translation: an Interlingua Approach", Center for Machine Translation, Carnegie Mellon University
[15] M. D. Okpor, "Machine Translation Approaches: Issues and Challenges", IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2, September 2014
[16] Vimal Mishra and R.B.Mishra, "ANN and Rule Based Model for English to Sanskrit Machine Translation", September 2017
[17] Mr. Krushna Belerao and Prof. V.S.Wadne, " Machine Translation Using Open NLP and Rules Based System 'English to Marathi Translator'", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 2, Issue: 6 , June 2014
[18] Latha R. Nair and David Peter S., "Machine Translation Systems for Indian Languages", International Journal of Computer Applications, Vol. 39– No.1, February 2012
[19] http://www.tdil-dc.in/undertaking/article/369153MT_Evaluation_Strategy_1.3.pdf
[20] https://www.cs.cmu.edu/~alavie/papers/GALE-book-Ch5.pdf
[21] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 311-318.
[22] Lucia Specia, Dhwaj Raj, Marco Turchi, "Machine translation evaluation versus quality estimation", Springer Science+Business Media B.V. 2010, Mach Translat (2010) 24:39–50
[23] https://www.slideshare.net/AaronHanLiFeng/lepor-an-augmented-machine translation-evaluation-metric-thesis-ppt