V.V.P. Engineering College

राष्ट्राय स्वाहा इदं न मम् ।

Machine Translation using Natural Language Processing

Adarsh Ojha, Nand Gondha, Darshil Bavishi

Adarsh ojha, Rajkot, Gujarat.
Nand Gondha, Rajkot, Gujarat.
Darshil Bavishi, Rajkot, Gujarat.

ARTICLE INFO

Keywords:
Natural Language Processing, Machine Translation, Example Based Machine Translation, Statistical Machine Translation.

ABSTRACT

Research in machine translation has come a long way since the idea of using computers to automate the translation process. The main approach is statistical machine translation (SMT). An alternative to SMT Example-based machine translation (EBMT). Under machine translation systems, traditional transformation techniques are somewhat difficult to construct because they essentially involve hard coding the particularities of both languages. The processing of natural languages concerns the processing of natural language. The language spoken by people in everyday life is nothing but the natural language. There are many different NLP applications, among which machine translation is one of the applications. In this paper, we describe machine translation based on examples using natural language processing. The proposed EBMT framework can be used for automatic text translation by reusing previous translation examples. This framework has three phases: matching, alignment and recombination.

1. Introduction:

Machine translation, sometimes referred to as MT (not to be confused with computer-assisted translation, computer-assisted human translation or interactive translation), is a sub-domain of computer linguistics that examines the use of computer-assisted translation. a text translation software or speech language to another.

Automatic translations, which alone can not normally produce a good translation of a text, because the recognition of whole sentences and their closest equivalents in the target language is required. The solution of this problem to the statistical and neural techniques of the corpus is a field in full growth which leads to better translations, to the management of linguistic typology differences, to the translation of idioms and to the isolation of anomalies. Current machine translation software often allows customization by domain or profession (for example, weather reports), improving production by limiting the scope of authorized replacements.

This technique is particularly effective in areas where a formal language or formula is used. As a result, automatic translation of government and legal documents is more likely to produce useful results than a conversation or less standardized texts. Machine translation (MT) re- search has come a long way since the idea to use computer to automate the translation process and the major approach is Statistical Machine Translation (SMT). An alternative to SMT is Example-based machine translation (EBMT) [1]. However, new translation problems have also been introduced, we will discuss them below.

1.1. *Example-based MT*

"Man does not translate a simple sentence into an in-depth linguistic analysis, instead human beings correctly translate an input sentence into certain sets of fragments and ultimately translate those fragments into a long sentence. of each set of sentences is carried out according to the principle of translation by analogy, with appropriate examples for reference. "- Nagao (1984)

EBMT is based on the idea of reproducing translations by imitating examples of similar sentences [Nagao 84]. In this type of translation system, a large number of bilingual / multilingual translation examples have been stored in a text database, and the input expressions are rendered in the target language by extracting the example from the database closest to the input.

MT based on examples (Nagao, 1984) considers a bilingual body as a database and retrieves similar examples to an input sentence. Then, a translation is generated by modifying the target part of the examples by referring to the translation dictionaries [2]. Most example-based TM systems use sentences or sentences as a unit for the examples, so that they can be translated using case relations or idiomatic expressions. However, if some examples conflict during the extraction, the example-based MT selects the best example obtained by the similarity between the input part and the source part of the example. This means that the sample-based MT does not check whether the translation of the specified input sentence is correct or not.

On the other hand statistical statistics IBM models (Brown et al., 1993) translate an input sentence by combining the word transfer and rearrange words. Therefore, when applied to a language pair in which the word order is very different (for example, English and Japanese, Figure 1) it will be difficult to find an optimal global solution Solution because of the huge search space (Watanabe and Sumita, 2003).

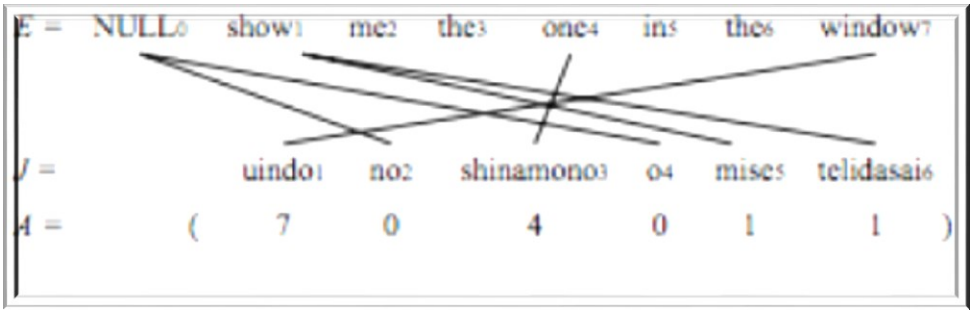

Figure 1.0 Example word alignment between English and Japanese.

1.2. *Classification Criteria*

In his seemingly preliminary conclusions on a EBMT definition, Somers (1999) discusses three increasingly specific criteria for the definition of the EBMT:
1. EBMT uses a bilingual corpus.
2. The EBMT relies mainly on a bilingual body
Based.
3. EBMT uses a bilingual corpus as main knowledge Base, at runtime.

Somers (1999) notes that the first two criteria also apply broadly, but argues that the third criterion may be too because it excludes, for example, statistical statistics, where all Probabilities based on corpora are calculated in advance.
While agreeing with Somers on the inadequacy of the first Two criteria that we would like to suggest that the third The criterion could also be too broad (not considered here) if it's too strict for the reasons at the same time ahead of Somers). In the following subsections, we become discuss the proposed criteria [3] gavagai Technology Incorporated.

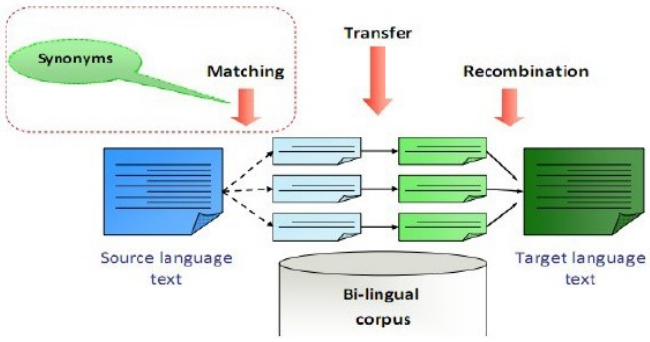

Figure 1.1 Example Based Machine Translation Work Flow [4].

1.3. *Statistical Machine Translation*

The most successful machine translation approach, called statistical machine translation. Statistical machine translation starts with a very large number of good translations. A body of texts (for example, United Nations documents) already translated into several languages that use these texts to automatically create a statistical translation model. This statistical model is then applied to new texts to guess a reasonable translation.

The first statistical machine translation ideas were introduced in 1949 by Warren Weaver, including ideas for the application of information theory by Claude Shannon. Statistical machine translation was reintroduced in the late 1980s and early 1990s by researchers at the IBM Thomas J. Watson research centre and contributed to a significant machine translation in recent years. years. Nowadays, it is the most studied machine translation method.

The idea of statistical machine translation comes from the theory of information. A document is translated according to the probability distribution display that a string T in the target language (for example English) is the translation of a string f (French)in the source language.

1.4. Why Statistical (or Least Empirical) Machine Translation?

We want to translate real-world documents. Thus, we should model real-world documents. A nice property: design the system once, and extend to new languages automatically by training on existing data.

Researchers in the field of machine translation (MT) largely agree that translation is a complex task that can not be solved by looking at words and their immediate neighbourhood. However, much of the existing work on MT is dependent on strong location assumptions for practical reasons. The first models of the modern statistical machine translation model (SMT) published in the early 1990s (Brown et al., 1990 and 1993) assume strong interdependence assumptions for the words of a sentence, and take into account only a very limited context, that of or two immediately preceding words in the target language for each word issued by the system.

Phrase- and syntax-based SMT, the two paradigms currently dominant, ease these assumptions of independence by taking into account a greater number of local dependencies, including in the source language. In a syntax-based MT, some far-reaching dependencies within the sentence can be taken into account.

However, even advanced MT systems still assume that texts can be translated sentence by sentence and sentences in a text are strictly independent of each other.

In SMT, it has just recently entered the focus of some research groups, and making use of discourse features to improve MT has turned out to be a daunting challenge.

We have used the latest NLP technology for empirical machine translation. Our Goal is two-fold. On the one hand we have investigated the problem of automatic MT evaluation. We have analyzed the major shortcomings of the current methodology and suggested some complementary ones Improvements. Our approach is based on the design of a heterogeneous set of automated metrics dedicated to capturing a variety of translation quality aspects at various linguistic levels the lexical, the syntactic, and the semantics.

We are also studying the possibility of Combine the scores provided by different metrics into a single quality standard. On the other hand, we have built an empirical MT system and analysed some of it Limitations. We have integrated language skills into the system to improve Overall quality of the translation.

In particular, we have dealt with the problem of lexical selection. We show that the use of linguistic information allows a better modelling of the translation context, effectively achieve improved translation quality. As a side issue, we also studied one of the main criticisms against empirical MT systems and empirical approaches for NLP in general their strong domain dependency. We show how its negative effects can be mitigated by appropriate measures Combining external knowledge sources when porting a system to a new domain.

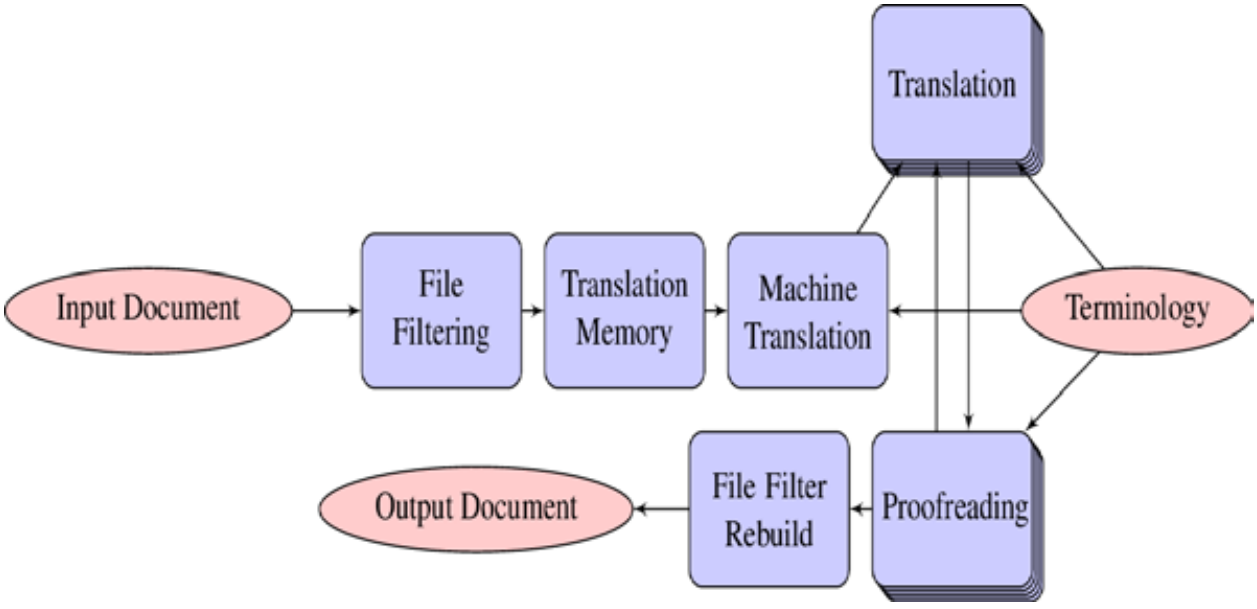

Fig. 2.1 Statically Machine Translation Equations.

2.0 Statically Machine Translation VS Example Based Machine Translation.

Statically Machine Translation	Example Based Machine Translation.
It uses different dictionary models and tables for best case result and provide more reliability.	It depends on bilingual corpus which are kind of data word library for translation.
SMT is better for User Generated Content and broad domain material such as patents.	EBMT is better for documentation and even software.
SMT is unpredictable but sentences are more fluid	EBMT is predictable: the sentences may not be pretty, but you know what you will get and will get the same result every time.
SMT has longer updating cycles (once or twice a year is typical)	EBMT is faster to update, maintain (can be done daily or more frequently).
Statically Machine Translation is open source and free to use and test its durability.	EBMT has expensive license to purchase its authentic version.
Statically Machine Translation is heavy on processing resource.	Example Based Machine Translation is heavy on linguistic.
Statically Machine Translation is more fluid.	While EBMT is less fluid.
SMT can handle over 50 languages out of the box (Google and Bing/Microsoft Translator).	EBMT can handle around 20 target languages out of the box.
SMT may need millions of bilingual and monolingual segments but engines may be pre-trained for a particular domain.	EBMT is ready to use off the shelf but needs customization on your domain and preferred terminology.

Table. 3.1 Statically Machine Translation VS Example Based Machine Translation.

3.1 Existence of Machine Translation

With the possible exception of the calculation of the artillery trajectory tables (Goldstine and Goldstine 1946). The machine translation has a strong claim to be the oldest established research discipline in computer science. Machine translation is a modern discipline, its success and very existence is based on the existence of modern computer hardware. Some basic ideas that ultimately serve as a basis for this new discipline has roots that precede the development of electronics digital computers[4].

This thesis examines the history of machine translation (MT) of its intellectual roots in the search for the universal language of the 17th century by its practical implementation in the late 20th and early 21st century. From the development of the first general purpose electronics. By the late 1940s, digital computers were investing heavily in the Internet Software development that can work fully automatically high quality automatic translation. We examine the main MT paradigms, including approaches based on transfer, interlingua and statistics.

The first systems for which there are detailed records have been developed independent in the early thirties in France and Russia. The French System developed by Georges Artsrouni (Corbé 1960, Hutchins 2004), seems to have been a general mechanical research and recovery System based on a strip of paper. The device was patented in 1933 and public Manifested at the Universal Exhibition of Paris in 1937, it was not a real machine translation system but could be configured as mechanical bilingualism Dictionary;[4] In this respect, Arstrouni's machine could be very Precursor of the subsequent paradigm of direct translation.

Different developers of machine translation have evolved over the years both a great pride and a great humility. Safe in the fifties MT researchers made unsubstantiated fully automated statements MT quality is easy to achieve in about five years. The failure of these extremely optimistic claims has contributed to this situation. Factor in ALPAC report and subsequent depriorisation of MT in the United States for a while. Other researchers, including Warren Weaver and Martin Kay advocated a more modest approach to machine translation modest and complementary skills, as well as the creation of user-centric interfaces where MT is one of the many tools for people he was built to serve[4].

4.0 Analysis of Machine Translation

Efforts to access documents in other languages have led to the development of a machine translation system integrating many heterogeneous functionalities and their implementations. Information professionals take advantage of machine translation to meet the needs of their users[5]. The methods of machine translation vary and each has its advantages and disadvantages. No translation tool can produce an exact version of the source language, but provides a main content of information that can be used to search for information contained in the source text. Sometimes, it is necessary to perform a post-processing via an internal linguistics after the translation output has been generated with the translation engine[5].

4.1 Analysis of Natural Language Processing

Natural Language Processing (NLP) Refers to the method of communicating with an intelligent systems using a natural language such as English. Processing of Natural Language is required when you want to hear from a dialogue based on a clinical expert system, etc[5].

The field of NLP plays the role of computers in the field of human use. The input and output of an NLP system can be -

Speech

Written Text

4.1.1 Steps in NLP

There are general five steps −

1) Lexical Analysis − It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.

2) Syntactic Analysis (Parsing) − It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as “The school goes to boy” is rejected by English syntactic analyzer.

3) Semantic Analysis - It draws the exact meaning or the meaning of the words. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyser disregards sentence as "hot ice-cream".

4) Discourse Integration - The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.

5) Pragmatic Analysis - During this, what was said is actually interpreted. It involves deriving those aspects of language which require real world knowledge.

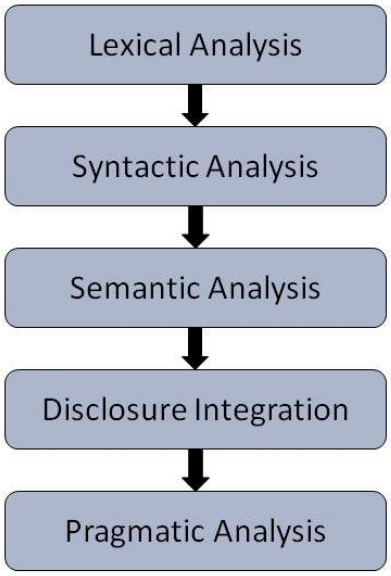

Fig. 3.1 Analysis steps of NLP.

5.1 Where Natural language processing is used?

No matter who you are and whatever you do, NLP can help you work more effectively. NLP is used in many areas, including business, sports, the arts, health, marketing, education and politics, wherever human activities are concerned[6].

NLP is widely used in business. He has quietly found his way into many courses on business, management, sales, presentations, planning and teambuilding (often without the name being mentioned).For HR professionals, NLP provides powerful tools for conflict resolution and management development.

NLP is widely used in sales and marketing. Some of them have given the discipline a bad name (but that shows how effective it is!) At NLP School, we insist on an ethical approach. We also want you to be aware of the tips that politicians, media and advertisers can play you! If you want to avoid pressing your buttons without your permission, you need to know how the process works.

NLP is an ideal tool to improve or change your career. In personal development NLP offers a wide range of applications, especially in combination with a coaching approach. You can use it to solve personal problems or help others to do the same. With NLP, people can examine and then change unnecessary thoughts, behaviours, and emotional patterns[6]. It has tools to control the effects of painful memories and mental images.

NLP can be of great help in setting health goals and achieving those goals. Many coach education programs are based on or use NLP as a basis. Our courses have attracted many trainers who want to deepen their professional understanding. The third book by our director, Robbie Steinhouse, is titled "How to train with NLP".

NLP is also used in sports to help athletes set the tone for excellence, mentally repeating their achievements and "going to the Zone".

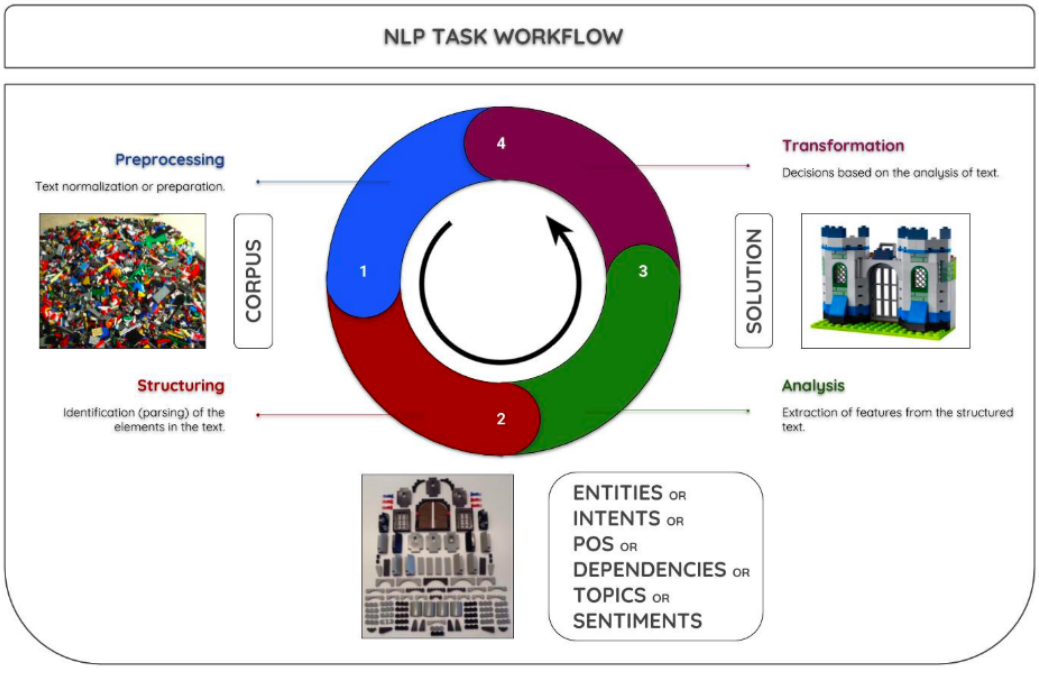

Fig. 4.1 NLP Task Flow.

6.1 Translators by Different Industries

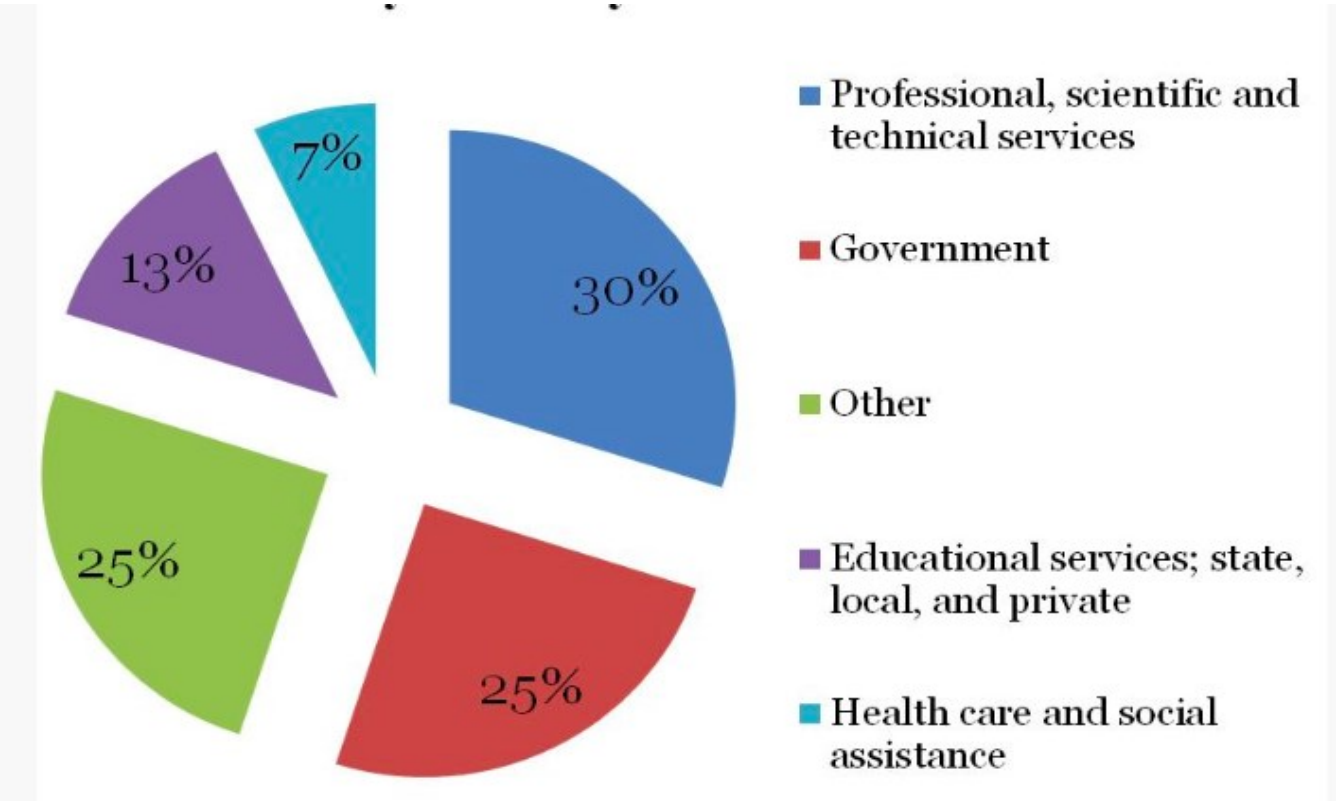

Fig. 5.1 Translator used y different Industries.

We live in a world where big cities have become a global community. People from several countries that live or visit the same geographical area are part of an interactive global community.

Sometimes communication or interaction becomes a major issue, so people can live their daily lives without problems. Translation services can help them simplify their lives.

Major industries that make products and services more accessible and useful to customers should use translation services to promote their products and services to their customers. Many large industries can benefit from translation services.

Conclusion:

Natural Language Processing is a field of research and application that explores how computers can be used to understand and manipulate natural language text or words to perform useful tasks. NLP researchers seek to understand how people understand and use language to develop appropriate tools and techniques that enable computer systems to understand and manipulate natural languages to perform the desired tasks.

An NLP application involves a number of areas of study, such as: For example, machine translation, natural language text processing and summarization, user interfaces, multilingual and multilingual information retrieval, voice recognition, artificial intelligence, and expert systems, etc.

An overview machine translation process reported in this paper show encouraging results. MT research has now reached a stage where The benefits can be enjoyed by people. A number of web search tools, Google, Lycos, Altavista and AOLoffer free MT facilities of web information resources. A number of companies also provide MT services commercially.

REFERENCES

1: Sunny Bhavan Sall I/C HOD (Department of Computer Engineering) page 1, *Example Based Machine Translation Using Natural Language Processing*, Journal of Scientific & Engineering Research.

2: Kenji Imamura, Hideo Okuma, Taro Watanabe, and Eiichiro Sumita, page 2, *Example-based Machine Translation Based on Syntactic Transfer with Statistical Models.*

3: Davide Turcato, Fred Popowich, *What is Example-Based Machine Translation ?,* gavagai Technology Incorporated.

4: Lane Schwartz, *The History and Promise of Machine Translation,* Department of Linguistics University of Illinois at Urbana-Champaign Urbana, USA.

5: https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.html.

6: https://www.nlpschool.com/what-is-nlp/where-is-nlp-used/

7: Figure 1.1 https://www.seminarsonly.com/computer science/Example-Based-Machine-Translation.php.

8: Figure 2.1 https://www.researchgate.net/figure/Typical-translation-workflow-in-SmartMATE_fig1_257225920.

9: Table. 3.1 Statically Machine Translation VS Example Based Machine Translation.

10: Fig. 3.1 Analysis steps of NLP.

11: Fig. 4.1 NLP Task Flow.

12: Fig. 5.1 Translator used y different Industries.