

Research Article

Consilience of Reductionism and Complexity Theory in Language Research: Adaptive Weight Model

Chao Zhang 

College of Chinese Language and Literature, Qufu Normal University, Qufu, China

Correspondence should be addressed to Chao Zhang; xiangyuezc@hotmail.com

Received 2 June 2021; Accepted 10 December 2021; Published 31 January 2022

Academic Editor: Rosa M. Benito

Copyright © 2022 Chao Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Reductionism and complexity theory are two paradigms frequently found in language research. There exist a number of conflicts in terms of concepts and methodologies between reductionism and complexity theory, which are not conducive to creating a unified language research framework. This paper starts by discussing the adaptability of complex dynamic systems and combines cognitive processing model and artificial neural networks to construct and verify an adaptive weight model, showing that the study of reductionism is induction of high-weight elements and the study of complexity theory is a discussion of system complexity from adaptability, meaning that there is a good fit between the two frameworks. The adaptive weight model is conducive to developing a unified interpretation of language research results.

1. Introduction

Experimental research has found a correlation between two sets of variables by controlling a certain item or set of factors. Therefore, experimental research has generally been the mainstream approach to language research. In recent decades, however, a series of theories based on the idea that language development is complex adaptive systems, such as connectionism, construction grammar, emergentism, complex dynamic system theory, embrace complexity, interconnectedness, and dynamism, has caused a revolution in theory and method [1–3], and the experimental approach has faced unprecedented challenges in the field of language research. In summary, it is debatable whether a certain factor potentially affecting language development can be studied separately and to what extent language development manifests as individual differences or group rules. In the final analysis, these disputes root in different ideas that vary between reductionism and complexity theory.

Recently, the machine learning model has demonstrated the potential to solve this problem. Machine learning has been proved to be able to simulate the language processing process and solve issues of language research [4]. Machine learning algorithms embrace interconnectedness and

dynamism [5], which embodies the philosophy of complexity theory. Moreover, the machine learning model can process a single element of context in language learning, such as sentiment analysis, which embodies reductionist thinking. Therefore, machine learning models have the potential to integrate complexity and reductionism in language research. Unfortunately, few studies have explored this issue using machine learning models.

In light of these considerations, the present research integrates the two theories with a theoretical model based on machine learning algorithms, aiming to trigger reflections on controversial issues in the academic field and provide useful insights into the construction of a unified language research system in the future. In the present study, firstly, we review the main controversies between the complexity theory and reductionism and introduce the concept of complexity and the idea that adaption builds complexity, after which the relationship between language cognitive system and artificial neural network was discussed. Based on the foregoing analysis, we elaborated on several machine learning models and discussed that the artificial neural network could address the research gap of the theoretical controversies of the two paradigms, after which the adaptive weight model was proposed, and the principle of the model

was discussed in the input, output, and rule layers. Furthermore, the methods for evaluating the adaptive weight model were introduced and quantitative evaluations were conducted to evaluate the model by text data. Moreover, the consilience of the reductionist approach and complexity theory were discussed with reference to language research, machine learning, and the adaptive weight model. We conclude with the discussion on the adaptive weight model that provides a comprehensive and adaptive view of language research to address the gaps in complexity theory and reductionism.

2. Reductionism and Complexity Theory

2.1. Main Controversies between Reductionism and Complexity. Reductionism is a philosophical perspective divided into three ways, namely, ontological reductionism, methodological reductionism, and theoretical reductionism [6]. Language research largely draws on methodological reductionism, which posits that the explanation of an object should be reduced to the smallest possible entity, i.e., the most accurate way of understanding an object is to decompose it to examine its parts [2, 7]. Reductionism focuses on the component related to language performance by controlling disturbance variables with an experimental design to investigate potential factors individually to conclude causality. In language research, the concept that establishes a causal relationship between a single or a group of factors and learners' language performance is reductionism.

The complexity theory in language research is a meta-theory, carrying a series of interrelated theoretical frameworks, namely connectionism, emergence theory, and constructionism [8], positing that language development is a complex dynamic system with interconnected elements. In this system, changing one element leads to changes in other elements across the system. These elements are intertwined and interacted, and the system behavior emerges through self-organization [2]. Therefore, the description of a subsystem of reductionism cannot describe language as a whole, and it is impossible to remove a factor to study its effects on language development alone.

Despite incompatible ideas between reductionism and complexity theory, a number of examples from language teaching may demonstrate that the two paradigms may be compatible. The complexity theory embraces nonlinearity because of interconnectedness, which posits that a change in one factor does not necessarily cause the system to change proportionally, meaning that predictions will be difficult [2, 9]. However, language teachers generally can predict the success of teaching and which learners will achieve higher scores in a test, meaning that they can distinguish different types of learners [10]. The orderly behavior of these complex language development systems is the concept of reductionism, showing that predictability and unpredictability, order and chaos, and simplicity and complexity are not absolutes.

2.2. Complexity. Complexity is a difficult concept to measure. Simon [11] believes that the most important characteristics of complex dynamic systems are hierarchy and

inseparability. Hierarchy means that the complex system is composed of subsystems, whereas inseparability means that the interactions of subsystems are inseparable. Therefore, Simon advocates the term "degree of hierarchy" to measure complexity. The more complex a complex dynamic system is, the more subsystems and mutual connections that it covers. Logic depth is a concept to measure how difficult a system is to construct, that is, highly ordered sequences have lower logic depth and vice versa [12]. Therefore, simple, specific measurement methods can be measured with binary operations on coding units. The more disorderly a concept and the more logical operations that it needs, the higher its complexity. Crutchfield and Yong [13] propose an indirect measure of complexity, i.e., the minimum amount of information needed to predict the future behavior of the system, in which complexity is connected to predictability. The more complex something is, the more difficult it is to predict. This scheme is similar to the "logical depth," which expresses the difficulty of the regularity of concepts that we understand by complexity.

Therefore, complexity is a construct that reflects how many levels of a system are nested, the degree of order and disorder, the ability to predict, and the attributes that can be used to understand the rules. It is straightforward to generalize laws from orderly phenomena, meaning that they can be defined as simple. On the other hand, for disordered phenomena, it is difficult to generalize laws, and these laws can be defined as complex. From the perspective of complexity theory, the complexity of these systems is because of the large number of participating factors and the existence of a large number of nonlinear relationships among factors. The more disorderly something is, the more complex it is, and the more complex it is, the more interconnected elements that it has, which makes it difficult to induce laws and make predictions. The essence of the conflict between static and dynamic, linear and nonlinear, individual differences and group laws in language research is the conflict between the concepts of reductionism and complexity theory, which is rooted in the conflict between the "simple" and "complex" elements of the language development system.

2.3. Adaptation Builds Complexity. The interconnectedness and nonlinear characteristics of complex dynamic systems are a direct cause of the inability to generalize rules [14, 15]. Since the system amplifies or reduces the impact of specific element changes on the system, various elements interweave to create the system's nonlinear characteristics. Holland [16] argues that adaptability builds complexity. The macroscopic behavior of a complex dynamic system emerges from an interaction of subsystems, i.e., the adaptive behavior of the system, which implies that humans intuitively believe that a complex dynamic system is controlled by complex behavior rules.

A learner's language development complex dynamic system is open, and the external environment itself is a complex adaptive system that changes constantly with the involvement of cognitive, social, and natural elements. However, if a certain element can be isolated, it will be linear.

For example, if a teacher changes the teaching strategy, the student will perceive this change and react. If only this element is involved in the process, the learner's behavior will be simple and predictable. However, a large number of elements are involved in the process of a learner's writing development, during which the learner adapts to changes to each element. Therefore, when the teacher changes the teaching strategy, the students may decide whether they are interested in learning topics in which past experience must be involved. Attention fluctuations and shifts, recent mood, and other predictable and unpredictable elements interact. It means that the cognitive language system adapts to each element and causes learner behavior, meaning that the impact of "teaching strategy changes" is complex and unpredictable.

3. Language Cognitive System and Artificial Neural Network

Assuming that changes to factors (such as teaching strategies) are inputs and the learners' language performance is an output, the interconnectedness of a system builds adaptation and complex processing rules. The input-output rule is the most basic rule of all complex dynamic systems, indicating that there is a potential logical relationship between the input and output [17]. The input information is output through complex adaptive rules. The complexity of cognition lies in the complexity of adaptive rules. The input may be any kind of information, such as a paragraph of text, the voice of the interlocutor, teaching strategies, changes in a social environment, feedback from another party, changes in temperature, etc. The output may be a muscle group that the brain can directly manipulate. Physical behaviors, such as controlling the vocal cords to make sounds, making a writing movement, and changing eye contact, may also be psychological changes. The output will become a new input and a new output, forming a dynamic cycle. The process of processing the input information in this dynamic cycle is the process of human cognition of the world, and its model expression is presented in Figure 1.

An information processing system (Figure 1) composed of inputs, rules, and outputs is suitable for all complex dynamic systems, representing the language cognitive process. The above process is the preliminary modeling of language cognitive processing. However, understanding the rules of brain information processing requires referring to the idea of connectionism or bionicsism, which aims to simulate brain processing mechanisms by machine learning models. Despite the cognitive model of the human brain still being controversial in the academic field, the modern machine learning model (e.g., deep learning model) derived from connectionism has surpassed the human brain in some aspects. For example, AlphaGo has defeated human Go players, and computer programs are far more efficient than human beings in natural language processing (NLP) and image recognition. Modern machine learning models can learn from a large number of inputs to enhance an algorithm and output more accurately. Therefore, machine learning

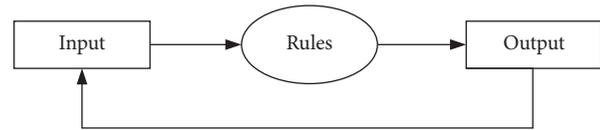


FIGURE 1: Information processing model of complex system [17].

models to explain the language cognition process has a theoretical and factual basis.

The algorithm model of machine learning is the artificial neural network, in which the information processing mechanism of the human brain as the antetype is simulated [18]. Figure 2 offers a comparison between a biological neural network and an artificial neural network. Hebb [20] found that synapses are plastic, and the connections between the neurons with higher activation frequency will be strengthened. The mainstream view in cognition science research is that neurons are the most basic units of biological neural networks, and the nerve cells generate action potentials to encode input signals with pulse frequencies. Inspired by neuron mechanisms, the artificial neural network uses artificial neurons as the operating unit and simulates the activation frequency of biological neural networks with the help of activation functions. The weight represents the strength of the connection among neurons. The iteration and combination of neurons construct an artificial neural network [21].

Cognition science and artificial intelligence research has found that there may be three working mechanisms of neural networks: (1) behaviorism, which refers to external stimuli, (2) symbolism, which is similar to computer logic operations, and (3) connectionism, which develops into various artificial neural networks. The connectionist model in the field of linguistics is different from the current "deep neural network," i.e., connectionism can explain nonlinear development using a parallel distributed processing model to simulate the distributed and parallel characteristics of language neural processing, however, it cannot simulate brain regions and circuits involved in different levels of language processing [22]. The current deep neural network originating from connectionism, involving mathematics, informatics, cognition neuroscience, logic, and other interdisciplinary science, has developed into a core model of artificial intelligence programs that can self-learn language to a certain extent. Neural networks emphasizing the interconnection of information processing are in line with the research results on semantic processing in neurolinguistics, in that language processing is the coordination of multiple brain regions rather than modular entities [23, 24], which conforms with DST from the comprehensive view of language development under the visual threshold that system behavior emerges from the interactions among various elements of the system.

The significance of understanding brain processing mechanisms (rule side) with an artificial neural network is that artificial neural networks are the simulations of brain research by neuroscientists because mathematics and logic-based methods often cannot be used to explain neural

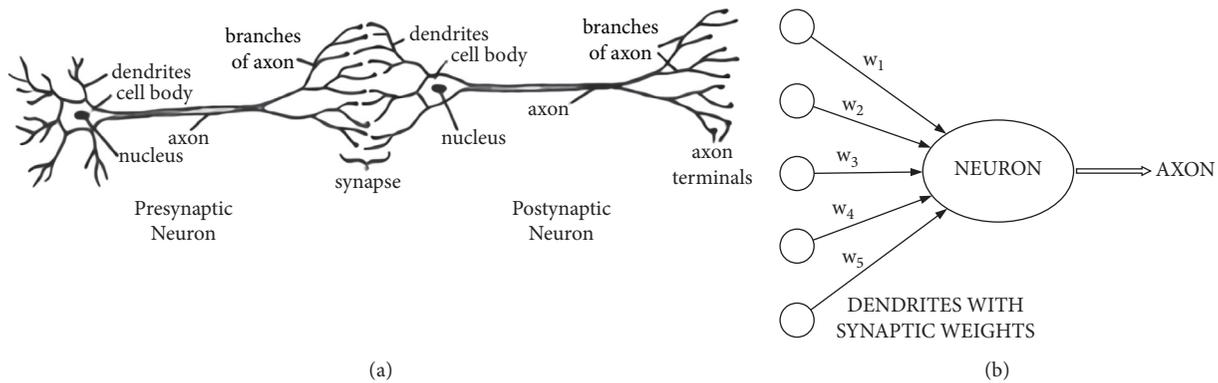


FIGURE 2: Biological neural network and artificial neural network [19].

networks that complement and rely on one another [25]. The origin of connectionism was using artificial neural networks to understand the intellectual activity, which has been supported by empirical findings. Researchers have found that the macaque brain's image recognition and the working mechanisms are similar to convolutional neural networks (an artificial neural network) [26]. However, though research is still far from fully simulating human cognition, artificial neural networks are an adequate model of the human brain mechanism that extracts human brain information adaptation and emergence mechanisms [27]. Therefore, this research uses an artificial neural network as the model for simulating brain information processing to understand the characteristics of the language development system and the process of language development. It proposes a thinking bridge for the convergence of simplification and complexity theories.

4. Adaptive Weight Model

4.1. Input and Output Layer. The input layer represents how the human brain processes the external world into information (neural impulse signals), whereas the output generates the behavior shown by the body and the changes of the output on the internal mental structure (such as memory) for language learning. Cognitive psychology research has suggested that the storage (memory) of input information is divided into three parts, namely short-term memory, working memory, and long-term memory. Three-level memory is a theoretical model of storage location rather than an algorithmic layer. The brain processes memory and language information using the strength of neuron connections. The neuron connection strength is derived from the concept of weight in computational psychology. There are two means of storing knowledge in artificial neural networks, namely long- and short-term memory and weight. Weight is divided into standard slow weights and fast weights. Slow weights correspond to long-term memory, in which slow learning decays slowly. Fast weights correspond to short-term memory, in which fast learning decays too quickly [21]. The priming effect in psycholinguistics can be explained using weights. Noise may interfere with the learner's recognition of a target word. However, the target

word can be recognized if the word has been exposed for a short time in that the target word recognition threshold is reduced, meaning that the weight given to the word increases. The processing mechanism of information in memory can also be explained by weight. An important event is less likely to be forgotten, which is because of the event's increased weight.

In terms of the input, the high-capacity weight matrix is used to store short-term memory and process information in a sequence of information. Compared to the long- and short-term memory stored in hidden neural units, the weight matrix information involves corresponding processing processes, i.e., the brain can determine the processing priority using a weight matrix. The weight of the input may explain why there are significant individual differences in the research finding when the same experimental intervention is provided. It can provide a preliminary explanation of the control parameters that are because of the different weights in the information processed by the brain. Attention can be interpreted as a quasi-weight participation process. In practical terms, students may be attracted by the interesting event, which is a high weight given to interest-related content. It generally involves involuntary attention. However, they may "filter" other information after the test start signal appears to concentrate on processing the information from the test paper, which is a conscious operation after analyzing the environmental information, generally involving voluntary attention.

In terms of output, language learners output the body activity and the changes after brain processing. Distributing weights among different patterns is the algorithm basis for specific information or patterns processing to be attractors in complex systems. The weight matrix is a memory storage configuration, which has a higher capacity and is more efficient than storing a set of information (or patterns) [21, 28].

4.2. Rule Layer. Brain adaptive processing rules are complex and involve many fields, such as neuroscience, computing science, and cognitive science. However, the processing rules of the brain can be understood from the algorithm layer, which is a neural network that processes input information with weights.

4.2.1. Hidden Markov Model. The hidden Markov model (HMM) is a mathematical model applied to time series analysis that describes a Markov model with unknown parameters. The core rule of the Markov model is that for a set of time series, the conditional probability distribution of its state at time t depends only on the state at time $t-1$ and has nothing to do with the state before $t-1$ [29], which can be expressed as follows:

$$P(X_t|X_{t-1}, X_{t-2}, X_{t-3}, \dots, X_1) = P(X_t|X_{t-1}). \quad (1)$$

To explain the Markov model, Figure 3 illustrates transition probability.

Assume that the language development state is a variable that can be observed. When we observe that language development is in state 1 at time t , the state of the learner at the next time $t+1$ may be in state 2 and the probability p_1 , in state 3 and the probability p_7 , in state 4 and the probability p_9 , or in state 1 itself and the probability p_{13} . Moreover, because of the limited state in the following stages, $p_1 + p_7 + p_9 + p_{13} = 1$. The *transition probability matrix* of a learner's second language writing ability development between different states can be expressed as follows:

$$\begin{bmatrix} p_1 & p_7 & p_9 & p_{13} \\ p_3 & p_{11} & p_2 & p_{14} \\ p_4 & p_{10} & p_5 & p_{16} \\ p_6 & p_8 & p_{12} & p_{15} \end{bmatrix}. \quad (2)$$

Assuming that the state is hidden and is difficult to directly observe, what we can directly observe is the explicit state that involves HMM. The hidden state is the unknown state parameter of HMM. It is important to infer the learner's hidden state from the explicit state to apply a corresponding teaching strategy. The state transition of HMM is expressed as Figure 4.

HMM argues that, when all the hidden and explicit states before time t are known, the probability of the hidden state at time t is related to the hidden state at time $t-1$ [29]. If the hidden state is represented by S and the explicit state is represented by X , this rule can be formulated as follows:

$$P(S_t|S_{t-1}, S_{t-2}, \dots, S_1, X_{t1}, X_{t-2}, \dots, X_1) = P(S_t|S_{t-1}). \quad (3)$$

When all of the hidden and explicit states before time t are known, the probability of the explicit state at time t is related to the probability of the hidden state at time t . This rule can be formulated as follows:

$$\begin{aligned} P(X_t|S_{t-1}, S_{t-2}, \dots, S_1, X_{t-1}, X_{t-2}, \dots, X_1) \\ = P(X_t|S_{t-1}). \end{aligned} \quad (4)$$

When the hidden state is known, all observations are independent of one another. The transition probability between the states in HMM corresponds to the weight on the input and output layers. HMM is a probabilistic model that infers subsequent states from previous states, also being able to predict time series. HMM provides the principle of time series processing. However, HMM is linear and contains a

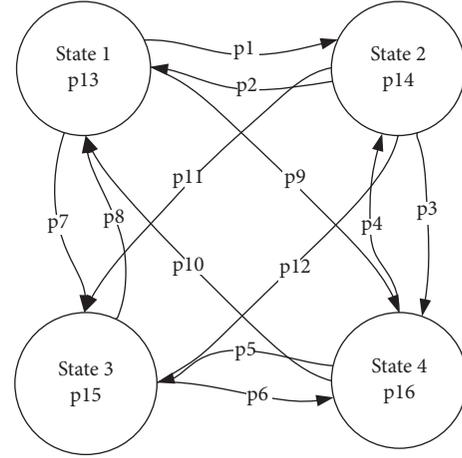


FIGURE 3: Markov model of language development states.

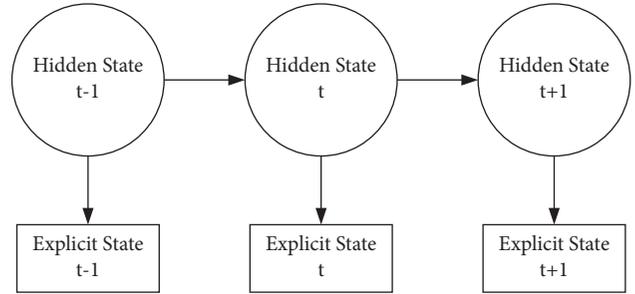


FIGURE 4: State transition of HMM.

few layers for human cognition, though the hidden states may be multiple and need “memory.” A higher-dimensional and more specific nonlinear models are needed to represent language cognitive processes, namely recurrent neural networks and deep neural networks.

4.2.2. Recurrent Neural Network. The recurrent neural network (RNN) was originally a neural network model of processing language sequence information now widely used in the fields of NLP, image processing, and deep learning. Figure 5(a) presents a schematic diagram of a recurrent neural network, in which x represents the input sequence vector (input layer), s represents a hidden layer, U represents the weight matrix from the input layer to the hidden layer, o represents the output vector, V represents the weight matrix from the hidden layer to the output layer, and W (weight) represents the weight, in that the value of the previous sequence of hidden layers is used as the weight of the current input. Figure 5(b) is an expansion of Figure 5(a), showing the relationship of regular operations at $t-1$, t , and $t+1$. The value of the hidden layer is not only affected by the current input sequence but also related to the previous input.

RNN can simulate a nonlinear adaptive system and process the input information based on the information sequence. Figure 6 is the RNN simulation of a nonlinear adaptive system, in which Φ is a nonlinear function and each input of information is related to the information from the

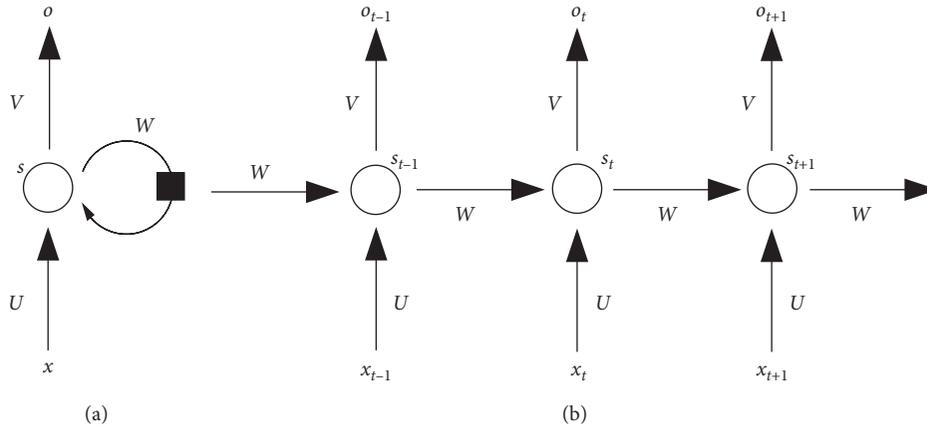


FIGURE 5: Schematic diagram and expanded diagram of cyclic neural network [27].

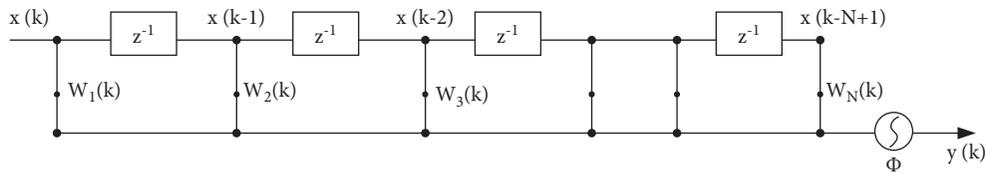


FIGURE 6: RNN simulation of nonlinear adaptive system [27].

previous moment. The input terminal x and output terminal y are fully connected through hidden layer z , and adaptation rules are embedded in the weight. The information processing in RNN has the characteristics of interconnectedness and adaptability.

RNN provides a feasible solution to the algorithm representation layer for the rule side according to the dynamic language view under DST, which posits that language development is an adaptive process of dynamic changes and interconnectedness among various elements. The input of RNN is based on the perception of the environment and the dynamically changing information sequence in memory, in which the hidden layer represents input information. The core rules of the sequence are affected by the previous experience, and they dynamically change with the input of information. By assigning weights, rules can adapt to new inputs.

RNN provides an algorithm for processing language and other time series information sequences. However, neuro-linguistics research has shown that the brain is a parallel multilevel processing system rather than a device used to process language and external information in the units of words or phrases [26]. To process language and perceive information more precisely, a hidden layer for multilevel superposition is required, which has been developed by deep neural networks (DNN).

4.2.3. Deep Neural Network. Figure 7 is a schematic diagram of a deep neural network. After the information sequence has been received in the input layer, the output layer is reached by the deep processing of the three-level hidden

layer for output. Each group of neurons has corresponding weights, indicating that all of the information is on all levels and is fully connected and interactive, meaning that there are differences because of different weights. DNN presents the process of “self-organization” and “soft assembly” in a complex dynamic system. Learners receive a large amount of information from feeling registration, which becomes an original source of the input layer. The information is interconnected in a deep neural network and the rules contained in every node. A certain piece of information given a low weight has little effect on the output layer result after hidden-layer processing. On the contrary, if a piece of information is given a higher weight, it will significantly affect the output layer result, and the information weight of repeated activation is also strengthened. A large number of elements in complex dynamic systems interact in DNN, and the results emerge in the output layer.

A set of deeply processed language information is processed at multiple levels in the brain similarly to DNN, and at different times, the domains are similar to RNN [22]. Therefore, the brain’s processing rules for language and information can be abstracted as an unfolded DNN on RNN (deep recurrent neural network (DRNN)). There are significant differences between human cognition and DRNN. The weight adaptation of the model derives from the fact that the weight is the adjustment of the system based on the output of the model. In the deep learning model, the commonly used algorithm for adjusting weight is the backpropagation algorithm. It is difficult to believe that the weight adjustment of human cognition is reverse derivation layer by layer. However, the learner compares the difference between his own output and the target language in language

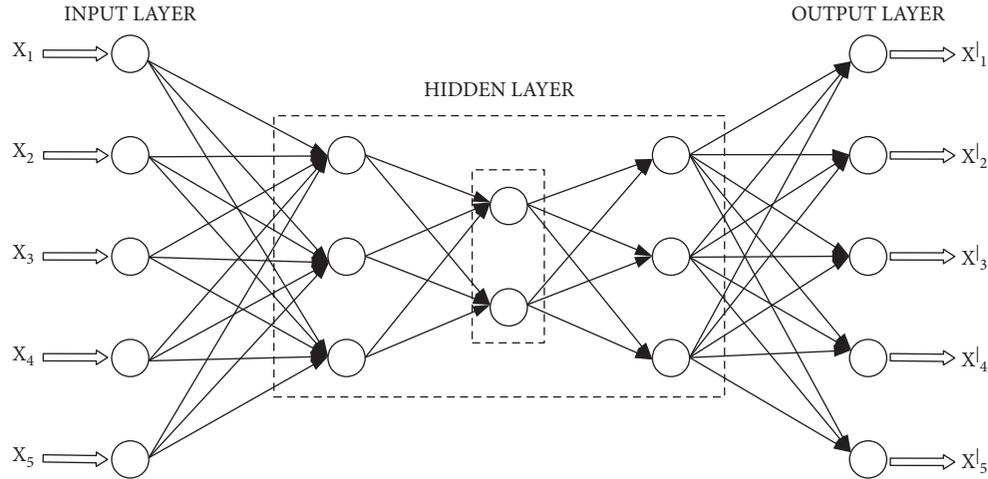


FIGURE 7: Deep neural network [18].

learning and causes psychological changes, which embodies the adaptation of a complex system. However, there has been no evidence that there are similar mechanisms in the human brain. The algorithm shows that the distribution rules of weights at each level are mathematical in artificial neural networks, whereas top-level rules for the connection strength of the quasi-weight neurons distributed by the human brain remain a mystery. Despite the unclear rules, artificial neural networks are excellent models in understanding the adaptive rules of the brain's complex adaptive systems [22, 25].

4.3. Adaptive Weight Model. Weight plays a key role in the algorithmic composition of the input, rule, and output of a complex dynamic system of language development. In terms of input, the weight determines which information enters the brain and becomes a sequence for processing. In terms of output, the weight matrix is an important form of knowledge memory. In terms of the rule, the weight is an important parameter in forming an adaptive rule. Combining the previous discussion on the input, output, and rules, an information processing model can be completed (Figure 8).

The adaptive weight model can be decomposed into the following parts:

- (1) **Input filtering:** input the information sequence to filter with quasi-weight information sequence, in which the quasi-weight matrix is assigned to the inputted information (e.g., attention).
- (2) **External information processing:** from the filtered sequence with quasi-weight information to the deep cyclic neural network, which is related to real-time processing.
- (3) **Storage:** retrieve the information containing the quasi-weight matrix from the storage to the quasi-deep cyclic neural network, and store the information processed by the quasi-deep cyclic neural network.

- (4) **External operation:** the process of outputting behavior and influencing the input information sequence.

Input filtering is the process from the input information sequence to the filtered information sequence with quasi-weight, in which the quasi-weight matrix filters information with lower weights. This process is similar to the attention mechanism, with which the brain may directly retrieve the stored quasi-weight matrix to complete or retrieve and store the current state through a quasi-deep recurrent neural network to deliberately filter information.

External information processing is the process of filtering the information sequence with quasi-weights to quasi-deep recurrent neural network. The information sequence with quasi-weight enters the adaptive DRNN in the brain for processing. The processing mechanism is similar to the deep cyclic neural network, and the information sequence with quasi-weight is processed based on the quasi-weight through the quasi-activation function, and the information may be achieved through integrated processing (DNN) or step-by-step processing based on the input time series (RNN). The stored information containing the quasi-weight matrix will also be involved in this process, and the input information will be adaptive to a new weight matrix for each processing level (input layer, hidden layer, and output layer).

The third component is storage. After the adaptive deep neural network processes the information, on the one hand, the system must store the weights and the adaptive changes of the activation function to form real-time information with processing rules. On the other hand, the information itself is stored in the form of knowledge containing the quasi-weight matrix. When the quasi-deep cyclic neural network is processed, both new and stored information are retrieved together or stored separately. The information containing the quasi-weight matrix is generated after processing and is stored for the next call. In practical terms, the brain may process the stored information separately. At this time, there is no new information input from the outside, which is referred to as "meditation." After meditation, cognition may

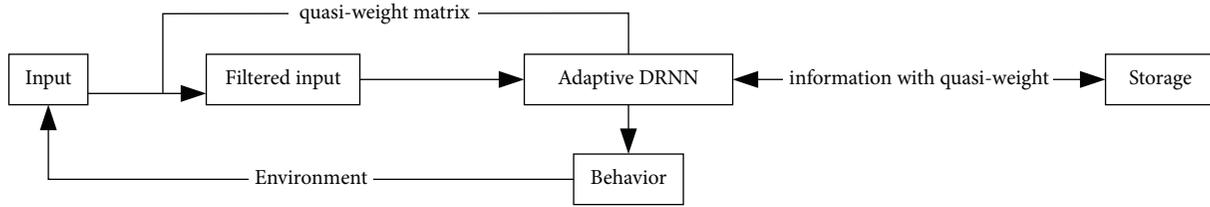


FIGURE 8: Adaptive weight model.

change. The knowledge (experience) generated by the information input in the past was, therefore, “meditation,” which is actually the brain’s self-organization process of storing information. The information containing the quasi-weight matrix continuously participates in the information processing, and the processed information enters the storage to complete the cycle between the deep cyclic neural network and storage.

The fourth component is the process in which the external behavior influences the next input of information through the environment. The individual participates in the self-organization as a subsystem of the complex dynamic system of the external environment to embody an individual’s subjective initiative. For example, the language dialogue controls muscle pronunciation, and writing changes the information input of the interlocutor in the external environment, making these sources new information received by the speaker after an interlocutor has responded.

5. Evaluation Methods for Adaptive Weight Model

To evaluate the model proposed in this study and the feasibility of using neural network models to study the complexity and reductionism issues, this study compiled a neural network model to verify the dependence of language features in adjacent time steps and the relationship between the long- and short-term memory and language performance.

The long short-term memory (LSTM) network is a variation of RNN. Conceptually, the LSTM recurrent unit can “remember” all the past knowledge of the network seen so far and “forget” irrelevant data [30], which is consistent with the design of the adaptive weight model. Therefore, the core algorithm of the adaptive weight model is LSTM, which has shown the application prospects in language identification and processing language sequences, such as sarcasm identification on text data. Accordingly, the model is based on the LSTM network, and the input data is designed as standardized time series data to prevent training divergence. The data was divided into two parts, one part is used for model training, and the other part is used for model verification. The model is designed for the data corresponding to each moment of the input sequence, and the LSTM network learns to predict the data characteristics of the next moment. The LSTM layer contains 240 fully connected hidden units (Hidden units) to match the complexity theory’s description of the “interconnectedness” of complex systems. The training adopts the adaptive moment estimation method

(Adam). The moving average specifies the exponential decay rate. To prevent the gradient from exploding, the gradient threshold is set to 0.9. The specified initial learning rate is set to 0.003, and the learning rate is reduced after 125 rounds of training.

The core of the model’s rule layer is interconnectedness, which is reflected in time iteration, i.e., a large number of elements in the learner’s environment interact with the learner to emerge the learner’s language characteristics. The concentrated expression of the process, i.e., the subsequent development state, is always rooted in the previous development state, and the language characteristics at each point in time reflect the interaction of a specific environment and individual elements. Therefore, the core of the rule end is whether the neural network model can find the development law of a certain language feature and make possible predictions. In view of this, this study selected 256 English majors from three normal universities in northern China to develop data on the accuracy of written syntax. The study collected 7 essays completed by the subjects in one academic year. The interval between each writing task was approximately four weeks. All essays were completed within a limited time in class, and 1932 original writing texts were collected. The accuracy of the evaluation only considers syntax errors and does not involve spelling errors. All texts were firstly judged by the writing teacher and then by a foreign teacher. The cross-consistency of the two teachers’ judgments was 0.903, and the inconsistencies were resolved through negotiation. The first reason for choosing accuracy is that its rise and fall are considered to be affected by the complexity of the task. When the complexity of the task increases, the language accuracy may be reduced because of the limitation of the learner’s processing ability. Therefore, if the model can predict the rise and fall of the language accuracy, it indicates that there are unknown rules. Secondly, accuracy has a single dimension than other language features (such as complexity and fluency). It is less controversial, has stronger objectivity, and is convenient for statistics.

6. Results

After training with the accuracy data shown by the subjects in the first six writing tasks, the model has been able to more accurately predict the syntactic accuracy rise and fall performance of the participants in the seventh writing task (Figure 9), but the root mean square error (RMSE) is 1.0847 (Figure 10), indicating that the model’s estimate of the data value is less biased. The analysis results show that the adaptive weight model based on LSTM neural network

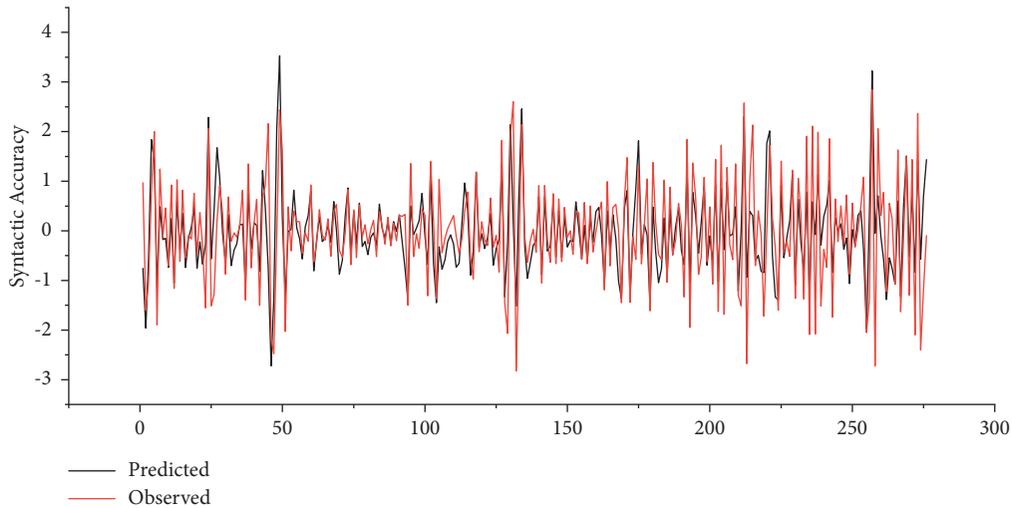


FIGURE 9: Comparison of the model's prediction curve and actual curve for the sentence accuracy of the test in the seventh writing task.

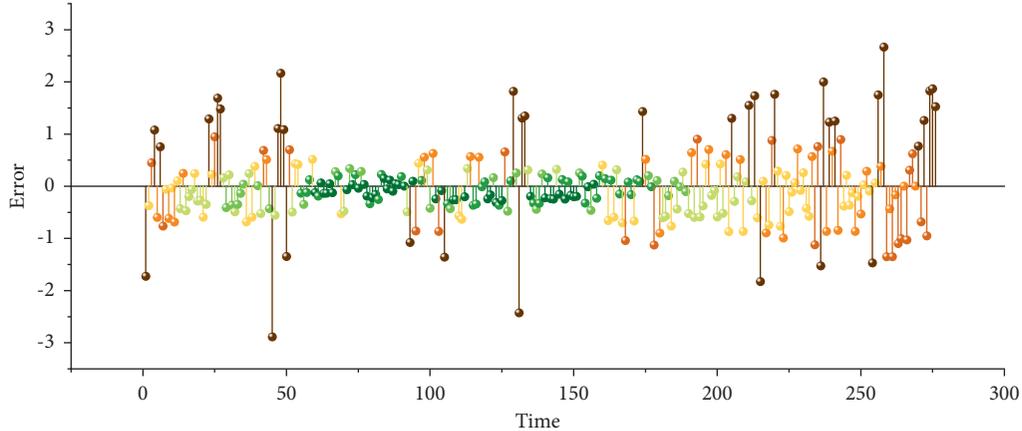


FIGURE 10: Model prediction error distribution and root mean square error.

can “learn” the tested syntactic accuracy development rules to a certain extent. The adaptive weight model simulates the rule end of language processing with potential rationality, whereas the rule contained in the weights between the units of the neural network is difficult to find through traditional statistical methods. Language development of learners has the “rules” that reductionist expect to discover, and shows the complexity and unpredictability of language development explained by complexity theories, that is, rules emerge from chaos, and predictability arises from complexity through self-organization of language development system.

7. Discussion

The adaptive weight model illustrates that reductionism is compatible with complexity theory in language research. Therefore, the linguistic research of reductionism is a summary of the important elements of high power, and the language research under complexity theory is a description of the blueprints of information processing.

7.1. Reductionist Approach. The adaptive weight model provides a basis for the reductionist approach. Language development is an iterative process of information over time, which is conducted on the basis of previous material and psychology at a specific time. Cognition creates an adaptive weight matrix in the process of processing information repeatedly, assigning different weights to elements. Elements with high weights have become significant factors in the development of the system. The reductionist approach can confirm the elements with high weight for most language learners by controlling the input, which is a process that involves inducing macro rules and determining causality. To determine the influence factors to the largest extent possible, experimental research based on reduction theory establishes the set of factors that influence language development by statistical methods, such as block analysis of variance, multiple regression analysis, and structural equation modeling.

The factors that have been under debate for a long time may not have a significant weight on the cognitive processing of most learners. Therefore, two conflicting

theoretical hypotheses that predict the effects of two groups of factors may be compatible. Compared to the hypothesis that is more correct, it may be more important to discover and promote conversion between hypotheses. The impact of task complexity on language performance mainly includes two hypotheses, namely the trade-off hypothesis [31] and the cognition hypothesis [32]. For much of history, academia has not reached consistent conclusions on the correctness of the two hypotheses.

Based on the adaptive weight model, it may be inferred that both hypotheses are a precise and incomplete depiction of language development. Therefore, task complexity is not a high-weight important element in the language development of the language learner group. The two hypotheses depict different periods or different learner groups. The assumptions of the adaptive weight model may also be extended to similar controversial hypotheses. It should be noted that this position is by no means skepticism, in that the adaptive weight model shows that the rules still exist and can be known. However, because of the complexity of the system's adaptive rules, the process of determining reality is complicated. Future reductionist research may need a diachronic design, and it may increase the granularity classification of learners.

7.2. Complexity Theory. Language development is not only an iterative process in time but also an adaptive process of parallel information processing, characterized by the complexity, dynamics, interconnectedness, and nonlinearity of elements of the cognitive system. The weight matrix is distributed to the information sequence containing language, which is processed in the adaptive deep neural network. The information of different weights is interconnected, and the brain processes information based on an unknown, complex adaption rule. The research based on complexity theory emphasizes the dynamic process of adaptability to create complexity, corresponding to the process that the quasi-weight matrix and storage will inevitably undergo adaptive changes following the input of new information. Studies that have used complexity theory have also emphasized the interconnectedness of various elements in language development, corresponding to interconnectedness during information processing of the adaptive deep neural network. Complexity theory research has focused on the nonlinearity of language development, corresponding to the complex quasi-weight matrix, and the input information is iterated and interconnected to enlarge or reduce the impact on the output. Therefore, the study of complexity theory is generally a case study, aiming to describe the microdetails of the complex adaptive process of language development. However, if a certain piece of information is of significant weight, the complex adaptive system may still exhibit significant causality predicted by reductionism. Therefore, the complexity theory discusses the process through which adaptability creates complexity.

In addition to interconnectedness, other points of discrepancy between reductionism and complexity theory are individual differences and group rules in language

development. The adaptive weight model indicates that learners have significant differences in the environment, input information sequence, quasi-weight matrix, storage, and other dimensions, i.e., the individual differences exist ubiquitously and objectively. However, the weight of specific information means that language development has group rules. Deep neural network research has found that too small or too large weights may cause the gradients to disappear or explode. A few key layers significantly influence the results of multilayer deep processing. The impact is that the two model information processing rules with similar key layers behave similarly despite the input information being different [27]. The attractor in the complexity theory is the embodiment of similar effects, which is a state or mode of system tendency, indicating that an order may emerge from a chaotic system. Attractors' tendency indicates a macroscopic rule of system development, which is what reductionism intends to establish. Attractors, as the direction of a group order, increase the dimension of distinguishing the language development type of learners, in that the group rule should be orderly first. Furthermore, learners of the same attractor type will have a greater probability of following the same rule. The system dynamics are so essential that it is reasonable to expect resonance in other situations. Although any case study will never claim to generalize beyond a specific situation, it may help us understand the core of system operations [10].

The adaptive weight model balances the complexity theory and reductionism, drawing on empirical research and theoretical experimental research from the complexity theory perspective. The complexity theory argues that the various elements of a language development system are interconnected and dynamic. As a result, the language development systems of different learners vary. Large sample studies based on the mean smooth out the differences between individuals. Because of the interconnectedness of a language development system's elements, it is impossible to isolate any individual factor. Therefore, the complexity theory conflicts with reductionism. Nonetheless, the adaptive weight model shows that the complexity theory is not completely incompatible with reductionism and the language development system can, to an extent, be simplified. If we insist on taking all participating elements into account in every investigation of language, we will fall into agnosticism. Appropriate simplification and reduction help us grasp the rules of a cognition system. According to the adaptive weight model, the large sample experimental research based on the mean value, under reductionism, can confirm the high-weight (i.e., the most impactful) factors in the cognitive systems of most learners. Identifying these high-weight factors is particularly important for the enrichment of the adaptive weight model and second language learning theories.

7.3. Adaptive Weight Model. The adaptive weight model is a convergence of several theoretical frameworks for second language acquisition. Compared with the previous theoretical model of language acquisition based on cognition, the adaptive weight model can take into account the dynamics,

complexity, nonlinearity, interconnectedness, predictability, and chaos of language development, along with the coexistence of group laws and individual differences. According to the classification of Mitchell, Myles, and Marsden [33], the adaptive weight model is a theoretical model based on processing and input. Compared with the efficiency-driven processor model [34] and processability theory [35], the adaptive weight model focuses on the dynamic changes of the external world and the interaction within language development systems. The adaptive weight model regards explicit learning as a general element of the input. Hence, the model is also compatible with skill acquisition theory [36], cognition hypothesis, trade-off hypothesis, and other theories and assumptions of explicit learning. Though the adaptive weight model is rooted in input-based theoretical models, such as complexity theory, emergentism, and constructionism, it is more realistic and operational in comparison.

Although the artificial neural network used in the rule layer of the adaptive weight model originated from the idea of connectionism in the field of language research, it has further developed. Firstly, the connectionist approach in language research focuses on modeling the connection between language input, output, and language training. The adaptive weight model emphasizes language input and the effect of various variables in the environment on the system. In other words, the adaptive weight model is not only a language training model but also a model of the language cognitive mechanism. Secondly, the connectionist approach in language research pays special attention to the computer realization of the model, in which the parameters are determined, the number of hidden layers is small, and there is a definite activation function. The adaptive weight model does not emphasize the reality of computer realization (though there is a possibility of computer realization). Rather, it emphasizes theoretical and explanatory attributes, thus avoiding many of the defects of the connectionism model in language research. For example, Robinson [37] believes that the connection model does not correspond to the human cognitive mechanism, and Williams and Kuribara [38] argue that language processing may only be similar to matching rather than grammatical induction. In addition, the theory cluster of connectionism, construction grammar, emergentism, and complexity theory indicates that the acquisition of the second language is a process in which all elements dynamically interact. The adaptive weight model proposed in this study is based on language development facts from the perspective of weight change. It reveals the mechanism of the interaction of elements and shows the dynamic connection between individual differences and group law, dynamics and complexity, and predictability and chaos.

8. Conclusion

The emergence of complexity theory has made a direct challenge to the authority of reductionism. However, there is a fundamental dispute between reductionism and complexity theory in language development. Based on the idea of connectionist cognitive simulation, this paper put forward

the adaptive weight model with deep neural networks that have achieved promising results in the field of artificial intelligence. The adaptive weight model provides a comprehensive and adaptive view of language research, showing that there is compatibility among language research, complexity theory, and reductionism, indicating that the research paradigm may be further expanded. Reductionist research can adopt a diachronic experimental design to explore the dynamics and complexity of potential factors to confirm which types of learners are more in line with the rules predicted by reductionism, whereas research based on complexity theory can adopt a large sample research design to focus on the group law while also focusing on the dynamics of individual differences [10].

This research provides recommendations for the dynamic modeling of second language development. In second language teaching and research, researchers have found many factors that affect the language development of learners. The factors that learners encounter in the process of language development involve many natural, social, physical, and psychological elements. However, incorporating all these elements into a language development model is unrealistic, and these elements do not all have an equivalent influence on language development. Therefore, the model assigns these factors different weights. The core factors of second language acquisition are age, cross-language influence, language environment, memory, consciousness, foreign language ability, motivation, emotion, social culture, and other limited factors. The number of core factors may vary. Although the model is simplified, only grasping the essence, it is necessary in scientific research. Such simplification is not reductionism because the elements have not been isolated, and the interconnectedness among them has not been ignored. These core factors generally have limited dimensions. Therefore, computer modeling based on adaptive weight models is feasible because neural network models with hundreds of millions of parameters are not uncommon. In addition, these factors are explicit and implicit. True and quasi-experiment research generally uses the explicit representation of hidden parameters, which is carried out by statistical methods, such as structural equation modeling. In comparison, the adaptive weight model can model each learner based on computer models such as HMM.

This study is a preliminary attempt to integrate the two paradigms in language research. The neuron connection rules urgently require the support of literature findings in cognitive science and artificial intelligence. Researchers may conduct experimental research to verify and enrich the model in the future.

Data Availability

The program code and data that support the findings of this study are available from Qufu Normal University (QFNU), Qufu, China, but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. Program code and data are, however, available from the authors upon reasonable request and with permission of QFNU.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the National Social Science Foundation of China (grant nos. 19BYY217 and 20BYY212) and in part by the National Social Science Foundation of Education of China (grant no. BEA180110).

References

- [1] K. De Bot, W. Lowie, and M. Verspoor, "Dynamic systems theory approach to second language acquisition," *Language Learning*, vol. 10, pp. 1237–1255, 2007.
- [2] D. Larsen-Freeman and L. Cameron, *Complex System and Applied Linguistics*, Oxford Univ. Press, Oxford, UK, 2008.
- [3] W. O'Grady, "The emergentist program," *Lingua*, vol. 118, no. 4, pp. 447–464, 2008.
- [4] J. P. Cárdenas, I. González, G. Vidal, and M. A. Fuentes, "Does network complexity help organize Babel's library?" *Physica A: Statistical Mechanics and Its Applications*, vol. 447, no. 4, pp. 188–198, 2016.
- [5] S. Marsland, *Machine Learning: An Algorithmic Perspective*, CRC Press, Boca Raton, Florida, FL, USA, 2014.
- [6] T. Honderich, *Oxford Companion to Philosophy*, Oxford University Press, Oxford, UK, 2005.
- [7] G. P. Montague, *Who am I? Who is she?, A Naturalistic, Holistic, Somatic Approach to Personal Identity*, Walter de Gruyter, Berlin, Germany, 2012.
- [8] N. C. Ellis, "Constructions, chunking, and connectionism: the emergence of second language structure," in *Blackwell Handbooks in Linguistics. The Handbook of Second Language Acquisition*, C. Doughty and M. H. Long, Eds., Blackwell Pub, Hoboken, New Jersey, NJ, USA, 2003.
- [9] D. Larsen-Freeman, "The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English," *Applied Linguistics*, vol. 27, no. 4, pp. 590–619, 2006.
- [10] Z. Dörnyei, "Researching complex dynamic systems: "Ret-rodicitive qualitative modelling" in the language classroom," *Language Teaching*, vol. 47, no. 1, pp. 80–91, 2014.
- [11] H. A. Simon, "The architecture of complexity," *Proceedings of the American Philosophical Society*, vol. 106, no. 96, pp. 467–482, 1962.
- [12] M. Mitchell, *Complexity: A Guided Tour*, Oxford University Press, Oxford, UK, 2009.
- [13] J. P. Crutchfield and K. Young, "Inferring statistical complexity," *Physical Review Letters*, vol. 63, no. 2, pp. 105–108, 1989.
- [14] J. C. Sprott, *Strange Attractors: Creating Patterns in Chaos*, M & T Books, New York, NJ, USA, 1993.
- [15] E. Thelen and L. B. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action*, The MIT Press, Cambridge, MA, USA, 1994.
- [16] J. H. Holland, *Emergence: From Chaos To Order*, Perseus Books, New York, NJ, USA, 1998.
- [17] J. H. Holland, *Hidden Order: How Adaptation Builds Complexity*, Addison-Wesley, Boston, MA, USA, 1995.
- [18] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*, Springer, New York, NJ, USA, 2018.
- [19] M. A. G. Cutter, *The Brain: Understanding Neurobiology through the Study of Addiction*, BSCS, Colorado Springs, 2000.
- [20] D. O. Hebb, *The Organization of Behavior: A Neuro-psychological Theory*, Lawrence Erlbaum Associates, Mahwah N.J, USA, 2002.
- [21] P. Cisek, T. Drew, and J. F. Kalaska, "Computational neuroscience: theoretical insights into brain function," *Progress in Brain Research*, Elsevier, Amsterdam, Netherlands, 2007.
- [22] M. A. Arbib and J. J. Bonaiuto, *From Neuron to Cognition via Computational Neuroscience. Computational Neuroscience*, The MIT Press, Cambridge, MA, USA, 2016.
- [23] A. D. Friederici, "The cortical language circuit: from auditory perception to sentence comprehension," *Trends in Cognitive Sciences*, vol. 16, no. 5, pp. 262–268, 2012.
- [24] F. Pulvermüller, "How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics," *Trends in Cognitive Sciences*, vol. 17, no. 9, pp. 458–470, 2013.
- [25] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.
- [26] L. Chang and D. Y. Tsao, "The code for facial identity in the primate brain," *Cell*, vol. 169, no. 6, pp. 1013–1028, 2017.
- [27] J. A. Chambers and D. P. Mandic, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures, and Stability*, John Wiley, New York, NJ, USA, 2010.
- [28] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645–6649, IEEE, Vancouver, BC, Canada, May 2013.
- [29] A. M. Fraser, *Hidden Markov Models and Dynamical Systems*, Society for Industrial & Applied Mathematics (SIAM), Philadelphia, PA, USA, 2008.
- [30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [31] P. Skehan, *A Cognitive Approach to Language Learning*, Oxford University Press, Oxford, UK, 1998.
- [32] P. Robinson, "The cognition hypothesis, task design, and adult task-based language learning," *Second Language Studies*, vol. 21, no. 2, pp. 45–105, 2003.
- [33] R. Mitchell, F. Myles, and E. Marsden, *Second Language Learning Theories*, Routledge, Oxfordshire, UK, 2013.
- [34] B. MacWhinney and W. O'Grady, "The handbook of language emergence," *Blackwell Handbooks in Linguistics*, Wiley Blackwell, Hoboken, NJ, USA, 2015.
- [35] M. Pienemann, "language processing and second language development: processability theory," *Studies in bilingualism*, pp. 0928–1533, John Benjamins, Amsterdam, Netherlands, 1998.
- [36] R. DeKeyser, "Skill acquisition theory," in *Routledge Studies in Germanic Linguistics. Theories in Second Language Acquisition: An Introduction*, B. van Patten and J. Williams, Eds., pp. 97–113, Routledge, an imprint of Taylor and Francis, Oxfordshire, UK, 2007.
- [37] P. Robinson, "Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA," *Studies in Second Language Acquisition*, vol. 27, pp. 235–268, 2005.
- [38] J. N. Williams and C. Kuribara, "Comparing a nativist and emergentist approach to the initial stage of SLA: an investigation of Japanese scrambling," *Lingua*, vol. 118, no. 4, pp. 522–553, 2008.