https://dx.doi.org/10.1590/1678-460X202238248453

Articles

A intersecção entre as Teorias linguísticas e a Linguística Computacional ao longo do tempo

The Intersection between Linguistic Theories and Computational Linguistics over time

Alexandra Moreira[1]
Alcione de Paiva Oliveira[2]
Maurílio de Araújo Possi[3]

ABSTRACT

Recent achievements have turned Computational linguistics into a dynamic research area, and an important field of application development that is being explored by leading technology companies. Despite the advances, there is still much room for improvement to allow humans to interact with computational devices, through natural language, in the same way that humans interact with native speakers of the same language. How to make computers understand or create new metaphors, metonymies or other figures of language? Is it possible to develop natural language systems capable of producing lyrics or poetry on the same level as humans? Can

1. Universidade Federal de Viçosa – UFV. Viçosa – Brasil. https://orcid.org/0000-0001-7459-1657. E-mail: xandramoreira@gmail.com.
2. Universidade Federal de Viçosa – UFV. Viçosa – Brasil. https://orcid.org/0000-0001-9129-8620. E-mail: alcione@ufv.br.
3. Universidade Federal de Viçosa – UFV. Viçosa – Brasil. https://orcid.org/0000-0001-8198-6761. E-mail: maurilio@ufv.br.

we produce systems capable of assigning new meanings to syntactic elements that are immediately perceived as coherent by humans? In this paper, we account for the evolution of computational linguistics, drawing a parallel with the evolution of linguistic theories and speculating about its research opportunities and foreseeable limitations.

Keywords: *linguistics; computational linguistics; natural language processing; evolution of computational linguistics.*

RESUMO

As conquistas recentes transformaram a Linguística Computacional em uma área de pesquisa dinâmica e em um importante campo de desenvolvimento de aplicativos que está sendo explorado pelas principais empresas de tecnologia. Apesar dos avanços, muita pesquisa ainda precisa ser realizada para permitir que os humanos possam interagir com dispositivos computacionais por meio da linguagem natural, da mesma maneira que os humanos interagem com falantes nativos da mesma língua. Como fazer com que os computadores entendam ou criem novas metáforas, metonímias ou outras figuras da linguagem? É possível desenvolver sistemas de linguagem natural capazes de produzir letras ou poesia no mesmo nível que os seres humanos? Podemos produzir sistemas capazes de atribuir novos significados a elementos sintáticos que são imediatamente percebidos como coerentes pelos seres humanos? Neste artigo, descrevemos a evolução da linguística computacional, traçando um paralelo com a evolução das teorias linguísticas e especulando sobre as oportunidades de pesquisa da área e suas limitações.

Palavras-chave: *linguística; linguística computacional; processamento de linguagem natural; evolução da linguística computacional.*

1. Introduction

Human language is a phenomenon that fascinates everyone who is able to perceive its versatility, flexibility, and expressive power. Anthropologists seek to understand how it evolved and what events made it possible. Writers, poets, and musicians increasingly expand the possibilities of natural language. Neuroscientists and linguists seek to discover how language-related cognitive processes take place. Computer scientists seek to establish formalism and techniques that

may allow computational devices to understand natural language for extracting information and create associated applications. In the latter case, the search for the development of computational systems capable to communicate and process natural language appeared nearly simultaneously with the emergence of the digital electronic computer. Motivated by the Cold War and the need to have access to sensitive content communicated in the languages of other nations, researchers from the 1940s and 1950s strove to develop automatic translation systems (Jones, 1994). Another great motivation was the need to create computational devices that could interact with humans through natural language. This form of communication would facilitate human-machine interaction and allow for the development of useful systems. There is also a motivation for the development of intelligent systems and, as the language used by humans is distinct from the forms of communication used by other animals, it helps to support the belief that intelligence is closely linked to human language.

Along the journey to achieve these goals, other goals were added and the area experienced periods of euphoria alternating with moments of discouragement. Over the past two decades, computational linguistics gained greater strength because of changes in hardware architecture, the availability of huge datasets, and changes in the theoretical approach to face the challenges of the area. The changes in the theoretical approach made it possible to establish a parallel with the alternations of the dominant theories in linguistics, from generative transformational grammar to cognitive linguistics. This parallelism in the evolution of computational linguistics and linguistics was due to mutual influences and natural collaboration between neighboring areas. Therefore, when one gives an account of the evolution of computational linguistics, one should point out the mutual contributions and influences of related areas.

In this paper, we aim to give an account of how computational linguistics has evolved, from its inception to the present day, seeking to establish a framework for the mutual contributions of related areas. We also try to identify the main challenges the area needs to overcome to continue evolving. Our approach differs from most works that describe the evolution of the area, as instead of dividing the stages of evolution into decades (Jones, 1994; Bates,1995) we show the evolution

of computational linguistics through the changes in the dominant paradigm, from formal systems to the connectionist paradigm, passing through the probabilistic paradigm.

In the next section we describe the emergence of computational linguistics and its first phase. Section 2 presents an account of its probabilistic phase. Section 3 presents the techniques used today. Finally, in section 4, we discuss the challenges that have to be faced for the area to evolve.

2. The Era of Formal Systems

The birth of the computational linguistics occurred not long after the emergence of electronic computers. The first general purpose electronic computing device built was a computer called Z3, developed by the German engineer Konrad Zuse, in 1941 (Rojas, 1997). Another important event identified with the rise of electronic computing was the construction of ENIAC (Electronic Numerical Integrator and Calculator), developed by J. Presper Eckert and John W. Mauchly, at the University of Pennsylvania in 1946 (Mauchly, 1980).

Three years later, on July 15th, 1949, Warren Weaver, who at the time was Director of the Rockefeller Foundation's Natural Sciences Division, issued a memo speculating on the possibility of using the newly invented digital computers to translate documents from one natural language to another (Weaver, 1955). One can establish the year of the writing of the memo as the year of the emergence of computational linguistics, because as stated by Hutchins:

> ... is perhaps the single most influential publication in the earliest days of machine translation. Written before most people had any idea of what computers might be capable of, it was the direct stimulus for the beginnings of research in the United States. (HUTCHINS, 2000, p.17)

Weaver's background in cryptography reinforced his belief that machine translation was just a special case of deciphering. In addition, Weaver assumed that machine translation could be accomplished through a formal system:

> A more general basis for hoping that a computer could be designed which would cope with a useful part of the problem of translation is to be found in a theorem which was proved in 1943 by McCulloch and Pitts. This theorem states that a robot (or a computer) constructed with regenerative loops of a certain formal character is capable of deducing any legitimate conclusion from a finite set of premises. (WEAVER, 1955, p. 9)

At the time, the formal systems approach was the main approach adopted to try to solve the machine translation problem as well as other related problems that have emerged in the context of natural language processing. The area gained momentum in the following decade and attracted several enthusiastic researchers. The first machine translation conference took place in 1952, at the MIT. In 1953, the University of Harvard began to research on machine translation. In 1954 a primitive English-Russian translation system was developed by IBM-Georgetown (Jones, 1994; Hutchins, 2000). In 1956 the second conference on machine translation took place. During the summer workshop at Dartmouth College, in 1956, the term "Artificial Intelligence" was coined and a research area that would cover all studies related to the development of intelligent systems – including systems that were able to deal with natural language – was created; computational linguistics. Such assimilation arose from the idea that the ability of computers to understand natural language was related to the ability of computers to exhibit intelligent behavior.

Perhaps the association of intelligence with the ability to understand natural language stems from Safir-Whorf's hypotheses (Whorf et al., 1956), which states that the structure of a language affects the worldview or cognition of its speakers and, therefore, people's perceptions are related to their spoken language. A reinforcement to such an association came in 1950, when Alan Turing (Turing, 1950) designed an experiment to verify whether the intelligence of a device was the same as that of a human, known as the "Turing test". The test was based on the capacity of a computational device to communicate using natural language without its interlocutor being able to tell whether he/she was talking to a human being or to a machine.

In addition to machine translation, computational linguistics started to look into other linguistic tasks as challenges to be overcome.

According to Jones (1994), at the 1958 International Conference on Scientific Information, held in Washington, natural language processing was associated with information retrieval. Peter Luhn (1958), for example, was able to generate automatically, by extracting segments of texts, abstracts for articles from one of the event sessions.

The techniques used to make computers capable of dealing with natural language had the same theoretical basis as the linguistic theories of the time. In the field of Cognitive Linguistics, the focus was on syntax. According to Jones (1994), this was due to the idea that the whole process was driven by syntax. Likewise, in linguistics, the predominant theory emphasized syntax over semantics. The behaviorist theory of language was being abandoned and the era when Chomsky's ideas were predominant had begun.

In 1957, Noam Chomsky (1957) laid the foundations for his linguistic theory (generative transformational grammar) and defined the independence of syntax and semantics. Chomsky's theory defined that a set of transformational rules was able to generate all the sentences in a language. Such process would happen in the mind and, thus, eliminate the behaviorist assumptions related to language use. This was the starting point of cognitive linguistics. This period in the history of linguistics was also known as the period of the rationalist approach, which postulates that a significant part of language knowledge is genetically inherited and not captured by the senses (Manning & Schütze, 1999, p.4).

Over time, applications for question-and-answer systems have become more important (Bates, 1995). To achieve this goal, the focus was more on semantics than syntax. The approach started to be strongly based on knowledge and several knowledge representation techniques were developed. At the same time, machine translation was losing space. According to Jones (1994), such a loss of space was mainly caused by the November 1966 report of the Automatic Language Processing Advisory Committee (ALPAC), sponsored by the US National Academy of Sciences (Hutchins, 2003). The report made a critical analysis of the results obtained so far in machine translation and in the final chapter (ALPAC, 1966, p. 32), stated that "...we have no useful automatic translation. Further, there is no immediate or

predictable prospect of useful machine translation". The effect of the report was the end of public funding for machine translation research in the United States for about twenty years.

As already mentioned, the move to a more semantic direction and to question-answer systems resulted in the development of several knowledge representation techniques. One of these representations was the semantic networks proposed by Ross Quillan, in 1966 (Quillian, 1966). This representation was intended to model the structure of human knowledge. In 1972, Schank proposed a representation model for natural language called conceptual dependency (Schank, 1972). It consisted of conceptual primitives that sought to capture the meaning of most events. Going a little further in the structuring of knowledge proposed by Quillian, Marvin Minsky (1974) proposed the model of representation called *frames* that attempted to represent stereotypical facts. The scripts proposed by Schank and Abelson (1975) represented a stereotyped scene, the classic example being the restaurant's script. In fact, the script is a specialization of Minsky's frames. While frames are general purpose structures for representing common clusters of facts, scripts are structures with the capacity to explore specific properties of a particular domain. One can say that the scripts are a combination of frames and conceptual dependency.

This period where the models had a "more semantic" character is being described together with the period which focused on formal systems, because we believe that human knowledge representations can be reduced to formal systems, despite the change in the level of representation. Russell and Norvig (2016) citation of Roger Schank is in line with our belief: "There is no such thing as syntax". In a way, such view is also shared by cognitive linguistics, which argues that the separation between syntax and semantics is arbitrary.

Still within this formal system approach, it is worth mentioning the launch, in 1978, of the ambitious project of the fifth-generation computer proposed by Japan (Moto-Oka & Stone, 1984; McCorduck, 1983; Warren, 1982). This project aimed to develop a highly parallel computer, capable of making "one gigalips", meaning one billion logical inferences per second (McCorduck, 1983). The basis of the project was logic programming, proposed by Robert Kowalski, in the

form of Horn clauses, implemented in the Prolog language. The Prolog language was widely used at the time for the development of expert systems and for the processing of natural language. The hope was that the construction of hardware that would support inferences in Prolog would allow for a greater advance in the development of systems written in that language. However, in 1992, the project – which had already cost over U$400 million dollars – was closed because it did not achieve its objectives (Pollack, 1992).

Although some researchers in computational linguistics believed that the field of linguistics would have nothing to contribute to the area (Jones, 1994: 4), linguistics itself was undergoing changes that were similar to those in computational linguistics. Such changes were driven by a crisis known as the Linguistics War. The Linguistics War was triggered by the emergence of the generative semantics theory, proposed by George Lakoff (1963), who shifted the focus of linguistics from syntax to semantics. This theory launched a series of attacks mainly between Chomsky and Lakoff (Harris, 1995), who defended their respective theories. Generative semantics ended up influencing the development of a whole set of theories in the field of cognitive linguistics, proposed by George Lakoff and Ronald Langacker. Unlike generative theory, which did not take semantics into account, cognitive linguistics had, as its main characteristic, an emphasis on semantics and meaning. Among the linguistic theories that emerged under cognitive linguistics, it is worth highlighting the semantic frames proposed by Charles Fillmore (1976). Despite its name, semantic frames are more similar to Schank's scripts than to Minsky's frames. The purpose of semantic frames is to define a conceptual construct where knowledge of all the elements involved, as well as the relationships among them, is necessary to understand the concept emerging from the scene being analyzed. Thus, we can see that the shift in focus in computational linguistics corresponded to the shift in focus in linguistics.

The slow advances seen in the development of useful natural language processing applications – that were restricted to specific niches and the development of graphical interfaces, which allowed an easy interaction with computational devices – caused a decrease in the pace of research in natural language processing. Thus, other events were needed to trigger a paradigm shift in computational linguistics.

3. The probabilistic phase

We can speculate that several events provided the conditions for a paradigm shift in computational linguistics. We do not intend to list them all, but some deserve a mention. The first event was the popularization of the Internet and the emergence of the Web. These worldwide communication platforms allowed for the creation of enormous textual bases (corpora). In 1990, the British National Corpus (BNC), composed of 100 million words, was launched (Davies, 2009). By 2015, according to the New York Times (Heyman, 2015), Google Books had already scanned 25 million books, including texts in 400 different languages from more than 100 countries. A text published on October 17, 2019, on the Google blog[4], claimed that by then the company had already scanned over 40 million books. This gigantic corpus was used to create another valuable textual dataset, the Google Books NGram (Michel et al., 2011). According to Bohannon (2011), this mega corpus, at the time, comprised books published between 1550 to 2008 and more than 500 billion words in French, Spanish, German, Chinese, Russian, Hebrew, and English. The latter language alone accounts for 361 billion words. Another mega corpus is Wikipedia. According to Wikipedia[5], on February 3, 2020, there were 6,008,537 articles in English, totaling over 3.5 billion words. Investigating corpora that comprise such a large number of words call for the application of statistical and stochastic methods. However, before these methods become the most popular ones other factors need to come into play.

Another important event that is present in almost every paradigm shift in computing, is the advancement in hardware performance. According to Koomey et al. (2011), computing power in the 1980s and 1990s reached between 107 to 108 computations per second. Also, at that time, new parallel hardware architectures were being introduced, which allowed the use of techniques that made intensive use of computational power.

A further motivator was the interest of large technology and electronic commerce companies in mining the information that was

4. https://www.blog.google/products/search/15-years-google-books/
5. https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

being exchanged on the Web, to discover user trends and preferences. Such interest led to greater investments in applications involving natural language processing. An example of a popular application that uses natural language processing is the Google translate. The first version of Google Translate, launched in 2006, embarked on the wave of probabilistic translation, with conditional probability extracted from a massive corpus of translations. The translations were done using a pivot language, which in most cases was English, as most of the translations are done into English. Thus, every translation from an L1 to an L2 was first translated into the pivot language (Bellos, 2011). When there was not a significant amount of language translations into English, another language was used as the pivot language. In the translation process, the apprehended probability distribution was used to determine the most likely translation. This translation method was replaced in 2016, when Google adopted translation based on deep neural networks.

These events – and others not mentioned here – helped to change the approach used to overcome the challenges in computational linguistics and allowed for the re-emergence of Natural Language Processing (NLP). Statistical and stochastic methods became state-of-the-art. Probabilistic language models based on n-grams were used to predict the next nth word from a sequence of n-1 words (Jurafsky & Martin, 2008). Another important model introduced in the area at that time were the hidden Markov models (HMM) (Baum & Petrie, 1966), which were trained to discover a hidden sequence of symbols given a known sequence of other symbols. A good example of the application of these models in NLP is text annotation. For example, when given a sentence (sequence of words), the annotation tool provides the corresponding sequence of grammatical classes (part-of-speech). The hidden Markov models can also be used in other types of annotations, such as the annotations of semantic roles or named entities. To train these models, a variation of dynamic programming algorithm, called Viterbi (Viterbi, 1967), is used.

Bayesian and HMM probabilistic models try to define the joint probability distribution p(x, y) of an entire sample space. Once they learn the distribution, they can generate new instances, and that is why they are called generative models. However, generative models need to learn all the joint probabilities, including characteristics that

are not relevant to a given event, so they may have greater difficulty in establishing the limits for classification in a dataset. On the other hand, discriminative models, such as logistic regression and Markov maximum entropy models, which directly calculate the conditional probability p(y|x), have the potential to obtain greater precision in the classification tasks. Both types of models became quite popular in the 2000s and are used in several types of NLP tasks, such as voice recognition (Levinson et al., 1986), sentiment analysis (Kang et al., 2012), syntactic and semantic annotation (Kupiec, 1992; Thompson et al., 2003), and named entities (Morwal et al., 2012).

Meanwhile, in the field of linguistics, Chomsky's generative program continued to suffer attacks from advocates of cognitive linguistics. The failure of symbolic theory to describe important phenomena in natural language, such as figures of speech (metaphors and metonymy), and to provide adequate support for the semantic and pragmatic aspects of language, shifted the focus of linguistic research to cognitive linguistics. Within this context, the idea that natural language can be modeled by statistical and stochastic models is aligned with the propositions of cognitive linguistics. According to Griffiths (2011), research in linguistics and cognitive linguistics shows that probabilistic and statistical theory partially explains how people produce and interpret sentences. The idea that probability and statistics can help us understand human language is not new. Zipf (1936) had already observed that the length of words is inversely proportional to their use. Wittgenstein (1953) had already pointed out that it would not be possible to describe language through formal systems, as the meaning of utterances would be established dynamically by the interaction between speakers, that is to say, the most likely meaning is context dependent. Despite such observations, the statistical approach was relegated to the background during the period dominated by the generative approach.

3. The Age of Complex Neural Networks

Despite the advances made as from the 1980s, some obstacles prevented further advances and the widespread adoption of natural language systems. Statistical language models performed much better

than the systems used in the symbolic era. However, they were still well below human performance and required a laborious and costly stage of feature engineering. The state-of-the-art represented by discriminative models, such as maximum entropy models and conditional random fields (Lafferty et al., 2001), took a long time to be trained, even with medium-sized corpora. To solve the problem of the feature engineering stage, others discriminative models, such as neural networks that can learn hidden structures in the training examples, were adopted. Nonetheless, until 2010, the use of neural networks on a complex data set was impracticable, as the number of layers and the number of neurons in each layer exponentially increased the number of parameters to be adjusted. In addition, an increase in the number of layers caused a vanishing gradient problem, because the information passed between the layers resulted from the multiplication of small value numbers.

One of the ways found to overcome problems like those described above was the popularization of convolutional networks. In the field of artificial intelligence, convolutional networks first appeared in the work of LeCun et al. (1998). However, this technique did not become popular until the work of Krizhevsky, Sutskever and Hinton (2017). Convolutional networks are used to circumvent the limitations of previous neural networks by adopting techniques such as local receptive fields, weight sharing, convolution operation and subsampling. These techniques were inspired by neuroscience findings on the functioning of the human brain visual cortex (Kuzovkin et al., 2018). Initially, they were created to classify images and used in many classification tasks in NLP (Kim, 2014). Convolutional networks showed an excellent performance for classification tasks. However, they proved to be inadequate for capturing long dependencies, as they were created to capture local dependencies. On the other hand, the understanding of natural language involves the recognition of dependencies between elements that occur distant from each other, either in a written text or temporarily in an oral statement. Therefore, several NLP tasks require that these dependencies be modeled, such as machine translation, coreference resolution and semantic annotation. For these situations Hochreiter and Schmidhuber (1997) proposed a variation of the recurring networks called long short-term memory (LSTM). These networks were able, like traditional recurring networks, to receive

sentences of arbitrary size as input, but they were also able to retain information about distant elements in the sequence.

In natural language processing, the success of architectures based on deep neural networks was reflected in the change in techniques used by technology giants to develop machine translation systems. In 2016, Microsoft and Google announced the adoption of neural networks in machine translation applications (Deng & Liu, 2018). Another important competitor that has recently emerged (2017) in machine translation is the DeepL Translator (DeepL) (www.deepl.com). Since its first version, the translator is based on deep neural networks and has surpassed its competitors in some published benchmarks (Macketanz et al., 2020; Tavosanis, 2019).

Much of the success of neural networks in NLP resulted from the adoption of semantic representation through dense vectors, also called Embeddings. In this type of representation, the element of natural language that one wants to represent (letters, word segments, words, sentences, etc.) in the vector semantics is replaced by large numerical dense vectors of size N. Typically, N varies from 100 to 1000, defining the position of this element in an N-dimensional space relative to the elements that occur in its vicinity. This idea was inspired by distributional linguistics and by Firth's (1957:11) famous statement "You shall know a word by the company it keeps". The vectors are generated through neural networks trained in large corporations to capture part of the contextual information. Several word embeddings techniques have been developed to date, capturing different levels of contextual information.

Mikolov et al. (2013) seminal work introduced the representation called Word2Vec, which helped to drive machine translation forward. Since then, several representations based on the initial ideas of Word2Vec – but with interesting advances – were developed. Pennington, Socher, and Manning (2014), proposed the GloVe representation (Global Vectors), which incorporated statistical information into the representation. Bojanowski et al. (2017) introduced the FastText representation, which aimed to circumvent the problem of dealing with unknown words through the vector representation of word segments. These and other forms of representation capture

information about elements of language within a given context. However, information about a particular element varies depending on the context in which it occurs. Peters et al. (2018) proposed the ELMo (Embeddings from Language Models) technique, which consists of a bi-directional LSTM network that computes a contextualized representation of words. Following the same line, other representations were proposed, e.g., BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019), GPT and GPT-2 (Generative Pre-Training) (Radford et al., 2019), and Transformer XL (meaning extra-long) (Dai et al., 2019).

Simultaneously, the architectures of neural networks that presented themselves as the state-of-the-art for NLP were being replaced by others that produced results that surpassed the previous ones. LSTM networks and their variations had some limitations, such as being slower to be trained than non-recurring networks due to the difficulty of parallelization. Also, the difficulty to retain information from distant relations remained, that is, in these networks greater weight is given to more recent relationships. The Transformer architecture (Vaswani et al., 2017), circumvented these difficulties, exploring the self-attention mechanism to focus on the relevant dependencies, and eliminated the use of recurring networks, making it easier to parallelize. Transformer established new performance standards to be beaten, raising the level of performance of natural language processing systems.

All these architecture evolutions were supported by the emergence of new hardware architectures that were derived from the evolution of the computational power of GPUs (Graphic Processing Units). Such GPUs are quite adequate for neural networks to execute their operations in massively parallel mode.

This phase of computational linguistics was, to a certain extent, in line with the evolution of cognitive linguistics theories. Neuroscience's findings impacted both areas of research. If, on the computational linguistics side, the discoveries inspired the elaboration of new architectures for artificial neural networks – as is the case with convolutional networks and attention mechanisms – on the cognitive linguistics side they allowed for a glimpse of how the human mind stores concepts and serializes, to transmit such concepts in the form of human language. Jerome Feldman (2008), developed a theory of how

the elements of human language, such as concepts and metaphors, are represented in the human brain in structured neural clusters. Talmy (2007) discusses the attentional system of language, postulating that the listener of an utterance does not focus his attention uniformly on all the elements of an utterance.

4. Conclusions

In this paper, we tried to show how the area of computational linguistics has evolved, from its emergence to the present day. We showed how the theories that support computer systems have been changing radically, in some cases. We tried to show that the changes in computational linguistics occur hand-in-hand with the changes in linguistics, as they mutually influence each other. We also tried to show that the different phases of both areas were correlated, starting with the formal systems phase, followed by the probabilistic phase, and ending with the complex neural networks phase. We speculate, with a fair amount of certainty, that computational linguistics and linguistics will continue to evolve in a parallel fashion and maintain mutual collaboration, as findings in one area provides insights into the other.

Recent advances in computational linguistics have been remarkable. Nowadays, it is common for us to interact with several computational devices using natural language. We are experiencing the excitement of the emergence of virtual assistants, with a variety of devices available for purchase in stores. Among the most popular virtual assistants are the Google assistant, Amazon Alexa, Microsoft's Cortana and Apple Siri (Tulshan & Dhage, 2018; Kepuska & Bohouta, 2018). These devices allow the user to control home appliances, shop online, select songs to listen to, listen to the news, consult the weather forecast or simply chat, through natural language communication. The success of virtual assistants can be measured by their sales data. Amazon© alone sold over 100 million Alexa devices as of January 2019 (Bohn, 2019). Virtual assistants are the most obvious examples of practical applications of research results in natural language processing. However, several other applications have benefited from recent research results. Several companies are replacing humans in customer service with software bots. Executives and politicians are making decisions based on the analysis

of texts produced by people on social networks. Texts are being written by automatic text generators (Lee & Hsiang, 2019).

In addition to such practical applications, computer systems are constantly capturing information expressed in natural language, to extract useful knowledge for decision-making. These results do not imply that there are no challenges to be overcome in computational linguistics. On the contrary, there are several challenges to be faced. Among the challenges to be met in the coming years we can mention the following: How to make computers understand or create new metaphors, metonymies, or other figures of language? Is it possible to develop natural language systems capable of producing lyrics or poetry as well as humans? Can we produce systems capable of assigning new meanings to syntactic elements that are immediately perceived as coherent by humans?

To be able to meet these challenges, it is necessary to create more energy efficient models. The great advance in the performance and precision of computational linguistics models using deep learning had its price: the ever-increasing demand for powerful computing resources, which requires large amounts of financial and energetic resources. The models with the best performance are those that demand greater computational power. Research in the field of computational linguistics is often limited by access to sufficient computational power to carry out experiments on new models and to compare them with current state-of-the-art models, because such experiments tend to take months (Ruder et al., 2019).

The current complexity of state-of-the-art models aimed at processing natural language was certainly the key to achieving the performance they demonstrate. However, the cost of such an advance has had a negative impact on its interpretability. According to Tenney et al. (2019), it is not easy to say whether a current model is really learning abstractions that represent natural language concepts or simply modeling complex statistical co-occurrences. This concern with the opacity of current models creates the need to develop new models that allow the extraction of information from their internal representations, to show that their abstractions are really representing language in a satisfactory way (Dalvi et al., 2019). Perhaps this limitation can be overcome by neuronal-symbolic integration, in the way that human

beings do, to facilitate the extraction of knowledge from neural representations.

Regarding conceivable applications, several areas can benefit from future results from research in computational linguistics, such as the area of law or medicine. In law, NLP applications would support the creation of associated products to help lawyers be more effective when researching processes, drafting petitions, and reconciliations. In the same line, one can imagine that natural language systems can help in the diagnosis of diseases, researching medical texts and dialoguing with patients. The potential of computational linguistics is huge and listing the benefits it can bring to several fields of knowledge could be the topic of a full paper. Nonetheless, it is important to keep in mind that there are also some risks related to the widespread use of this technology. One of these risks is the possibility of using text generation to produce fake statements that can be used to manipulate people.

Finally, since human language is the main form of communication and transmission of knowledge, it will continue to be a very active and challenging multidisciplinary research area, absorbing contributions from different areas of knowledge and leveraging society's progress.

Acknowledgements

This work was carried out with the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Financing Code 001 and the Fundação de Amparo à Pesquisa do Estado de Minas Gerais - Brazil (FAPEMIG).

Conflicts of Interest Statement

The authors declare they have no conflict of interest.

Credit Author Statement

We, Alexandra Moreira, Alcione de Paiva Oliveira e Maurílio de Araújo Possi, hereby declare, for all due purposes, that we all participated in the study conceptualization, methodology, design, and article writing as well

as the final edition of the article submitted and approved for publication in Delta Magazine. As such, we approve of the final version of the manuscript and are responsible for all aspects of it.

References

Bates, M. (1995). Models of natural language understanding. In *Proceedings of the National Academy of Sciences*, *92*(22), 9977-9982. http://doi.org/10.1073/pnas.92.22.9977.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, v. 37, n. 6, p. 1554-1563. http://doi.org/10.1214/aoms/1177699147.

Bellos, D. (2011). *Is that a fish in your ear?: Translation and the meaning of everything*. Penguin Books.

Bohannon, J. (2011, January 14). Google Books, Wikipedia, and the future of culturomics. *Science*, *331*(6014), 135. http://doi.org/10.1126/science.331.6014.135.

Bohn, D. (2019, January, 4). Amazon says 100 million Alexa devices have been sold. *The Verge*. https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp (accessed November 22, 2021).

Bojanowski, P., Grave, E., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135-146. https://doi.org/10.1162/tacl_a_00051.

Chomsky, N. (1957). *Syntactic structures*. Paris.

Dai, Z. et al. (2019). Transformer-xl: Attentive language models beyond a fixed-length context. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 2978-2988). http://doi.org/10.18653/v1/P19-1285.

Dalvi, F. et al. (2019). What is one grain of sand in the desert? analyzing individual neurons in deep NLP models. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence: AAAI 19*, *31*(01), 6309-6317. https://ojs.aaai.org/index.php/AAAI/issue/view/246 (accessed November 22, 2021).

Davies, M. (2009). The 385+ million-word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159-190. https://doi.org/10.1075/ijcl.14.2.02dav.

Deng, L., & Liu, Y. (Ed.). (2018). *Deep learning in natural language processing*. Springer.

Devlin, J. et al. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1,* 4171-4186. http://doi.org/10.18653/v1/N19-1423.

Feldman, J. (2008). *From molecule to metaphor: A neural theory of language*. MIT press.

Fillmore, C. J. (1976). Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, *280*, 20-32. https://www1.icsi.berkeley.edu/pubs/ai/framesemantics76.pdf (accessed November 22, 2021).

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. In J. R. Firth et al. (Eds.), *Studies in Linguistic Analysis* (pp. 1-32). Blackwell.

Griffiths, T. L. (2011). Rethinking language: How probabilities shape the words we use. In *Proceedings of the National Academy of Sciences*, *108*(10), 3825-3826. https://doi.org/10.1073/iti1011108.

Harris, R. A. (1995). *The linguistics wars*. Oxford University Press on Demand.

Heyman S. (2015). Google books: A complex and controversial experiment. *The New York Times*. https://www.nytimes.com/2015/10/29/arts/international/google-books-a-complex-and-controversial-experiment.html (accessed November 22, 2021).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hutchins, W. J. (Ed.). (2000). *Early years in machine translation: Memoirs and biographies of pioneers*. John Benjamins Publishing.

Hutchins, J. (2003). ALPAC: the (in)famous report. In S. Nirenburg, H. L. Somers, & Y. A. Wilks (Eds.), *Readings in machine translation* (pp. 131-135). MIT Press. 1 https://doi.org/0.7551/mitpress/5779.001.0001.

Jones, K. S. (1994). Natural language processing: a historical review. In A. Zampolli, N. Calzolari, & M. Palmer (Eds.), *Current issues in computational linguistics: In honour of Don Walker* (pp. 3-16). Springer, Dordrecht.

Jurafsky, D. M., & James, H. (2008). *Speech and language processing: An introduction to speech recognition, computational linguistics, and natural language processing*. Prentice Hall.

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, *39* (5), 6000-6010. https://doi.org/10.1016/j.eswa.2011.11.107.

Kepuska, V., & Bohouta, G. (2018). Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In *8th Annual Computing and Communication Workshop and Conference* (pp. 99-103). IEEE. https://doi.org/10.1109/CCWC.2018.8301638.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1746-1751). https://doi.org/10.3115/v1/D14-1181.

Koomey, J. et al. (2010). Implications of historical trends in the electrical efficiency of computing. In *IEEE Annals of the History of Computing*, *33*(3), 46-54.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. In Communications of the Association of Computing Machinery, *60*(6), 84-90. https://doi.org/10.1145/3065386.

Kupiec, J. (1992). Robust part-of-speech tagging using a hidden Markov model. *Computer Speech & Language*, *6*(3), 225-242. https://doi.org/10.1016/0885-2308(92)90019-Z.

Kuzovkin, I. et al. (2018). Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. *Communications Biology*, *1*, 107. https://doi.org/10.1038/s42003-018-0110-y.

Lafferty, J., Mccallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In C. E. Brodley, & A. P. Danyluk (Eds.), *ICML '01: Proceedings of the Eighteenth International Conference on Machine Learnin*g (pp. 282-289). Morgan Kaufmann Publishers Inc. https://dl.acm.org/doi/proceedings/10.5555/645530 (accessed November 22, 2021).

Lakoff, G. (1963). *Toward generative semantics*. Technical report. UC Berkeley. https://escholarship.org/uc/item/64m2z2b1 (accessed November 22, 2021).

Lecun, Y. et al. (1998). Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, *86*(11), 2278-2324. https://doi.org/10.1109/5.726791.

Lee, J.-S., & Hsiang, J. (2020). Patent Claim Generation by Fine-Tuning OpenAI GPT-2. *World Patent Information*, *62*, September 2020,

101983. https://arxiv.org/pdf/1907.02052.pdf (accessed November 22, 2021).

Levinson, S. E., Rabiner, L. R., & Sondhi, M. M. (1986). *Hidden Markov model speech recognition arrangement* (U.S. Patent n. 4, 587, 670). U.S. Patent and Trademark Office. https://patentimages.storage.googleapis.com/84/c0/08/af2eacbc2df545/US4587670.pdf (accessed November 22, 2021).

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, *2*(2), 159-165. https://doi.org/10.1147/rd.22.0159.

Macketanz, V., Burchardt, A., Uszkoreit, H. (2020). *Tq-autotest: Novel Analytical Quality Measure Confirms That Deepl Is Better Than Google Translate.* Technical report. The Globalization and Localization Association. https://www.dfki.de/fileadmin/user_upload/import/10174_TQ-AutoTest_Novel_analytical_quality_measure_confirms_that_DeepL_is_better_than_Google_Translate.pdf (accessed November 22, 2021).

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

Mauchly, J. W. (1980). The ENIAC. In N. Metropolis, J. Howlett, & G.-C. Rota (Eds.), *A History of Computing in the Twentieth Century* (pp. 541-550). Academic Press.

Mccorduck, P. (1983). Introduction to the fifth generation. *Communications of the ACM*, *26*(9)*,* 629-630. https://doi.org/10.1145/358172.358177.

Michel, J.-B. et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176-182. https://doi.org/10.1126/science.1199644.

Mikolov, T. et al. (2013). Efficient estimation of word representations in vector space. In *International Conference on Learning Representations*, Arizona. https://www.arxiv-vanity.com/papers/1301.3781/ (accessed November 22, 2021).

Minsky, M. (1974). *A Framework for Representing Knowledge*. Technical report. Massachusetts Institute of Technology. https://doi.org/10.1016/B978-1-4832-1446-7.50018-2.

Morwal, S., Jahan, N., & Chopra, D. (2012). Named entity recognition using hidden Markov model (HMM). *International Journal on Natural Language Computing 1*(4)*,* 15-23. https://doi.org/10.5121/ijnlc.2012.1402.

Moto-oka, T., Stone, H. S. (1984). Fifth-generation computer systems: A Japanese project. *Computer*, *3*, 6-13. https://doi.org/10.1109/MC.1984.1659076.

National Research Council (US). (1966). Automatic Language Processing Advisory Committee (ALPAC). *Language and machines: computers in translation and linguistics: A report*. National Academies.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1532-1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1.

Peters, M. E. et al. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 2227-2237. https://doi.org/10.18653/v1/N18-1202.

Pollack A. (1992). 'Fifth Generation' Became Japan's Lost Generation. *The New York Times*. https://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html (accessed November 22, 2021).

Quillan, M. R. (1966). *Semantic memory*. Bolt, Beranak & Newman.

Radford, A. et al. (2019). Language models are unsupervised multitask learners. OpenAI Blog, *1*(8), 9. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (accessed November 22, 2021).

Rojas, R. (1997). Konrad Zuse's legacy: the architecture of the Z1 and Z3. In *IEEE Annals of the History of Computing*, *19*(2), 5-16. https://doi.org/10.1109/85.586067.

Ruder, S. et al. (2019). Transfer learning in natural language processing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials* (pp. 15-18). https://aclanthology.org/N19-5004.pdf (accessed November 22, 2021).

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. 3rd edition. Pearson Education Limited.

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, *3*(4), 552-631. https://doi.org/10.1016/0010-0285(72)90022-9.

Schank, R. C., & Abelson, R. P. (1975). Scripts, plans, and knowledge. In *Proceedings of the Fourth* International Joint Conference on Artificial Intelligence, 151-157. https://www.ijcai.org/Proceedings/75/Papers/021.pdf (accessed November 22, 2021).

Talmy, L. (2007). Attention phenomena. In D. Geeraerts, & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics*. Oxford university Press. https://doi.org/10.1093/oxfordhb/9780199738632.001.0001.

Tavosanis, M. (2019). Valutazione umana di Google Traduttore e DeepL per le traduzioni di testi giornalistici dall'inglese verso l'italiano. In R. Bernardi, R. Navigli, & G. Semeraro (Eds.), *CLiC-it 2019 – Proceedings of the Sixth Italian Conference on Computational Linguistics, Machine Translation*, *2481*, 494-525. http://ceur-ws.org/Vol-2481/paper70.pdf (accessed November 22, 2021).

Tenney, I, Das, D., & Pavlick, E. (2019). Bert rediscovers the classical NLP pipeline. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4593-4601). Association for Computational Linguistics. https://arxiv.org/pdf/1905.05950.pdf (accessed November 22, 2021).

Thompson, C. A., Levy, R., & Manning, C. D. (2003). A generative model for semantic role labeling. In N. Lavrač et al. (Eds.) *Machine Learning: 14th European Conference on Machine Learning Cavtat-Dubrovnik, Croatia* (pp. 397-408). Springer-Verlag. https://link.springer.com/book/10.1007%2Fb13633 (accessed November 22, 2021).

Tulshan, A. S., & Dhage, S. N. (2018). Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa. In S. M. Thampi et al. (Eds.), *International Symposium on Signal Processing and Intelligent Recognition Systems: 4th International Symposium SIRS 2018, Bangalore, India* (p.p. 190-201). Springer. https://link.springer.com/book/10.1007/978-981-13-5758-9#toc (accessed November 22, 2021).

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *lix*(236), 433-464. https://doi.org/10.1093/mind/LIX.236.433.

Vaswani, A. et al. (2017). Attention is all you need. In U. von Luxburg, & I. Guyon (Eds.), *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 6000-6010). Curran Associates Inc. https://dl.acm.org/doi/10.5555/3295222.3295349 (accessed November 22, 2021).

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13(2),* 260-269. https://doi.org/10.1109/TIT.1967.1054010.

Warren, D.H.D. (1982). A view of the Fifth Generation and its impact. *AI Magazine*, *3*(4), 34-34. https://doi.org/10.1609/aimag.v3i4.380.

Weaver, W. (1955). Translation. In W. N. Locke, & A. D. Booth (Eds.), *Machine Translation of Languages: Fourteen Essays* (pp. 15-23). MIT Press. https://repositorio.ul.pt/bitstream/10451/10945/2/ulfl155512_tm_2.pdf (accessed November 22, 2021).

Whorf, B. L., Carroll, J. B., & Chase, S. (Eds.) (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MIT press.
Wittgenstein, L. (1953). *Philosophical investigations*. John Wiley & Sons.
Zipf, G. K. (1946). The psychology of language. In P. L. Harriman (Ed.), *Encyclopedia of psychology* (p.p. 332-341). Philosophical Library.

Recebido em: 13/05/2020
Aprovado em: 08/05/2021