Understanding Terminologies of CAT Tools and Machine Translation Applications

Author's Details:
[(1)]**Zaki, Muhammad Zayyanu-**French Department,Faculty of Arts,Usmanu Danfodiyo University, Sokoto – Nigeria muhammad.zayyanu@udusok.edu.ng https://orcid.org/0000-0003-3785-2488
[(2)]**Nwanjoku, Chukwunonye Anthony-**Department of Foreign Languages and Literary Studies, Faculty of Arts, University of Nigeria, Nsukka anthony.nwanjoku.@unn.edu.ng https://orcid.org/0000-0002-0189-3010 [(3)]**John Ogboji-**Department of General Studies, Benue State Polytechnic, Ugbokolo
Obgojijoh2014@gmail.com

Abstract
The research tends to make a clear description of Machine Translation for users to be familiar with terminologies related to Machine Learning and precisely CAT tools (translation applications). The purpose of this paper is to broaden the scope of the use of translation technology and to explore its application in translation. There have been issues relating to Machine Translation. Since most translators are not familiar with translation aids, it is, therefore, necessary for translators to explore the translation applications. Translation can be achieved in different ways i.e. human and machine. Consequently, human translation lacks speed and accuracy hence, translation applications must come into play. One can find a variety of such applications online. The software assists human incapabilities and limitations in different ways especially for translation memory which assists the translator to be consistent and coherent in using terminologies. This paper applied analytical and descriptive approaches of translation. The study finally observed that each translation software has its unique identity and most of the sophisticated ones are not free. The researcher traced the processes or systems that support CAT tools such as machine language, machine learning, and finally features and metrics of evaluating Machine Translation. The researchers, therefore, concluded that translation without machine technology will definitely be difficult especially when translation involves huge tasks and different languages.
Keywords: *Translation, Machine Learning, Machine Translation, Analytical approach, CAT Tools*

Abbreviations/Terminologies

AI	-	Artificial Intelligence
ALPAC	-	Automatic Language Processing Advisory Committee
AMTE	-	Automatic Machine Translation Evaluation
ASCII	-	American Standard Code for Information Interchange
BLEU	-	Bilingual Evaluation Understudy
CAT	-	Computer Assisted Translation
DL	-	Deep Learning
DTP	-	Desktop Publishing
GT	-	Google Translate
HMTE	-	Human Machine Translation Evaluation
HMT	-	Hybrid Machine Translation
ML	-	Machine Learning
MT	-	Machine Translation
NER	-	Named Entity Recognition
NLG	-	Natural Language Generation
NLP	-	Natural Language Processing
NLTK	-	Natural Language Tool Kit
NLU	-	Natural Language Understanding
NN	-	Neural Networks
POS	-	Part of Speech
RBMT	-	Rule-Based Machine Translation
SMT	-	Statistical Machine Translation
TB	-	Termbase
TM	-	Translation Memory
TMA	-	Terminology Management Applications

TU - Translation Unit

1. Introduction - Translational Background

At the moment, there are approximately 8,143 languages across the globe. Among these languages, 40% are endangered, some of which have less than 1,000 speakers, Mauris, (135), Anderson, (2). Meanwhile, only 23 languages have more than half of the world's population. But the problem now is, how an exchange of information between all these different languages can be possible. The possibility seems to be translation. Translation has several definitions; generally, translation can be defined as the transfer of thoughts, notions, or ideas from one language to another, whether the language is in written, oral, or in a sign form. In other words, it is a duplicate of the original. Poibeau (15) maintains that "the translation of a text should be faithful to the original text: it should respect the main characteristics of the original text, the tone and style, the details of the ideas as well its overall structure". The demand for translation is increasing rapidly due to the economic, political and academic needs around the globe. Thanks to translation theorists and experts like James Holmes, Gideon Toury, Jean-Paul Vinay and Jean Darbelnet, Albert Nida, Danica Seleskovitch, Marianne Lederer, Jean Delisle, Mona Baker, Jeremy Munday, Anthony Pym among other translators experts that have helped us to have a clear notion of Translation Studies as discipline.

This paper explores the manner in which people embrace the use of technology in translation. This technology is what helps a translator to achieve his task easily. It involves the use of machine technology in translation, as technology is inevitable these days. This brings about Natural Language Processing which is a combination of computational linguistics and computer science. Natural Language Processing (NLP) is the ability of computer systems to comprehend natural language. A computer can analyze, understand, alter and regenerate its language called Machine Language. "Natural languages are what humans use to share information with each other" and "a natural language processing system often involves several stages of processing where natural language flows in one end and the processed output flows out the other end", Lane et al., (4). For Lane et al.,:

> Natural language processing is an area of research in computer science and artificial intelligence (AI) concerned with processing natural languages such as English […]. This processing generally involves translating natural language into data (numbers) that a computer can use to learn about the world. And this understanding of the world is sometimes used to generate natural language text that reflects understanding, (4).

Machine language is the only language a computer understands (not words and symbols) as in the case of humans. Therefore, human language, (speech/text in particular) serves as an input to be processed, converted and analyzed as data by the computer which the system recognizes.

Besides, Machine Translation, Speech Recognition and Sentimental Analysis are sub-systems used in the online process considering different situations. Neural Networks (NNs) on the other hand, act as neurons in the human system serve as links with other applications like Translation applications. According to Lane et al., (156), Neural Networks "are electrical signals that flow into the cell through the *dendrites* into the nucleus, an electric charge begins to build up. When the cell reaches a certain level of charge, it *fires*, sending an electrical signal out through the *axon"*. So, Natural Language Understanding (NLU) and Natural Language Generation (NLG) are two divisions of Natural Language Processing. Natural Language Generation studies the understanding and production of natural language by machines. However, NLU takes a lot of time and processes to understand the natural language of a human being while Natural Language Processing is the most challenging task by machines. So, NLP enables machines to analyze and understand humans' imperfect way of writing or speaking while NLG is a mechanical process that can generate natural language text and human speech from pre-defined data using it to translate data. This constitutes the nucleus of this paper. Therefore, technology has enormous potential for activities in translation like CAT tools in Google and Microsoft.

Moreover, there is a need for an understanding of the steps involved in the process. *Tokenization* is the first step in understanding language, this is a process of breaking strings into tokens which in turn become small structures or units that can be used in the process. The second is *stemming* which normalizes words into their base forms or root forms. The third is *Lemmatization* which takes into consideration the morphological analysis of a word so that the algorithm can link back to its original root word. It has *Lemma* that groups different infected forms of a word. Its output is a proper word. The fourth is *Part of Speech*

(POS) *Tags* that indicate how a word functions in meaning as well as in grammar in a sentence. A word can have more than one part of speech based on the context it is used. The fifth is *Named Entity Recognition* (NER) which is the process of detecting the name entities. Finally, *chunking* is picking up pieces of information and grouping them into bigger pieces to have meaningful information. In all, *Python* and *Natural Language Tool Kit* (NLTK) serve as a library that processes and execute the above-mentioned processes. Note that NLTK is used in natural language processing and text analysis.

Researches conducted by experts in the field of NLP and MT such as Khurana (n.d.) Sidirov et al., (2012), Shetty (2018) and Zong, et al., (2018) have made practical contributions to Translation Applications. For example, Khurana et al., (n.d.) highlighted that NLP has recently gained much attention for representing and analysing human language computationally. It has spread its applications in various fields such as machine translation, email spam detection, information extraction, summarization, medical, and question answering, etc. The study distinguishes four phases of NLP and components of NLG followed by a presentation of the history and evolution of NLP and the various applications of NLP, its current trends and challenges. Our interest here is human language analysis and its application in MT.

Also, Sidirov et al., (2012) introduced and discussed the concept of *syntactic n-grams* (sn-grams). *Sn-grams* differ from traditional n-grams in the manner of how we construct them, i.e., what elements are considered neighbours. In the case of *sngrams,* the neighbours are taken by following syntactic relations in syntactic trees, and not by taking words as they appear in a text, i.e., *sn-grams* are constructed by following paths in syntactic trees. In this manner, *sn-grams* allow bringing syntactic knowledge into machine learning methods; still, the previous parsing is necessary for their construction. *Sn-grams* can be applied in any NLP task where traditional *n-grams* are used. The researchers describe how *sn-grams* are applied to authorship attribution and used on baseline traditional *n-grams* of words. Using *Parts of Speech tags* and characters; three classifiers were applied: Support Vector Machine (*SVM), Naïve Bayes (NB), J48 algorithm*. *Sn-grams* give better results with the SVM classifier. The study discusses parsing using *sn-gram* which allows syntactic knowledge in Machine Language.

Shetty (2018) research on *Natural Language Processing(NLP) for Machine Learning* shows that NLP can help computers analyze text easily i.e detect spam emails, autocorrect, and show how NLP tasks are carried out for understanding human language. The researcher further explained NLP in real life, how to install NLTK, reading and exploring the dataset, processes involved in NLP, and the features of engineering. The idea of NLP is the understanding of human language and how computers analyse the text as a result of translation application interest us too.

Zong et al., (2018) research on NLP and its development process are almost the same as machine translation, and the two complements each other. The researchers compared the natural language processing of statistical corpora with neural machine translation and conclude that natural language processing: Neural machine translation has the advantage of deep learning, which is very suitable for dealing with the high dimension, label-free and big data of natural language. Therefore, its application is more general and reflects the power of big data and big data thinking.

2. The objective of the Study

The objective of this research paper is to explore the terminologies involved in the application of language in technology using Natural Language Processing, computational linguistics, Machine Translation, and Computer-Assisted Translation tools translation applications and their uses. It also informs and sensitizes learners and researchers in the field of language to diversify their methods and techniques by using technology.

3: Research Methodology

The study explores the sophistication in machine terminologies and their significance in translation applications. The researchers applied a descriptive approach and analytical theory of translation.

4: Machine Language

The language of a machine is the language computer understands. Machine language as defined by Schmit (15) is “the language understood by a computer but it is the only thing that the computer can work with”. It is very difficult to be understood by humans. All programs and programming languages eventually generate programs in machine language. Machine language is made up of instructions and data that are all binary numbers. It is normally displayed in hexadecimal form so that it is a little bit easier to read.

Therefore, Machine language or code is a low-level language that includes binary digits (0s and 1s). Computers are machines that recognize only binary data. CPU as the powerhouse processes data, programmes, videos, images, and text that are represented in the Machine Language which the computer accepts as input. The resulting output is sent to an operating system that will be displayed visually on the screen. For instance, the American Standard Code for Information Interchange (ASCII) value for the letter "A" is 01000001 in machine language but "A" is viewed on the screen as data received. A file of a photo may have millions of binary values that will determine the colour of each pixel. Machine code comprises 0s and 1s as different processor architectures use different machine codes. A compiler must compile high-level source code for the correct processor architecture for a program to run correctly. According to Deitel (2) today programmers rarely write programs in machine language; instead, they use clearer assembly languages or high-level languages. These languages are partly responsible for the current widespread use of computers. Programmers, burdened by machine language programming, began using English-like abbreviations for the various machine language instructions called mnemonics (memory aids).

5: An Overview of Natural Language Processing

Machine Learning is a broad field and a subset of Artificial Intelligence (henceforth, AI) that focuses mainly on the designing of systems thereby allowing them to learn and make predictions based on some experiences which are data in machines. As there is a lot of data today not only generated by people but also from phones, computers, and other devices. Data now is everywhere in the world which includes pictures, music, speech, videos and texts. This is maintained by Lane et al., (5) that "Google's index of natural language documents is at least 100 million gigabytes" which is incomplete. Currently, the actual size of natural language content online must exceed 100 billion gigabytes and almost 1000 times that of Google. And Machine Learning (ML) brings the idea of driving meaning from data. As Arthur C. Clarke clearly said, "any sufficient technology is indistinguishable from magic". ML is not magic but a set of tools and technology that can be used to answer questions with your data. Traditionally, humans have analyzed data and adapted systems to the changes in their patterns. However, as the volume of data surpasses the ability for humans to make sense out of it and manually write those rules, users will turn increasingly to automated systems. Users move to automated systems that can learn from the data and importantly, make changes to the data to adapt to the ever-changing landscape. It is observed that ML is in all the product devices we use today.

We try to make a clear distinction between the two terms, as Millstein (9) elucidated that on one hand, ML uses our computers to run predictive models, which are capable of learning from already existing data to forecast future outcomes, behaviours, and trends. On the other hand, DL is an important subfield of ML in which algorithms or models are inspired by how the human brain works. Both ML and DL are connected to AI. The AI concerns computer systems, which replicate human intelligence, while the wider field of ML allows the machine to learn completely by itself.

Some examples of ML include Tagging people and objects in pictures with a camera and also knowing the selected results to show in Google search engine. The ML is used to make human activities faster, safer, better and easier than ever before. It assists translators in achieving tasks they could not achieve on their own within a limited time. The researchers conclude by saying that ML is using data to answer questions. This is divided into two parts. The first part is training by using sufficient data and the second part is prediction or inference to answer questions. In fact, data is the key in ML and a life without ML will be difficult in today's world as every part of our lives involve ML.

The ML has features that use the data to detect patterns in a dataset and adjust programme actions accordingly and focus on the development of computer programmes that can teach themselves to grow and change when exposed to new data. It enables computers to find hidden insights using iterative algorithms without being explicitly programmed.

Steps involve in Machine Learning

Human beings learn from their past experiences while machine follows instructions given by humans. Humans now train the machines to do what machines can do much faster. This is where ML comes into the picture. It is far more than learning as it involves understanding and reasoning. ML model works with input given which then gives the output according to the algorithm applied. If it is right, the user takes the output as a final result and else provides feedback to the trimming model asking it to predict and let's learn. Model is generated through training and ML creates a model that provides accurate answers to the questions.

Then, in order to train a model, the machines need to collect data to train them. The steps involve in ML are: (1) collecting data, (2) data wrangling constitutes discovery, structure, clear and enrich data, (3) choosing a model, (4) training algorithm, (5) analyse data using patterns, 6) test algorithm using a checklist and (7) prediction that is the result.

Types of Machine Learning

ML is classified based on data processing and output.

Supervised ML

In this type, oversee a certain activity and make sure it is done correctly. The machine learns under guidance. The machine learns by feeding it with labeled data explicitly telling it this is the input and this is exactly how the output must look labeled. An example of supervised learning is a child that needs proper care and guidance.

Unsupervised ML

This means that the machine has to learn and work without supervision or anybody's direction, as it discovers heading patterns and trends in the data. The data here is not labeled and there is no guide. The machine has to figure out the hidden patterns itself and make predictions about the output. An example of unsupervised learning is an adult that does not need supervision and guidance. An individual can do things by himself.

Reinforcement ML

The machine needs encouragement or the establishment of certain patterns of behaviours. The machine here learns to adapt by applying hit and trial concepts through learning by experience. It discovers an agent that is put in an unknown environment where it explores the environment through a transition from one stage to the other for maximum results.

6: Concept of Machine Translation

Machine Translation also abbreviated MT, is derived from the intersection between NLP and ML; as a computer, it is trained to understand human language. For accurate MT translation in NLP, there are word embeddings. Word embedding may include thousands of information and the machine needs millions of words to truly learn the embeddings. It is a form of Computational Linguistics and Language Engineering which uses software to translate text or speech from one language to another. The two most common engines are Rule-Based Machine Translation (RBMT) and Statistical Machine Translation (SMT). These engines are different in the way that they process language content. They are often combined within the same system known as Hybrid Machine Translation (henceforth, HMT).

Besides, RBMT produces a more predictable output for terminology and grammar through the use of a customized terminology list to find through the engine. It uses linguistic rules to break down the content. It also has the ability to correct every error with the targeted rule. Rule-Based Engine (henceforth, RBE) does not need large instruction of text such as bilingual corpora to create the translational system. The SMT uses statistical models to generate the translation from the source content. It does not analyze text based on language rules; instead, it is built by analysis of bilingual corpus which required appropriate bilingual content to translate data.

Also, MT is circling since 2013; Google and Microsoft began to explore the possibilities to use Neural Networks (NNs). The NNs are statistical learning models that are used in speech and image technology. It is an application in MT that enables MT engines to train themselves using a process that is similar to how the human brain works through trial and error. This process is referred to as Deep Learning (DL). DL according to Michelucci (31) "is based on large and complex networks made up of a large number of simple computational units". It deals with gigantic databases. And MT comes from the principle which was established through the implementation of big data analytics. It improves the translation in most cases as it looks more fluid and more human. The level of language in Google Translate (GT) and Microsoft

Translator is already switching to Neural MT. The researchers observed multiple connections to many of the Machine Translation Systems such as Systran.

How would you know the right machine?

As each engine generates and processes data differently, the engine chosen for the project depends on Target Language (TL) and available reference materials for source files. In most cases, MT is used in combination with Translation Memory matches to target new segments through one of the engines. These segments include new or heavily modified content. The raw MT is a post edited by experienced linguists either accepted or modified before being inserted back into the translated documents. These combined methods produce the best result considering the best quality and time efficiency. Technical contexts such as user assistance content, customer support, user documentation are considered the most suitable type of content for MT. The suitability of content if the content is optimized by the MT.

So, the translator tries as much as possible to be a mediator with the help of NLP that breaks language barricades with automatic translation but still struggles to preserve meaning to handle figurative language. As reinforced by Douglas (29) that "the frustrating slowness of translation […] fuel dreams of machine translation: just as computers can do calculations in nanoseconds that it would take humans hours, days, weeks to do so […] ideal translation machine translate in minutes a text that took five people two weeks to write". Accuracy and time are some of the factors that contribute to the development of how the translator can achieve his task within a limited time frame using MT.

The MT works best with content that is repetitive and simple where the same content is reused and synonyms were minimized. You can expect significantly higher products activity by better-structured sentences that are shorter and less complex; by creating a glossary or terminology list that is clearly defined before the translation process is initiated. Furthermore, there is a content management system for quality translated content that is reused for future projects. Having this system in place, you can reduce translation costs and increase consistency across the project. These qualities are suitable for better raw MT output; that's in turn higher posted in productivity.

MT per average produces 8,000 translated words a day as compared to 2,500 by human translation. Benefits of MT for users include, among other things: increasing the overall efficiency, and productivity, reducing time-to-market, and translation costs, it is as well increases consistency.

An Approach to MT includes 3 stages: pilot project, integration, and engine tuning. Stage 1 is a *pilot project*: this processes workflow to create the MT engine, processing raw MT output, post-editing and results. Stage 2 is *integration*, which deals with segments not leveraged by translation machines to determine work productivity and future estimation for MT using the same combination as RBMT configure the engine. Once translated content is post edited. Stage 3 is *Engine turning*. This uses Fine-tuning the engine by adding newly created bilingual files in the corpus and retraining the engine, for Rule-based engines, updating terminology rules, setting rules and exceptions for feedback.

By using millions of translated texts from human translators, the MT generates automatic translations. Microsoft Translator and GT are among the most widely freely used site at https://www.translate.google.com. According to Douglas (38), MT is "an online Statistical Machine Translation (SMT) system whose reliability has improved to the point where some translators, in some language pairs, find it cost-effective to create a *first draft* with GT and then edit it into professional form". GT is also among the most powerful personal digital Assistant that translates files into different languages of the world. So,

> GT pipeline (or any similar machine translation system) relies on a deep tree of feature extractors, decision trees, and knowledge graphs connecting bits of knowledge about the world. Sometimes these feature extractors, decision trees, and knowledge graphs are explicitly programmed into the system, Lane et al., (27)

According to Poibeau (128), "Google's translation system also integrates terminologies and semantic resources when available and has recently begun to deploy a new generation of systems based on the deep learning approach". The DL is completely about creating more complex models and behaviour by adding additional processing layers to the conventional two-layer machine learning model architecture of feature extraction followed by modeling, Lane et al., (19). And also reinforced by Munday (289) that "the production of multiple TL versions (e.g. software localized for distribution worldwide in the local languages) modifies the 'simple' model of ST-TT transfer". The viewpoint is applied to "internationalization" which leads to the adaptation of accepted communication models. Human translators take too long to explore computer solutions, Munday (131) added that "software should fulfill the function of instructing the TT receiver in the same way as the ST does for the ST reader". Applying this will ease the task of the translator with MT in terms of speed, consistency, and accuracy.

As maintained by Douglas (33), "translators are usually, and understandably, hostile toward machine-translation systems, which promise clients enormous increases in speed at a fraction of the cost of human translation". Translators are familiar with the MT environment that helps them process the task of a translator within a limited time. This involves, therefore, the inclusion of ST and TT text into automatic procedure using natural language to software that serves as an intermediary between text inputs. At the end of the procedure, a translator is needed for some adjustments in terms of human intervention. In support of the above, Munday (291) insisted that "the rapid evolution of computer systems meant that it was possible to create an electronic 'corpus' (plural 'corpora') of naturally occurring texts (texts which had been written for a real communicative context and not artificially invented by the language researcher) that could then be processed and analyzed with translation software to investigate the use and patterns of the word-forms it contained". This certainly proves that translation applications work faster than human translators in terms of consistency and speed.

7: What are CAT Tools?

Computer-Assisted Translation (CAT) is software used by translators and linguists. It supports the translation process and allows a translator to edit, store and manage translation. Somers (6) made emphasis that "the translator would be provided with the software and other computer-based facilities to assist in the task of translation, which remained under the control of the human: CAT**".** CAT does not translate text and it is not a Machine Translation. What it does is that it helps you to do your translation with ease. The key components of CAT tools are Translation Memory (TM) which stores source sentences and their translations. It stores what the translator translates manually. Translations in the TM can be used for future use. It helps to identify a text being translated and provides its stored translation for reuse. Termbase (terminology) is similar to dictionaries as it stores single words or words and expressions. It is used for organizations or to produce specific terms and dictionaries for retrieving words, checking spellings and fragment matching in Machine Translation Engines (MTE) and Desktop Publishing (DTP). CAT tool enables you to open a lot of different files. DTP helps in extracting different files so that you can work on the translation that you do not have to do a lot of work.

With CAT tools, you can integrate and open files of PowerPoint, Word, Hypertext Multimedia Language (HTML), extract the texts so that you can work on the translation. A translator does not want to spend a lot of time adjusting text, realigning it. He wants to translate and spend much of his time possible translating. And a CAT tool allows you to do that as it takes all the files while you work on the text. The beauty of it is that it will put all the text files back in their original formats. A translator of course has to do some adjustments as languages are tricky. This is because some languages are longer than others.

Also, CAT tools are software that helps linguists or translators' speed, consistency and efficiency in translation. Some examples are OmegaT and Felix which are free and open access applications while others like WordFast, Déjàvu, MetaTexts, Lilt, MemoQ, SDL Trados are paid services translation applications. It is said to be a super-intelligent translation assistant which uses previous decisions of translations in memory on the current task. It is software on the translator's side that offers massive benefits for clients. The software forwards the source text in an organized way to reduce inefficiencies. It links virtually with other translation sources like MT and web translations.

On this note, it is important to differentiate between two similar machines: MT and CAT tools. MT is software that retrieves word terms, sentences, a segment from a really prepared dictionary and then puts it into a target language translation. The goal it has is to have a usable translation in the smallest time possible. However, the computer will not be able to deal with syntax grammar the way a linguist will do. According to Somers (79), a clear distinction is drawn that "MT tries to replace a translator to a certain extent, CAT tools support the translator by preventing repetitive work, automating terminology lookup activities, and re-using previously translated texts". And CAT tools have one specific goal and that is to facilitate the work of a translator. It does not replace a translator. CAT tools are not MT. CAT tools are a result of globalization and larger volumes of translator's work. It is essential for every professional translator.

7.1: Features of CAT Tools Application Software

Software is a complex process that involves many actors and occurs through successive or parallel steps. The features of translation application software include translation memories and terminology database. In this study, we will briefly explain the terms thus:

Translation Memories

TMs are language pair databases that store segments of text which have been previously translated so they may be recalled for later use. In this context, a segment can be a phrase, a sentence, a heading, or a paragraph but phrases and sentences are the useful segments that ease the translation process. There is a Translation Unit (henceforth, TU), which is the building block as TM stores the source text and translation on TU. Identical segments can be used from the TM database for future translation.

So, in a simpler form Somers (80) stated that "a TM system is no more than a database which stores translated sentences". The TM has some advantages in that it never translates the same sentence twice, it stores all previous translations and it helps the translator translates faster. TM can store more than two thousand words a day for future use. The CAT applies the TM to find any matches following the procedure, the translator used the suggestions to amend and adapt. They are context match, 100% match, fuzzy and fragment match.

Termbase

Termbase (TB) is a database. TB is a key to your consistent use of terminologies, and they are created and used from different accounts. The terminology works with Terminology Management Applications (TMA) which is linked with automatic terminology lookup. Somers (81) elucidated that "Automatic terminology lookup means that terms in the source text, which are found in the dictionary or terminology database, are automatically displayed with their translations".

7.2: Performance Metrics for Measuring Qualities in Evaluating MT

As stated by Poibeau (112) that "it is clearly difficult to evaluate the quality of a translation since any evaluation involves some degree of subjectivity and strongly depends on the needs and point of view of the user". The following metrics help to evaluate the suitability of MT to your content:

1. F-Measure deals with the overall quality performance of an engine,
2. Bilingual Evaluation Understudy (BLEU) Scores evaluates the fluency of the translation correlation with human evaluation,
3. Terminology (Ter) Scores estimates the amount of post-editing effort required, and
4. Post-editing efficiency measures the amount of engine efficiency of post-editing work.

According to Lopez (14), the first campaign for the Automatic Language Processing Advisory Committee (ALPAC) already revealed the wide gap between human and machine translators. MTs are still evaluated as if they were human translations. In fact, two of the most important MT quality items- target language fluency and fidelity to the original- are the outcomes of human capacities: language proficiency and understanding. The Human MT Evaluation (HMTE) especially the criteria for scoring these items, are very subjective and vary among evaluators. The procedures to obtain as much reliable and objective data as possible have increased the costs of HMTE in time and money. As an alternative, Automatic MT evaluation (AMTE) performs methods that have lower costs and establish metrics whose values are not subjective perceptions.

There is an assumption that is stated in terms of MT, it is as better as it is similar to the human translation of the same original. The cost has been reduced, development cycles of MT systems have been accelerated but the reliability of these assumptions has been critically analysed.

8: Conclusion

With the advancement in technology, MT revolves with significant improvements in availability, increased ability to support the traditional translation process, new technology – NMT, managing the quality with controlled English, the three-step process to determine the suitability and leveraging with Translation

Memories. Technology (Internet) in this regard, has become a central part of modern society. Every day, billions of users across the globe access machines on the web for different reasons. This provides abundant data that make translation applications like CAT tools use multilingual databases in the future.

In natural language processing, ambiguity is the most inescapable issue and applies to most types of words, which makes ambiguity a much superior issue than primarily thought. NLP is an interdisciplinary field *par excellence.* It draws its sources from a wide range of related disciplines such as computer science, Artificial Intelligence, linguistics, cognitive psychology and neuroscience. CAT tools make it possible, without mastering many languages, to communicate and translate text in different languages. In the field of telecommunications and information technology, this technology is supported by several commercial and free translation applications.

Works Cited/Consulted

i. *Catford, J. C., A Linguistic Theory of Translation – An Essay in Applied Linguistics. London: Oxford University Press, 1965.*

ii. *Czulo O., et al., Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation: TC3 II. (Translation and Multilingual Natural Language Processing 4). Berlin: Language Science Press. 2017.*

iii. *Deitel, H. M., et al., 'Programming Languages' in An Introduction to Information Processing (1986) 2*

iv. *Dershowitz, N., et al, Language, Culture, Computation – Computational Linguistics and Linguistics. Festschrift: Springer. 2014.*

v. *Douglas, R., Becoming a Translator: An Introduction to the Theory and Practice of Translation. Third Edition. London: Routledge, 2012.*

vi. *Hassan, A. et al., "Exploring use of Google speech recognition API, assisting teaching-learning of French language with special reference to UDUS". International Journal of Multidisciplinary Research and Growth Evaluation (2021) Vol. 2 (4) pages 146-151*

vii. *Khurana et al., 'Natural Language Processing: State of The Art, Current Trends and Challenges'.(n.d.)*

viii. *Kurdi, M. Z. Natural Language Processing and Computational Linguistics 2 - Semantics, Discourse and Applications. London: John Wiley Inc., 2017.*

ix. *Lane, H., et al., Natural Language Processing in Action - Understanding, Analyzing, and Generating text with Python. Shelter Island: Manning Publications Co. 2019.*

x. *Lopez, J. M., "Machine Translationness: a Concept for Machine Translation Evaluation andDetection". Diss. Universitat Oberta de Catalunya, April 2015.*

xi. *Michelucci, U., Applied Deep Learning A Case-Based Approach to Understanding Deep Neural Networks. Dübendorf: A press, 2018.*

xii. *Millstein, F., Deep Learning with Keras - Beginner's Guide to Deep Learning. Amazon: Frank Millstein Publishers, 2018.*

xiii. *Moorkens, J. et al., Translation Quality Assessment from Principles to Practice. Machine Translation: Technologies and Applications. Cham: Springer International Publishing, 2018.*

xiv. *Munday, J., Introducing Translation Studies: Theories and Applications. London: Fourth Edition. London: Routledge - Taylor and Francis Group, 2016.*

xv. *Poibeau, T., Machine Translation. Cambridge, Massachusetts: The MIT Press, 2017.*

xvi. *Schmit, M. L., 'Machine Language' in Pentium™ Processor. Science Direct. (1995) 1.*

xvii. *Shetty, B., 'Natural Language Processing (NLP) for Machine Learning'. Towards Data Science. (2018) Online Tutor.*

xviii. *Sidirov, G. et al., 'Syntactic N-grams as Machine Learning Features for Natural Language Processing'. (2012) https://www.cic.ipn.mx/~sidorov/*

xix. *Somers, H., Computers and Translation: A Translator's Guide. Amsterdam: Benjamin Translation Library, 2003.*

xx. *Yakasai, S. A., Jagoran Ilmin Walwalar Harshe. Kaduna: Amal Printing Press, 2020.*

xxi. *_______, Sanin Makamar Fassara. Kaduna: Amal Printing Press, 2019.*

xxii. *Zaki, M.Z, et al., ''Appreciating Online Software-based Machine Translation: Google Translator''. International Journal of Multidisciplinary Academic Research. (2021) Vol. 2 (2) pages 1-7*

xxiii. *Zong et al., 'On Application of Natural Language Processing in Machine Translation' Research Gate (2018) 1 – 6*