JOURNAL OF FOREIGN LANGUAGE RESEARCH
PRINT ISSN: 2588-4123 ONLINE ISSN: 2588-7521
ww.Jflr.ut.ac.ir

The Fundamental and Strategic Linguistic Data, the Key for the Reciprocal Complementation of Technology and Artificial Intelligence in Machine Translation

Alireza Valipour*
Professor of the Faculty of Foreign Languages & Literatures, University of Tehran, Tehran, Iran
Email: alreva@ut.ac.ir

ABSTRACT

This question about the possibility of inventing a system more capable than humans in performing behaviors has always been on the table, even before the advent of powerful computers and artificial intelligence algorithms. Some linguists and philologists believed this process to be impossible. Triumph over the obstacles of linguistic barriers is one of the most important goals and tasks for developments in communication technologies. The idea of constructing an artificial communicative language, such as Esperanto, or the attempt at unifying the language for exchanging information has proven to be abortive due different reasons; therefore, the notion of automating translation process and using Machine Translation (MT) systems continues to be of consideration and reference. Machine Translation is one of the important achievements of artificial intelligence and computational linguistics. This field of knowledge contains the latest and greatest theories of applied linguistics and computer science.

Today, we confide in the fact that the idea of exploiting artificial intelligence, and facilitating its application in multiple areas of science, has been a necessity; however, the implementations of this achievement were not provided by mechanics. This field was not a proper context for creating the reciprocal complementation of linguistics, logic, statistics, and mathematics to revolve around it and could not provide the required agility to employ its achievements in accomplishing Machine Translation. On the other hand, with the advent of a phenomenon known as digital, not only did the procedures of formal logic turn into mathematical logic and that of the mathematics turned into modern mathematics, but also new trends were established in this regard and developed rapidly in linguistic researches properly relevant to digital contexts. After a brief introduction to the topic and its theoretical framework, this paper tries to determine how these achievements have been accomplished and intends to elaborate upon the future prospects with respect to the highlighted methods. While presenting his personal experiences in this area, the author of this paper emphasizes establishing a diverse institutional resource in modern linguistics. In his view, future achievements and accomplishments are due to developing the latest and fundamentally required data in linguistics as an interdisciplinary field of knowledge in order to gain practical achievements.

DOI: 10.22059/JFLR.2021.331469.900

ARTICLE INFO
Article history:
Received: October 3, 2021
Accepted: November 3, 2021
Available online: Autumn2021

Keywords:
Comparative grammar, Data-processing, Informatics, Machine Translation (MT), Relevant technology, Structural grammar

valipour, A. (2021). The Fundamental and Strategic Linguistic Data, the Key for the Reciprocal Complementation of Technology and Artificial Intelligence in Machine Translation. Foreign Language Research Journal, 11 (3), 541-555.

* Alireza Valipour, professor of Russian department at the Faculty of Foreign Languages and Literatures, University of Tehran, graduated from the Comparative, Comparative and Historical Linguistics Department of Moscow State University. Areas of interest and activity: Research in translation of ancient Russian texts, translation problems, machine translation and artificial intelligence, simultaneous and contemporary Russian phonetics, historical and comparative linguistics and Russian literature, and special linguistics of Russian language and lexicography.

پژوهش‌های زبان‌شناختی در زبان‌های خارجی، دوره ۱۱، شماره ۳، پاییز ۱۴۰۰، از صفحه ۵۴۱ تا ۵۵۵

JOURNAL OF FOREIGN LANGUAGE RESEARCH
PRINT ISSN: 2588-4123 ONLINE ISSN: 2588-7521
ww.Jflr.ut.ac.ir

The Fundamental and Strategic Linguistic Data, the Key for the Reciprocal Complementation of Technology and Artificial Intelligence in Machine Translation

Alireza Valipour*
Professor of the Faculty of Foreign Languages & Literatures, University of Tehran, Tehran, Iran
Email: alreva@ut.ac.ir

ARTICLE INFO
Article history:
Received: October 3, 2021
Accepted: November 3, 2021
Available online: Autumn2021

Keywords:
Comparative grammar, Data-processing, Informatics, Machine Translation (MT), Relevant technology, Structural grammar

ABSTRACT
This question about the possibility of inventing a system more capable than humans in performing behaviors has always been on the table, even before the advent of powerful computers and artificial intelligence algorithms. Some linguists and philologists believed this process to be impossible. Triumph over the obstacles of linguistic barriers is one of the most important goals and tasks for developments in communication technologies. The idea of constructing an artificial communicative language, such as Esperanto, or the attempt at unifying the language for exchanging information has proven to be abortive due different reasons; therefore, the notion of automating translation process and using Machine Translation (MT) systems continues to be of consideration and reference. Machine Translation is one of the important achievements of artificial intelligence and computational linguistics. This field of knowledge contains the latest and greatest theories of applied linguistics and computer science.
Today, we confide in the fact that the idea of exploiting artificial intelligence, and facilitating its application in multiple areas of science, has been a necessity; however, the implementations of this achievement were not provided by mechanics. This field was not a proper context for creating the reciprocal complementation of linguistics, logic, statistics, and mathematics to revolve around it and could not provide the required agility to employ its achievements in accomplishing Machine Translation. On the other hand, with the advent of a phenomenon known as digital, not only did the procedures of formal logic turn into mathematical logic and that of the mathematics turned into modern mathematics, but also new trends were established in this regard and developed rapidly in linguistic researches properly relevant to digital contexts. After a brief introduction to the topic and its theoretical framework, this paper tries to determine how these achievements have been accomplished and intends to elaborate upon the future prospects with respect to the highlighted methods. While presenting his personal experiences in this area, the author of this paper emphasizes establishing a diverse institutional resource in modern linguistics. In his view, future achievements and accomplishments are due to developing the latest and fundamentally required data in linguistics as an interdisciplinary field of knowledge in order to gain practical achievements.

DOI: 10.22059/JFLR.2021.331469.900

valipour, A. (2021). The Fundamental and Strategic Linguistic Data, the Key for the Reciprocal Complementation of Technology and Artificial Intelligence in Machine Translation. Foreign Language Research Journal, 11 (3), 541-555.

* Alireza Valipour, professor of Russian department at the Faculty of Foreign Languages and Literatures, University of Tehran, graduated from the Comparative, Comparative and Historical Linguistics Department of Moscow State University. Areas of interest and activity: Research in translation of ancient Russian texts, translation problems, machine translation and artificial intelligence, simultaneous and contemporary Russian phonetics, historical and comparative linguistics and Russian literature, and special linguistics of Russian language and lexicography.

Preface

A while ago, one of the M. A. students visited me to receive guidelines and consults; back then, I was his consultant, and my colleague in the department was his mentor. Somewhere in the course of our discussion, he started complaining, questioning the practicality of his thesis' probable outcomes and sharing his doubt on that whether his comparison of two Russian prepositions with parallel Persian prepositions could be of any use; he wondered if anyone could ever put such an outcome into practice! I managed to reassure him that our scientific records are very weak; therefore, our frustration over their practicality is not our current matter of concern but that of the future scholars. I went on to explain that it is essential for the national scientific archives of every country to be overwhelmed by data output, results, and achievements in every field of science and that considering scientific productivity simply due to its current state is not a thoughtful mindset. Things whose due expirations are in the near future would vanish too early, and a short-term scientific database would be doomed and subject to frail. The ideas of Razi, Mulla Sadra, Kant, and Einstein have survived throughout history due to their long-term practicality.

To clarify the importance of basic achievements in modern linguistics, I may share a personal experience so that the reader better understands the crucial importance of fundamental data. Long ago, I used to cooperate with Russian specialists in a phase of a project for developing a Russian-Persian MT system.

This Russian-Persian MT system project was a national project that started to proceed by the private sector and with the participation and cooperation of a large company after several mutual negotiations. However, the progress of the project was slow due to multiple reasons and it was finally doomed to failure. The most important problem was that we, the Iranian party in this cooperation, could not provide the essential linguistic requirements, which included a variety of frequency dictionaries and textual and syntactic dictionaries, expecting the Russians to compile them for us and encode them to be employed for programming computer systems!

The Iranian manager was not satisfied with the process, believing that the agency in charge was demanding a range of unnecessary tasks, intending to elicit extra wages. However, in this air of distrust, the agency demanded another requirement, exacerbating the tensions caused by mistrust and doubt more than ever.

The secretary in charge of providing linguistic data announced that considering the very specific and professional propensity which was remarked and ordered for the system of machine translation, before compiling multiple frequency dictionaries and textual and syntactic dictionaries, the first issue of consideration in the agenda was the necessity of making a comprehensive textual dictionary. The Russian agency demanded access over all the texts of interest for translation so that the textual dictionary could be compiled by taking account of its contained words. The reason for this demand was due to two upcoming issues which could interfere in the next stage of compiling a frequency dictionary: the first issue was that without the demanded data, the number of words and phrases could not be estimated, and as a natural consequence, the weight value for the output data would be even and non-gradable, and following this false strategy, the applied parser (syntactic analyzer) would have been subject to numerous errors while translating the keywords and key terms.

Everything came to an end after a few months and progress of 40%, almost when the first version of the system was about to appear on the screen monitor. Microsoft made contacts to the Arsenal company in Russia, the other party on our contract, which by that time, had become a branch of Microsoft, making it clear that due to the sanction laws, Arsenal is not allowed to transfer its technology to Iranian NGOs, or it would risk losing the support and companionship provided by Microsoft. Referring to a clause in its contract with Iran, Arsenal indicated to Microsoft that the ongoing contributions with Iran had been originally that of a “know how” and Arsenal was actually selling a product of its technology rather than transferring it. Nonetheless, Microsoft rejected Arsenal’s explanations, Arsenal gave up, and the project was canceled.

We need to accept that our scientific strategy in the course of this project was amateurish. We have entered a shared battlefield, ignorant of its beginning and conclusion. We could have done well in compiling basic textual, frequency, and syntactic dictionaries beforehand; hence, asking an agency to handle this section in a compressed time period would not have been a matter of concern. I may remark that the well-known proverb claiming “when there is a wheel, there is a way” proves to be true only if the “will” is implemented by wisdom and well-provided strength. On the other hand, adopting appropriate strategies in carrying out joint activities is of critical importance, especially with foreign parties and in the current situations of imposed sanctions. In the above-mentioned experience, even though the Iranian party of this contract with Arsenal was a private company and not the Iranian government, from the beginning, it would have been better to regard the section about compiling dictionaries as a general and unprofessional subject; we have mistakenly attracted the attention of those ill-willed associates by showing interest in a very specialized and professional subject. We could have simply clarified our intentions for receiving the technology in order to develop an MT system, directly communicating our demands. Wrong strategy!

I mentioned this experience in the introduction of this study aiming to emphasize that the scientific data and resources are valuable hoards to be used in their due time. This expectation that a good, a service, or scientific product is supposed to be used immediately and in its best shape is not reasonable. As the Russians indicate, “Moscow was not built in a day”; hence, the only thing matters is that in every nation, the construction procedures continue to remain

پژوهش‌های زبانشناختی در زبان‌های خارجی، دوره ۱۱، شماره ۳، پاییز ۱۴۰۰، از صفحه ۵۴۱ تا ۵۵۵

active in the most vital domains. I believe that Iranian's modern sense of morality is stained with a tinge of impatience and hurry; we do not seek long-term programs. Though patience and tolerance are strongly recommended in our culture, and while we constantly quote the prover that "others planted and we consumed, we shall plant for other to consume", and despite this awareness, little is done with regard to taking effective actions in crucial and demanded domains, especially that of developing new and fundamental data-processing conducts; hence, the absence of a general scientific mapping in language research is still quite tangible.

Discussion

Machine Translation and artificial intelligence have always been among the most controversial subjects in different branches of modern linguistics and computer sciences. The question about the possibility of designing a machine with the ability to perform smarter human behavior has been of debate even before the advent of powerful computers and advanced algorithms. Many philologists truly believed this to be impossible.

However, the technology of artificial intelligence proved its potentials to overcome those presupposed limitations over time. Now the present question is how far the progress of artificial intelligence could go? We may admit that today, no one offers such pessimistic responses to these questions, and as we have mentioned before, the cautious implication of "We don't know" might seem a more proper answer.

The study of intelligence is one of the most ancient subjects in the history of science, and for a long time, philosophers have been struggling to understand the process through which seeing, learning, remembering, and reasoning is accomplished; however, with the advent of computers in the early 1950s, the actual possibility of converting mental power into empirical and theoretical rules became available. Many people had the impression that the new "electronic mega-minds" have an unlimited potential for intelligence. However, as the process of inventing a machine for producing artificial intelligence entities proceeded, it became more evident that employing artificial intelligence was much more difficult than what was initially imagined to be; the new ideas appeared to seem more creative and more interesting than the early ones.

This question about the possibility of inventing a machine more capable than humans in performing behaviors has always been on the table, even before the advent of powerful computers and artificial intelligence algorithms. Some linguists and philologists believed this to be impossible.

On the other hand, some scientists in the field of artificial intelligence believe that "from a rational perspective and according to the statistical and mathematical relations, and the subset and possibility of interchanging the alphabets, a simple machine can merge binary, multiple, etc., elements, forming words and phrases which might be meaningful, meaningless, or ambiguous in every language. Certainly, if this machine works permanently, it may manage to produce the best poems of most

پژوهش‌های زبان‌شناختی در زبان‌های خارجی، دوره ۱۱، شماره ۳، پاییز ۱۴۰۰، از صفحه ۵۴۱ تا ۵۵۵

famous poets or the maxims of the most well-known writers, and might even compose poetry better than the ones in hand." (Popov 1990, 201). Although the machine might perform this task, the output waste might be remarkable since the invested input data is still very insufficient. Nevertheless, if the invested intelligence is upgraded and the combinations are defined more accurately, the amount of waste might reduce; since humans are the only being blessed with creative intelligence, and no computer could ever enjoy such a potential. By the way, human beings are capable of providing the suitable conditions to achieve artificial intelligence and then allow many automatic procedures to be conducted by computers. Availability of relevant software and hardware has paved the way for this achievement. Only one aspect remains that has not yet developed sufficiently in different automation areas, including Machine Translation; however, it is still subject to growingly rapid progress: the crucial problem of designing an optimized method for transferring human intelligence to machines.

Additionally, the records of Iranian linguists on the notion of Machine Translation have been mostly false and sometimes inaccurate or wrong. In one of the Iranian linguistic works, which belongs to three decades ago, it is mentioned that "For now, Machine Translation could not be available to the extents of commercial scales because the output is neither satisfactory nor economical." (Bateni 1976, 107) However, at the very same time, many MT systems, including AMPAR in Russia and SYSTRAN in Canada, have been utilized professionally and taking their course of progress with accelerated growth.

Surprisingly, after a long while, the very same assumption about Machine Translation was also included in the first edition of the Dictionary of Linguistic Terms, even though such an assessment seemed truly vague, overtly generalized, and originally flawed!

The reader may find a sense of falsehood in the following paragraphs. Sagharvanian, in the Dictionary of Linguistic Terms, claims that "Translation Machine, or in other words, electronic computer(!) is a machine that transforms a range of icons in accordance with the formerly received instructions, changing them into a range of new icons", and following this vague and general explanation he adds that "editing the text, both before and after the process, helps improve the quality of the Machine Translation." (Sagharvaninan 1900, 77). However, this issue was developed to some extent in the first period and in many models; then, it was canceled due to its lack of efficiency, especially in the case of pre-processing edition. Sagharvanian has quoted an interior reference, repeating that "Machine Translation could not yet be available in terms of commercial concepts and scales due to the fact that the quality of its output is neither satisfactory nor economical!" (Sagharvaninan 1900, 76).

In a general scope, however, we need to confirm that today, the experts and professionals of linguistics are not concerned with responding to these questions, and their assessment of the future prospects in technology is mainly positive.

پژوهش های زبانشناختی در زبانهای خارجی، دوره ۱۱، شماره ۳، پاییز ۱۴۰۰، از صفحه ۵۴۱ تا ۵۵۵

Gottfried Wilhelm Leibniz (1646 -1716) was the first person who invented a mechanical device for processing simple translations. However, with regard to mechanics, the formulation of his logic was too weak, and his concept processing machine could not process language data for achieving satisfactory results (Russell 2004, 8).

Progress in this field was halted for more than a century until it occurred to Charles Babbage (1871-1792) that tables of logarithms could be computed by machinery. To do so, he designed his idea to include the notions of memory address, stored programs, and conditional jumps for the first time. However, the idea did not develop much in operation and technology due to its failure in mechanical aspects.

To find the reason for this failure, we need to reclaim the inevitable necessity of a rich array of research records for conducting practical sciences and achieving remarkable goals. Similarly, in the case of computational linguistics, it is required for associated preliminary records to be prepared and accessible. There have always been cases where linguistics, as an interdisciplinary field of science, was deprived of such important records for defining wordnets, textual dictionaries, and linguistic corpora.

This slow and somewhat disappointing process in artificial intelligence lived on until the 40s, when, with the invention of the computer, its progress started to race forward in an amazingly accelerated speed This process is still going on rapidly and steadily.

The speed and potentials of hardware increase every day alongside the decrease in prices. The advances made in computer science have proved to be extraordinarily successful, and it seems that this incremental growth knows no limits. Parallel computers are capable of performing truly complex and complicated computations whence receiving proper inputs and using computational algorithms to obtain better outputs. Artificial intelligence shares a growing range of achievements with other fields of science, including agriculture and Aerospace engineering, resulting in many great revolutions. Language and linguistics were also required to adhere to this positive trend, and relevant transformations, investments, and revolutions, or enhancements in language research should have started at some point where ensured productive progress was the certain outcome.

As it was mentioned, after the 1940s, and by the invention of the computers, an astonishing velocity also enhanced the development and progress of MT systems during different periods. After the previous enhancements, in every following period, specific features have remarkably developed. The language was encoded, and then specialized dictionaries were compiled for this professional treat. Then came the stage where the first algorithms and the first software program were developed.

However, it is acknowledged that the MT systems which were offered to this period lacked the required professional functionality and merely provided for the commercial prospects and representational aspects of the whole project in a stereotypical manner. In other words, it was so evident that the functionality of these machines was extremely limited; the input sentences were required to be very simple in order not to trouble the machine for the analyzing process, and then, the output still required editing.

After this period, the specialized features of MT systems reached an acceptable level of development; now, in the mid-60s, the essential demand for designing MT systems with algorithms as their fundamental functioning elements had been experimentally proved; according to statistical prospects, the systems came to be recognized capable of providing acceptable productivity and success. Additionally, it was acknowledged that the new designs needed to include explanations and clarifications about the syntactical elements suitable for being automated by machines in which the words were classified in accordance with new measures as distinguished with the traditional forms in syntax. These new measures included the following features:

1- The focus shifted from lexicon to syntax. 2- New ideas were proposed concerning a style of semantic analysis specialized for machine automation. 3- Syntactic and semantic algorithms were employed in MT systems. 4- Those methods were considered in which translation would be possible from a language into many different languages. 5- Operational application of Machine Translation started in many organizations, however, on limited scales. 6- The need to include other fields of science such as math and programming and to extend the algorithms became tangible. 7- Frequency dictionaries were compiled, and new plans were scheduled in accordance with associated statistics. 8- Industrial translation proved to be absolutely successful. 9- With the enhancement of computer processing speed, the possibility of using smart programming seemed available.

Finally, arrived the third period, which began in the early seventies, with informatics starting to grow. “The rapid growth of computer technology provided computer programmers with better facilities; however, the work process was slow until 1976” (Marchuk 1983, 20).

The remarkable point is that in this third period, the style of formalist programmers all around the world could be divided into two categories: those based on Chomsky’s practices as opposed to Melchuk’s initiative style. Hence, we need to point out that with regard to actual experience and based on objective realities, Melchuk’s meaning-text theory has proven to be more practical in developing MT systems in comparison to Chomsky’s theory of universal grammar.

Different types of dictionaries perform multiple and important roles in the MT system; prioritization and valuation of the processing stages regarding syntax and lexicon would turn out to be false and abortive without compiling frequency dictionaries. In order to avoid the speed declination in machine analysis programs due to managing search operation in an extended vocabulary volume, these words are required to be encoded by numbers. It means that the morphe, which is repeated more frequently, will be assigned by No 1, etc. This numbering process is required to be in accordance with the data obtained from the frequency dictionary to ensure its definite accuracy (Valipour 1173, 2007)

As the grammar required in this system is different from the traditional grammar and includes many extensive complexities, the MT system dictionaries are also distinguished from ordinary ones in many different aspects. These differences need to be considered for extracting word entries, processing, and encoding them and also for employing them for non-human applications such as MT systems. Dictionaries used in MT systems have basic differences with those used for general applications beyond Machine Translation: The first distinction is that the MT system dictionaries are encoded and will be stored and restored in their coded formats; The second point is associated with this quality that they might be employed respectively and distinctly during the different processing stages of the MT system. The encoding and compiling of machine dictionaries are required to be designed so efficiently that they bear no damage due to the MT system's constant references, and their main quality of high-speed operation needs to be guaranteed.

One of the most significant obstacles on the way of constructing a Machine Translation and at the lexicon level is the issue with multiple meanings of the same words, which is the most well-known characteristic of natural languages (Apresyan 1987, 47), and it is the outcome of a linguistic principle famously referred to as "economy language". A translator may identify and choose the true equivalent of the word by relying on his language competence and with regard to the context and subject of text while dealing with a term that bears multiple meanings; however, the computer does not have such capabilities. Therefore, the machine designer is supposed to take some preliminary actions and provides some algorithmic arrangements in order to alleviate the effect of this remarkable gap. These arrangements are required to initiate from the earliest processing stages, i.e., from the construction of lexicon and syntax, and to proceed until the last stage, i.e., the semantic analysis of the text. But the problem of words. Nevertheless, the problem concerning the words with multiple meanings had not yet been resolved. Additionally, the semantic aspects of the machine had not improved to a satisfactory level despite the many excessive efforts for encoding and constructing a proper grammatical structure to include both inflection and syntax; even in the case of human interference with machine operation for deciding over the meaning choices, the task was believed to be a professional deed, and not capable for public use. Hence, in the conference held in Luxembourg on the notion of machine translation in 1977, it was concluded that both humane methods and mechanical systems are required to be included in the process of translation.

Finally, since 1977, practitioners of MT systems agreed on limiting the procedure; therefore, as was the case in the earlier stages where poetry and novel have been

پژوهش‌های زبان‌شناختی در زبان‌های خارجی، دوره ۱۱، شماره ۳، پاییز ۱۴۰۰، از صفحه ۵۴۱ تا ۵۵۵

considered to be too grave subjects for translation, the notion of general translation was also considered to be not suitable for this system. This achievement was postponed to the near future[1]. On May 26th, 1978, in New York, the United Nations held an assembly for experts on Machine Translation in which systems with high-performing systems were introduced and presented. World Translation of Canada presented its MT system named SYSTRAN, originally designed at Georgetown University in 1974 and then developed in cooperation with other research institutions. Many linguists were positive about using the notions of transformational grammar, and syntactic structures introduced those theorists like Chomsky and believe that "word to word translation is not possible and practical for designing a computer capable of translating a text from one language to another". (Dabir Moghaddam 2004, 120).

There is no doubt that scientific progress is largely dependent on the translation and dissemination of information through this procedure. In the Russian translation center of the early 70s, about 6 thousand pages were translated yearly, and in the late 70s, this number reached 100 thousand pages per year, yet, this demand is still growing.

Eight official languages are acknowledged in the European Economic Community, and all the documents and certificates are required to be translated into these languages. According to the reported statistics, in 1978, the number of linguists and translators working in the organization were 1,300 people who have translated 600,000 pages in the course of one year, and currently, the number of translators has increased to 2,500 people though not sufficient to fill the gaps, and still the shortcomings are tangible (Marchuk 1985, 5).

Beyond the current demands of modern industrial communities for information access, eliminating linguistics barriers in inter-humanistic communications has always been a universal desire for humanity. Thus, there are and always have been sufficient, strong, and even permanent incentives to accomplish this goal. There is no individual or government free from the need to gain information on extensive scales; today, information exchange has proven to be more profitable than ever, and this trend intends to proceed. In Russia and western countries, the history of efforts for building MT systems extends beyond half a century; however, the procedure of these researches and experiences has always been associated with many vicissitudes due to adopting different procedures. On the other hand, the science of Machine Translation, which, same as artificial intelligence, has a history of more than 50 years of practical and functional application, is an intersection

[1]. There is no doubt that in a reasonable and logical viewpoint, by reinforcing the achievements of artificial intelligence, enhancing computer speed and memory, and other factors, machine translation will successfully replace human translation as was the case with automatic bakery or automatic surgical services. Although right now it may seem absurd, this possibility is not so farfetched that Popev's Jules Verne-like imaginations come true, and besides machine translators, literary machines may start to develop with the capability to produce poetry and fiction.

پژوهش های زبانشناختی در زبانهای خارجی، دوره ۱۱، شماره ۳، پاییز ۱۴۰۰، از صفحه ۵۴۱ تا ۵۵۵

for informatics, especially software engineering to be mixed and entangled with linguistics. This integration was so profound the boundaries between these fields of science were removed that in some cases, and some linguists or software engineers intended to achieve this goal on their own and independent from the other group (Valipour 157, 2007).

Nevertheless, the samples of successful experiences in this work indicate the importance of cooperation between linguists and engineers; besides, being aware of the specialized methods in both of the relevant fields and professional work compliance may facilitate and promote the work procedure within the context of smart systems for employing Machine Translation.

Today we are aware of the fact that the idea of exploiting artificial intelligence, facilitating the application in different areas of science has proven to be both right and necessary; however, the means for achieving this aim was not mechanics. This field was not a proper context for creating the reciprocal complementation of linguistics, logic, statistics, and mathematics to revolve around mechanics and could not provide the required agility to employ its achievements in accomplishing Machine Translation. On the other hand, with the advent of a phenomenon known as digital, not only did the procedures of formal logic turn into mathematical logic and that of the mathematics turned into modern mathematics, but also new trends were established in this regard and developed rapidly in linguistic researches properly relevant to digital contexts. (Valipour 122, 2007).

Consequently, concerning contemporary linguistics and due to the growing advances in Machine Translation and the widespread demand for this technology, the necessity of conducting modern research for collecting and processing language data seems inevitable. “The processing of words and inflectional analysis of lexicon in each language is associated with specific characteristics and problems. Some of these problems, which are common in all languages and are not assigned to a specific language, include: having multiple meanings, determining the styles and limits, as it is natural in translating one sentence into another language where the number of inflectional and consequential icons do not match equally on both sides, the meaning of words also differ from one language to another language and are not definitely parallel.” (Barkhudarov 1995, 122) Let us, for instance, assume that an infinite list of English phrases is provided for and English language speaker together with the syntactic and grammatical description of all those phrases. By using an English language dictionary and relying on his own linguistic knowledge, that person has the ability to extract all the potential meanings of each phrase in that infinite list. He would be able to identify the different meanings of a single phrase, and also to recognize whether a sentence is meaningless or meaningful, and remark those sentences in the listed with similar meanings. Now, compare the above-mentioned example with the following alternative: Let us assume that we have constructed a machine that is mechanically capable of recording the many different meanings of each morpheme in each sentence on the mentioned infinite list by employing a dictionary. Apparently, the number of meanings associated with each morpheme in a dictionary, is significantly more extensive than the meaning of that morpheme in a single phrase. However, our hypothetical system cannot achieve the true meaning of the morphemes in one phrase unless the meaning is determined by syntactic information associated with that morpheme. For example:

Those phrases like “the fish was bought by the cook” and “the fish was bought by the river” represent the inefficiency of a machine to decipher the meaning in similar contexts; however, by designing a proper logarithm, the machine would be capable of recognizing the distinction between the word “by” as a verb and “by” as an indicator of a passive voice. Also, in another example mentioned below, the system would be capable of deciding over the correct meaning of the word “seal” in a phrase like “seal the letter”, provided that it has been instructed about the role of this word as a verb in this phrase. Similarly, the system would be able to gather the correct meaning of the word “seal” in a phrase like “The seal is on the letter”, given that it has been instructed to recognize it as a noun. Nevertheless, the system is unable to distinguish the correct meaning of the word “seal” in a phrase like “one of the oil seals in my car is leaking” from its other meanings, which are irrelevant to this context and even being instructed that this word is a noun would not be a significant clue since all the indicated meanings already mentioned for the word “seal” in the above examples are nouns too. What our device fails to operate upon is exploiting the semantic relations between the different morphemes in a single phrase; hence, the meaning or meanings of a sentence could not be extracted clearly. Due to the same incompetence, our system would not recognize those phrases which are meaningless from those which are meaningful. For instance, this machine could not distinguish between “The wall is covered with silent paint” and “The wall is covered with fresh paint”. The final assumption is that this system can only recognize those phrases with exactly similar syntactic structures where the constructing words of each phrase are either exactly parallel or at least synonymous! (Dabir Moghaddam 2004, 120) Therefore, besides word order, word sequence may also seem non-grammatical due to the violation of syntactic rules while selecting relevant vocabulary items. To more clarification on this point, see how example “a” seems grammatical, while example “b” sounds non-grammatical:

a. The cake is good.

b. The expire is good.

The problem with phrase “b” is that the syntactic rules of the English language demand a particular class of words before “the”; this class is named as “noun” in traditional grammar. In phrase “a” the word "cake" is a noun, so the whole phrase seems grammatically correct, but in phrase “b” the word “expire” is a verb.

The word "cake" belongs to a class different from that of a word like "expire". Different names have been designed over time for naming these classes; the first word above represents nouns, and the second represents verbs. The knowledge concerning word classes is associated with grammatical measures, which enables the speakers to produce and understand grammatically correct sentences, and simultaneously identify non-grammatical sentences, hence to avoiding them. Therefore, the syntax description of the English language is required to include word classes too. Thus, it is clear that due to somehow similar reasons, these problems with Machine Translation may remain so until their incremental improvement. Since the notions of syntax and meaning, which are even considered the same by some scholars include the most confusing aspects of each language, some contemporary scholars like Bloomfield have regarded the accurate formulization of syntax and meaning for practical applications as an impossible abortive task.

We may remark that when the adherents of structural linguists began to study syntax, those discovery procedures that have seemingly proved to be effective in the case of sounds and words could not provide any outcome. The analysis of sentence structure, which is more complex than those of sounds and words, demanded more complex research principles and methods besides analysis. Structural linguistics had a small share of our understanding of the meanings (Julia S, Falk 1998, p 30); therefore, their achievement did not contribute to designing proper algorithms for Machine Translation. The author of this article believes that the evaluation of the results achieved by accurate formulization of syntax has always been a subject of extremity in practical applications. Experience has shown that neither the ignorance of Bloomfield's structuralism toward this problem nor the simplified generalizations of Chomsky's practices and others who have offered new patterns could not contribute to promoting developments in Machine Translation. The brief history of applied linguistics shows that ignorant simplification of the revolutionary achievements has hardly contributed to the development of this field and even has impeded its actual progress. This process demands great care and accuracy; hence, it would be naive to consider its development and progress only due to promoting new linguistic theories. For instance, after the publication of Chomsky's book name *Syntactic Structures*, Robert Lindsay has positively elaborated upon the possibility of using this theory in "Machine Translation" and its great significance. He has considered the current methods of that period (i.e., word to word translating) inapplicable for designing a computer that would be able to translate a text from one language into another language. Instead, he has suggested that building a set of three separate machines, each corresponding to one of the three linguistic levels proposed in Chomsky 's theory, would be a practical and approachable possibility regarding Machine Translation (Dabir Moghaddam 2004, p 120).

Further advances in linguistics showed that the problem was significantly more complex than it was supposed to be. Language has

پژوهش‌های زبان‌شناختی در زبان‌های خارجی، دوره ۱۱، شماره ۳، پاییز ۱۴۰۰، از صفحه ۵۴۱ تا ۵۵۵

many ambiguous and unspoked aspects; hence, understanding the language demands understanding the associated topics and relevant contexts, rather than merely identifying the sentence structure. (Russell 2004, 16)

In his book, Chomsky has mentioned some works in which generative grammar was considered to be an outcome of the efforts made to employ computers for different purposes, including Machine Translation; however, he denies this view and regards the origins of his grammar to be traditional linguistics. (Dabir Moghaddam 2004, p 179). Chomsky has yet to refrain from expressing his motivation for introducing this grammar system. Neither did the documents he and his followers had provided for the superiority of the generative grammar over the other systems could prove its applicability in Machine Translation. Chomsky had long held on this belief that a comparison between the core sentences in different languages would reveal that, despite the apparent differences and of the surface structures in these languages, there is a lot of similarity between them in the underlying level, hoping that by universal decoding of deep structures in all the sentences, the proper context might be provided for constructing suitable algorithms in accordance with the common methods in transformational grammar (Valipour 119, 2007).

Today, due to various reasons, some of whom are referred to in this article, this mindset has revealed itself to be unwise and excessively optimistic since practically it has never come true. Since, as it was mentioned in the introduction, the problems in Machine Translation are mainly due to the inefficiency of the system's procedure in identifying the semantic relations between the morphemes in a phrase, and therefore, it could not extract the meaning or meanings of that phrase successfully. Besides, the various generative transformations which have so far represented the transformation of deep structures to surface structures do not cover all existing syntactic structures, and there are still many exceptions and ambiguous points. On the other hand, scientific evaluation of human and Machine Translation is not a simple task, and currently, one of the most significant issues about the function of Machine Translation is adopting a means for a proper evaluation of translation results. Evaluating MT systems by giving them some input sentences to process and comparing the output with a good human translation is the simplest method, applicable for non-specialist too; however, such way of assessment is not accurate and professional; hence, it might be used mainly as a corresponding and complementary piece of research.

The prevailing MT systems in linguistics include research and investigation of word classes, inflection and syntax, and semantics. The three components mentioned above are in correspondence with the three levels that modern linguistics considers as inherent for each natural and living language. The applicable methods for studying and representing the research achievements in these fields are different from the common classic and traditional methods. It can be inferred that there are no limitations in using linguistic theories for producing Machine Translation. In order for compiling textual and grammatical dictionaries and designing syntactic analyzers, the achievements of comparative grammar, contrastive grammar, expectance grammar, pivot grammar, structural grammar, phrase structure grammar, transformational-generative grammar, or any new method or theory suitable for the procedure have been employed so far in different MT systems. For instance, in the Russian language, the method for identifying the base form of a word or a verb in the MT system is different from its classic and traditional definitions; thus, despite the fact that textual dictionaries could be applicable in solving the problem of multiple meanings in the MT systems, this method does not always prove to be efficient. There are some cases in which other procedures are demanded. All of these aspects indicate the broadness of this field for further research and study and the tremendous demand for constructing infrastructures with multiple dimensions as a necessity for contemporary linguistics on both national and global scales.

Conclusion

MT systems include multi-dimensional procedures, which are based on problem-solving competence and other initiatives in multiple areas. This task also includes a remarkable dose of interdisciplinary knowledge. Artificial intelligence strategies could provide a proper form of language suitable for programming and transmission to machines by employing mathematical principles, statistics, and logic. In order to make advances in Machine Translation, mastery over data analysis, defining and designing proper language models for smart applications, developing and employing relevant software for this task, and also employing the technology of artificial intelligence and programming. Taking account of the distinction between these three fields, this paper elaborated upon some linguistic dimensions of MT systems in accordance with experienced approaches.

Due to the increasing progress in Machine Translation and the widening demand for this technology, conducting new pieces of research for collecting and processing language data is an inevitable necessity. Although the history of designing MT systems now goes beyond fifty years, and despite the numerous attempts at making compatibilities in lexical and inflectional paradigms, little progress has ever been made in syntactic and grammatical areas (excluding lexicons). Evidently, traditional descriptions of the traditional grammar at the syntax level do not resolve the many ambiguities tangible in the process of comparative translation. According to some approaches approved by hermeneutics practitioners such as Chomsky, all languages share the same underlying structure, and languages differ only in their selection of hermeneutics. This theory is still not prescribed as a certain rule due to its inherent ambiguities in practice.

If this theory were absolutely true, it would have ensured great and practical reflections in Machine Translation; the syntactic algorithm of mechanical procedure would have been concentrated solely on this subject so that after identifying the hermeneutics of the source language and the target language, and the removal of the surface structures, an underlying structure of universally shared sentences could have been obtained. Finally, it would have been possible to apply the hermeneutics of the target language to this obtained underlying structure in order to reach a proper translation. Although these experiments belong to a different period, the results of accurate formulization of syntax have always been a subject of extremity in practical applications; experience has shown that neither the ignorance of Bloomfield's structuralism toward this problem nor the simplified generalizations in the practices of Chomsky and others who have offered new patterns could not contribute to promoting developments in Machine Translation.

Therefore, the only remaining solution in order to improve these mechanical procedures would be exploring and extracting syntactic relations on a large scale and taking the essential accuracy required in syntax and semantics into consideration. The success of Machine Translation in the future depends on trying to understand and extract semantic relations and syntactic structures and then classifying them for practical application. Our experience in this study also confirmed that the more well provided the potential syntactic algorithms for the system's procedures, the more successful the extracted translations would be. The accuracy of this claim has been proved in experimental operations and evaluations performed on the procedures of Machine Translation. Partial achievement to the desired outcomes provides a fundamental reinforcement for modern linguistic data and would pave the way for its application in technology.

Resources

1. Falk, Julia S. (1998). Linguistics and language. Trans. Gholam Ali Zadeh, k. Ed. 5th . Astan Print.

2. Dabir Moghaddam, M. (۲۰۰٤). Theoretical linguistics. Tehran: Samt.

3. Russell. Robert. (2004). Artificial intelligence. Tr. Rahnamoun, R. Tehran: Naghoos Press.

4. Valipour, A. An examination of the linguistic dimension of machine translation and a brief introduction of relevant sciences. Foreign Languages Research, No 37, Spring 2007.

5. Valipour, A. A comparison and review of some Russian and Western machine translation systems. Foreign Languages Research, No. 42, Winter 2007.

6. Bateni, M. (1976). Modern linguistic issues, Tehran: Agah.

7. Dabir Moghaddam, M. (2004). Theoretical linguistics, Tehran: Samt.

8. Deitel, Harvey M. (2006). XML How to Program. Tr. Pashaie, B. Ed. 11th. Tehran: Jahanno.

9. Russell, J. S. (2004). Artificial intelligence, modern approach. Tr. Rahnamoun, R and A, Hamavandi. Ed. 3rd. Tehran: Naghoos Press.

10. Sagharvanian, S. J. (1990). The dictionary of linguistic terms, Mashhad: Nama.

11. Apresyan, Yu.D. (1987). *Linguistic support in the third generation automatic translation system. USSR Academy of Sciences*, preprint of the scientific council on “the problem of Cybernetics”. M.

12. Yngve, V. (1965). The value of machine translation research. – ITY, M.

13. Marchuk, Yu.N. (1983). *Machine translation problems*. M.: Наука.

14. Melchuk., I.A. (1989). *To the construction of a working model of the language* “Sense-Text”. M.

15. Popov, E. V. (1990). E.V. (1990). *Artificial intelligence book 1 radio and communication*. M.

16. Barkhudarov T.I. New lexicography. — M .: VTsP, 1995.—183 c.