ARTIFICIAL INTELLIGENCE AND LANGUAGE ACQUISITION

Ćalušić, Mirko

Sveučilište u Splitu

Filozofski fakultet

Odsjek za engleski jezik i književnost

Mirko Ćalušić

ARTIFICIAL INTELLIGENCE AND LANGUAGE ACQUISITION

Diplomski rad

Split, 2021.

University of Split

Faculty of Humanities and Social Sciences

Department of English Language and Literature

ARTIFICIAL INTELLIGENCE AND LANGUAGE ACQUISITION

Master's thesis

Student: Supervisor:

Mirko Ćalušić Danijela Šegedin Borovina, PhD, Assistant Professor

Split, 2021

Contents

1. Introduction 1
2. Artificial intelligence 4
 2.1. Theoretical background of artificial intelligence 4
  2.1.1. Computational theory of mind 4
  2.1.2. Connectionist theory of mind 5
  2.1.3. Classicism versus connectionism 7
 2.2. History of learning machines 8
3. Artificial intelligence research and linguistics 10
 3.1. Artificial intelligence language modelling 10
  3.1.1. Quillian's semantic model 10
 3.2. Generative linguistics and artificial intelligence 12
  3.2.1. The connectionist-generativist debate 14
  3.2.2. Generative-connectionist synthesis 16
4. Practical applications of artificial intelligence for language learning and teaching 19
 4.1. Measuring language learnability 20
 4.2. Error analysis and correction 22
  4.2.1. Pronunciation correction 25
 4.3. Adaptive learning and difficulty adjustment 26
 4.4. Classroom observation 28
 4.5. Predicting learner performance 29
 4.6. Flipped classroom aided by deep neural networks 31
 4.7. Impact of AI on teachers and students 33
5. Discussion 38
6. Conclusion 41
7. Bibliography 42
Abstract 45
Sažetak 46

List of figures and tables

Figure 1: *A Finite State Grammar generating an unbounded set of grammatical English sentences* ……………………………………………………………………………………..13

Figure 2: *The U-shaped acquisition of the correct forms of the past tense*…………………..15

Figure 3: *The structure of a Simple Recurrent Neural Network*……………….…………….17

Table 1: *The seven major error types recognized by the Intelligent Tutor*…………….…….23

1. Introduction

At some stage therefore we should have to expect the machines to take control.

–Alan Turing, *Intelligent Machinery, A Heretical Theory*, 1951

From the beginnings of artificial intelligence (AI) development, the research was closely tied to linguistics. One of the first avenues of interdisciplinary AI research opened with Chomsky's (1957) theory of universal grammar, as shall be expounded upon below. Additionally, a major focus of AI research and development has been the creation of a computer intelligence capable of producing natural speech. It should come as no surprise that computer scientists and mathematicians working on AI took a keen interest in linguistics from the very beginning. The relationship was, at least in its inception, one-sided. Linguists did not see much use in AI's theoretical or practical use (Rosenberg, 1975). Now that the research is advancing at a fast rate and AI already sees vast and far-reaching practical applications, it is time to re-examine what this field of study means for linguists and language teachers. The theoretical and practical uses of AI in the field of language acquisition will be the focus of this paper.

Advanced artificial intelligence promises to replace humans in performing various complex tasks. From driverless cars to virtual teachers, the economy of the future will be significantly impacted by fast-advancing AI technology (Szczepański, 2019). That is not to say that language teachers will be replaced by machines any time soon, but rather that they should get ahead of the game by developing an understanding of AI, what it is, what it can do, and how it can, and will, affect their work in the future. Moreover, I believe it is necessary to keep track of technological changes as they emerge because that is the best way to harness them to the benefit of both teachers and learners. Rapid technological change can bring uncertainty and even anxiety (Toffler, 1970), but I believe, once sufficiently analysed and understood, it can become a source of opportunity. This idea has been the impetus behind the topic of this paper.

Within the context of language acquisition, there are two relevant aspects to the study of artificial intelligence. The first is the study of how AIs, particularly neural networks (NNs), acquire language, and how this relates to the acquisition of language by humans. The biological plausibility of neural networks may indicate a similarity to biological brains – and indeed, many common language learning phenomena arise during an NNs "learning" of a language – like L2 interference. The study of neural networks may thus find a useful application in the fields of neuroscience and linguistics.

The second aspect relates to the practical application of AI research in the classroom. In the first instance, theoretical findings in research may be used to improve language learning. By making the deep and unconscious language acquisition processes explicitly known, we may be able to advance teaching methods beyond what we might come up with using conventional or "common-sense" methods. Just like how previous advances in theory have affected language teaching methods in the classroom, it can be reasonably assumed that the cutting-edge study of neural networks will one day be practically applied. Secondly, with the proliferation of advanced NN language models, teachers will have access to technology which can – almost independently - produce teaching material in the target language: texts for cloze tests, dictations, reading material suited for the learners' language level, accurate placement tests, and so on. AI could also be used to track a class's progress far more efficiently and accurately than a single teacher. For learners, an advanced language model can provide a speaking partner for individual learning. It can also be used to give personalized learning materials tailor-made for a specific student's needs. Another practical use of neural networks may be the preservation of dying languages.[1] By training a neural network on such a language, it becomes preserved effectively forever (given it is securely stored), and this may prove invaluable in preserving world cultural heritage (some far future learner may too learn this language from the preserved language model). The COVID-19 pandemic (ongoing as of the writing of this paper) caused disruptions worldwide for teachers and learners at all levels. With some already predicting long-lasting changes as a result, with the potential normalisation of remote work and learning, novel applications of technology in the learning process may well be the defining feature of the classroom in this unprecedented age. Some learning institutions have already taken bold, if questionable, leaps forward, such as when Concordia University students discovered that their art history professor had died in 2019, and the university had been using his pre-recorded video lectures to teach art history under his name.[2]

The thesis will proceed with an overview of the history of AI research, outlining its theoretical background. This is a broad topic, cutting across multiple disciplines, such as mathematics, neuroscience, and philosophy. The overview here must necessarily be brief and will focus mainly on those aspects of AI research that are relevant to the topic at hand. Then a description and history of the two most common types of artificial intelligence will be provided. Once the basic theory and concepts have been explained, we shall consider the implications of

[1] https://www.fairplanet.org/story/embracing-artificial-intelligence-to-preserve-dying-languages/
[2] https://slate.com/technology/2021/01/dead-professor-teaching-online-class.html

AI research within linguistics. Section 4 of this paper deals with the practical uses of artificial intelligence as it relates to language learning. This is followed by a discussion of the current state of the field, and considerations about its future.

2. Artificial intelligence

The history of AI research has its roots in Turing's theoretical contribution describing a Turing machine (Turing, 1937), an idealized model of a computing device that is able to carry out any formalized set of instructions. Building on the basis of this work, researchers (Rosenblatt, 1957; Minsky, 1958; Quillian, 1969) attempted to construct computational models of mental processes. The research split into two main branches; the first was artificial intelligence, which sought to engineer thinking machines. The second was computational psychology, which aimed to construct computational models of human mental activity (Rescorla, 2019).

2.1. Theoretical background of artificial intelligence

The fields of AI and computational psychology are tightly interconnected. The first theory of "mind as computer" describes it as a linear algorithmic machine processing a symbolic language, much akin to a standard digital computer, or a Turing machine. This view of mental computation, known as *the computational theory of mind,* was popular from the 1960s to the early 1980s (Rescorla, 2019). A theory which emerged later describes the mind as a network of interconnected nodes. The two theories correspond to the two types of AI: algorithmic AI and neural networks, respectively.

2.1.1. Computational theory of mind

The computational theory of mind (CTM) holds that intentional states of the mind are relations between the thinker and symbolic representations of the content of the states (Horst, 2003). For example, "to believe that there is a cat on the mat means to be in a particular functional relation" to a symbolic mental representation with the semantic value "there is a cat on the mat" (p. 2). These representations have semantic and syntactic properties, and reasoning processes are performed only using the syntax of the symbols – the semantics are irrelevant to the process. This is known as formal symbol manipulation and can be defined as a form of computation. Since the semantic properties of symbols can be formalized, that is to say, represented through syntactic relations, according to this theory, they can also be represented mechanically. Anything that can be formalized can also be executed by a Turing machine. The computer could conceivably duplicate what the human mind is doing, which is an idea Turing (1980) himself suggested. Other writers have taken a more conservative approach, however, suggesting that what the computer does is merely a simulation of the human computer (Horst, 2003). Since this formulation of CTM is based on classical (algorithmic) computation, it is also known as the *classical computational theory of mind* (CCTM) (Rescorla, 2019).

CTM was from the very beginning in an interrelationship with AI research. On the one hand, CTM provided a general theory of the mind as a computer that lent credence to the idea that human-like AI was feasible; on the other, successes of AI research in modelling reasoning, language and perception in machine intelligence lent credence to the theory of CTM (Horst, 2003).

In empirical research, the connection to Chomsky's (1957) generative linguistics is of historical importance. Chomsky rejected the behaviourist account of language acquisition, claiming that the general principles of classical and operant conditioning did not adequately explain how a child is able to acquire language so efficiently early in life. He suggested a mental mechanism optimized for language learning. A child's efforts in understanding grammar were often described in terms of forming and confirming hypotheses. This would require an inner language of thought (Horst, 2003). The idea of the mind operating on the basis of an inner language (often called *Mentalese*) is a central proposition in the language of thought hypothesis (LOTH), which was advanced by Fodor in his seminal work *Language of Thought* (1975). Fodor considers CTM a necessary component of LOTH (Rescorla, 2019). The systematic and productive features of language competence are also features of thought. Language is systematic and productive because it is an expression of a mind that already possesses these features, and because the mind is a syntactically structured representational system (Horst, 2003).

2.1.2. Connectionist theory of mind

In cognitive science, the connectionist theory of mind, or connectionism, is a movement that seeks to explain human intelligence by the use of neural networks to model the workings of the brain. Neural networks are composed of a large number of nodes or units, which are analogous to individual neurons in the brain. Units are connected to each other, modelling synapses between neurons, and the strength of the connection between each pair of nodes is determined by weights, which model the strength of the synaptic link. We shall look at neural networks in more detail in section 3.2.2.

Connectionism provides a challenge to the classical computational theory of mind (Buckner and Garson, 2019). The strength of connectionism as a framework for understanding the nature of the mind and brain lies in its biological plausibility. The brain actually is a neural network consisting of a myriad of neurons and synapses. Additionally, the way neural networks process data is close to natural cognitive processing. While traditional computers usually fail in their task when circuits are lost, or the data signal becomes noisy, the neural network can accomplish its task successfully – albeit with less accuracy -- even with lost circuitry and an imperfect signal, and the accuracy of the neural network degrades gradually as these impedances become more extreme. The neural network is accordingly better poised to take on the challenges of the real world (ibid.) where perfect conditions rarely exist. In other words, neural networks are able to accomplish tasks under conditions similar to those of biological brains.

Connectionism may also provide an answer to an old philosophical problem: the seeming inability to provide a definition of a thing with necessary and sufficient conditions (Buckner and Garson, 2019). For example, let us take the concept of a “chair”. Even though it would be difficult, perhaps impossible, to give a definition of “chair” which would include all objects which we perceive as chairs, while excluding all others which we do not (such as sofas and ottomans) a person can still easily determine whether a given object is a chair without such a working definition. There is no explicit intellectual process involved; we recognize chairs by intuition. Connectionism can provide a good explanation for the flexibility of the human intellect. Without the need for hard and fast rules, NNs are well suited for differentiating subtle statistical patterns in a way that rigid forms of symbolic representation are not (ibid.). And keeping in mind that CCTM relies on symbolic representation, the neural network appears to be a more powerful and accurate way to model how biological brains operate, and consequently how human minds perceive the world.

Despite these advantages, some arguments against connectionism can be made. Firstly, current neural networks are merely abstractions which ignore some important features of the brain, such as the variety of neuron typology and the effects of neurotransmitters and hormones. Secondly, supporters of the classical model raise the objection that neural networks are not well suited to the kind of rule-based processing (which classical computers can perform very well) that is thought to be the basis of language and reasoning (Pinker and Prince, 1988, in Buckner and Garson, 2019).

2.1.3. Classicism versus connectionism

To sum up, the classical view is that human cognition is analogous to digital computation. Per the classical account, mental information can be represented as strings of symbols and cognition similar to digital data processing, where symbols are processed sequentially according to the instructions of a program (an algorithm). On the other hand, the connectionist model describes information as being stored in the connections between nodes of a neural network, with cognition being described as the dynamic evolution of activity in the network (Buckner and Garson, 2019).

Some connectionists have sought to unify these disparate views by arguing that the mind is a neural net but that, at a more abstract level and at higher levels of cognition, it also functions as a symbolic processor. Others have outright rejected symbolic processing as a function of the mind at any level, pointing out the failure of classical computing to match the effectiveness and flexibility of human intellect (ibid.).

The conflict between classicists and connectionists is closely related to the innateness debate. The big question is: are higher-level abilities such as language and reasoning innate, or are they learned? The ability of neural networks to learn complex tasks starting from a randomly chosen state lends credence to the empiricists who claim the infant is able to construct the higher level features of the intellect (such as language) solely from perceptual input and a simple learning mechanism. On the other side of the debate, nativists claim that the poverty of the stimulus argument demands the existence of a specific mechanism in the brain specifically suited for acquiring language (the Language Acquisition Device, Chomsky, 1965). This, in turn, could support the thesis that all other higher cognitive abilities are acquired through similar genetically determined mechanisms. A synthesis of the two views can arguably be made: the ongoing learning of the neural network could be interpreted as the refinement of genetic characteristics across the generations of a species. The idea that genetically determined knowledge in the brain can be represented in the connectionist model by biasing the starting weights of a neural network to make a specific task trivial to learn (Buckner and Garson, 2019). On the other hand, connectionism could in the future make the case against the poverty of the stimulus argument by providing a working language model running on a neural network, using simple learning mechanisms and linguistics inputs available to humans (ibid.). The conflict between connectionism and generative grammar will be explored in more detail in section 3.2.

2.2. History of learning machines

According to Turing (1980), the seeds of the idea of artificial intelligence sprouted from the computing machine Charles Babbage designed in the 19th century. However, the foundations for modern artificial intelligence research were laid in the mid-20th century, coming out of discussions about providing mathematical solutions to decision-making problems (Pace-Sigge, 2018). As previously mentioned, Turing's description of a powerful digital computer (1937) was the basis both of all modern computer science, as well as AI research. It was Turing himself (1980) who posed the question: "Can machines think?" In order to determine whether a machine did indeed possess a human-like intelligence, Turing proposed a test. The Turing test, also known as the imitation game, takes the form of a question-and-answer interrogation. An interrogator communicates via keyboard and on-screen text to two conversation partners; one is a man, the other is a machine. It is the task of the interrogator to determine which is which. If a machine can produce answers that would convince its interrogator that it was actually a person, that is, if the machine could understand linguistic input, and produce sufficiently human-like output, that would be sufficient proof that the machine is truly intelligent. Turing, by this point, had moved away from the static computational device he had described in 1937: he imagined learning machines which would be able to use "fuzzy logic", predicting important features of neural networks (Pace-Sigge, 2018). The earliest recognized form of artificial intelligence was presented in 1952 by Marvin Minsky. Called the Stochastic Neural Analog Reinforcement Computer (SNARC), it was developed from conceptual models of neural networks, as an attempt to artificially recreate a biological neural network (Minsky, 1958, in Pace-Sigge, 2018). Echoing Turing, Minsky emphasized that an artificial intelligence needs to learn from experience. Minsky also noted that a crucial task for the learning of artificial intelligence was pattern recognition (ibid.).

It was in 1957, the same year Chomsky published *Syntactic Structures*, that Frank Rosenblatt published his work on perceptrons, which we now know as neural networks. The two young men (both were 28 at the time) laid the foundations for two diverging lines of research in cognitive science. Chomsky and Rosenblatt came at the problem of cognition from opposite ends. Chomsky started with language, a high-level phenomenon of cognition, and attempted to show that computational machinery was not able to represent it, before proposing a more powerful model (the LAD) that could. Rosenblatt started out from very basic machinery representing neurons and synapses and attempted to show that they could represent low-level cognitive processes, and that they could learn algorithmically (Pater, 2019).

Both Chomsky and Rosenblatt interacted with researchers who were working on what was dubbed artificial intelligence. Chomsky's arguments about the complexity of language were made in context of what was being explored at the time in AI research, namely finite state machines and probabilistic Markov chains[3] (ibid.).

Although early AI research focused on neural networks, by the mid-to-late sixties, thanks to technological and theoretical barriers at the time, as well as Minsky and Papert's (1988, in Pater, 2019) criticism of existing neural network models, the field shifted to logic-based AI, which used algebraic manipulation of symbols (this type of AI is based on the previously mentioned CCTM). It would not be until the 1980s that neural networks would see an upsurge in practical research. By that time, they would also become a core part of cognitive science. The third wave of neural network research began in the 2010s and is an ongoing field of research (Pater, 2019).

[3] A Markov chain is a stochastic model of a possible sequence of events where the probability of each event is determined only by the preceding event in the chain (Gagniuc, 2017).

3. Artificial intelligence research and linguistics

Starting in the late 1950s, researchers began to consider the possibility of building machines which could learn human language. In order to tackle the complexity of this task, computer scientists began to collaborate with linguists (Pace-Sigge, 2018). Contemporary information technologies allow for trillions of words from various sources to be collated and processed. This allows for a new, fully empirical vision of language. Moreover, this collected data can be used to form models for machines to mimic natural speech. The ability of artificial intelligence to process language is based on linguistic knowledge. As this technology becomes more advanced, machines will come closer and closer to creating an identical model to human understanding, processing and production of speech (ibid.). Consequently, AI provides opportunities for linguistic theories to be proven experimentally, while undermining opposing theories, because "if a form of AI works, this can be seen as a result of turning one theory into practice" (Pace-Sigge, 2018, p. 2).

3.1. Artificial intelligence language modelling

In simple terms, AI language modelling involves teaching an artificial intelligence to recognize inputs of a specific language and produce output in that language resembling as close as possible that of the human speaker. According to Pace-Sigge (2018), the whole field of AI language processing can be traced back to M. Ross Quillian's theoretical work on the *Teachable Language Comprehender* (TLC) (Quillian, 1969, in Pace-Sigge, 2018). Quillian's machine is a simulation of the human mind learning a language. The machine is trained on and learns through inputs provided. Quillian proposed a set of 20 short texts for the machine to learn, an ambitious number in 1969. Today, it is possible to process a virtually unlimited amount of data; the big problem is finding suitable texts, cleaning them up and making them machine-readable (Pace-Sigge, 2018).

3.1.1. Quillian's semantic model

Quillian described the TLC as being able to comprehend a text because it would learn from its training input. The input for the TLC consisted of 20 short children's books about firefighters to let the machine learn basic information about them. This, for Quillian, would be a digital representation of human language development: "we assume that there is a common core process that underlies the reading of all text – newspapers, children's fiction, or whatever – and it is this core process that TLC attempts to model" (1969, p. 461, in Pace-Sigge, 2018). According to Quillian, natural language is communicated by causing the mind to recall concepts it already knows and connect them to other concepts. The TLC would learn not by working on big structures, but piece by piece, developing the structure over time (Pace-Sigge, 2018). Crucially, Quillian's model was semantic, and he proposed resolving polysemies by exploiting semantic clues in the text. Take, for example, the following two sentences:

1) "He reached the bank."
2) "He got a loan from the bank."

The first sentence is clearly ambiguous, but the latter has sufficient clues in order for a reader to understand the meaning of "bank" (ibid.). In the TLC, words are connected semantically in what is called a *semantic web.* "Bank" would thus be connected to concepts such as "money", "loan", "robbery" and so on. Quillian actively rejected Chomsky's generative linguistics (ibid.). The issue of resolving polysemantic ambiguities is, per Quillian, not satisfactorily resolved by generative linguistics; using only grammatical features and their locations does not paint a full picture of human understanding:

> That human beings do not so limit themselves, but also utilize semantic clues extensively, would appear obvious from the fact that people are able to understand language that is full of grammatical and syntactical errors (1962, in Pace-Sigge, 2018).

Quillian (1966, in Pace-Sigge, 2018) pointed out that language users employ recognition memory as opposed to recall memory. The difference between these can be illustrated by example: if a reader were to be told that the word "the" can mean "her", the reader might not be able to immediately recall a linguistic context where this is true. Now one must imagine a situation where the reader is presented by the sentence "I took my wife by the hand". Having recall of the word *the*, the reader would possibly define it as a grammatical word with little or no semantic content. The reader will however use recognition memory to realize that "the" in that sentence could be replaced by "her", as it is a back referential to "my wife".

In Quillian's model, the meaning of a word can be deduced from other words connected to it in the semantic web – each word is a single node, and the connections between them are called *associative links.* These associative links are not based in grammatical structures, and they are enormously flexible. For example, the word "mistress" would have a completely different network of associations in the 19th century, when the word meant "lady of the house", perhaps strongly linking to words such as "house", "of", and "master". In the 20th century, the word took on a different primary meaning, and its semantic web changed (Pace-Sigge, 2019).

To put the hypothesis to the test, Collins and Quillian (1969, 1970, 1971, 1972a, 1972b, in Pace-Sigge, 2018) conducted experiments testing reaction times of test subjects to find that they react faster to true sentences (e.g., *tennis is a game*) than false sentences (e.g., *football is a lottery*). These tests became the basis for experiments by psycholinguists, including Meyer and Schvaneveldt (1971, in Pace-Sigge, 2018) who expanded on the experiment by Collins and Quillian. Meyers and Schvaneveldt showed that reaction times were faster when a word was preceded by another semantically related word (e.g., *nurse – doctor*) than by a semantically unrelated one (e.g., *nurse – bread*). This effect is known in psychology as *priming,* and it is interesting to note that it was an attempt to create a thinking machine that led to the recognition of this facet of human cognition.

Pace-Sigge (2018) stated that AI language technology not only has enormous practical applications in terms of productivity gains, but that it presents novel opportunities "to understand both language and characteristics of human society, in particular the nature of human discourses." (p. 15)

3.2. Generative linguistics and artificial intelligence

According to Chomsky (1957), the problem of language analysis is one of classification: the aim of language analysis is to separate the grammatical sequences which belong to said language, from grammatical sequences which do not. For example, *The lion sleeps* is a grammatical sentence, while *Sleeps lion the* is not (Pater, 2019). To determine whether a sentence is grammatical is a question of determining if it has been generated by a grammar, hence the name *generative linguistics* (ibid.).

A way to encode the difference between these sentences is via allowable transitions between words, employing a model (a finite state machine) which specifies a set of states, and allowable transitions between states. A finite state grammar (FSG) allows for an infinite number of sentences to be generated from limited resources. This is possible because the finite state model allows for looping.

As seen in the example below (Figure 1), the finite state machine generates *The man comes*, *The old man comes, The old, old man comes,* etc. The loop allows for an infinitely large set of sentences.

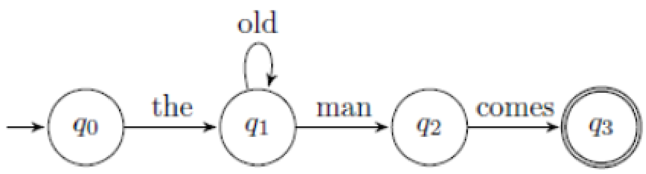

Figure 1: A Finite State Grammar generating an unbounded set of grammatical English sentences (Pater, 2019)

Chomsky (1957, in Pater, 2019) argues that finite state grammar cannot represent the full complexity of sentence structure in English. In particular FSGs cannot account for centre embedded clauses, such as *The girl that hit the balls that flew over the fence is cheering* (Pater, 2019). An adequate English grammar model must go beyond sequential restrictions and be able to represent the hierarchical structure of a sentence. Chomsky goes beyond the surface structure of a language and proposes a transformational grammar whereby a deep structure derivationally generates the surface structure (which are the actual spoken sentences). These two kinds of structures are postulated in generative analyses of aspects of a language other than syntax, such as phonology (ibid.). According to Pater (2019), a common example is the postulated abstract /ai/ in words such as *writer* and *title* in Canadian English, in which they are pronounced with a raised diphthong [ʌi] in contrast to the surface [ai] in words such as *bridle* and *rider*. The two phonemes in the surface structure can be generated from a single deep structure phoneme based on the assumption that *title* and *writer* have an underlying /t/ which conditions raising.

The derivational depth of deep structures has been controversial (Anderson 1985; Dresher 1981, in Pater, 2019) because it is assumed that it presents difficulties for learning. The linguistic structure that is not apparent to the learner is called “hidden” and poses specific learning challenges (Tesar and Smolensky, 2000, in Pater, 2019). Here Pater (2019) draws a connection to hidden layers of neural networks, which can possess features that are not present in the surface input nodes. Like derivations and hierarchy of a generative language analysis, the activation of the hidden layer must be inferred by the learner.

Pater (2019) identifies two ways in which drawing parallels between hidden language structures and hidden layers of neural networks may be useful for AI language. First, techniques for learning with hidden layers could be of use in learning with hidden linguistic structure that have been explicitly encoded. Second, hidden layers can be used to learn representations that replace such hidden structures.

By postulating an abstract structure of language, generative linguistics was differentiated from mainstream AI which was based on generalized learning mechanisms of connectionist neural networks but bore similarity to the early work done with algebraic AI (ibid.).

3.2.1. The connectionist-generativist debate

The two threads of cognitive research, generative linguistics and connectionism, which started with Chomsky and Rosenblatt, respectively, encountered each other for the first time in the 1980s. By that time, learning had become central in generative linguistics, but it differed dramatically from the neural network conception of learning. Chomsky (1980, in Pater, 2019) argued for a framework of describing the syntax of a natural language as being defined in terms of a finite set of principles common for all human languages and a finite set of parameters which are binary switches that are set on or off for a particular language. The so-called principles and parameters theory endows the LAD with significant capability. The LAD is able to rapidly induce the rules of a grammar by being exposed to sentences from the relevant language. The principles and parameters theory of universal grammar is restrictive to explain how a child's mind is able to, depending on its linguistic environment, learn any of the thousands of human languages that exist. Guided by this principle of restriction, a common critique of connectionism by generative linguists is that neural networks are able to produce patterns not attested in human language. Neural network research, for its part, is largely emergentist[4], and researchers of connectionist linguistics contrast it with innatism in the tradition of Chomsky (Pater, 2019).

The cause of the connectionist-generativist debate was a neural network that was trained to output the past tense of English verbs from an input of uninflected forms of English verbs (Rumelhart and McClelland, 1986, in Pater, 2019). The simulation was set up to challenge generative linguistics, in particular the views that the rules of language are stored explicitly as propositions, and that there is an innate learning mechanism that knows the possible range of natural languages. In Rumelhart and McClelland's model, there were no rules that added a past tense morpheme to the uninflected stem, and neither were there rules that determined the shape

[4] Emergentism is the belief that properties of a system are emergent after a certain threshold of complexity and are qualitatively distinct from its constituent parts.

of that morpheme. Learning was accomplished through weight adjustment as errors were found in the output. This neural network modelled the nonlinear, U-shaped acquisition of the past tense (see Figure 2 below) which is in line with how we observe children learn, from initial accurate prediction of irregular forms to over-regularization (e.g. *holded, readed*), and finally to the correct target form (Pater, 2019).

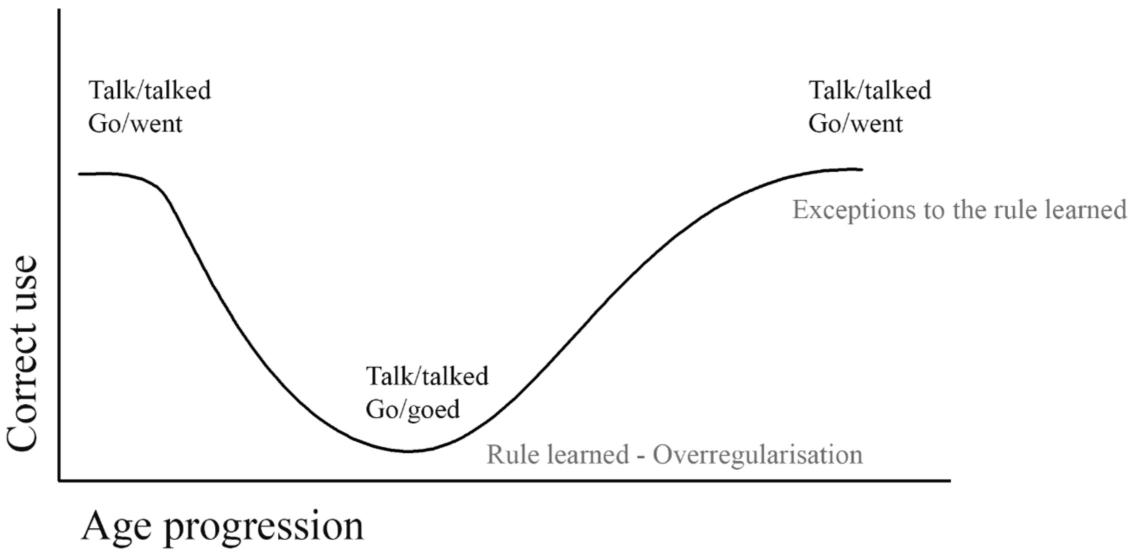

Figure 2: The U-shaped acquisition of the correct forms of the past tense (author's own work)

Pinker and Prince (1988, in Pater, 2019) offered a critique of this model based on several aspects of its representation of the past tense. Most notably, they claimed that the regular and irregular forms were created by separate systems, rather than a single cognitive feature. According to Pinker and Prince, irregular verb formations exhibit certain rules or regularities that are the product of a memory system, for example [ei] to [ʊ] with a final [k], as in *shake/shook* and *take/took*. These may be formalized in a connectionist model because anything that looks like a rule-based regularity is merely an artifact of how the irregular past tenses are lexically stored (Pater, 2019). On the other hand, according to Pinker and Prince (1988, in Pater, 2019) the regular pattern is produced by morphosyntactic rules which add the *-ed* morpheme, and phonological rules of voicing assimilation and vowel epenthesis which produce appropriate forms.

According to Pater (2019) the debate between connectionists and generativists often turned on a definition of what it meant for a model to be rules-based. Some have argued

(Lachter and Bever, 1988, in Pater, 2019) that the past tense neural network model incorporated rules through particular configurations of nodes in the neural network. Pater (2019) noted that "[some] of the back-and-forth in the past tense debate can be tiring precisely because it consists of one side accusing the other of not being true to its principles in incorporating aspects of the first sides' theory," (p. 16) but that, ultimately, this is the most fruitful part of the debate because it shows that there is space for integration of connectionist and generative models. There is nothing about the connectionist account that prohibits the use of symbols and variables; connectionist models are not fully emergent because parts of their structure need to be specified, and generative models are not fully innatist because parameters must be set by experience.

3.2.2. Generative-connectionist synthesis

As we have seen in the previous sections, early connectionist language researchers (Quillian, 1966, in Pace-Sigge, 2018; Rumelhart and McClelland, 1986, in Pater, 2019) actively contrasted their efforts to generative linguistic models. However, beginning with the late 1980s, some researchers (Lakoff, 1988; Hare, Corina & Cottrell, 1989; Legendre, Miyata & Smolensky, 1990; Goldsmith, 1993; Lakoff, 1993; Wheeler & Touretzky, 1993; Gupta & Touretzky 1994, in Pater, 2019) began to integrate generativist assumptions into connectionist models of language, and a synthesis began to develop. An especially prolific generative-connectionist fusion is Prince and Smolenksy's (2004, in Pater, 2019) Optimality Theory (OT). OT bears similarity to Chomsky's (1980) principles and parameters theory; instead of binary parameters, OT has constraints whose ranking must be determined. Closely related to OT is Harmonic Grammar (HG; Legendre, Miyata & Smolensky 1990; Smolensky & Legendre 2006, in Pater, 2019) which, rather than rank its constraints, weighs them numerically.

Parameters in the classical theory may be activated or not, but once active they are inviolable in grammatical language. By contrast, the constraints of OT/HG theory are violable through constraint interaction. A lower ranked/less weighted constraint will be overridden by a higher ranked/more weighted one. This allows for non-uniform constraint application. A consequence of this is that, for language typology, it is possible to have generalised rules which still account for details of individual languages. Within parametric theory, observing a surface violation of a constraint leads to changing a constraint into a set of constraints which can remain inviolable. (Pater, 2019) This fine-tuning of theory to fit linguistic data was criticised by Prince and Smolensky (2004, in Pater, 2019)

The project of generative-connectionist synthesis, as well as the debate surrounding it, are still ongoing. A contemporary focus of the debate is a particular type of neural networks know as Recurrent Neural Networks (RNNs), which do exceptionally well in modelling aspects of natural language syntax. The debate centres on whether RNNs can fully learn syntactic regularities without resorting to hierarchical representations typical of generative linguistics. RNNs process a sequence of elements (words, phones or letters) one element at a time. The network predicts the next element of the sequence, and the weights are updated based on that. When the network moves on to the next layer in the sequence, the current hidden layer is copied into a context layer, which is used as an additional input for the next set of hidden layer computations. In simple terms this means that the RNN has a form of sequential memory (Pater, 2019). Figure 2 below shows a representation of the structure of an RNN.

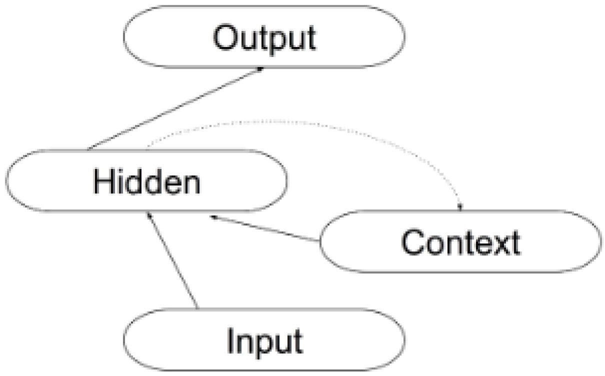

Figure 3: The structure of a Simple Recurrent Neural Network (adapted from Lewis & Elman 2001; Pater, 2019)

In recent years, neural networks, and RNNs in particular, have been applied to a number of broad language tasks with considerable success. In AI translation, RNNs can map from one language to another without the use of intermediate linguistic structures and can do so as well or better than earlier models which did use intermediate structures such as phrases. The success of modern RNNs is due to advanced architecture, training methods and increased computational power. The networks are large, with many layers, and they are trained on massive datasets. Studies (Frank et. al., 2013; Linzen et al., 2016; Bernardy and Lappin, 2017, in Pater, 2019) done on the capabilities of RNNs to learn natural language syntax without explicit linguistic structure show that current models are partially successful even at long-distance dependencies, which suggest that they learned something like structural analysis, but that they also tend to assume incorrect linear regularities. The cause of these errors is not clear. RNNs in theory should be able to produce the correct structures, but whether a specific network, with a specific learning method, and with specific training data will be able to do so is another matter. Research into connection weights and activation patterns of neural networks can reveal how they internally represent the data they process, and most recently it was shown (Palangi et al., 2017, in Pater, 2019) that a type of neural network known as Tensor Product Recurrent Networks (TPRN) does in fact produce representations that can be interpreted as syntax. TPRNs sit somewhere in-between emergentism and innatism, since they are given structural building blocks but must learn its configuration (Pater, 2019). Exactly how much explicit linguistic structure is needed in AI models of language is still a matter of debate (Pater, 2019).

When it comes to generative linguistics, the field could profit immensely from integrating the connectionist account of learning. Statistical learning theory could provide a rigorous method to test claims about universal grammar, and it may be possible to assess to what extent grammatical structures are learned, and which aspects of grammar are innate. Theories integrating the insights from neural network data may in turn be used to create even more advanced neural networks (Pater, 2019).

4. Practical applications of artificial intelligence for language learning and teaching

Computer-assisted language learning (CALL) is not a new concept. The application of information and communications technology to the language classroom goes as far back as the 1960s (Marty, 1981). CALL encompasses everything from the use of multimedia in the classroom to the use of interactive whiteboards, distance learning, and virtual worlds. The use of at least some of these methods has become integral to everyday language teaching, and the usefulness of CALL has been punctuated by the necessity of distant learning owing to the ongoing COVID-19 pandemic. Intelligent Computer-Assisted Language Learning (ICALL) combines CALL with artificial intelligence. This field began to develop in the late 1970s (Heift, 2017) and it is not yet full formed, because it relies on currently emerging technologies, as well as the enormous complexity of the task, for which standard algorithmic AI is not well suited, but which benefits greatly from the developing field of neural networks and deep learning.

CALL technologies have been available to second language (L2) learners for decades. The CALL software Rosetta Stone was first released 29 years ago (Swad, 1992). Since then, applications like Babbel and Duolingo appeared, as well as a slew of others. Learning apps like these, which assist learners, and even allow some degree of interaction through chatbots[5], raise the issue of whether language teachers will be needed in the future (Dargan, 2019). Despite the optimism of some neural network researchers, the technology is not yet advanced enough to replace teachers, but it can be used to augment their abilities, allowing them to do more within the limits of their time and resources.

Digital assistants such as Apple's *Siri* and Amazon's *Alexa* employ speech recognition and synthesis to assist users in everyday tasks. Aside from simple commands, AI can be employed in more complex scenarios, such as in a learning environment (Lotze, 2018). The problem with most current computer-assisted platforms, like the above-mentioned Duolingo, is that they are largely based on outdated concepts, for example the translation method. Some apps employ artificial tutors which are chatbots that help the learner communicate in the target language. Though this technology is not widely used in foreign language classrooms, it has seen some success in university teaching, for example the e-learning content for German linguistics,

[5] A chatbot is a software application designed to convincingly simulate a human conversational partner.

grammar and orthography at the Leibniz University Hannover, featuring the artificial tutor *El Lingo* (Lotze, 2018).

With the projected growth of the number of English as a second language (ESL) students in the world in the coming years, there is a real need to offload teachers' work and provide students with a greater quality of learning, particularly in areas where language teachers and learning materials are scarce. AI is already being applied in real world learning scenarios, with China being the biggest market player currently ($568m were spent on AI-assisted education in China in 2017, with projections to surpass $26bn in 2022) – this is undoubtedly driven by a lack of English teachers in most Chinese schools (Sejnowski, 2020). Recent advances in neural networks, combined with huge datasets, provide a much clearer picture of how learners advance in a certain language, which parts of a language are harder to learn than others, which are liable to be forgotten, etc. This, at least in theory, allows for highly personalized and effective teaching materials to be created and allows for the creation of more accurate placement tests.

What follows is an overview of studies that have been done on the practical applications of artificial intelligence as it pertains to language learning and classroom assistance.

4.1. Measuring language learnability

Learning a second language is usually easier if a learner's L2 is similar to their first language (L1) (Ellis, 2015). However, the similarity between the two languages is problematic to quantify, so it is difficult to ascertain its effect on learnability. Research methods used to determine the effects of similarity on L2 acquisition are usually experimental, typically proceeding in one of two ways. The first method is to take one group of L2 learners, with the same L1 and compare their acquisition of various structures in the L2, such that one structure is similar to L1 and the other is different. The second type of experimental approach holds constant the target structures to be learned and instead compares the acquisition of those structures across learners with different L1s. Both approaches rely on binary same/different evaluations at a feature-by-feature level (Cohen et al., 2020).

For their study, Cohen et al. (2020) adopt a different approach to measure language similarity and learnability: they built a series of neural network models for a set of five artificial languages with established degrees of similarity between each pair of languages. These languages represented L1 speakers learning second languages. Observing the change in the activity of the cells between the L1-speaker model and any of the remaining artificial languages, they estimated how much change was needed for the model to learn the new language, then compared it for each L1/L2 bilingual model. Their findings showed that this approach could find the facilitative effect of similarity on L2 acquisition and "offer new insights into differential effects across different domains of similarity" (ibid.).

They predicted that the model would need to change less to learn L2s that were like L1, than second languages that were different from L1. For each pair, they coded the degree of overlap as 0, 1, or 2. So, each of the languages they created had less overlap than the previous one, with the fifth language having no overlap with the L1 whatsoever. The languages were not only coded by degrees of overlap, but for the domain of overlap, as well, to "explore whether that affected the amount of cell activity change produced by learning a second language" (ibid., p. 6). To explore the impacts of domains of overlap, for example, syntax vs morphology for language pairs with one degree of overlap, or syntax/morphology vs vocabulary/morphology for language pairs with two degrees of overlap. Cohen et al. (2020) built a mixed effect linear regression model, using the same software as in their degree analysis. They found that shared syntax has a negligible effect in reducing cell activity differences between L1 and L2. Language pairs which overlap only in syntax showed no meaningful difference in cell activation change from language pairs which do not overlap at all. Moreover, language pairs which overlap in vocabulary and syntax show similar reductions in cell activity as language pairs which overlap in vocabulary alone. Shared morphology seems to be the most beneficial in reducing changes in cell activity between L1 and L2 - especially when it is combined with a second degree of overlap. Languages which overlap in morphology alone show lower differences in cell activity than languages with no overlap, or which overlap only in syntax; and languages which overlap in morphology and another domain have the lowest differences in cell activation (ibid.).

Their results are especially interesting because they offer insights into which elements of linguistic similarity, and which linguistic structures seem to be most problematic for the learner during second language acquisition. The authors note that a shared morphological system between L1 and L2 especially appeared to facilitate learning. On the other hand, shared syntactic structures did not appear to make a significant difference. Additionally, function words are seemingly harder to learn in comparison with content words (ibid.).

The authors of this project believe that this approach can be generalised to natural languages which would allow language acquisition researchers to make verifiable predictions about how difficult second languages might be for speakers of various mother tongues, because deep learning systems are sophisticated enough to learn natural languages as well as artificial ones (Cohen et al., 2020). However, the first step will be to apply the methods they used to natural languages to see whether the patterns they found in a simulation can be generalised. For instance, Cohen et al. (2020) did not consider phonological similarity or the role of semantics in the artificial languages they constructed, which rendered it impossible to explore or model the effects of cognates or false friends as a domain of language similarity. Additionally, these effects have a complex interaction not only with the general linguistic context, but also individual cognitive abilities of the speaker. Neural network models could, in theory, be altered to represent the capacities of individual speakers — for example, differences in working memory capacity can be modelled. All models they used in this research had identical internal structures, simulating one learner in different language learning situations. The authors state that this approach, even though it is still in its early stages, has the potential to democratise language learning, by predicting which languages can be easier for which speakers and identifying which domains of grammar could potentially be most challenging (Cohen et al., 2020). This project used artificial language learners and artificial languages to test a method of researching L2 acquisition that has substantial potential for development. By building artificial languages, they were able to avoid the problem of defining how similar two languages are. They also assessed the learnability of a language by concentrating on changes within the generative machine itself instead of the output it produces. While the authors are careful not to overstate the value of their findings, they believe the research stands as a proof of concept (ibid.).

4.2. Error analysis and correction

Although the literature on second language acquisition (SLA) varies in terms of theoretical approaches to error correction, cognitive theorists agree that such correction is advantageous and contributes to learning (Dodigovic, 2007). While some emphasise the importance of the communicative context in which the correction occurs (e.g. Doughty & Williams, 1998; Long & Robinson, 1998, in Dodigovic, 2007), others emphasise the importance of raising awareness (R. Ellis, 1997; James, 1998, in Dodigovic, 2007). Dodigovic's (2007) main thesis is that error remediation is definitely beneficial and that artificial intelligence can be a useful tool in this respect.

Dodigovic (2007) proposed implementing artificial intelligence in the second language error remediation phase in an effort to reconcile the two approaches. The AI program used in the research was the Intelligent Tutor; it identifies some common errors made by university students studying English as a second language in their writing. The research question explored in the study was, “Does exposure to the Intelligent Tutor (i.e., systematic error correction) have an impact on learning?” The sample included 266 university students from three countries (Taiwan, Australia and the United Arab Emirates) who studied English as a second language. The subjects were aged 19-21; 107 were located in the UAE, 83 in Australia, and 77 in Taiwan. The study lasted for several months in 2004 and 2005.

The Intelligent Tutor's systematic error correction, according to the hypothesis, has a significant impact on learning outcomes. As a result, the analysis was confined to one group and one treatment protocol, which was preceded by a pre-test and followed by a post-test, both of which were linked to seven common structural errors (shown in Table 1) and a few morphological errors.

Table 1: The seven major error types recognized by the Intelligent Tutor (Dodigovic, 2017)

Error type	***Example***
Pseudo-passive	Malaria can find all over the world.
Ergative construction	The immune system can be failed.
Tough movement[6]	More difficult to be realized…
Existential construction	There is a new problem occur
Malformed expressions of feelings/ reactions/states	The disease had* dominant over human.
Missing copula	Secondly, communities* affected.
Finite/nonfinite verb construction	It will caused death of both mother and baby.

The pre-test consisted of 12 multiple-choice questions. The students were asked to judge the grammaticality of utterances. Their proficiency with this task was taken as a reflection of their competence with these structures. The post-test was a short answer test, asking the

[6] *Tough movement* is a term used for sentences with certain superordinate predicates in which an object or adverbial is extracted from an extraposed infinitival clause and replaces anticipatory *it* (Mair, 2008).

students to construct their own sentences. The design of the post-test was intentionally different from the pre-test, in order to prevent the learning from the pre-test to influence results.

The treatment procedure consisted of individual work with the Intelligent Tutor. The texts were adapted to suit the students' level of proficiency in English (TOEFL 500-550).

The Intelligent Tutor's ability to detect and correct certain learner errors depends on the frequency, gravity, and communicational importance of the errors detected in a learner corpus, as well as the target learners' exposure to the structures in question and their particular needs. One of the benefits of the Intelligent Tutor is that it can accommodate individual learners to some degree. These seem to vary in many respects. ICALL has not yet attempted to support individual differences at this stage, despite the fact that intelligence, language aptitude, and affective factors appear to be very important in individual learning success. The Intelligent Tutor made a modest effort to accommodate different learning styles; it employed Willing's (1988, 1989, in Dodigovic, 2007) learner types approach.[7] Specifically, the Intelligent Tutor mimics the communicative language learning approach, which focuses on increasing students' autonomy and control over the language learning process. The tutor, on the other hand, approached analytical learners in the manner that they wanted to be approached: by giving them problems to solve, assisting them in understanding the essence of their own errors, and providing opportunities to learn grammar (Willing, 1989, in Dodigovic, 2007). The student was still free to choose their preferred course of action, though it was expected that their decision would be affected by their learning style. The tutor gave the learner three options after diagnosing an error: try again, get a hint, or see the answer. The 'try again' and 'get a hint' options catered to theoretical and communicative learners, respectively, while the 'solution' catered to concrete and authority-oriented learners.

According to Dodigovic (2007), the concrete learner might interpret the correction as a recast, while the authority-oriented learner may recognize it as the authority's solution. Having a hint can be part of a communicative technique for the communicative learner. An analytical

[7] A basic distinction among learners is between those who are "analytical" or left-brained, and those who are "concrete" or right-brained. The analytical learner processes knowledge in a linear, sequential, logical, objective, abstract, verbal, and mathematical manner, with a strong emphasis on description, reflective and careful thought, and selective, low-intensity stimuli. On the other hand, the concrete learner processes knowledge in a pattern-seeking, spatial, intuitive, subjective, emotional, and visual manner, concentrating on overall impression while being impulsive, needing rich and varied input. Communicative learners are another type of learner who is more likely to use computer-mediated communication or human-computer communication which imitates human-human interaction (Willing, 1989, in Dodigovic, 2007).

learner could enter the correct version and obtain a parse tree[8] that provided analysis, which was exactly what this learner category required. The parse tree was also likely to reinforce correct language while increasing structural understanding. In other words, the analytical learner's need to understand not just what is accurate but also why it is accurate would be thoroughly satisfied. As a result, the Intelligent Tutor was built to promote consciousness-raising on the one hand, by providing a remedy for what the student does not know, and awareness-raising on the other by providing clues about what the student was supposed to know. A further step in raising knowledge was taken by showing the parse tree upon successful completion of the assignment, allowing the student the opportunity to learn something explicitly that they might already know implicitly.

The post-test demonstrated an average reduction in error rate of 83 percent across the three student samples as compared to the pre-test (Taiwan, Australia, and the Emirates). Taiwanese students had the best results (94 percent error reduction rate), followed by Australian students (85 percent error reduction rate), UAE students (79 percent error reduction rate), and finally, international English language learners in Australia (73 percent error reduction rate).

This research also indicated that artificial intelligence may help learners of English as a second language correct their L2 errors. This is a major step toward a deeper understanding of the mechanisms of second language acquisition and teaching.

4.2.1. Pronunciation correction

Computer-Aided Pronunciation Training (CAPT) is a subfield of CALL which assists learners with their pronunciation. A platform employing CAPT performs tasks such as speech recognition and pronunciation error detection. Phonemic errors – substitutions of a phoneme with another, similar phoneme – are quite easy to categorize. Prosodic errors present more difficulty due to the frequently subtle nature of the error. Artificial intelligence has been employed to assist with these tasks, with different approaches being developed and trialled (Nazir et al., 2019).

Traditionally, posterior-probability[9] methods have been used for mispronunciation detection (ibid.). One such likelihood-based strategy is the Goodness of Pronunciation (GOP)

[8] A parse tree is a diagrammatic representation of the structure of a sentence. Outside of the field of computational linguistics, the term syntax tree is more commonly used.
[9] In Bayesian statistic, posterior probability is the revised probability of an event A occurring after event B has occurred (Brownlee, 2019).

(Witt and Young, 2000) measure, which uses a known orthographic transcription and a set of hidden Markov models (HMMs)[10] to determine the likelihood of the acoustic segment $O^{(q)}$ corresponding to a phone q. The GOP scores are calculated for each phone based on these statistical methods, and finally a threshold is applied to each score to reject mispronounced phonemes. The height of the threshold is determined by the required level of strictness (ibid.).

Although they recognize pronunciation quality well, the downside of probability-based methods is that they cannot determine the nature or correct location of the error (Nazir et al., 2019). Other statistical methods were developed for this purpose; however, most require human intervention to extract acoustic features. By contrast, neural networks can extract features without human input and can do it more accurately. Nazir et al. (2019) showed that a deep convolutional neural network outperforms handcrafted methods to detect mispronunciation.

4.3. Adaptive learning and difficulty adjustment

Language learners, and learners in general, learn best when the learning material they use is tailored specifically to their skills, language level and other factors, such as age, goals, styles and affective states (Pandarova et al., 2019). However, this presents a lot of work for the teacher. Not only do they have to put in the effort to understand what a student knows, but they also have to keep constant track of how well they are advancing. Most language classes are too big, and teachers can rarely put in the time to adapt course material for every individual student. Another issue is that teachers' judgements about their students' knowledge and skills, can be too subjective to be accurate, even with the best of intentions. In recent years, there has been a push to develop technologies that use AI to dynamically adjust learning material to suit the student's abilities, including artificial-intelligence-powered intelligent language tutoring systems (ILTSs), which can acquire and analyse data to make adjustments to the learning process (Shute and Zapata-Rivera, 2012; Slavuj et al., 2016, in Pandarova et al., 2019). A part of this technology deals with error analysis and error-specific feedback, such as the previously mentioned Intelligent Tutor (Dodigovic, 2007). However, this technology can be applied to long-term performance analysis and take into account other variables, such as the learner's age and goals, and even affective factors (Brusilovsky and Millan, 2007; Slavuj et al., 2016, in Pandarova et al., 2019).

[10] A good primer on Hidden Markov Models and their use in speech recognition is Lawrence Rabiner's *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition* (1989).

Pandarova et al. (2019) conducted a pilot study in preparation of an intelligent language tutoring system for practicing tenses in English. The study focused on the selection and sequencing of learning materials which are sensitive to the learner's developing language ability. The learning materials for this study consisted of cued gap-filling items (CGFIs) targeting English tenses. Each item consists of a sentence with a one-word gap and a bracketed cue after the gap, for example:

The Taj Mahal____________ (build) around 1640.

The adaptation to individual learner needs was achieved by dynamically matching the difficulty of the learning content to the learner's current level of ability. Pandarova et al. (2019) refer to this process as dynamic difficulty adaptation (DDA).

The goal of DDA is not merely assessment but also the promotion of learning and motivation. If the material is too easy, no learning will occur, and the student may become bored and lose motivation. Conversely, if the student finds the material too difficult, they will not be able to learn anything and may be discouraged as a result. Therefore, the optimal learning material, and one which will motivate the student to continue learning, is that which challenges them but allows them to succeed. This is not a new idea in learning theory and relates to Vygotsky's (1978) zone of proximal development, and Krashen's (1985) input hypothesis (Pandarova et al., 2019). Studies (Fritts and Marszalek, 2010; Martin and Lazendic, 2018, in Pandarova et al., 2019, p. 344) showed that "DDA can lead to higher achievement, test-relevant motivation and engagement, as well as to more positive subjective test experiences and lower anxiety levels than non-adaptive tests".

However, Pandarova et al. (2019) noted that studies on the precise effects of DDA in digital learning environments have so far been scarce and have shown mixed results, possibly indicating that DDA's effectiveness varies between learning domains and task types. They also indicated the urgent need for more empirical research into optimal DDA algorithms that target multidimensional learning domains (like the tense system in English). DDA has so far been applied mostly in intelligent language tutoring systems focusing on vocabulary and learning skills (Heilman et al., 2010; Sandberg et al., 2014; Sung and Wu, 2017, in Pandarova et al., 2019) and less so in the field of grammar, with Pandarova et al. (2019) pointing out just two examples (Moeyaert et al., 2016; Zapata-Rivera et al., 2017) which focus on formal accuracy only.

A basic prerequisite for DDA is that the materials used have a known difficulty level. Materials are often rated by humans (educators, learners, and so on), which has the potential downside of subjectivity or bias. Other times, items are objectively rated using real learner performance, but this may be infeasible when a large number of exercise items is required. Pandarova et al. (2019) proposed an alternative approach, employing machine learning to predict difficulty objectively. They gave 9th and 10th grade students (14-15 years of age) in Germany a written test consisting of cued gap-filling items. The test produced data points which were categorized into "correct" and "incorrect" answers. The linguistic features tested were extracted digitally. Machine learning was then employed to analyse the difficulty of the extracted linguistic features. The results showed that, as was expected, the type of verb/semi-modal required had a large impact on the difficulty of the task. The effect of epiphenomenal[11] features was also significant, and the interaction between tenses/semi-modals and epiphenomenal features. Additionally, the frequency of the lexical verb form, lemma or type also had an effect. In the future, the ability to correctly predict item difficulty based on these features would eliminate the need for subjective ratings or expensive pilot testing (ibid.).

4.4. Classroom observation

Classroom observation is the process of observing a teacher in the classroom and providing structured feedback on their work. However, according to Khan and Zualkernan (2020), classroom observation is costly and may suffer from human observer bias, and it is unreliable because teachers may change their behaviour if they know they are being observed. Partial automation of this process would go a long way towards increasing the quantity and consistency of feedback provided to teachers, which could contribute to improved education. Khan and Zualkernan's paper (2020) describes the design and implementation of a Convolution Neural Network (CNN) that categorised classroom activities based on the Stallings Classroom Snapshot using audio data from class observations. Professor Jane Stallings invented the "Stallings Classroom Snapshot" instrument in the late 1970s to perform research on the performance and quality of teachers at the primary school level in the United States. In this method, a classroom observer records his observations using a standardised coding sheet after taking a 15-second 360 view of the classroom every 5 minutes or so for a 50-minute lesson. The coding is mainly focused on the activity in which a teacher is participating, the resources that are being used, and the size of student groups in which the teacher is working. The method

[11] Pandarova et al. (2019) list voice, polarity, subject-verb agreement, word order, adverb placement and (ir)regular lexical verb morphology as epiphenomenal grammatical features which were targeted in the analysis.

assesses the teacher's use of instructional time, materials, including information and communication technologies, core pedagogical strategies, and student's engagement.

The data consisted of 121 videos, which were divided into approximately 1-minute-long audio episodes, taken from different schools. The data was used to train a CNN to sort each audio episode into one of six Stalling categories: Classwork, Classroom Management, Lecture/Demonstration, Practice & Drill, Discussion/Question and Answer, and Reading Aloud. The highest accuracy achieved by the model was 89.97 percent. Their results show that a Convolution Neural Network can be trained to perform practical classification for classroom observations of teacher activities. Despite the fact that the model outperformed previous models, the authors note that some of the Stalling categories need better discrimination, and that the data set should be more balanced. Nonetheless, these CNN models can be used on cell phones to provide quick response times, which can assist teachers providing them with feedback.

4.5. Predicting learner performance

Performance prediction is important for the study of the learning process, and as a way to improve curriculum designs and student outcomes. It could be used to, for example, identify students who are at risk of academic failure. Additionally, being able to understand which variables have the biggest impact on student performance could be a significant help to improve current approaches to teaching (Musso et al., 2013). Predicting students' academic performance involves a number of factors from various theories on learning, and typically a traditional statistical approach – such as discriminant analysis and multiple linear regressions -- to account for these factors (ibid.). Predicting performance purely by statistical methods has been criticised as inaccurate (Everson, 1995; Garson, 1998, in Musso et al., 2013). Therefore a different approach using artificial neural networks can be found in recent literature, and preliminary research showed that it boasts increased accuracy (Everson et al., 1994; Hardgrave et al., 1994; Perkins, Gupta, Tammana, 1995; Weiss & Kulikowski, 1991, in Musso et al. 1993). ANNs can be used to assess large amounts of data and provide a continuous evaluation of student performance.

The factors involved in the complex problem of academic performance are not clearly understood and they often interact with each other in a non-linear fashion. ANNs have proven to be able to deal with these problems very well, allowing researchers to work with a large number of variables and exploit their relationships without the typical parametric constraints.

This would enable researchers to gain a deeper understanding of the factors that contribute to academic results, allowing for the identification of students most in need of support (Musso & Cascallar, 2009a; Boekaerts & Cascallar, 2011, in Musso et al., 2013).

One use of performance prediction is the identification of gifted students. A study conducted in Croatia (Pavlin-Bernardić, Ravić & Matić, 2016) investigated the usefulness of neural network application to such a usage scenario. The authors point out the difficulties in identifying gifted children in Croatian schools: many schools do not systematically identify gifted children, and many often do not have a school psychologist. Further, sometimes the teachers' assessments are used in the identification process, although the objectivity of this has been brought into question (Gagné, 1994, in Pavlin-Bernardić, Ravić & Matić, 2016).

Pavlin-Bernardić, Ravić & Matić (2016) explored how ANNs could be used to identify gifted children when not all of the recommended data is available, since neural networks can work from incomplete and missing data (refer back to Section 2.1.2). They sampled data from 221 elementary school students from an elementary school in Zagreb who were tested in the fourth grade. The categories used as input variables for ANNs were as follows:

1. Teacher's nominations, based on a giftedness scale from Koren (1989, in Pavlin-Bernardić, Ravić & Matić, 2016) consisting of six subscales: general intellectual abilities, creative abilities, specific school abilities, management, artistic and psychomotor abilities.
2. Peers' nominations, using a scale consisting of the six subscales listed above. The children had to nominate three of their peers for each category.
3. School readiness, from the assessment by the school's professional team obtained before the first grade.
4. Grades, obtained for Croatian language, mathematics, natural science and foreign language.
5. Parents' level of education.

In total, there were 23 input variables. As output variables, the children's results on Standard Progressive Matrices (Raven, 1995, in Pavlin-Bernardić, Ravić & Matić, 2016) were used. The researchers used two different criteria to classify students as gifted or not gifted: the first, stricter criterion placed children in the gifted category if they scored in the 95th centile or above; the less strict criterion categorized children as gifted if their results were in the 90th centile or above. The reason for this was to see how the neural network performed with a broader criterion (Pavlin-Bernardić, Ravić & Matić, 2016).

The sample was randomly divided into two subsamples: one dataset was used to train the neural network, and the other was used to test it. For the 95th centile criterion, the best model the researchers were able to obtain produced a 100% hit rate for non-gifted children, and a 75% hit rate for gifted children. For the 90th centile criterion, the hit rates were 94.7% for non-gifted children, and 66.7% for gifted children. Thus, the function for the stricter criterion proved to be more accurate.

The researchers also performed sensitivity analyses to determine which input variables were most important in determining the outcome. The five most sensitive variables were:

- foreign language grade (end of the school year)
- children's nominations: school abilities
- children's nominations: general intellectual abilities
- mathematics grade (1st semester)
- mother's education

Additionally, Pavlin-Bernardić, Ravić & Matić (2016) performed a traditional statistical discriminative analysis to compare its accuracy to the neural network method. The hit rate of the discriminative analysis for the 95th centile criterion was 96.4% for non-gifted and 42.3% for gifted children; for the 90th centile criterion, it was 96.6% for non-gifted and 40% for gifted children. Accordingly, the neural network was determined to be more accurate (ibid.).

The authors found the results encouraging but noted that further tests needed to be performed before the results could be generalized and used for the practical purpose of helping identify gifted students. They also noted that their neural network model was in no way meant to replace intelligence tests or school psychologists but that it could be a potentially useful tool to assist in this area (ibid.).

4.6. Flipped classroom aided by deep neural networks

There are several issues with the way English in universities is usually taught: the students are largely inactive, passively taking notes, and there are few opportunities to actively use the language (Chang, 2021). The flipped classroom model presents a potential solution, but it has not been broadly applied, and there is still a lack of empirical research on its effectiveness.

Chang & Nazir (2021) presented a method to assist flipped classroom college English teaching using big data[12] and deep neural networks. Flipped classrooms are a relatively new concept of teaching, overturning the typical classroom model. The term was introduced in 2000 by Rach Pratt (in Chang & Nazir, 2021), and it refers to the reversal of the typical classroom model whereby knowledge is taught in class and internalised after class through homework and practice. In the flipped classroom model the knowledge is taught before class and the class is used for the internalisation of knowledge.

More specifically, the flipped classroom process that was the object of this study went as follows: learners used digital resources for autonomous learning before class and conducted interactive activities between each other and the teachers. This allows the students to gain an understanding of the subject and discover any problem areas or deficiencies they may have. They also helped each other solve problems and exchange opinions. For flipped classroom education, the production of digital materials is very important; in this particular case micro-videos were used. These are short video clips summarizing the key aspects to be learned.

The digital materials for individual learners can be personalised to optimise learning thanks to big data collected and processed by neural networks. There are certain attributes of interactive learning materials, such as resource styles and interaction methods that may have a relationship with certain attributes of learners, such as cognitive levels, gender, learning goals, etc. The object of the study (2021) was to find which of these attributes correlate strongly based on collected data and make predictions on which materials would be most preferable for every student. Later the recommendations outputted by the neural network were verified by experiments using several pools of data, including public datasets consisting of course data, learner information and learner behaviour data, as well as a sample of 230 students from two classes of second-year English majors from a Chinese university. Data was collected on all students to analyse their autonomous learning abilities and English academic performance. Experimental results showed that the machine learning algorithm was feasible and effective in predicting personalised materials. Chang & Nazir (2021) concluded that the trained neural network can help improve the teaching quality of flipped classrooms for college English.

[12] Big data are datasets that are too large or diverse to be managed and processed by traditional databases (Snijders, Matzat & Reips, 2012).

4.7. Impact of AI on teachers and students

With the proliferation of advanced technologies in the classroom, it is important to understand their impact on teachers and students. While advanced neural networks are still very much an active area of research and most use cases are in the early experimental phases, the answer to this question may yet be difficult to find empirically. We can, however, turn to studies conducted on the impact of less advanced computer technologies and rudimentary AI on the classroom.

When Schofield, Evans-Rhodes and Huber conducted their study (1990), microcomputers had proliferated at an incredible rate in both elementary and secondary schools during the preceding decade. Although the incredible speed with which microcomputers were installed in schools was clear, the impact on students and teachers was less so. In fact, our understanding of how this change affected classroom organization and function was extremely limited, and some research indicated that instructional software's impact might differ significantly from what its creators had expected.

One of the purposes of the study was to see how using microcomputers as intelligent instructors affected classroom structure and function. Intelligent computer-based tutors should be able to follow what a student is trying to do, identify the difficulties the student is having, and provide guidance that is relevant to those difficulties, offering individually tailored learning experiences that progress at a pace determined by the student's abilities (Anderson, 1984, in Schofield, Evans-Rhodes & Huber, 1990).

The data for this study was obtained during a two-year period (1985-1987) at an urban high school in the United States with 1300 pupils from various backgrounds. Intensive qualitative classroom observations and frequent interviews with teachers and students were the two main techniques of data collection. The artificially intelligent geometry proof tutor that was used in this research was called GPTUTOR, which consisted of three parts. The knowledge required to create geometry proofs was covered in the first part. The second part was the tutor, which contained material for teaching students, and the third part was the interface, which allowed students to communicate with the computer through keyboard or mouse. When a student was working with a tutor, several aid and review choices were accessible, whether at the student's request or when the student made too many errors.

The use of computer tutors seemed to alter the amount of attention provided to students of various abilities. In a traditional classroom setting, the teachers tended to call on the brighter students to solve the problems on the board or answer questions. It enhanced the amount of time dedicated to individuals who were having difficulties. Not surprisingly, in these situations teachers tended to call disproportionately on the more advanced students. This saved considerable time, raised the probability of a correct answer, and saved the poorer students the embarrassment of making mistakes in public. When employing the computer tutor, the slower students were frequently given far more attention than the stronger students. Because students working on their assignments were often unaware of who the teacher was working with, such attention was unlikely to be humiliating (ibid.).

In the computer tutor classes, there was a second change in the teachers' behaviour. The instructors, in particular, were less authoritative figures and more collaborators than they had been earlier. Instead of addressing the entire class in a formal manner, the teacher tended to work with students one-on-one (ibid.).

Finally, the use of computer tutors resulted in adjustments in the grading methods of the teachers. Teachers placed a greater focus on effort when utilizing the computer tutor than they did previously. Because one of the primary benefits of the computer tutor was that it allowed students to work at their own pace, evaluating everyone against the same level of achievement no longer appeared reasonable (ibid.).

With the introduction of the computer tutor, there were also substantial behavioural changes in students. The increase in student participation and effort was one of the most noticeable changes in the classrooms utilizing the GPTUTOR. This shift was reflected in increased task duration, obvious increases in perceived focus, and other indicators. In fact, when asked how utilizing the computer tutor affected their behaviour in post-use interviews, the most prevalent response was an increase in effort.

In many classrooms that did not utilize computer tutors, it was normal for a significant number of students to spend ten to fifteen minutes every period conversing about topics unrelated to the lesson. This took up a significant amount of the 45-minute class sessions. Many students began working on their proofs almost immediately after starting to use the GPTUTOR, a scenario that was almost never witnessed in the control and comparison classes. Furthermore, they regularly worked after the bell rang, which is quite unusual in other geometry classes.

When asked why they started arriving to class early after the computer tutors arrived, about 40% of the students spontaneously stated that it was because of the competitiveness. Furthermore, when asked explicitly if the introduction of the tutors had increased the level of competition in the classroom, the majority of students said yes. Ironically, the tutors were meant to allow students to progress at their own pace, but produced a climate that encouraged student rivalry.

Students in typical geometry classrooms never have the opportunity to move far ahead of or behind one another, as they may in a class with computer tutors. The students' ability to communicate simply and clearly about how far they had progressed was aided by the fact that the tasks were numbered, and they were sat near enough to each other that they could chat (ibid.). Because they enjoyed using the computers, students may have been inspired to work more when using the computer tutor. In interviews, the vast majority of students stated that using the GPTUTOR was more enjoyable than traditional learning, and some even related this greater enjoyment to improved motivation. When working on the computer, several kids reported delight in their relative independence from direct adult supervision. Furthermore, many students had a strong association between computers and playing video games, which predisposed them to prefer working on computers and drove many of them to work in a productive yet joyful manner despite the competition. A "success sound" that happened when pupils completed a proof was one element of the program that may have contributed to this sensation of playing while working. Much more significant proved to be the pupils' sense of personal challenge, which is a common element of many games.[13] Several students mentioned this sense of challenge in their interviews. Furthermore, pupils reacted favourably to the computers because they felt free to express their frustration to them in a way that they couldn't do with a teacher without breaching strict rules (ibid.).

Another element contributing to students' interest in computer work and their perception that it was enjoyable was most likely a feeling of being more at ease in computer tutor classes due to a reduced fear of humiliation. Working on the computer tutors, students agreed, was less likely to be humiliating than doing geometry the usual manner. Students are frequently called upon to perform in front of others in traditional geometry classrooms, as the teacher has them do board work, answer questions at their seats, and so on. This can be humiliating for students who are behind or just lost, as their problems are typically made public (ibid.).

[13] The introduction of elements similar to games into other areas of activity is known as gamification and it is a founding feature of some language learning apps, such as Duolingo. Although the novelty of using computers has certainly worn off for students who use smartphones more powerful than the most powerful supercomputers in 1990, gamified elements are still considered very effective motivators.

Because the computers were set up so that pupils couldn't see each other's displays, mistakes on the computer tutor were more likely to be private.

Furthermore, while professors circulated and commented on students' computer work, the fact that students were facing different directions and working on separate issues made it less probable that others were seeing these comments. Moreover, because the tutor featured a number of support tools that allowed students to ask the computer for a review of prior material or ideas on how to solve a problem, students who were wary of appearing to require assistance had a non-human source of aid easily accessible (ibid.).

There is a lack of similar studies on the effect of digital technology and AI on language learners. An argument could be made that, since the GPTUTOR assisted students in learning geometry, we should not expect the same results to carry over for language learning. However, although the subject is very different, the learning environment is similar enough that we can infer that at least some of the results would apply to language learning as well.

It should be noted that this study is very dated, being over three decades old at this point. AI has advanced leaps and bounds since then. However, it does offer us several important insights. Firstly, any new technology introduced in the classroom is likely to have both unexpected and unintended effects. Secondly, it shows us that, despite the technology now being old, and despite evidence of it being beneficial, it has not, in the decades that followed, exactly changed the paradigm when it comes to the classroom environment. Remote learning due to COVID-19 notwithstanding, the classroom of today looks very much like the classrooms of the pre-digital era, with digital learning materials making up very little of the curricula in typical schools[14], and even so, the digital materials are often used in a way that does not take full advantage of the format. For example, a multiple-choice exam filled out on Google Forms is not very different from a written exam. Essentially, it could be argued that that the educational system is slow to adapt or resistant to changes, especially those that might present a paradigmatic upheaval in methodology.

The study of intelligent tutoring is the study of a potentially revolutionary educational breakthrough that is on the verge of becoming a technological and financial reality, and it is a breakthrough that we have evidently been chasing for decades.

[14] https://www.rand.org/blog/2020/05/new-teacher-survey-shows-that-digital-materials-were.html

However, it is crucial to realize that the effective use of artificially intelligent tutors may result in or necessitate significant behavioural changes in both teachers and pupils. Intelligent tutoring, for example, is likely to involve a role change on the part of teachers than the use of traditional drill and practice applications. However, we have limited information on what these adjustments could include.

Such information appears to be crucial for two reasons. First and foremost, it may be beneficial to those seeking to train instructors to use computers in the classroom as effectively as possible. Second, it may reveal unintended effects of computer use, allowing educators to determine whether and how to utilize computers with a better understanding of the full scope of their decisions.

5. Discussion

AI research is a decades-old field of study, yet it still holds an incredible amount of potential. It promises to transform our entire world, from the way we interface with technology to how we work and even how we learn. For linguistics, we showed that AI shows great potential for use as a tool to prove linguistic theories experimentally, as well as a starting point for the understanding of some aspects of human cognition. As an example, we can look at the way Quillian's TLC was the origin for the research into the concept of priming in cognitive science and psycholinguistics. On the theoretical side, the insights gained from connectionist accounts of language acquisition could be integrated into generative linguistics to investigate universal grammar. Advanced language models and enormous amounts of data allow us to construct a fully empirical account of language.

Potential practical applications of artificial intelligence are varied: from language learnability analysis, to error correction, to dynamic difficulty adjustment, to performance prediction, artificial intelligence and neural networks in particular have demonstrated that they can find their place in the learning process.

Contemporary neural networks show an uncanny ability to learn a language, and to map one language onto another. The question is not: can AI assist in language acquisition? Or even: can AI replace human teachers? As long as current trends hold, and we have no reason to believe they will not, AI will certainly prove capable. And if the research we have discussed in this paper is anything to go by, artificial intelligence could be worked into almost every step of the learning process. Yet, there are reasons for us to be sceptical. As was already noted in this paper, computers have been used in classrooms for decades now. However, the classroom of today certainly bears a lot of resemblance to the classrooms of the past, even with tablets and smartboards. We still teach and learn in much the same way as we did a decade or two ago. The technology that has found its way in the classroom has been of use, certainly. However, it has not resulted in a radical transformation of the learning process. The essence of how we learn and teach has not changed. Historically there has been a gap between predicted "revolutionary" implementations of technology and the reality of the situation (Schiff, 2021). Schiff (2021) identifies several causes for this in the educational context. Firstly, there are failures to consider implementation: cost, teacher training and development, complementary curricular and pedagogical support, necessary structural changes etc. Secondly, at the root of technological predictions, there is the faulty assumption that new technologies will be used in the way their creators intended, but, as Schiff (2021) notes, teachers and students are bound to strategically adapt and repurpose tools in creative and surprising ways. The last point is the value of teacher-student interaction (ibid.). It seems improbable that human teachers will be taken out of the equation entirely, because they can stimulate the affective aspects of learning in a way that a machine simply cannot.

However, the lack of ESL teachers in some countries (such as China) reveals a real need for new technologies and methodologies and presents an avenue for AI to prove itself in large-scale practical use. Additionally, the disruptions caused by the COVID-19 pandemic have made distance education a mainstay in many students' lives. Distance education radically re-envisions both the phenomenological experience of education as well as the teacher-student ratio (ibid.). It seems probable that the introduction of artificial intelligence in the learning process will, over time, lead to language teachers managing a large number of students with the help of AI, and perhaps even distance education will take over as the defining educational paradigm of the future, providing that educational institutions – and society – will embrace it.

It has been shown that it is quite possible to train AI to predict learners' performance, identify problematic students, and conversely, detect which students are gifted, all with an uncanny accuracy that matches or exceeds other methods. It is clear that AI does not suffer from the all too human fault of bias. But this requires us to place our trust in advanced algorithms we may not fully understand. Apart from rating students on their performance, both past and predicted, neural networks are capable of observing teachers as they work – and perhaps in the future keeping constant track of their performance as well. Any new technology is also a source of anxiety and doubt. As teachers and educators defer a part of their responsibility to machines, they also lose a measure of control.

The author of this paper had the opportunity to listen to a speech by Christian Heilmann (Shift Conference, 2017), principal program manager for Microsoft, during the 2017 Shift Conference in Split entitled *The Soul in the Machine: Developing for Humans*. In the speech, Heilmann presented a vision for a humanist AI-powered future, built on the condition that AI would not marginalise human labour, or take over the jobs that humans are good at but would do the rote tasks that humans find boring and unchallenging. Any implementation of AI technology in the educational process must be built on a humanist basis. AI should be created to assist teachers, not to replace them. For learners it should allow for more options on how they want to learn, instead of forcing them down a specific path. All this can be accomplished if these systems are designed by consulting the needs of educational institutions and the communities they serve.

6. Conclusion

This work has traced the historical development of AI, especially as it pertains to the fields of linguistics and language acquisition. Artificial intelligence has advanced considerably over the years, and modern neural networks are able to process large amounts of data very quickly and efficiently, in many cases more effectively than traditional statistical methods. This makes AI a good candidate for many tasks related to language learning. It has shown potential in many use cases, such as error correction, personalised learning materials, difficulty adaptation, and performance prediction.

However, despite its potential it is not without its problems. In the first instance, there are institutional barriers to any large-scale adaptation of new technology in education. The cost of the technology itself must be considered and its impacts on teachers, learners, and the community. Any new technology is bound to change the behaviour of people in unexpected ways, and there is a possibility of unforeseen negative effects. As we have noted previously, the teacher-student relationship is incredibly important, and the introduction of AI into the learning process may introduce feelings of alienation, particularly in the context of distance education. Artificial intelligence, and any kind of new technology for that matter, should not be implemented haphazardly. We must put the needs of human beings first, and AI should be implemented in a way that allows us to retain all that makes us human.

7. Bibliography

Brownlee, J. (2019). *A Gentle Introduction to Bayes Theorem for Machine Learning.* https://machinelearningmastery.com/bayes-theorem-for-machine-learning/. Accessed on September 11, 2021.

Buckner, C., & Garson, J. (2019). Connectionism. In: Edward N. Zalta (ed.), *The Stanford Encyclopaedia of Philosophy (Fall 2019 Edition)*.

Chang, H., & Nazir, S. (2021). College English Flipped Classroom Teaching Model Based on Big Data and Deep Neural Networks. *Sci. Program., 2021*. doi:10.1155/2021/9918433

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Cohen, C., Higham, C. F. & Nabi, S. W. (2020). Deep Learnability: Using Neural Networks to Quantify Language Similarity and Learnability. *Frontiers in Artificial Intelligence*, 3, 43. doi:10.3389/frai.2020.00043

Dargan, J. (2019). *Artificial Intelligence: The Angel of Death for Foreign Language Teachers.* https://chatbotslife.com/artificial-intelligence-the-angel-of-death-for-foreign-language-teachers-cbff644a4967. Accessed on April 20, 2021.

Dodigovic, M. (05 2007). Artificial Intelligence and Second Language Learning: An Efficient Approach to Error Remediation. *Language Awareness - LANG AWARE*, 16, 99–113. doi:10.2167/la416.0

Du Boulay, B. (2016). Artificial Intelligence as an Effective Classroom Assistant. *IEEE Intelligent Systems*, 31, 76-81. doi:10.1109/MIS.2016.93.

Ellis, R. (2015). *Understanding Second Language Acquisition* (2nd Edition). Oxford: Oxford University Press.

Gagniuc, P. (2017). *Markov Chains: From Theory to Implementation and Experimentation.* doi:10.1002/9781119387596.

Heift, T. (2017). History and Key Developments in Intelligent Computer-Assisted Language Learning (ICALL). In S. L. Thorne & S. May (Reds), *Language, Education and Technology* (pp. 289–300). doi:10.1007/978-3-319-02237-6

Horst, S. (2003). The Computational Theory of Mind. In: Edward N. Zalta (ed.) *The Stanford Encyclopaedia of Philosophy (Summer 2019 Edition).*

Lotze, N. (2018). *Goodbye to classroom teaching? Artificial intelligence in language learning.* https://www.goethe.de/en/m /spr/eng/gls/2129062 9.html. Accessed on June 14, 2021.

Mair, C. (2008). Tough-Movement In Present-Day British English - A Corpus-Based Study. *Studia Linguistica, 41*, 59 - 71. doi:10.1111/j.1467-9582.1987.tb00773.x.

Marty, F (1981). Reflections on the use of computers in second language acquisition. *System,* 9 (2), 85–98.

Musso, M. F., Kyndt, E., Cascallar, E. S. & Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research,* 1, 42-31. doi: 10.14786/flr.v1i1.13

Nazir, F., Majeed, N., Ghazanfar, M. & Maqsood, M. (2019). Mispronunciation Detection Using Deep Convolutional Neural Network Features and Transfer Learning-Based Model for Arabic Phonemes. *IEEE Access*. 52589 - 52608. doi: 10.1109/ACCESS.2019.2912648.

Pace-Sigge, M. (2018). *Spreading Activation, Lexical Priming and the Semantic Web*. doi:10.1007/978-3-319-90719-2

Pandarova, I., Schmidt, T., Hartig, J., Boubekki, A., Jones, R.D. & Brefeld, U. (2019). Predicting the Difficulty of Exercise Items for Dynamic Difficulty Adaptation in Adaptive Language Tutoring. *International Journal of Artificial Intelligence in Education,* 29(3), 342-367. doi:10.1007/s40593-019-00180-4

Pater, J. (2019). Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language, 95* (1), e41 - e74.

Pavlin-Bernardić, N., Ravić, S. & Matić, I.P. (2016). The application of artificial neural networks in predicting children's giftedness. *Suvremena psihologija*, *19* (1), 49-58. doi:10.21465/2016-SP-191-04

Rescorla, M. (2019). The Language of Thought Hypothesis. In: Edward N. Zalta (ed.), *The Stanford Encyclopaedia of Philosophy (Summer 2019 Edition).*

Rosenberg, R. S. (1975). Artificial Intelligence and Linguistics: A Brief History of a One-way Relationship. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, *1*, 379 – 392.

Schiff, D. (2021). Out of the laboratory and into the classroom: the future of artificial intelligence in education. *AI & Soc 36*, 331-348. doi:10.1007/s00146-020-01033-8

Schofield, J., Evans-Rhodes, D. & Huber, B. (1990). Artificial Intelligence in the Classroom: The Impact of a Computer-Based Tutor on Teachers and Students. *Social Science Computer Review - SOC SCI COMPUT REV*, *8*. 24-41. doi:10.1177/089443939000800104.

Sejnowski, Terrence J. (2020). The unreasonable effectiveness of deep learning in artificial intelligence. *Proceedings of the National Academy of Sciences Dec 2020*, 117 (48). 30033-30038. doi:10.1073/pnas.1907373117

Shift Conference. (2017, June 9). *Shift 2017: The Soul in The Machine - Developing for Humans - Christian Heilmann (Microsoft)* [Video]. YouTube. https://youtu.be/BEcTRqPW4MM

Snijders, C. , Matzat, U., Reips, U. (2012). "Big Data": Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science, 7*. 1-5.

Swad, S. (1992). *Rosetta Stone* [Software]. Arlington, Virginia, USA: Fairfield Language Technologies.

Szczepański, M. (2019). *Economic impacts of artificial intelligence.* European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637967/EPRS_BRI(2019)637967_EN.pdf. Accessed on September 29, 2021.

Toffler, A. (1970). *Future shock*. New York: Random House.

Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, s2-42* (1), 230–265, doi:10.1112/plms/s2-42.1.230

Turing, A. M. (1980). Computing Machinery and Intelligence. *Creative Computing 6*, 44-53.

Witt, S. & Young, S. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, *30*, 95-108. doi: 10.1016/S0167-6393(99)00044-8.

Abstract

This paper discusses artificial intelligence within the context of language acquisition. There are two relevant aspects to this topic. Firstly, there is the matter of the theoretical and practical applications of artificial intelligence in building and empirically verifying linguistic theories. The second aspect relates to the use of artificial intelligence in the language classroom. This paper presents an overview of some current and experimental applications of AI to assist learners and teachers and discusses the benefits, drawbacks and potential future implications of this technology.

Key words: artificial intelligence, neural networks, language acquisition, language teaching, technology in education

Sažetak

Ovaj rad se bavi umjetnom inteligencijom u kontekstu usvajanja jezika. Dva su relevantna aspekta ove teme. Prvo je pitanje teoretskih i praktičnih primjena umjetne inteligencije u stvaranju i empirijskoj provjeri lingvističkih teorija. Drugi aspekt se odnosi na uporabu umjetne inteligencije u učenju jezika. Ovaj rad predstavlja pregled nekih postojećih i eksperimentalnih primjena umjetne inteligencije u pomaganju učenicima i učiteljima u učenju i poučavanju drugog jezika, te raspravlja o prednostima, manama i potencijalnim budućim implikacijama ove tehnologije

Ključne riječi: umjetna inteligencija, neuronske mreže, usvajanje jezika, podučavanje jezika, tehnologija u obrazovanju

Obrazac A.Č.

SVEUČILIŠTE U SPLITU

FILOZOFSKI FAKULTET

IZJAVA O AKADEMSKOJ ČESTITOSTI

kojom ja Mirko Ćalušić, kao pristupnik/pristupnica za stjecanje zvanja magistra/magistrice anglistike i filozofije, izjavljujem da je ovaj diplomski rad rezultat isključivo mojega vlastitoga rada, da se temelji na mojim istraživanjima i oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio diplomskoga rada nije napisan na nedopušten način, odnosno da nije prepisan iz necitiranoga rada, pa tako ne krši ničija autorska prava. Također izjavljujem da nijedan dio ovoga diplomskoga rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Split, 24.11.2021.

Potpis

OBRAZAC I.P.

IZJAVA O POHRANI ZAVRŠNOG / DIPLOMSKOG RADA U DIGITALNI REPOZITORIJ FILOZOFSKOG FAKULTETA U SPLITU

STUDENT/ICA	Mirko Ćalušić
NASLOV RADA	Artificial intelligence and language acquisition
VRSTA RADA	Diplomski rad
ZNANSTVENO PODRUČJE	Humanističke znanosti
ZNANSTVENO POLJE	Usvajanje jezika
MENTOR/ICA (ime, prezime, zvanje)	doc.dr.sc. Danijela Šegedin Borovina
KOMENTOR/ICA (ime, prezime, zvanje)	
ČLANOVI POVJERENSTVA (ime, prezime, zvanje)	1. doc.dr.sc. Mirjana Semren 2. doc.dr.sc. Danijela Šegedin Borovina 3. doc.dr.sc. Ivana Petrović

Ovom izjavom potvrđujem da sam autor/ica predanog završnog/diplomskog rada (zaokružiti odgovarajuće) i da sadržaj njegove elektroničke inačice u potpunosti odgovara sadržaju obranjenog i nakon obrane uređenog rada. Slažem se da taj rad, koji će biti trajno pohranjen u Digitalnom repozitoriju Filozofskog fakulteta Sveučilišta u Splitu i javno dostupnom repozitoriju Nacionalne i sveučilišne knjižnice u Zagrebu (u skladu s odredbama *Zakona o znanstvenoj djelatnosti i visokom obrazovanju*, NN br. 123/03, 198/03, 105/04, 174/04, 02/07, 45/09, 63/11, 94/13, 139/13, 101/14, 60/15, 131/17), bude (zaokružiti odgovarajuće):

a.) u otvorenom pristupu

b.) rad dostupan studentima i djelatnicima Filozofskog fakulteta u Splitu

c.) rad dostupan široj javnosti, ali nakon proteka 6/12/24 mjeseci (zaokružiti odgovarajući broj mjeseci)

U slučaju potrebe dodatnog ograničavanja pristupa Vašem ocjenskom radu, podnosi se obrazloženi zahtjev nadležnom tijelu u ustanovi.

Split, 24.11.2021.

mjesto, datum

potpis studenta/ice