

# Use of Machine Translation in translating new scientific Terms

Shahrazed NEKKAA

Translation Institute University Oran 1, Ahmed Ben Bella, Algeria

## **Introduction:**

Talking about translation of scientific texts, with all their specificities, far from technology, is inconceivable. From the beginning, researchers concentrated almost exclusively on the translation of scientific and technical documents, where the difficulties of cultural differences and variable contexts are less acute than in the more ‘culture-bound’ translation of literature, legal texts, and many areas of sociology. In science and technology, the demand for translation has almost always exceeded the capacity of the translation profession, and these demands are growing rapidly. In addition, the coming of the Internet has created a demand for immediate online translations, which human translators cannot possibly meet. Basically there are two types of demand. There is the traditional need for translations of ‘publishable’ quality, particularly the production of multilingual documentation for large corporations. Here the output of MT systems can save time and costs by providing draft translations which are then edited for publication – this mode of use being called human-aided machine translation (HAMT). However, what is often needed is not a ‘perfectly’ accurate rendition but something that can be produced quickly (sometimes immediately) conveying the essence of the original text, however grammatically imperfect, lexically awkward and stylistically crude. The latter is often called ‘machine translation for assimilation’ in contrast to the production of publishable-quality translations, called ‘machine translation for dissemination’. More recently a third application has been identified where MT is used in social interchange (electronic mail, chatrooms, etc.) where again high-quality is not essential: ‘machine translation for communication’.

Over the years, translators were becoming familiar with the benefits of computers for their work – word processing, creation of individual glossaries, facilities for on-line access and transmission of documents. They were not, however, satisfied with the quality of output of MT systems as such. It was clear already that translators wanted to have computer aids where they are in control of processes, and not to be ‘slaves’ of automatic systems. Many tools were developed, notably for concordancing, dictionary creation, terminology management, and document transmission. In the early 1990s, however, came the most significant development, the marketing of integrated tools in the ‘translator’s workstation’ (or ‘workbench’). Translation workstations combine multilingual word processing, OCR facilities, terminology management software, facilities for concordancing, and in particular ‘translation memories’. The latter facility enables translators to store original texts and their translated versions side by side, i.e. so that corresponding sentences of the source and target are aligned.<sup>1</sup>

---

<sup>1</sup> W. John Hutchins, Machine translation: a concise history, [Website: <http://ourworld.compuserve.com/homepages/WJHutchins>]

The translator can thus search for phrases or even full sentences in one language in the translation memory and have displayed corresponding phrases in the other language, either exact matches or approximations. In addition, translation workstations often provide full MT programs (for translating segments, paragraphs or whole texts), to be used or adapted by translators as appropriate. <sup>2</sup>

There are now many vendors of workstations. The earliest were Trados (Translator's Workbench), STAR AG (Transit), Atril (Déjà Vu), SDL (the SDLX system), Xerox (XMS), Terminotix (LogiTerm), MultiCorpora (MultiTrans), Champollion (WordFast), MetaTaxis, and ProMemoria. And many other workstations. These latter, has revolutionized the use of computers by translators; especially for translating sciences and technical texts. Trying to enable them to deal constantly with the huge flow of terminology. In this paper we will try to know at what extent this aim is achieved.

## **Machine Translation :**

Machine Translation (MT) is the process of translation from one language to another by the use of computing devices. MT is an automatic process in which all the translation jobs are done with the help of programming languages and software. All materials need to be translated, this includes commercial, business and scientific documents, instruction manuals, text books, World Wide Webs.

Machine Translators are serving for multiple languages. Quality of MT also varies with language pairs. Two languages for instance English to French may have high quality of translation, some other pair may have translation of low quality. Currently none of the language pairs are translated accurately. The main difficulty in automated translation of one language to another is varied written scripts and multiple lexical choices for a single idea. A single word may have various meanings; in different situations it is used for different purposes. At the current level of research, a single level of representation of every language is almost impossible. Every language needs to be considered separately with its scripting choices for automatic translation <sup>3</sup>

The dominant framework of MT research until the end of the 1980s was based on essentially linguistic rules of various kinds: rules for syntactic analysis, lexical transfer, syntactic generation, morphology, lexical rules, etc. However, since 1989 the dominance of the rule-based approach has been broken by the emergence of new methods and strategies which are now loosely called 'corpusbased' methods.

1. **Corpus-based approaches:** The most dramatic development has been the revival of the statistics-based approaches – seen as a return to the 'empiricism' of the first decade, and a challenge to the previously dominant rulebased 'rationalism' of the 1970s and 1980s. With the success of stochastic techniques in speech

---

<sup>2</sup> W. John Hutchins, Machine translation: a concise history, [Website: <http://ourworld.compuserve.com/homepages/WJHutchins>]

<sup>3</sup> Corpus Based Machine Translation for Scientific Text, Chapter- May 2018, DOI: 10.1007/978-3-319-92007-8\_17

recognition, a group at IBM (Yorktown Heights, New York) began to look again at their application to MT. The distinctive feature of their Candide system was that statistical methods were used as virtually the sole means of analysis and generation; no linguistic rules were applied. The IBM research was based on the corpus of French and English texts contained in the reports of Canadian parliamentary debates (the Canadian Hansard). The essence of the method was first to align phrases, word groups and individual words of the parallel texts, and then to calculate the probabilities that any one word in a sentence of one language corresponds to a word or words in the translated sentence with which it is aligned in the other language (a 'translation model'). The outputs were then checked and rearranged according to word-to-word transition frequencies in the target language, derived from the corpus of bilingual texts (a 'language model'). What surprised most researchers brought up in linguistics-based methods was that the results were so acceptable: almost half the phrases translated either matched exactly the translations in the corpus, or expressed the same sense in slightly different words, or offered other equally legitimate translations.

Since this time statistical machine translation (SMT) has become the major focus of many research groups, based primarily on the IBM model, but with many subsequent refinements. The original emphasis on word correlations between source and target languages has been replaced by correlations between 'phrases' (i.e. sequences of words, not necessarily 'traditional' noun phrases, verb phrases or prepositional phrases), by the inclusion of morphological and syntactic information, and by the use of dictionary and thesaurus resources. There has been a vast increase in the sizes of aligned bilingual databases and of monolingual corpora used in 'language models'; and the SMT approach is being applied to an ever widening range of language pairs. The main centers for SMT research are the universities of Aachen and Southern California, and they have been recently joined by the Google Corporation. The second major 'corpus-based' approach – benefiting likewise from improved rapid access to large databanks of text corpora – was what is known as the 'example-based' (or 'memory-based') approach. Although first proposed in 1981 by Makoto Nagao, it was only towards the end of the 1980s that experiments began, initially in some Japanese groups and during the DLT project. The underlying hypothesis of example-based machine translation (EBMT) is that translation often involves the finding or recalling of analogous examples, i.e. how a particular expression or some similar phrase has been translated before. The approach is founded on processes of extracting and selecting equivalent phrases or word groups from a databank of parallel bilingual texts, which have been aligned either by statistical methods (similar perhaps to those used in SMT) or by more traditional rule-based methods. For calculating matches, some groups use semantic methods, e.g. a semantic network or a hierarchy (thesaurus) of domain terms, other groups use statistical information about lexical frequencies in the target language. A major problem is the re-combination of selected target language examples (generally short phrases) in order to produce fluent and grammatical output. Nevertheless, the main advantage of the approach (in comparison with rule-based approaches) is that since the texts have been extracted from

databanks of actual translations produced by professional translators there is an assurance that the results will be idiomatic. Unlike SMT, there is little agreement on what might be a ‘typical’ EBMT model, and most research is devoted to example-based methods applicable in any MT system. Although SMT is now the dominant framework for MT research, it is recognized that the two corpus-based approaches are converging in many respects in so far as SMT is making more use of phrase-based alignments and of linguistic data, and EBMT is making wider use of statistical techniques. As a consequence, it is becoming more difficult to isolate distinctive features of the two models. Since these corpus-based approaches to MT research rely to a great extent on the availability of bilingual and multilingual text corpora (and indeed also monolingual corpora), there has been a major focus in the last decade or more on the collection and evaluation of text databases. And since statistical corpus-based approaches have also become dominant in the wider field of computational linguistics and natural language processing, linguistic resources are now central to both MT and NLP, with the collateral consequence that the MT is now returning to the ‘mainstream’ of computational linguistics – a position which it lost after ALPAC report (see section 4 above), and which is reflected in the coverage of the already-mentioned Linguistic Data Consortium, the European Language Resources Association, the biennial Coling conferences, and the biennial Language Resources and Evaluation Conferences (LREC).

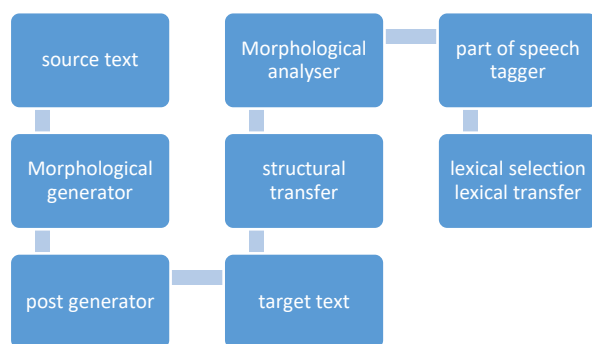
- 2. Rule-based approaches:** Although the main innovation since 1990 has been the growth of corpus-based approaches, rulebased research continued in both transfer and interlingua systems. A number of researchers involved in Eurotra worked within the theoretical approach developed, e.g. the CAT system at Saarbrücken; and one of the fruits of Eurotra research was the PaTrans transfer-based system developed in Denmark for Danish/English translation of patents. Another example of the linguistics-based transfer approach has been the LMT project, which had begun under Michael McCord in the mid-1980s, at a number of IBM research centers in Germany, Spain, Israel and the USA. Implemented in Prolog, LMT (‘Logic-programming Machine Translation’) has the traditional four stages: lexical analysis; syntactic analysis of source texts, producing representations of both surface and deep (logical) relations; transfer, involving both isomorphic structural transfer and restructuring transformations; and morphological generation of target texts. The interlingua approach continued at Carnegie-Mellon University (CMU). In 1992, it began a collaboration with the Caterpillar company on a large-scale knowledge-based and controlled-language system CATALYST for multilingual translation of technical manuals.<sup>4</sup>

Towards the end of the decade, the knowledge-based approach at CMU was combined with developments in statistical analysis of text corpora for the rapid prototyping and implementation of special-purpose systems (DIPLOMAT), e.g. for translation of Serbo-Croatian in military operations.

---

<sup>4</sup> Muhammad Irfan, Machine Translation, Department of Computer Science, Bahria University Islamabad.

In the mid 1990s other interlingua-based systems were started, e.g., the ULTRA system at the New Mexico State University developed by Sergei Nirenburg, the UNITRAN system based on the linguistic theory of Principles and Parameters, developed at the University of Maryland by Bonnie J. Dorr (1993), and the Pangloss project, a collaborative project involving the universities of Southern California, New Mexico State and Carnegie-Mellon. Pangloss itself was one of three MT projects supported by ARPA (Advanced Research Projects Agency), the others being the IBM statistics-based project mentioned above, and a system developed by Dragon Systems. The restitution of US government support for MT research signalled the end of the damaging impact of the ALPAC report (section 5 above). Finally, at the end of the 1990s, the Institute of Advanced Studies of the United Nations University (Tokyo) began its multinational interlingua based MT project – based on a ‘standardised’ intermediary language, UNL (Universal Networking Language), for initially the six official languages of the United Nations and six other widely spoken languages (Arabic, Chinese, English, French, German, Hindi, Indonesian, Italian, Japanese, Portuguese, Russian, and Spanish) – involving groups in some 15 countries.



### Rule-based Machine Translation<sup>5</sup>

3. **Speech translation** One of the most significant developments since the late 1980s has been the growing interest in spoken language translation, presenting the formidable challenges of combining speech recognition and synthesis, interpretation of conversations and dialogues, semantic analysis, and sensitivity to social contexts and situations. British Telecom did some experiments in the late 1980s on a spoken language phrasebook type system. However, the first long-standing group was established in 1986 at ATR Interpreting Telecommunications Research Laboratories (based at Nara, near Osaka, Japan). ATR has been developing a system for telephone registrations at international conferences and for telephone booking of hotel accommodation. Slightly later came the JANUS project, under Alex Waibel at Carnegie-Mellon University, and subsequently collaborating with the University of Karlsruhe (Germany) and with ATR in a consortium

<sup>5</sup> Muhammad Irfan, Machine Translation, Department of Computer Science, Bahria University Islamabad.

C-STAR (Consortium for Speech Translation Advanced Research). The JANUS project has also focused on travel planning, but the system is designed to be readily expandable to other domains. Both the ATR and the C-STAR projects are still continuing. A third shorter-lived group was set up by SRI (Cambridge, UK) as part of its Core Language project (Alshawi 1992), and investigated Swedish-English translation via ‘quasi-logical forms’. On a much larger scale has been fourth spoken language project, Verbmobil, directed by Wolfgang Wahlster and funded from 1993 until 2000 by the German government at a number of universities. The aim of Verbmobil was the development of a transportable aid for face to face English-language commercial negotiations by Germans and Japanese who do not know English fluently. As with Eurotra, although the basic goal was not achieved the development of efficient methodologies for dialogue and speech translation and the establishment of top-class research groups in Germany in the field were considered notable successes.<sup>28</sup> More recently, spoken language projects have been set up for doctor-patient communication between English, French and Japanese (MedSLT) – based on the above-mentioned SRI research in Cambridge and Sweden – and for commercial exchange and tourism between English, French, German and Italian (NESPOLE!), also involving Catalan and Spanish (in its FAME ancillary project). Simpler in conception is the ‘voice translator’ developed for the US military (Phraselator), a kind of phrasebook with spoken output in a number of languages, Hindi, Thai, Indonesian, Pashto, Arabic, etc. Loosely related to this interest in spoken language translation are a number of systems developed for the translation of television captions (or subtitles) – from English into Spanish and Portuguese, from English into Korean, English into French and Greek, etc. The particular constraints of the task include not just dealing with spoken language (transcribed and then reduced) but also the space limitations on screens and the fact that comprehension of written texts can be slower than that of speech.

#### 9.4. Hybrid systems

The expansion of methodologies in the past decade and the introduction of new applications for automated translation processes have highlighted the limitations of adopting one single approach to the problems of translation. In the past, many MT projects were begun by researchers who saw MT as a testbed for a particular theory or particular method, with results that were either inconclusive or of limited application. It is now widely recognized that there can be no single method for achieving good-quality automatic translation, and that future models will be ‘hybrids’, combining the best of rule-based, statistics based and example-based methods. One approach is the idea of running parallel MT systems and combining the outputs: the so-called ‘multi-engine’ system – the group at Carnegie-Mellon University has investigated combinations of knowledge-based and example-based systems. More commonly, hybrids are currently envisaged as systems combining the statistical methods of SMT or EBMT with some linguistics-based methods (from rule-based approaches), particularly for morphological and syntactic analysis. An example is the research at Microsoft (Dolan et al. 2002). However, there are other ways to combine corpus-based and rule-based methods, as the research at the National Tsing-Hua University (Taiwan) (Chang & Su 1997) illustrates: lexical and syntactic rules can be statistically derived from corpus data and optimized by feedback from output.

#### 9.5. Evaluation

MT evaluation has become a major and vigorous

area of research activity. In the 1990s there were numerous workshops dedicated specifically to the problems of evaluating MT, e.g. Falkedal 1994, Vasconcellos 1994, and the workshops attached to many MT conferences. The methodologies developed by Japan Electronic Industry Development Association (JEIDA 1992) and those designed for the evaluation of ARPA (later DARPA) supported projects were particularly influential (ARPA 1994), and MT evaluation proved to have significant implications for evaluation in other areas of computational linguistics and other applications of natural language processing.<sup>29</sup> Initially, most measures of MT quality were performed by human assessments of such factors as comprehensibility, intelligibility, fluency, accuracy and appropriateness.<sup>30</sup> But such methods of evaluation are expensive in time and effort and so efforts have been made, particularly since 2000, to develop automatic (or semi-automatic) methods. One important consequence of the development of the statistics-based MT models (SMT, section 9.1 above) has in fact been the application of statistical analysis to the automatic evaluation of MT systems. The first metric was BLEU from the IBM group, followed later by the NIST (National Institute for Standards and Techniques)<sup>31</sup>. Both measures are based on the availability of human produced translations (called ‘reference texts’). The output from an MT system is compared with one of more ‘reference texts’; MT texts which are identical or very close to the ‘reference’ in terms of word sequences score highly, MT texts which differ greatly either in individual word occurrences or in word sequences score poorly. The metrics tend to rank rule-based systems lower than SMT systems even though the former are often more acceptable to human readers. Consequently, there is no denying their value for monitoring whether a particular system (SMT or EBMT) has or has not improved over time, but there is much doubt about the general value of these metrics for comparative MT evaluations, and search for more suitable and sensitive metrics has intensified.<sup>6</sup>

## **Terminology in Scientific Texts**

Introduction Scientific translation is an important channel of knowledge dissemination; i.e. it is a means of seeking knowledge to fulfill scientific research needs. Translation of English scientific texts into Arabic is extremely scarce and is not keeping pace with global knowledge explosion. Scientific translation is an important step to acquire new technologies. The system of terminology in scientific language is not closed and constant in the age of globalization. It is in constant development as the new scientific disciplines emerge and develop.

Terminology is one of the problems that translators face in translating scientific English texts into Arabic. A good scientific translator should guarantee an accurate rendering of terminology. Actually, Arabic suffers a serious

---

<sup>6</sup> Terminology and Machine Translation, Article: January 2001

shortage of scientific terminology. Language purists argue that translators should find original Arabic terminology which is better than loanwords. Selection of an equivalent scientific term in the target language also complies with the requirement of precise transmitting of thoughts and ideas. Therefore, terminology must be carefully chosen in order to be unequivocal. First, translators should have scientific knowledge. The challenges for the scientific translator are to be able to research subjects and to have scientific knowledge in a particular field. Moreover, translation of science poses a huge linguistic challenge. The key feature of scientific texts is terminology which is considered a key principle in scientific knowledge. Terminology and science have been interwoven for a long period of time. The relation between terminology and science is even more strongly understood by seeing an intrinsic dependency of the development of language and the development of science.

Newmark (1988; 1991) suggests that the issue is a conflict of loyalty to the source language or the target language. His semantic translation method attempts to be close to the phonetic, morphological and lexical structure of the source. It is similar to Eugene Nida's formal equivalence. His communicative translation method, also, attempts to produce a similar effect on the target language. It resembles Nida's dynamic equivalence. Thus, the main questions that we could ask here are: How often are transliteration, calque, gloss translation and communication translation used in the translation of scientific terms? And what are the factors which play role in the choice of these strategies?

There are many different translation strategies, most of which are based on equivalence. The concept of equivalence is a central and controversial issue in translation. It has been studied by various theorists (cf. Jakobson 1959, Catford 1965, House 1977, Nida and Taber 1982, Newmark 1988, Vinay and Darbelnet 1995 and Baker 1992). Since the introduction of semantic and communicative translation theories by Newmark (1988; 1991), there have been various studies on them. Some studies compare them; others analyze them in some specific literature texts or in the translation of news reports. Scientific translation has been discussed by many scholars. Gerzymisch-Arbogast (1993) explains translation problems in technical and scientific translation. Gommlich (1993) utilizes text typology in scientific and technical translation. Franco Aixela (2004) surveys the history of technical and scientific translation. Few studies deal with how scientific terms are translated into Arabic. El-Shami (2010) discusses the translation of prefixes and suffixes in the scientific terms. El-Khoury (2010) examines the translation of medical terms into Arabic.

Data Collection and Methodology Data in the study are scientific terms collected from the translated Arabic version of the famous Scientific American (Arabic Edition) and Nature (Arabic Edition). I depended on a descriptive analytic method: I have gone through the two magazines in the period from 1993 to 2016. Three hundred scientific terms have been chosen from each magazine.

In these examples scientific terms from miscellaneous fields, including astronomy, physics, computer studies, medicine ... etc. we restrict choices in translating these terms; trying to identify their exact meanings using Oxford Dictionary of Science to find out meanings.

Alzheimer الزهايمر 2. Microwave ميكروويف 3. SARS سارس Transliteration follows the phonetic rules of the target language. Transliterated words are often naturalized to assimilate the structure of the target language. Translators introduce some phonetic and morphological changes to the foreign term (Ghazzala 1995). Naturalization of loanwords refers to the addition of some affixes to the foreign words without changing their roots. The affixes are added to suit the nature of the Arabic language: تكنولوجيا → Technology تكنولوجي → Technological → تكنولوجيا → Technologically ليزرة → Lasering Ghazzala (1995) is in favor of using pure Arabic terms. For example, the word "تقنية" can be used to render the word "technology". According to Baker (1987) transcription and naturalization are rejected because they threaten the position of Arabic. In fact, these strategies do not explain the meaning of the words in Arabic. Notice, for example, how the following two words (ballistic and cruise) are transliterated: صواريخ باليستية missile Ballistic صواريخ كروز missile Cruise The Arabic reader may not know the meaning of "باليستية" and "كروز". Another example is the translation "الأمن السيبراني" for the English phrase "cyber security". Using transliteration does not enrich Arabic vocabulary. In fact, it implies that Arab translators are not able to create new pure Arabic terms and prefer the easiest way to translate scientific terms.

As to Semantic Translation at the Morphological Level (Calque): Calque or loan translation is a kind of semantic translation. This strategy renders a phrase borrowed from another language, keeping the source language structure. Anaerobic (لا هوائي) Scientific American 2015, Vol.7-8, p.74 Gasohol (بنزحول) Scientific American 1994, Vol.12, p.58) Some calque words are not natural. Consider the following words: Spacetime (زمكان) Nature 2016, Vol.43, p.12) Electromagnetic (كهروطيسية) Scientific American 2000, Vol.14, no.3

For Communicative Translation of Scientific Texts On the contrary, communicative translation aims at influencing the reader. It is reader-centered and effect-oriented. Newmark proposes the principle of "equivalent effect". According to Newmark (1988), it focuses on making the target language readers understand the source language author's thoughts. It is used in genres which are communicative in nature such as news report, textbooks, and public announcement. The differences between communicative and semantic translation is based on their different emphasis. In semantic translation, the focus is on the original's formal properties. However, communicative translation does not adhere to the source language text. Then, it attempts to eliminate any exoticism and to look natural; smooth translation. While communicative translation attempts to produce on its readers an effect close to that of the original, semantic translation aims to render the semantic and syntactic structures of the original (Newmark 1991: 11). Semantic translation [is] usually more awkward, more detailed, more concentrated, but briefer.... Communicative translation [is] easy reading, more natural, smoother, simpler,

clearer, more direct, [and] more conventional... (1991: 11). As pointed out by Newmark, Communicative and semantic translation may well coincide.... There is no one communicative or one semantic method of translating a text—these are in fact widely overlapping bands of methods. A translation can be more, or less, semantic—more, or less, communicative—even a particular section or sentence can be treated more communicatively or less semantically (1991: 10) 7

It has been used in almost 50 % out of 300 terms. Examples of transliterated words are: Faxes (المثالات) (الفاكسات) Scientific American 1996, Vol.12, no.7-8, p.40) Profile (بروفایل) (لاحه / سيماء) Scientific American 2011, Vol.27, no.9-10, p.50) Some transliterated words are naturalized by using derivational processes. Automata (أتوماتيكية / ) (أتمتة) (أتوماتا) Scientific American 2015, Vol.8, no.3-4, p.26) Chemokine (كيموكينة) Scientific American 1997, no.12, p.19) ISSN 2039-2117 (online) ISSN 2039-9340 (print) Mediterranean Journal of Social Sciences MCSER Publishing, Rome-Italy Vol 8 No 2 March 2017 187 Chlorinated (مكلورة) Scientific American 1994, no.10-11, p.40) Galvanization (غلفنة) Scientific American 1997, no.1, p.52) Gasification (تغويز) Scientific American 1993, no.7-8, p.26) Ultrafiltrating (فائقة فلتره) Scientific American 1999, no.8-9, p.95) Automated flight (مؤتمت تحليق) Scientific American 2015, no.11-12, p.28) It is noticed that while Nature (Arabic Edition) utilizes gloss translation, Scientific American (Arabic Edition) uses transliteration. The following two terms are transliterated in Scientific American (Arabic Edition); DNA (الدنا) Scientific American 1995, no.1, p.70) RNA (الرنا).<sup>7</sup>

## II. PRINCIPLES OF SCIENTIFIC TERM TRANSLATION

From the previous examples, we can deduce that the two principles guiding the translation of scientific terms include:

### A. Recombination of Semantic Components:

words, complex words, and phrases may not be translated as equivalent of some single words, complex words, and phrases. However, On the contrary , the recombination of the semantic components of the terms may occur in the target language.

### B. Functional Equivalence Principle

Translating a scientific term entails seeking functional equivalence “between the term in the source language and its translation in the target language” (Luu, 2011). Arntz & Picht (1991, p. 160) suggest four levels of functional equivalence:

---

<sup>7</sup> Bahaa-eddin Hassan, Terminology and Machine Translation, Mediterranean Journal of Social Sciences, Vol 8 No 2 March 2017 Article· MCSER Publishing, Rome-Italy

1) Complete conceptual equivalence

2) Partial equivalence

3) Inclusion, a term merges into another

4) No conceptual equivalence

### **Machine translation in the service of scientific terms**

The number of scientific texts, other than English keeps increasing quickly as compared to past, as the scientific communities in non-English countries grow. However, majority of high impact journals are published in English. For translating scientific text, considering only the semantic representation is not enough. Different terminologies have various meanings in multiple domains. A single word may give a total different concept in various fields. Terms are different in every subject, while translating from one language to another these terms should be considered according to the scenario. For the true sense of data all of the terms should be translated with true meaning of the domain. None of the translator is working on the translation of scientific text in true meaning. While translating scientific text accuracy is a major issue. We can get more advantage by using MT for translating these scientific text in local language or from local language to English language.

Problems in Automated Translation of Scientific Text Translation of scientific text is not as simple as it seems. There are approx. main branches of science and each branch is followed by many sub branches so approx. there are more than 100 fields of sciences. Each field has different terms and meaning of terms, so a single generic system for all these sciences is not easy. Translation of scientific text requires domain and translator expert. These texts are written by using Languages for Special Purposes (LSP). Translation of scientific text not only requires the knowhow of the language but it also requires the deep understanding of the field. For the translation of scientific text both of the skills: translation skill and domain skill are compulsory. Although there are many translators doing the job but translation of scientific text is still ambiguous. Translators are not trained for domain skills. Translators are working only with the translation. There is need to develop a MT system which can be a benchmark for translating scientific text while considering domain of the text.

Shortfalls in Existing MT Techniques for Scientific Text Various terminologies are overlapped in multiple fields but their meaning is different in ever field. Such as the word "monitor" is a device in the field of Computer Science (CS) , but in classroom environment it is used for a student. The word "Python", in CS is the name of a programming language, outside of the CS it is considered a snake. Both examples shows generic words may have different meanings in various field. Existing translators are translating data in generic meanings. They are not able to translate scientific text in real sense. These systems do not distinguish between domains. Above mentioned are the few samples of basics sentences of CS. Translating a whole book or a research paper is more pathetic, humorous. So far no such benchmark is available for translating scientific text. People have to make

extra effort in understanding scientific text because have to understand the language also. So there is need of a study, to identify different techniques which should be used for better translation of scientific text. It is to decrease the effort in learning language skills or translating manually. Existing work do not bridge this gap to translate the scientific text with correct sense. There is a need to develop a customized translator which can effectively bridge this gap. Domain specific translators are tend to give better results as compared to the translation of generic systems.

Methodology Terms of any field plays an important role in the translation of the text. It enables to understand the meaning or idea correctly. For quality translation a term tagger, can play an important role in translation. It is complex to create single term tagger for all the fields together. The development of a quality translator is a challenging and tricky task in MT research community. This is mainly due to the diversity of the languages.

## **Conclusion**

It is now clear that different types of MT systems are required to meet widely differing translation needs. Those identified so far include the traditional MT systems for large organizations, usually within a restricted domain; the translation tools and workstations (with MT modules as options) designed for professional translators; the cheap PC systems for occasional translations; the use of systems to obtain equivalences for the purposes of surveillance or information gathering; the use of MT for translating electronic mail and Web pages; systems for monolinguals to translate standard messages into unknown languages; systems for speech translation in restricted domains. Some of these needs are being met or are the subject of active research, but there are many other possibilities, in particular combining MT with other applications of language technology (information retrieval, information extraction, summarization, etc.) As MT systems of many varieties become more widely known and used, the range of possible translation needs and of possible types of MT systems will become more apparent and will stimulate research and development in directions not yet envisioned.

It is also important to notice that Scientific terms contribute to academic style of a science text. Translating a scientific term, whether inside or outside a context, involves identification of translation units, functional equivalence degree, and how semantic components are recombined.

## **References:**

- Bahaa-eddin Hassan, Terminology and Machine Translation, Mediterranean Journal of Social Sciences, Vol 8 No 2 March 2017 Article· MCSER Publishing, Rome-Italy
- Bahaa-eddin Hassan, Translating Scientific Terminology: Examples from the Arabic versions of Two International Magazines

□ Corpus Based Machine Translation for Scientific Text, Chapter· May 2018, DOI: 10.1007/978-3-319-92007-8\_17

4 authors, including: Ghulam Rasool Tahir and Allama Iqbal Open University

□ W. John Hutchins, Machine translation: a concise history [Website: <http://ourworld.compuserve.com/homepages/WJHutchins>]

□ Muhammad Irfan Machine Translation Department of Computer Science Bahria University Islamabad  
Email: [muhammad.Irfan@outlook.com](mailto:muhammad.Irfan@outlook.com)