AUTOMATIC TRANSLATION: HISTORY AND FUTURE CHALLENGES

Alberto Maria LANGELLA, Ph.D. in computational linguistics,
Research Fellow, University of Salerno, Italy

Abstract: *This article will retrace the main stages in the history of machine translation*[1]*, from the dawn of pioneering research in the 1940s and 1950s to the promising techniques of today's statistical machine translation. We will evaluate the main characteristics of the different approaches used over time, and we will focus on the differences between the different types of machine translation practiced in seventy years of research. In conclusion, we will evaluate how the increased computing capacity of modern processors has profoundly influenced machine translation, and how today's artificial intelligence techniques have succeeded in automatically processing some of the more complex aspects of natural language.*

1. The first phase of machine translation: from the end of the 1940s to the mid-1980s

We are all used to using free electronic dictionaries and automatic translators available on the web. The level of machine translations has greatly improved today, even compared to just a few years ago, not to mention the very unsatisfactory performance of machine translators in the 1960s and 70s. The reason for this improvement is both theoretical and technological. Theoretical for two reasons: a better understanding of how natural language works by linguists; the adoption of innovative statistical techniques for the automatic analysis of language. Technological to the extent that processors have undergone a very rapid evolution in recent years, with a significant increase in computing capacity following the microchip miniaturization processes. But before these improvements in

[1] I recommend the text Translation as an introduction to the complexity and pervasiveness of the translation activity. A Very Short Introduction by Matthew Reynolds.

recent years, machine translation has come a long way that roughly coincides with that of information technology tout court. In this paragraph we retrace the main historical stages of automatic translation systems, from the precursors to the first real automatic translators of the 1950s, to finally arrive at today's innovative methods of statistical analysis.

1.1 The forerunners

The need to translate is ancient and has always represented the desire to overcome the constraints of incommunicability, to overcome linguistic barriers. We speak several languages, today only about 7000. Some are successful languages, for political and cultural reasons, others, and many, risk extinction under the centrifugal impact of a few dominant linguistic traditions (English, Spanish, etc.).

Translating has always been an activity carried out by men for other men. Today we are witnessing something new from this point of view. Translating can also be an activity performed by a machine, from which everyone can benefit. This presupposes, of course, the existence of calculating machines, before which there was no automatic translation. However, in the past some important reflections have anticipated and stimulated the lively contemporary interest in this research sector. Leibniz and Descartes were among the first to be interested in the possibility of conceiving a universal language, within a tradition of philosophical thought inspired by the search for an alleged Adamic language. Leibniz thought that a language of this type could solve problems of a philosophical, legal and moral nature, facilitating communication between people of different nationalities. In the same years Descartes was interested in the same problem as evidenced by a letter of 1629 to Mersenne:

> If [someone] put into [his] dictionary a single symbol corresponding to *aymer*, *amare*, *philein* and each of the synonyms, a book written in such symbols could be translated by all who possessed the dictionary. (POIBEAU 2017: 40-41)

Descartes therefore thought of a unique and universal numerical code for all languages; in this idea he was far-sighted, anticipating the inspiring principles of modern computers by many centuries. For the latter, in fact, words like any other data are nothing more than binary numbers. During the seventeenth century, this reflection inspired the research of many other researchers who attempted to create Descartes' universal numerical dictionary: Cave Beck in 1657, Johann Joachim Becher in 1661, Athanasius Kircher in 1663 and John Wilkins in 1668. In 1811 Joseph de Maimieux and Arman-Charles-Daniel de Firmas Périés created a numerical dictionary for military communication needs.

During the second half of the 19th century, Johann Martin Schleyer invented the artificial language Volapuk, while Ludwik Lejzer Zamenhof invented Esperanto. The idea of both was to facilitate cooperation and peace between peoples.

In the first half of the 20th century we have two other important contributions: the mechanical brain of Georges Artsrouni and the assisted translation machine of Smirnov-Trojansky. Two prototypes of the first were built between 1932 and 1935, and it received a prize at the Universal Exhibition in Paris in 1937. Neither one, however, managed to influence the research of the 1940s and 1950s, whose spirit instead it was affected by the advent, in the same years, of the first calculators. The mechanical brain of Artsrouni was conceptually superseded in favor of electronic devices, while fully automatic translation projects were preferred to Smirnov-Trojansky's computer-assisted translation machine.

1.2 The beginning of machine translation: rule-based systems and dictionaries

After the Second World War, the first computers appeared, and machine translation was immediately one of the applications of greatest theoretical interest. In addition to the theoretical interest in the functioning of languages, the first researchers in this sector were also inspired by the specific need, originating from the Cold War, to translate from Russian into English. Among the pioneers of those very early years we remember the Englishman Andrew Booth of Birbeck College in London and Warren Weaver[2]. Booth pioneered speech recognition studies and invented an

[2] Weaver was together with Claude E. Shannon the inventor of the mathematical theory of communication. In 1949, the two published a book entitled The Mathematical Theory of Communication. This contribution had a great impact on the research of those years and influenced the course of studies in this research sector in the following decades.

automatic technique for morphological analysis called stemming[3], still used today by search engines. Stemming is particularly effective for English and arises from the need to simplify morphological analysis. If an automatic system encounters, for example, the word running, it proceeds to progressively eliminate the final letters until it finds a word contained in its dictionaries: in this case it stops at the word run. In this way it is possible to bring a large number of morphological variants back to a small set of basic lemmas. Weaver's research was influenced by his mathematical theory of communication based on a scheme in which there is an emitter who encodes a message and transmits it through one of the potential channels to a receiver who proceeds to decode it. This theory was the basis of subsequent studies on cryptography. Weaver conceived of machine translation as analogous to decrypting a message. In a 1947 letter to the American mathematician and statistician Norbert Wiener he wrote:

> One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode." (POIBEAU 2017: 53)

Wiener believed that the major problem facing machine translation research was the polysemy of words. This aspect, common to all natural-historical languages, made any attempt at automatic word-by-word translation problematic, in the impossibility of establishing between them, in the relations between one language and another, a biunique correspondence relationship:

> As to the problem of mechanical translation, I frankly am afraid that the boundaries of words in differemernt languages are too vague […] to make any quasi-mechanical translation scheme very hopeful. (POIBEAU 2017: 53)

Such vagueness and polysemy of languages would have required automatic translation systems to disambiguate words on the basis of the co-text[4], a possibility beyond the reach of the processors of those years,

[3] This technique was popularized in the 1980s by Martin Porter and is known today as "the Porter stemming algorithm".

[4] I use the term co-text in this article with a different meaning than context. The first refers to words that co-occur with a given word within a text, the second to the pragmatic situation that surrounds the act of enunciation.

which were still very limited in terms of computing capacity. Moreover, the research on artificial intelligence, from which a valid approach to deal with this type of problem would have arisen, was still in an embryonic stage.

Despite these objections Weaver wrote a memorandum on machine translation in 1949, historically considered the beginning of research in this area. The main points of the memorandum were four:

1. The correct semantic interpretation of a word must be performed by analyzing its co-text. Words that need to be disambiguated belong to three classes: nouns, verbs and adjectives.

2. It is possible to determine a set of logical and universal rules that solve the machine translation problem.

3. Cryptography is a valid theoretical and methodological model to conform to.

4. Instead of translating directly between two languages it is necessary to think of some abstract representation model that can facilitate the task.

The first point anticipated the functioning of current statistical machine translation systems by many decades. The second instead stimulated linguistic research on the concept of formal grammar. The third point reaffirmed the importance, as for cryptography, of statistics in the study of language. Finally, the fourth proposed the possibility of using an intermediate language, called interlanguage or pivot language, in the process of automatic translation. Each of these points had profound repercussions for research for years to come.

The first machine translation experiment was that of the Georgetown University research group in collaboration with IBM. The system translated 49 sentences from Russian to English using a dictionary of 250 words with the addition of 6 grammar rules; had a considerable media coverage, resulting in a substantial increase in public funding in favor of this research sector. This demonstration was made two years after the first conference on machine translation, held at MIT on the initiative of the linguist and philosopher of language Bar-Hillel[5]. The results of this experiment, very

[5] Bar-Hillel was a very influential figure during the early period of machine translation. Of Israeli origins, he had a post-doc fellowship at MIT, where he spent two years, from 1951 to 1953, under the guidance of Rudolf Carnap. Carnap worked in those years on a logical syntax of natural language and was a leading figure in the field of formal logic studies. Bar-Hillel would return to the United States

positive for that time, had the merit of stimulating the start of research on machine translation in other countries of the world, where research groups would be born within a few years that could benefit from substantial government funding.

All machine translation systems of those years were based on transfer rules and bilingual dictionaries, or alternatively on an interlanguage. In the first case it was a simple word-for-word translation[6] between the source and target languages with a subsequent rearrangement of the words in accordance with transfer rules. These rules had the task of checking that the translated sentences respected the syntax of the target language. This approach presented considerable difficulties, to the extent that it assumed a good knowledge of the syntactic rules of the two languages, in a historical moment in which linguistics was taking its first steps and did not yet possess sufficient control of many aspects of historical-natural languages. Furthermore, word-by-word translation required a continuous multiplication of the inputs of electronic dictionaries, since each word can have multiple meanings depending on the co-text as well as being subject to a continuous and inexorable diachronic evolution; finally, in an electronic dictionary, in addition to the terms, all the morphological variants must be listed. This disambiguation work, which speakers carry out without major problems, determined that the systems of the time were forced to handle dictionaries of considerable size, at a time when computers had very limited calculation capacity and memories.

The systems based on an interlanguage, on the other hand, tried to tackle the problem linked to the translation between genetically and typologically distant languages, a particularly thorny one since it required programmers to have a good knowledge of languages that had not yet been sufficiently studied at the time. English (pivot language) was often used, which performed the function of the intermediate and best studied

in the late 1950s, eventually becoming one of machine translation's greatest detractors.

[6] As already pointed out, in the context of the Cold War these first pioneering US research groups worked exclusively on translation from Russian to English. In the second half of the 1950s, work began on machine translation between other languages and English in other countries. The prestige of the latter marked the entire history of machine translation to some extent, especially in the first period, at a time when machine translation systems benefited from the linguistic knowledge that the various US research groups were acquiring.

language in which one translated from the source language, and then moved on to translating from English to the target language. However, with systems based on a pivot language, we collided with all the difficulties and limitations that we have seen characterizing systems based on rules and transfer rules.

These difficulties, which the Georgetown University system did not take into account limiting itself to the translation of a few sentences in a specific linguistic domain, would have led within a few years to harsh criticisms on the performance of those first automatic translators, and finally to doubts relating to the possibility that machine translation could never in the future provide satisfactory results.

1.3 The first criticisms: the article by Bar-Hillel and the ALPAC report

Translatologists agree that word-for-word translation is very unsatisfactory and the first machine translation systems were no exception to this rule. Being based on transfer rules and extensive electronic dictionaries, they also met the limit of the limited memory and calculation capacities of the processors of those years. The multiplication of the meanings of words, according to the many and very numerous contexts, proved to be an insurmountable obstacle for those first systems based on rules and dictionaries or on a pivot language. The dictionaries began to assume considerable dimensions becoming, for the reasons just explained, difficult to implement. Furthermore, the idea that electronic dictionaries could collect all the meanings that speakers naturally decode encountered theoretical as well as practical limits.

Bar-Hillel was the first to point out the limitations of those early systems in a February 1959 report commissioned by the U.S. Office of Naval Research entitled "Report on the State of Machine Translation in the United States and Great Britain". He noted that the syntactical knowledge necessary for satisfactory machine translation, especially between genetically distant languages, was very complex and still to be acquired from the linguistic research of the time.

In addition, he made some considerations on semantics in an appendix to this report entitled "A demonstration of the non-feasibility of fully automatic, high-quality translation"[7], in which he underlined some interpretation problems related to semantic ambiguity as in the example following: John was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy. The ambiguity consists in the interpretation of pen as a space where children play and not as a pen, an interpretation that not even the analysis of the context can suggest but only the knowledge of the real world possessed by the speakers. This represented an insurmountable theoretical limit for a fully automatic translation. Bar-Hillel therefore suggested assisted translation as a possible and most useful research direction: an automatic translator provides translation suggestions that can be corrected (post-editing) and used by professional translators.

This report had the result of dampening the enthusiasm of the first years of research and emphasizing the difficulties associated more generally with the automatic processing of natural language. The idea of the Georgetown researchers to commercialize an automatic translation system quickly proved unfeasible, and incidentally, the 1954 demonstration of these pioneers was based on an automatic translator that was too rudimentary (a few sentences from Russian to English and a very limited lexicon). After Bar-Hillel's criticisms, many researchers began to migrate to related sectors such as computer science or linguistics and on the government front the idea of entrusting themselves to a commission of experts matured. In 1964 this commission, called the Automatic Language Processing Advisory Committee (ALPAC), was formed. The direction was entrusted to a theoretical computer scientist, John R. Pierce, and was composed of machine translation experts, artificial intelligence specialists, linguists and a psychologist. The report was published in 1966 with the title "Languages and Machines: Computers in Translation and Linguistics" and considered the research and results conducted up to that point highly negative both on a theoretical and practical level. The ALPAC report accelerated the flight of many machine translation research pioneers to other research fields, and also resulted in a significant reduction in government agency funding. According to the commission, the automatic translators were completely unsatisfactory and expectations for the future were anything but optimistic.

[7] https://aclanthology.org/www.mt-archive.info/Bar-Hillel-1960-App3.pdf

1.4 From the mid-60s to the mid-80s: a long hibernation

Following the criticisms of Bar-Hillel and the ALPAC commission, machine translation entered a phase of hibernation, especially in the United States and the UK. This disinterest lasted twenty years, continuing throughout the first half of the 1980s. However, elsewhere, during these 20 years, research was stimulated by quite peculiar linguistic conditions. In Canada, for example, bilingualism created the need to use machine translation systems for official documents. The Montreal research group, called Traduction Automatique de l'Université de Montréal (TAUM), directed first by Alain Colmerauer and then by Charles Chandioux, designed an automatic translation system which turned out to be the first successful translator, with satisfactory results albeit with limited application: the TAUM-Météo system, later called simply Météo, and operational from 1977 to 2002. This was capable of translating weather forecasts, both nationally and for all Canadian provinces, plus times during the day. The results were appreciable thanks also to the recurring phraseology and to a very limited lexicon, typical of the meteorological language. In another multilingual context, that of the European Community, the need to translate a significant amount of parliamentary acts and other institutional documents into the many languages of the member countries led to collaboration with the American company Systran. Founded in 1968 by Peter Toma, a member of the Georgetown University research group in previous years, it was one of the first companies in the world to deal with machine translation. The collaboration between the European Community and Systran started in 1975 and ended in the early 2000s in a not entirely idyllic way, with a lawsuit for copyright infringement brought and won by the American company in 2010.

Also in this case the level of automatic translations turned out to be appreciable, and as in the case of Météo, this was largely attributable to the sectorial nature of the purely legal language of the European Commission. Despite these first two successful experiments, TAUM in Canada and Systran in Europe, it must be said that the machine translation of those years never managed to get out of the barriers of sectoral languages: the systems worked if dedicated to limited subsets of the lexicon and to a subset of - language (meteorological, legal, etc.) with a standardized phraseology.

During these twenty years of disillusionment with the United States and Great Britain, countries that had been leaders in this research sector, various research groups in various countries, as well as in Canada and Europe, had the merit of taking up the baton of research on translation automatic. In Japan, for example, various companies in those years started research on automatic translation systems between English and Japanese, and then moved on to dealing with other Asian languages such as Korean and Chinese as well. Thanks to TAUM, Systran and other players in the sector, the hopes linked to machine translation were not entirely abandoned and survived until the mid-1980s, when important technological innovations and significant theoretical reflections were about to usher in new and innovative methods.

2. Today's revolution of intelligent algorithms: Machine Learning and natural language

Starting from the mid-1980s, a series of decisive theoretical insights, the increased computing capacity of computers and the availability of digital texts changed the face of machine translation, improving its performance and filling the gaps that we have seen characterize the first systems based on rules and dictionaries. In this paragraph we will review the decisive contributions of those years and conclude with an examination and some reflections on the most innovative machine translation techniques of our day.

2.1 Machine translation based on examples

The first significant innovation can be attributed to the great Japanese computer scientist Makoto Nagao (1984). Nagao started from the consideration that rule-based systems, still the dominant approach in those years, tend to become more complex with each successive implementation of new transfer rules and with the addition of lemmas to dictionaries. These early systems had a further limitation for Nagao, that of having to carry out an analysis of the entire sentence before proceeding to translate it; any difficulty encountered in analyzing the sentences caused the system to freeze and no translation was possible. Furthermore, Nagao observed how professional translators do not carry out a preliminary analysis of the entire sentence to be translated but work starting from sentence fragments, translated separately and then recombined. He therefore thought of using the discrete availability, starting from those years, of parallel digital corpora, and of using entire segments of sentences in the machine translation process rather than conducting a word-by-word translation. Parallel corpora are bilingual (bitesti) or multilingual (multitext) texts that represent one translation of the other. Through a process called alignment in computer science, an automatic translation system learns to match each sentence of a text with a certain sentence of the other or of the other texts. Nagao therefore thought that a computer could learn and store in its memory the translation of many sentence fragments starting from the parallel corpora, and then recombine them when necessary according to the new sentences to be translated. For example, starting from the following sentences of an English/Italian bittext:

1. To complain is not always a good way of living in society
   Protestare non è sempre un buon modo *di vivere in società*
2. To complain is not always a good way of relating to others
   Protestare non è sempre un buon modo *di relazionarsi con gli altri*
3. I have always taught my students the importance of understanding each other
   Ho sempre insegnato ai miei alunni l'importanza ***di capirsi l'un l'altro***
4. There are a lot of different ways of understanding each other
   Ci sono tanti modi diversi ***di capirsi l'un l'altro,***

the automatic translation system would have learned, from the first two examples, to translate *To complain is not always a good way* with the Italian sequence ***Protestare non è sempre un buon modo*** and, from the last two examples, to translate of *understanding each other* with ***di capirsi l'un l'altro***. If the system had to translate the sentence *To complain is not always a good way of understanding each other* it would have known how to translate it through the simple combination of the translation segments stored starting from the previous sentences, automatically producing the following translation: ***Protestare non è sempre un buon modo di capirsi l'un l'altro***. It would thus have been possible to have a huge translation memory with thousands of examples or segments of sentences of a bittext to be recombined in the automatic translation phase.

2.2 Machine Learning

Nagao's intuition was followed, in the late 1980s and early 1990s, by a series of important articles by the IBM research group in Yorktown Heights, New York. These articles theorized the possibility of applying statistical techniques originally used for speech transcription to machine translation. These articles assumed that a program could be taught to statistically associate the recurring presence of a word and its translatant within parallel corpora. Starting from a word Ms, in the source text, it was possible on purely statistical bases to teach the system to associate it and translate it correctly with the word Mt in the target text. This was done by associating a probability value between 0 and 1 to each potential translating Mt in the target sentence. This value, reiterating the analyses, approximates to 1 in the case of the most probable translations and to 0 for the highly improbable ones.

The statistical methods developed by the IBM team dominated the machine translation sector at least until the early 2000s, when a new and revolutionary approach to machine translation[8], based on machine learning algorithms, began to take hold. This approach, which has evolved in the last 10 years in deep learning, originated in the second half of the 1950s from the pioneering reflections of some important artificial intelligence theorists. Computer scientist Frank Rosenblatt of Cornell University came up with the concept of the perceptron in 1957, based on previous research resulting from the collaboration between neurophysiologist Warren McCulloch and mathematician Walter Pitts. The Rosenblatt percepton is the basic unit of analysis of a neural network. The latter represents an algorithm conceived in analogy to the network of the human brain. Just as our brain is made up of communicating neurons, a neural network is the result of the coordinated activity of a certain number of perceptrons.

Rosenblatt's ideas remained confined to the theoretical dimension until a few years ago, when the increased computational capabilities of the new processors made their practical application (implementation) possible.

[8] A good introduction to machine learning for those without a math background is Machine Learning for Dummies.

The French computer scientist and linguist Thierry Poibeau summed up well the novelty represented by neural networks for machine translation:

> Deep learning achieved its first success in image recognition. Rather than using a group of predefined characteristics, deep learning generally operates from a very large set of examples (hundreds of thousands of images of faces, for example) to automatically extract the most relevant characteristics (called *feauters* in machine learning). Learning is hierarchical, since it starts with basic elements (pixels in the case of an image, characters or words in the case of a language) in order to identify more complex structures (segments or lines in an image; sequences of words or phrases in the case of a language) until it obtains an overall analysis of the object to be analyzed (a form, a sentence). (POIBEAU 2017: 183)

Learning systems based on deep learning consist of a preparatory and decisive training phase, during which they are able to learn on a statistical basis to recognize linguistic patterns starting from a large number of aligned bilingual texts. The results obtained are far superior to those obtained with the old machine translation methods and the future prospects seem very promising. It is an algorithmic method that is now fundamental to all research on artificial intelligence. The difficulties of implementing Rosenblatt's ideas, as already anticipated, were linked precisely to the training phase, in which the processor must possess a high calculation capacity to identify linguistic patterns and learn to translate correctly from the thousands of pages of analyzed translations. A phase that even with the latest processors still requires a few days of training. The training phase is followed by a testing phase, in which the automatic translator is put to the test in translating texts. The quality of the results depends on the quality of the training, and this ultimately depends on the quantity of texts on which the training is conducted. The reason why the quantity can also influence the quality of the translations is that it has been observed that translation errors, or even not very happy translations, are amply balanced by good translations provided that the analyzed corpora are of significant size. In this regard, the observation of Robert Mercer, one of the pioneers of machine translation, was prophetic, who realizing the revolutionary significance of machine learning on a statistical basis, peremptorily stated «There is no data like more data»: the best data are the data of which is available in abundance. In the following paragraph, I will propose some reflections on the future prospects of this innovative machine translation method.

2.3 Future prospects

Almost all machine translation systems today adopt the new machine learning techniques discussed above. Google, Bing, Facebook, Systran and many others are rapidly adopting the innovative methods of artificial intelligence based on deep learning. The linguistic fields in which these methods find application are many: speech recognition, simultaneous specch-to-speech translation, simultaneous text-to-speech translation and many others.

But what does the innovative scope of deep learning really consist of? First of all in its algorithmic architecture. Rosenblatt conceived of how it works in analogy with the way our nervous system processes information. Neural networks are analogous to the nervous system, at least as far as we know today about this extraordinarily complex biological machine that is our brain[9].

The interest and enthusiasm that deep learning arouses are mainly related to the learning possibilities it gives to processors. There is in this case a precise analogy with the way in which children learn. In fact, in the case of natural language for example, these learn in the very first years of life through simple exposure to the use that their parents and other adults around them make of their mother tongues. This learning takes place on a statistical and inductive basis, in the sense that children infer the meaning of words from their frequent repetition by adults and from the examples they receive. This occurs long before schooling can introduce these young speakers to the grammatical rules of their language. You learn mainly through direct experience of language. Algorithms based on deep learning allow processors to learn in the same way, starting from language examples (training) and without using grammar rules. The analogy can be pushed even further, to the extent that any processor can learn today to translate on corpora representative of one or more contemporary languages, and can be retrained in the future to record changes in the body

[9] Much remains to be discovered of this fascinating organ made up of approximately 87 billion neurons, each of which is connected on average to another 1,000 neurons, for a total of almost 100,000 billion synapses. The mapping project of all brain connections, undertaken in recent years by some pioneers of neurological research, is very interesting. A good introduction to the topic is the book *Connectome* by Sebastian Seung.

of each language[10] over the years, precisely how children in each generation will master languages different, albeit slightly, from those learned by their peers of the previous and the following generation.

This approach has many advantages over those traditionally used for machine translation until the second half of the 1980s. First of all, the old systems, mainly based on transfer rules and dictionaries, collided with the problem of having to know in depth the grammatical behavior of each pair of languages to be subjected to the translation process. For known and genetically related languages, such as those of the Indo-European family, this problem had been addressed with some success, although it required a considerable effort on the part of the linguists and computer scientists involved in writing the programs. For genealogically distant languages the difficulties of formalizing rewriting rules became much more complicated.

In addition, the dictionaries had to be continuously updated to keep up with the constant changes in the lexicons of the languages in question. Another limitation of this approach was polysemy, a particularly problematic aspect of historical-natural languages. Linguistics has identified co-textual variations, in the context of co-occurring words, as the main cause of the semantic variations of each lexical entry. This, in the case of the old rule-based and dictionary-based systems, required an enormous multiplication of the lemmas of electronic dictionaries, since each meaning of the same lexical entry required the addition of a new lemma. The dictionaries soon reached excessive dimensions which significantly slowed down the computation. On these three problematic aspects, namely genealogical distance, dictionary updating and headword multiplication, deep learning provides satisfactory and computationally advantageous answers. On the first problematic aspect, the genealogical distance of the languages, the great advantage represented by training on aligned bilingual corpora should be underlined, which can ultimately be interpreted as an automatic process of normalization of the morpho-syntactic differences between the languages being analysed. Through it the processors learn to automatically identify homologous morpho-syntactic structures for each pair of languages

[10] We know from studies of historical-comparative linguistics how languages change incessantly and how many words undergo changes in form and semantics over time. Many new words can enter the lexicon of any language through neologisms or through loanwords from other languages; other words may instead become obsolete and eventually disappear completely. The strength of every historical-natural language lies precisely in its ability to change over time and to adapt to the new expression needs of the society that adopts it.

even if genetically distant. As for updating dictionaries, it is clear how training on recent corpora provides statistical machine translation systems with a snapshot of all the most relevant lexical changes, without the need to continuously update dictionaries. The traditionally problematic third aspect, the variation of the meaning of words according to the co-text, is resolved by the ability that deep learning has to identify recurring linguistic patterns and to learn to translate them correctly. This occurs to the extent that the system has available corpora of significant size for training, a prerequisite guaranteed today by the great abundance of digital texts available on the web, at least for the most widely used languages.

This last aspect is in my opinion one of the limits and future challenges for machine translation based on deep learning. The capabilities of these systems rely solely on the abundance of bilingual or multilingual texts, and where these are lacking, the older rules-based and dictionary-based systems prove to be much more effective. It is a matter of understanding whether or not the evolution of the web will move in the direction of linguistic democratization. If there are many aligned corpora on which to train these systems, the results will certainly be satisfactory. Perhaps we will never completely replace the figure of the professional translator, but at least we will be able to provide everyone with greater ease of access to foreign languages, and industry professionals an acceptable basis on which to work in the post-editing process. always keeping in mind Robert Mercer's observation: «There is no data like more data».

Conclusions

We have seen how machine translation, even before becoming a concrete possibility through the great availability of digital texts with the advent of the web, was an aspiration with ancient roots that date back to the debated question of the existence of an original Adamic language . In the twentieth century, the problem of a machine capable of translating was inextricably linked to the idea of being able to design an intelligent machine, a computer capable of thinking. It was the English mathematician Turing who first proposed a criterion by which to measure the degree of intelligence of a computer.

The evolution of calculating machines starting from the first half of the 1950s was slow, and the evolution of the first automatic translation systems moved slowly together. Despite this, the possibility of being able to delegate to a computer the demanding task of overcoming the obstacle of language barriers aroused great enthusiasm right from the start, which was followed by twenty years, from the mid-60s to the mid-80s, of strong disillusionment, especially in the country that had inaugurated research on machine translation: the United States.

In the mid-1980s, the growing availability of digital texts, which coincided with a rapid evolution of processors, stimulated renewed theoretical interest. Between the end of the 1980s and the beginning of the 1990s, a series of seminal articles by the IBM research team in New York arose from that climate of renewed enthusiasm. In these articles a new type of machine translation was theorized no longer based on complex rewriting rules and hypertrophic electronic dictionaries, but on machine learning principles based on the identification of recurring linguistic patterns within the texts.

This led many researchers to rededicate themselves to machine translation research, an intellectual ferment in which computer scientists, statisticians, mathematicians, engineers and linguists converged. With the further and rapid evolution of computers in the 1990s it was possible to experiment with a new type of machine learning, with theoretical roots in the research conducted since the 1940s on neural networks and which culminated in the theorization of the perceptron by Frank Rosenblatt in 1957. Today, neural networks are the basis of deep learning, a method of machine learning that is providing promising results in areas as diverse as image recognition, speech recognition, machine translation and many others.

The renewed enthusiasm and important funding of governments and companies in the machine learning research sector are coupled with a very rapid evolution of the computing capabilities of computers, linked to the continuous process of miniaturization of microchips. Today it seems possible to finally realize what the pioneers of artificial intelligence research could only formulate in theoretical terms. Machines appear capable of learning a number of skills previously thought to be unique to humans. The probable advent of quantum computers in the years to come could give a further and strong impetus to research related to neural networks and learning on statistical bases, re-actualizing the ancient and never extinct Promethean desire to be able to dominate the most peculiar aspect of our species: symbolic intelligence.

Bibliography

ALPAC, 1966, *Languages and Machines: Computers in Translation and Linguistics*, Washington, National Academy of Sciences.

HILLEL B., 1949, *Report on the State of Machine Translation in the United States and Great Britain*, Gerusalemme, Hebrew University Press.

HILLEL B., 1959, *A demonstration of the non-feasibility of fully automatic, high-quality translation*, https://aclanthology.org/www.mt-archive.info/Bar-Hillel-1959-App4.pdf

MASSARON M., MUELLER J. P., 2021, *Machine Learning for Dummies*, Hoboken, JohnWiley & Sons.

POIBEAU T., 2017, *Machine Translation*, Cambridge, MIT Press.

REYNOLDS M., 2016, *Translation. A Very Short Introduction*, Oxford, Oxford University Press.

SEUNG S., 2016, *Connettoma*, Torino, Edizioni Codice.

SHANNON C., WEAVER W., 1949, *The Mathematical Theory of Communication,* Urbana, University of Illinois Press.

Alberto Maria Langella, Ph.D. in computational linguistics,
Research Fellow, University of Salerno, Italy