Do Machines Learn Language Like Us?

Comparing Connectionist Approaches to Language Learning With Human Cognition

Natalie Yu

University of Toronto

COG250Y1

Dr. John Vervaeke

April 8, 2024

Abstract

This paper argues that connectionist models and human brains learn and approach the productivity and systematicity of language similarly. A famous argument against the connectionist model by Fodor and Pylyshyn argues that the productivity and systematicity of language are unable to be computed due to their inborn nature. Through exploring the similarities between productivity, systematicity, and learning in both humans and machines, it becomes evident that both systems exhibit comparable abilities in processing and generating language. While the connectionist model has its limitations when it comes to language, as it struggles to compute the innateness, the intentional meaning, and the structural ambiguity of language, it's important to note that these perceived "challenges" are also present in the diverse ways human cognition operates. Therefore, despite its imperfections, the model can still be considered a form of cognition as it reflects the diversity of cognitive processes observed in humans. This paper highlights the parallels between language in connectionist models and human cognition, shedding light on the complexity and diversity of cognitive processes in both artificial and natural intelligence.

Visualizing language as nothing more than links between small, simple parts appears too simplistic to account for language's limitless potential. Nonetheless, connectionism, a method of cognitive modelling, can be used to compute language. With this approach, artificial neurons, or nodes, make up structures known as neural networks through simple, weighted connections. When these models interact in parallel, they are capable of computing language via higher-level processing (Maclennan, 2015). However, Fodor and Pylyshyn strongly oppose the connectionist model, stating that human cognition is language with their language of thought hypothesis (LOTH) and therefore cannot be computed using a neural network model. Additionally, they can't do unsupervised learning. This essay argues that connectionist models and human brains learn and approach productivity and systematicity in a similar manner.

Connectionism

A prominent view of artificial intelligence (AI) is connectionism, which is the principle that mental processes can be described by interconnected networks. The nodes that are interconnected are modelled after the neurons in the brain; there are different signal strengths, inhibitory vs. excitatory firing rates, and the weight of each connection (J. Vervaeke, personal communication, February 12, 2024). Knowledge in a connectionist model is represented as large patterns of numerical data, which allows us to use categorization and concepts similarly to how human cognition works. This works by parallel processing, which is the idea that multiple interconnected nodes, such as a category, are lit up or activated by a single node or input. The architecture of neural networks reflects this parallel processing, consisting of layouts and layers similar to the brain's interconnected structure. These models learn by supervised learning; they are given a data set, and someone takes the outputs and adjusts the weight, or strength, of connections. When trained with a large amount of data, connectionist models create a larger

amount of connections, and with increased exposure to the world, they become increasingly intelligent (Joanisse & McClelland, 2015).

Language of Thought Hypothesis

A very important argument against the connectionist model is the language of thought hypothesis (LOTH). Argued by Fodor and Pylyshyn, the LOTH is the hypothesis that thought is language; or thought takes place within a mental language. Most notably, they claim that human thought is productive and systematic like language, which is not possible in a connectionist model (J. Vervaeke, personal communication, March 11, 2024).

Productivity

Productivity can be defined as the ability of a cognitive system to produce and understand an infinite number of novel and meaningful combinations from a finite set of basic elements. However, these connectionist models cannot make these connections unless they have been seen before. For instance, there are a finite amount of words in the English language, however, humans can create an infinite number of sentences and thoughts. For instance, the sentence “His mother’s mother's mother’s mother’s...” can be continued infinitely, even though we are bound by biological limits, such as memory and processing capacity. Contrastly, typically connectionist models struggle with generalizing from learned examples to novel situations, as connectionist models only know the data that they have been trained with. Additionally, they may struggle with creativity or understanding abstract concepts.

Systematicity

Systematicity is the property of a cognitive system whereby if it can perform some operations on certain kinds of representations, then it can also perform similar operations on similar representations. For example, in the sentence “John loves Mary," the word “John”

denotes John, the word “loves” denotes the action of loving, and “Mary” denotes Mary. The three words make the exact same contribution to the sentence “Mary loves John," but the sentences have different meanings. As humans, if we can consider the thought “John loves Mary," we can also consider the thought "Mary loves John." However, in a machine, the underlying representations of these two phrases in a connectionist model are composite, share identical nodes, and have meanings that are dictated by the individual nodes and their arrangement; therefore, a linguistically structured system is what explains the ability to consider the second sentence. According to Fodor and Pylyshyn, connectionist models frequently fail to exhibit systematicity since their learning is dependent on slow modification of connection weights and pattern recognition, which may not generalize well to all relevant cases.

Unsupervised Learning

Another issue faced by connectionist models is the inability to learn unsupervised. Connectionist models learn by training with a data set. With a fixed data set, someone must feed the set into the system and determine the weightings of the connections based on the inputs. Unsupervised learning is very helpful for noticing previously unseen patterns in data and processes, such as creativity, which is believed to be the key to real AI-generated decisions (Wang & Biljecki, 2022). Connectionist models cannot learn unsupervised, as they require someone to adjust connections and correct outputs. This is a big argument against neural networks–while human cognition is able to find patterns in data that maybe have not been seen previously, machines require prior knowledge and training of the patterns to generalize.

Supervised Learning, Productivity, and Systematicity in Neural Networks and Humans

When these neural networks are pre-trained with datasets large enough through the usage of supervised learning, the machine can produce a seemingly infinite variety of responses. This

productivity stems from the model's ability to capture the underlying structure of language and generate text that is contextually appropriate and coherent. In essence, productivity and systematicity work together to enable neural network models to effectively process and generate language, allowing them to be flexible, adaptable, and capable of understanding and generating a wide range of linguistic expressions.

Supervised Learning

An argument against the connectionist model as a form of AI is the inability to learn unsupervised and without human input. The inability to learn unsupervised in a machine, however, is very comparable to supervised linguistic development in humans. For instance, if a young child were to use the incorrect past tense of a verb, they would not know that this is incorrect until someone, such as a parent or teacher, corrects the child and teaches them the correct verb tense (Maclennan, 2015). When learning a language or with new words that are introduced, external factors, such as a dictionary or another person, must confirm the usage of a specific word. Both machines and humans are able to assume the meaning of a word through context and prior knowledge, but there is no way of confirming if the assumption is true or false.

Additionally, supervised learning in a machine actually appears very similar to the way children learn language. For example, learning English is difficult due to its irregularities. When learning the past tenses of verbs, both connectionist models and children learning past tenses can struggle with irregular verbs, which do not follow the typical "-ed" pattern like 87% of verbs (Joanisse & McClelland, 2015). Connectionist models may initially find it challenging to generalize from regular verbs to irregular verbs, as the latter often require memorization of specific forms. Similarly, children may initially overapply regular patterns to irregular verbs (e.g., saying "runned" instead of "ran") before mastering the correct forms through exposure and

practice. In both cases, when the vocabulary is still small, machines and children make mistakes when applying these generalized patterns to irregular verbs. As the vocabulary increases, the ratio of irregular to regular verbs decreases, and with further training, the machine becomes highly accurate (Walmsley, 2016). A similar challenge is found in reading in English when considering words that are spelled similarly. The majority of the time, words spelled similarly are audibly similar. However, there are irregularities, such as trough, rough, and drought, which require training in both machines and children and follow the same sort of trajectory as learning past tenses.

Systematicity and Productivity in Neural Network Models

Categorization and parallel processing are crucial in connectionist systems to combat the LOTH, in particular the argument for systematicity and productivity. A study by Bhatia and Richie (2024) tried to compute the categorization of human thought in multiple types of neural networks, with concepts including animals, foods, tools, and geographic locations. Certain networks failed for different reasons, showing that semantic cognition relies on more than just the thematic content of the sentence, the linguistic probability of the sentence, or the similarity of the concept and feature in the sentence. However, the model that succeeded had been pre-trained with 3.3 billion tokens and over 500,000 true or false statements. It contained 12 transformer layers (layers that process input data) and 768 hidden layers (which allow computation of an exclusive-OR function), which clearly portrays the importance of parallel processing and categorization in connectionist models (Walmsley, 2016). While these steps are obviously not realizable or realistic in a human brain, it is likely that the pretraining and fine-tuning steps in this network model may also be at play in human learning. Previous knowledge of concepts and language structure allows children to judge new sentences with unknown truth values. The

process of categorization involves understanding the underlying structure of the data and applying learned rules or patterns to classify new examples. This demonstrates systematicity in the model's ability to generalize its learning and apply it in a structured manner to new instances, as well as productivity in its ability to generalize to new instances and classify them into known categories.

Solving Riddles

A simple way to describe supervised learning, productivity, and systematicity in language is through the use of riddles. Let's use the popular riddle, "What gets wet from drying?" "A towel". Riddles can be seen as a form of language creativity, as they often involve wordplay, puns, and unconventional uses of language. Solvers must be able to generate creative, novel interpretations (productivity) of the language in the riddle, alongside its structured manner (systematicity), to arrive at the answer.

Supervised Learning. To understand this riddle, the solver, whether it be a human or a neural network, requires the knowledge that a towel can be the subject of the verb "to dry." If a neural network does not have previous knowledge that a towel could dry something else or that the verb "dry" could be used in this context, it processes and creates a new connection between nodes—one that signifies this novel verb. With the feedback from supervised learning, the machine will be able to incorporate the whole sentence, which in turn processes the "meaning" of the word. This works no differently for a human, who is unable to understand the riddle unless knowledge of the new meaning of the verb "dry" has been acquired.

Productivity. A machine can demonstrate productivity by generating a wide range of outputs based on a set of input instructions or examples. In a neural network, if a machine has more knowledge or previous training with multiple definitions of words, it is likely to be able to

solve this riddle with the knowledge it has obtained. Even if this machine has never seen this riddle before, if they have the previous knowledge that a towel can dry, the neural network should be able to produce the correct answer, potentially among a range of possible answers due to the ambiguity of the riddle. Based on previously learned associations and the context of the riddle, a word will be chosen. Similarly, in children, they exhibit productivity by producing a novel interpretation based on the clues provided. Riddle solving in children has been linked to higher cognitive function in comparison to children who do not tell or solve any riddles, which may be similar to the amount of pretraining in a neural network (Bowes, 1981). This example demonstrates the productivity of language and the ability of individuals and machines to generate new meanings and interpretations based on existing

Systematicity. Systematicity can be seen in riddles, as they often follow a specific structure or pattern that requires the solver to understand and apply a set of rules or conventions. For instance, when a child is first introduced to a riddle, they are likely unable to use a less salient definition of a word to solve it and will likely fail. However, with the aforementioned riddle, one may realize that a riddle similar to this one may require a less salient definition of a word to solve a problem. When introduced to another riddle of a similar form, such as “What has many rings but no fingers? A telephone!” A child may realize that this riddle may require a less salient definition of "rings," such as the verb “to ring” rather than the plural noun. This is akin to the structured nature of riddles, where solvers must apply systematic reasoning to understand the clues and arrive at the answer. Similarly to the human brain, neural networks can learn to recognize patterns in riddles and apply these patterns to solve novel riddles.

Neural Network Shortcomings

Structural Ambiguity

While the outputs of a neural network may be the same as those of the human brain in certain situations, Bever et al. argue that neural network models are very poor models of mind and brain for language organization (2023). They argue that neural networks are unable to understand structural ambiguity. For example, in this article, the authors input the sentence “The chicken is ready to eat.” into ChatGPT. When ChatGPT is then asked to replace “chicken” with "children," ChatGPT replaces the word; however, the machine does not change that this word was originally the object of the verb "eat." When asked the meaning of this new sentence, the machine thinks the word “children” is the object of the word “eat” and then returns disturbing content relating to cannibalism, such as the story of Hansel and Gretel.

The authors of the aforementioned article argue against neural network machine learning models by using ChatGPT as an example. When I input the exact previous ambiguous sentence into ChatGPT 3.5, the machine returns the correct meaning of the second sentence, stating it is “a straightforward statement indicating that the children are ready to start eating” (OpenAI, 2024). Additionally, with our current knowledge of how neural networks interpret ambiguous sentences, such as in riddles, machines interpret ambiguous sentences differently depending on context and the model’s training data. In my case, ChatGPT chose the more common or default interpretation for this type of sentence in the context it was given. This article does not specify the exact phrases used as input or the version of ChatGPT used, but research in artificial neural networks seems to disprove the argument. Connecting back to the thesis, the human brain process is similar to a machine’s when reading the aforementioned sentence, “The children are ready to eat." While children being the object of the word “eat” is a syntactically correct possible statement, it is highly unlikely that this truly is the meaning of the sentence, as it is a less common or default interpretation of the sentence. Furthermore, individuals with autism spectrum

disorder (ASD) have difficulty interpreting statements where words are given an uncommon interpretation. Participants in a study were given the two statements “Clare was robbed by the river.” and “The bank was the scene of the robbery” and then asked, “Was the robbery in a village bank, a river bank, or a bank?” Individuals with ASD have trouble interpreting that a robbery may occur on a river bank because robberies are normally associated with financial establishments (Jolliffe and Baron-Cohen, 1999). These individuals consider the most common interpretation of this sentence, similarly to a neural network, which also has a stronger connection weight based on previous training. Incorporating principles of neurodiversity into AI research can lead to the development of systems that are more attuned to the diverse ways in which humans think and understand the world, ultimately advancing the field of AI and cognitive science.

Nature vs. Nurture

In the same article, Bever et al. argue that human language is almost entirely created by the human brain (2023). Decades of research show evidence of a universal inborn grammar rather than a learned language from any sort of induction process. Additionally, while humans learn language through primarily external input, such as a parent or teacher, there are limitations set by nature that cannot be replicated in a neural network, such as processing limitations (He, 2022).

While this neural network model does not consider any sort of innate characteristics or neurodiversity, the vast differences in human cognition have yet to be fully studied, which makes it difficult to compute a neural network system without full comprehension. However, this model could provide insight into the differences by investigating models that struggle in similar contexts as some individuals. The inspiration behind the neural network architecture was the

human brain, where neurons or nodes are connected on multiple layers. Conversely, researchers in cognitive science and AI can gain valuable insights into the limitations of current models and work towards developing more inclusive and comprehensive AI systems.

Understanding Intention

Connectionist models may be able to understand syntax and language, but the meaning or intention in the sentences is not apparent in syntax. The same sentence could be said two different times, and the tone or word stress could be the difference between two entirely different meanings, such as how sarcasm is conveyed. For instance, if the phrase "I never said she stole my car" is repeated seven times, each with a different word that is stressed, the meaning of the sentence changes while being syntactically and grammatically identical. The connectionist model previously proposed and argued for a form of language organization that only considers grammatical and syntactical structures.

However, a previous study suggests that speech contains reliable prosodic markers for word meaning and that listeners use these prosodic cues to differentiate meanings (Nygaard et. al., 2009). In other words, the meaning of a sentence or a word can be realized through changes in stress and intonation. Machines can accurately recognize changes in prosody as acoustic-related information, including the pitch (fundamental frequency), loudness (intensity), duration, and spectral tilt (Kakouros and Räsänen, 2015). With enough previous training, understanding changes in the prosody of words is no different than any other sort of pattern recognition available in a neural network model. Additionally, with this information, it is plausible that the system responsible for computing language could be entirely different from the system that is responsible for recognizing changes in acoustic-related information. Comparing this model to the vast amount of diversity in cognition, individuals with autism most notably

struggle with prosody, despite a complete understanding of the syntactical structure of a sentence (Patel et. al., 2023). From a connectionist model, we could hypothesize that the parts of the system that impact communicative interactions in individuals with autism could possibly not be at all connected to the system that processes language.

Conclusion

In conclusion, this essay has argued that connectionist models and human brains learn and approach productivity and systematicity in a similar manner. Through exploring the similarities between productivity, systematicity, and learning in humans and machines, it becomes evident that both systems exhibit comparable abilities in processing and generating language. The parallels between connectionist models and human cognition offer valuable insights into the nature of learning and intelligence, advancing our understanding of both artificial and natural intelligence. However, it's important to note that a singular model of cognition may not be sufficient to capture the full complexity of human cognition, considering the diverse variations of cognition across individuals.

References

Bever, T. G., Chomsky, N., Fong, S., & Piattelli-Palmarini, M. (2023). Even deeper problems with neural network models of language. *Behavioral and Brain Sciences*, *46*, e387. https://doi.org/10.1017/S0140525X23001619

Bhatia, S., & Richie, R. (2024). Transformer networks of human conceptual knowledge. *Psychological Review*, *131*(1), 271–306. https://doi.org/10.1037/rev0000319

Bowes, J. (1981). Some cognitive and social correlates of children's fluency in riddle-telling. *Current Psychology*, *1*(1), 9–19. https://doi.org/10.1007/BF02684421

He, A. X. (2022). Optimal input for language development: Tailor nurture to nature. *Infant and Child Development*, *31*(1), e2269. https://doi.org/10.1002/icd.2269

Joanisse, M. F., & McClelland, J. L. (2015). Connectionist perspectives on language learning, representation and processing. *WIREs Cognitive Science*, *6*(3), 235–247. https://doi.org/10.1002/wcs.1340

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: Linguistic processing in high-functioning adults with autism or Asperger syndrome: is local coherence impaired? *Cognition*, *71*(2), 149–185. https://doi.org/10.1016/S0010-0277(99)00022-0

Kakouros, S., & Räsänen, O. (2016). Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features. *Cognitive Science*, *40*(7), 1739–1774. https://doi.org/10.1111/cogs.12306

MacLennan, B. (2015). Cognitive modeling: Connectionist approaches. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition) (pp. 84–89). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.43021-7

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*(1), 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x

OpenAI. (2024). ChatGPT-3.5 (Apr 2024 version) [Large language model]. https://chat.openai.com/chat

Patel, S. P., Landau, E., Martin, G. E., Rayburn, C., Elahi, S., Fragnito, G., & Losh, M. (2023). A profile of prosodic speech differences in individuals with autism spectrum disorder and first-degree relatives. *Journal of Communication Disorders*, *102*, 106313. https://doi.org/10.1016/j.jcomdis.2023.106313

Rescorla, M. (2023). The language of thought hypothesis. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/language-thought/

Srinivasan, M., & Barner, D. (2013). The Amelia Bedelia effect: World knowledge and the goal bias in language acquisition. *Cognition*, *128*(3), 431–450. https://doi.org/10.1016/j.cognition.2013.05.005

Walmsley, J. (2016). *Mind and machine.* Palgrave Macmillan London.

Wang, J., & Biljecki, F. (2022). Unsupervised machine learning in urban studies: A systematic review of applications. *Cities, 129*, 103925. https://doi.org/10.1016/j.cities.2022.103925