

REVIEW ARTICLE OPEN ACCESS

The Roles of Neural Networks in Language Acquisition

Eva Portelance^{1,2}  | Masoud Jasbi³¹Department of Decision Sciences, HEC Montréal, Montreal, Canada | ²Mila - Quebec Artificial Intelligence Institute, Montreal, Canada | ³Department of Linguistics, University of California, Davis, California, USA**Correspondence:** Eva Portelance (eva.portelance@hec.ca) | Masoud Jasbi (jasbi@ucdavis.edu)**Received:** 14 November 2023 | **Revised:** 30 August 2024 | **Accepted:** 9 September 2024**Keywords:** cognitive models | instrumentalism | language acquisition | language models | neural networks | proof of concept | realism

ABSTRACT

How can modern neural networks like language models be useful to the field of language acquisition, and more broadly cognitive science, if they are not a priori designed to be cognitive models? As developments towards natural language understanding and generation have improved leaps and bounds, with models like GPT-4, the question of how they can inform our understanding of human language acquisition has re-emerged. As such, it is critical to examine how in practice linking hypotheses between models and human learners can be safely established. To address these questions, we propose a model taxonomy, including four modelling approaches, each having differing goals, from exploratory hypothesis generation to hypothesis differentiation and testing. We show how the goals of these approaches align with the overarching goals of science and linguistics by connecting our taxonomy to the realist versus instrumentalist approaches in philosophy of science. We survey recent work having adopted each of our modelling approaches and address the importance of computational modelling in language acquisition studies.

1 | Introduction

In recent years, language models, like large language models such as GPT-4 or Llama-2 (OpenAI 2023; Touvron et al. 2023), have become omnipresent in our daily lives. Their ability to generate natural sounding text and to answer queries has meant that they are now used across the world to help power search engines, as creative tools for productivity, and as generators of information content we are exposed to. Whether these contemporary neural network models truly understand language (Bommasani et al. 2022) or whether they can truly reason with it (Mahowald et al. 2024) are hotly debated topics. However, it is undeniable that these models seem to have learnt many aspects of language. As a consequence, a new subfield within computational linguistics with the goal of understanding what type of linguistic knowledge neural language models encode has flourished. These studies use ‘probing’ experiments, involving post-hoc testing of models’ knowledge representations, and have included an impressive array of grammatical

phenomena (Linzen, Dupoux, and Goldberg 2016; Lake and Baroni 2018; Futrell et al. 2019; Manning et al. 2020; Hu et al. 2020) and even grammatical acceptability judgements tests (Lau, Clark, and Lappin 2017; Sprouse et al. 2018; Warstadt, Singh, and Bowman 2019). However, there is still much to learn about how these models learn and represent the meanings of words and whether they share any resemblance with early language learning in children (Huebner and Willits 2018; Haley 2020; Huebner et al. 2021; T. A. Chang and Bergen 2022; Vong and Lake 2022; Portelance, Frank, and Jurafsky 2024; Lavechin et al. 2023; Li et al. 2024). Though the claim is often made that these models could be useful in addressing linguistics hypotheses (e.g., *the poverty of the stimulus* hypothesis, Baroni 2022), we have yet to properly reckon with the impacts that recent advances in NLP research may have on our understanding of language learning in children (Bommasani et al. 2022). In this paper, we address the following question: **how can contemporary neural networks like language models be useful to the field of language acquisition, and**

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Language and Linguistics Compass* published by John Wiley & Sons Ltd.

more broadly cognitive science, if they are not a priori designed to be cognitive models? We will argue that there are four ways that such models can serve us in reaching our overall scientific goals.

1.1 | A Novel Modelling Paradigm

When ‘traditional’ versus ‘novel’ modelling approaches are referred to in cognitive science, they usually are contrasting symbolic models with neural network models. Since the scope of this paper is limited to the ways in which neural networks can be used in the field of language acquisition, we instead use ‘traditional’ models to refer to connectionist or neural network-based cognitive models, which we discuss further in Section 2.1. Such models have been used to study language learning since the mid-80s (McClelland, Rumelhart, and PDP Research Group 1986) and the advent of connectionism, a philosophy in psychology which proposed that computational neural networks could be studied as simplified models of the neural processes in the mind. Traditional models are thus designed to emulate human learning and processing, making them useful for testing hypotheses about human language learning specifically.

The powerful language models we have seen in recent years (Devlin et al. 2019; Brown et al. 2020), unlike cognitive models, are not designed to model human processing, yet they seem to reproduce many of our linguistic behaviours. The question of how these models might connect to human language learning has therefore gained significant interest (Linzen 2019; Lake, Linzen, and Baroni 2019; Warstadt and Bowman 2022; Mahowald et al. 2024). In parallel, the idea that these contemporary neural network models should be thought of as formal theories has also been argued (Guest and Martin 2021; Baroni 2022; Piantadosi 2023). Unlike the traditional approach, these newer models are not designed to be cognitive models or to instantiate theories about human learning. As a field, we therefore have to address what their purpose is and how they can align themselves with our scientific goals in studying human language acquisition. In Sections 2.2 through 2.4 we describe three novel ways in which these models can and have been used to measure, define, and generate new hypotheses. In the next section we discuss overarching approaches in philosophy of science and linguistics to show how these novel neural network modelling approaches align with them.

1.2 | Realism and Instrumentalism in Science

The modelling approaches presented in this paper are in many ways related to the discussions of realism and instrumentalism in philosophy of science (Popper 1962; Lakatos 1978; Van Fraassen 1980; Putnam 1982; Psillos 1999, 2000; Rowbottom 2019; Chakravartty 2017). Scientific realism refers to approaches that consider the goal of science to be discovering the truth, albeit approximately. Theoretical (unobservable) entities, constructs, their properties, and relations are expected to reflect the reality of the subject matter under investigation. Scientific language is taken more literally and scientific progress is seen as increasing ‘verisimilitude’ or truth-likeness. Scientific

instrumentalism, on the other hand, considers the main goal of science to be accurate description and prediction of observations. Science is an instrument for understanding the relations among the observable data. Theoretical (unobservable) entities, constructs, properties, and relations are meant to be useful tools for making predictions and do not have to reflect reality. Scientific language, therefore, is taken to be largely figurative. In its strong form, instrumentalism argues that science is only a tool for understanding and predicting observable phenomena and scientists need not and should not commit to the reality of unobservables. Here we consider a weaker form of instrumentalism which allows for realist commitments as well. We consider them as two goals that members of the scientific community can aim to fulfil to different degrees separately or simultaneously.

Figure 1 represents these two goals along two axes, creating a space of possible theories, hypotheses, or models. Here we focus on models as specific implementations of theories or hypotheses. We also assume that the goal of modelling is to describe, predict, and explain human linguistic behaviour. The x-axis represents the degree to which a model is aimed at being instrumental; that is, it is proposed to describe observations regarding the external behaviour of a linguistic system and make accurate predictions. The researcher is not committed to the unobservable internal components or mechanisms of the model as real representations or descriptions of human linguistic cognition. The y-axis, on the other hand, represents the degree to which a model is committed to represent the reality of human cognition. A model low on the y-axis has little to no commitment to exhibit ‘verisimilitude’ with respect to its assumptions, components, or mechanisms. A model high on the y-

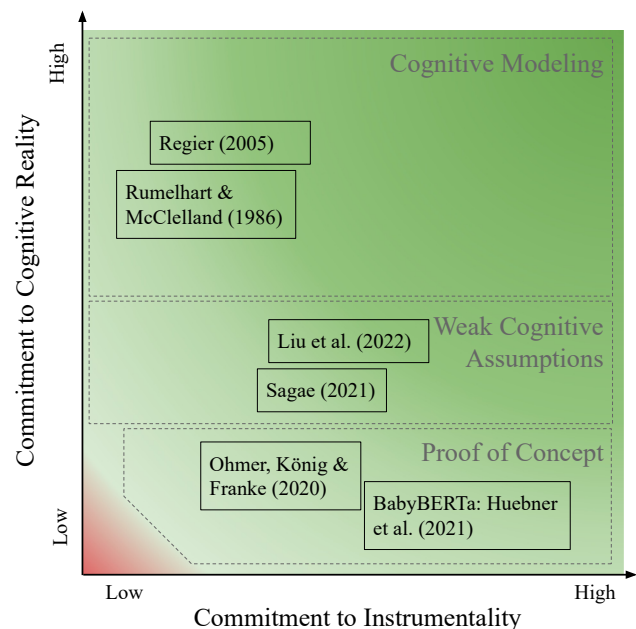


FIGURE 1 | Schematic depiction of realism and instrumentalism as two goals of modelling in cognitive science. The x-axis represents a model’s commitment to represent cognitive reality and the y-axis the model’s commitment to provide accurate prediction of linguistic behaviour, regardless of how much the model corresponds to the reality of human cognition.

axis is committed to representing entities, processes, or mechanisms that are cognitively real. The area around the point of origin (shaded red) is undesirable because models in that area satisfy neither goal. They are neither useful for prediction nor committed to be cognitively real. At the start, the challenge for any theory or model is to move away from this point of origin. Some researchers may prefer to create models that move right along the *x*-axis and predict observations accurately and reliably regardless of how realistic they are cognitively (i.e., their placement on the *y*-axis). Other researchers may care more about models moving up the *y*-axis, so that they can represent the internal structure of the human mind realistically regardless of how accurately they predict all observations. The top right region of this graph is very desirable because the models in that region are both committed to be cognitively real and useful for predicting linguistic behaviour.

Our model taxonomy and the discussion of scientific realism is also related to Chomsky's three levels of observational, descriptive, and explanatory adequacy (Chomsky 1957, 1964, 1965). A grammar (or a model of language) is observationally adequate if it correctly generates the sentences of a language. A grammar is descriptively adequate if in addition to being observationally adequate, it manages to capture the intuitions of native speakers and assign the right structural descriptions to the sentences of the language. Finally, a grammar that has explanatory adequacy selects the correct grammar from all possible grammars given the data available to a child. According to Chomsky, observational adequacy is concerned with the external linguistic behaviour, while descriptive and explanatory adequacy are concerned with mind-internal representations and processes. Therefore, observational adequacy is more aligned with an instrumentalist approach to linguistic science while descriptive and explanatory adequacy are more aligned with realist approaches.

We can think of model building as an act of moving around in the space of Figure 1. Different types of empirical evidence may help us move right or up in the space. More importantly, models themselves may guide us with respect to the type of empirical data needed to move in each direction. Of course each model is complicated and results from choices regarding data, assumptions, learning mechanisms, computational processes, storage, representation, ...that can contribute in different ways to a model's success in being instrumental or cognitively real. So one model may be realistic regarding the data or representations used, but unrealistic with respect to the learning mechanisms. Figure 1 should be considered as a 2D projection of a much more complex space in which different model components also vary along these dimensions.

2 | A Taxonomy of Computational Modelling Approaches

In this section, we describe four ways that neural networks can be useful tools in answering scientific questions about language learning and cognition more broadly. In addition to reviewing current research which exemplifies each of these approaches, we offer clear definitions of the set of necessary assumptions

linking our models to our hypotheses about language acquisition. We finally situate each of these approaches in the space delimited by the trade-off between scientific aspirations towards realism versus instrumentalism and discuss their limitations.

2.1 | Cognitive Models

The first approach, which we referred to in the introduction as the 'traditional' approach is to define a neural network model as a cognitive model (Phillips and Pearl 2015; S. L. Frank, Monaghan, and Tsoukala 2019). By cognitive modelling, we intend to mean that explicit linking hypotheses are established about the learning mechanisms, inputs, and learning outcomes present in both the model and human learners. In other words, the neural network model represents the cognitive processes we hypothesise could explain some linguistic learning outcome in children. By defining such a model, researchers make explicit their assumptions about how human processing may work and how these processes are emulated in their model. This is the approach first taken in the connectionist tradition (McClelland, Rumelhart, and PDP Research Group 1986, vol. 1, Elman et al. 1996).

In connectionism, neural network models were seen as simplified models of the mind, where networks of synthetic neurons governed by connections and weights could be used to model cognitive behaviours such as language learning. Though these models were intended to be biologically plausible, they did not necessarily argue that their specific network connections mirrored those in the brain, but were approximations, or models, that learnt and used both knowledge representations and learning mechanisms that emulated human processing.

Before the most recent rise of deep learning in the mid 2010s, neural network models were often described as cognitive models and used to study multiple language acquisition phenomena, to name a few: speech segmentation and word identification (Cairns et al. 1997; Christiansen, Allen, and Seidenberg 1998), learning the meanings of words (Rogers and McClelland 2004; Regier 2005; Mayor and Plunkett 2010; McMurray, Horst, and Samuelson 2012; Samuelson, Kucker, and Spencer 2017), and syntactic acquisition (Elman 1990, 1991; F. Chang, Dell, and Bock 2006; Clair, Monaghan, and Christiansen 2010; Christiansen, Dale, and Real 2010). These studies all used models that were explicitly designed to instantiate hypotheses researchers had about language learning in humans. As such they also made testable predictions that could be studied in parallel experimentally.

The cognitive modelling approach typically has high commitment to the cognitive and developmental reality of its models. To put it in terms of Figure 1, the goal is to propose models that increase verisimilitude on the *y*-axis. The proposed models often focus on modelling specific and limited linguistic phenomena such as children's developmental trajectory in learning the English past tense (Rumelhart and McClelland 1986) or developmental patterns in children's word learning (Regier 2005). Model design is often informed by evidence about human cognition that motivates specific internal components, structures, or processes

to the model. In fact, the focus of cognitive modelling is the internal structure of the model and the human mind rather than the external linguistic behaviour. For example, Regier (2005) focuses on simulating children's word learning but for word representations, the model uses one-hot vectors not tied to any specific words. This is because the goal of such modelling is not to cover more observational grounds regarding external linguistic behaviour but rather discovering the internal constructs and mechanisms of the human mind using computational modelling.

2.1.1 | Key Insight

Cognitive models can be used to help test and confirm hypotheses we have about the processes behind language learning outcomes. Though they make strong assumptions, they also offer the possibility to make strong conclusions, the strongest of all the approaches discussed in this paper.

2.2 | Models With Weaker Cognitive Assumptions

Taking the very strong position that contemporary neural networks are models of human processing is often disfavoured since most are not designed to be cognitive models (S. L. Frank et al. 2019; Mahowald et al. 2024). It is still however possible to take a cognitive modelling approach to using neural networks, albeit with softer assumptions. Our second approach expects explicit linking assumptions between the inputs, learning outcomes, and some subset of the learning processes involved in language acquisition for both the models and children. By a subset of the learning processes, we intend to mean that the model may instantiate prior learning biases, intermediate representations, or intermediate input transformations that we hypothesise to be available in human learners as well (see Johns, Jamieson, and Jones 2023 for further discussion of separation between representation and processing in cognitive models). For example, we could assume that there exists a clear link between a models' intermediate word meaning representations and those of humans, without assuming that these word representations are achieved using the same processing steps in humans and machines. We could then use the word embeddings of a language model trained on realistic data as an approximation for our underlying word representations. This assumption is implicitly made whenever latent semantic representations are used in studies of human linguistic behaviour, such as using them to measure semantic similarity (S. L. Frank and Willems 2017; Hollenstein et al. 2019; Michaelov et al. 2021).

In language acquisition, this approach has been recently used to study the relations between language model sequential predictability and novel word acquisition (Portelance, Degen, and Frank 2020; Portelance et al. 2023), language model sentence representations and syntactic development (Sagae 2021), as well as model representations of previous linguistic and auditory contexts and backchannel signalling use in conversation (Liu et al. 2022).¹ In all of these cases, the models are trained on realistic input, such as corpora of child directed utterances and recordings of child and caregiver interactions, and are used to

predict some real learning outcome, such as the average age of acquisition of words, the syntactic development of a child, or the uses of backchannel signalling by children. None however commit to their language models being cognitive models, only that the intermediate representations learnt by these models may be useful proxies for those used by human learners. For example, Liu et al. (2022) use a weakly cognitive modelling approach to help answer the question: when do children develop adult-proficiency in their use of backchannel signals in conversation? Some previous work suggested that children in middle childhood (6–12 year olds) still have not developed this skill completely (Hess and Johnston 1988; Dittmann 1972), while other work found that they produce backchannel cues at a similar frequency to adults (Bodur et al. 2023). Liu et al. help settle the question by showing that these children not only produce cues at the same rate as adults, but do so in the same conversational contexts, showing mastery of conversational dynamics. To do so, they use neural networks trained on previous contextual information to predict whether or not adults or children produce backchannel signals in such contexts. Using video recordings of child-caregiver interactions and minimally different adult-adult interactions (Bodur et al. 2023), they extract previous acoustic, visual, and verbal cues, some of which are measured using intermediary neural network model representations. One such case is final word surprisal, where higher final word surprisal in a context—less predictable or newer information—might be expected to elicit backchannelling from the listener. They measure surprisal using the intermediary representations of a large language model trained on and to produce text conversations, assuming that this model's word surprisal in such contexts is a good proxy for a listener's surprisal in the same contexts.² They find that indeed children's backchannel productions are just as predictable as adults using a mixture of visual, verbal, and acoustic cues.

Models with weaker cognitive assumptions have the goal of capturing aspects of human cognition but they do not commit to the reality of their constructs and mechanisms as much as cognitive models do. In terms of realist-instrumentalist goals and Figure 1, models with weak cognitive assumptions rank lower on the y-axis while they vary on where they land on the x-axis. The weaker cognitive assumptions often allows such models to be more flexible with their internal structure and cover more ground regarding their external linguistic behaviour. Weak cognitive models talk about their representations non-literally and typically suggest that aspects of their models capture the information that the mind represents or processes but perhaps not in the same format and using the same mechanisms as used in their models.

2.2.1 | Key Insight

Weakly cognitive models still boast some amount of cognitive plausibility, as not only do they reproduce predictive input to learning outcome behaviour, but they additionally assume that some amount of the model's internal learning mechanisms is a good approximation of relevant cognitive representations or processes. This approach can be useful for developing quantitative metrics which are grounded in these models as a way to

measure or predict a variety of word learning phenomena in children.

2.3 | Models as Proofs-of-Concepts

This third approach considers models as independent learners from humans—in other words like a new ‘species’ of language learners—that can be leveraged to implement ‘proofs of concepts’ (Lappin 2021, ch. 1.2, Tsuji, Cristia, and Dupoux 2021; Warstadt and Bowman 2022; Pearl 2023). A proof of concept can show us what is possible ‘in practice’ for models and ‘in principle’ for humans. If a model can learn to reproduce certain human learning outcomes given the presence of specific input information and learning strategies, then a human could ‘in principle’ also reach these outcomes with the help of that same input information and possibly similar learning strategies. In order to connect these model learners to humans learners, we must also weigh in on what leads models to learn a given outcome in the first place, and acknowledge that these proposed prerequisites for learning must also be available to kids acquiring language (Baroni 2022).

Let us consider a concrete example: whether the mutual exclusivity bias (Markman and Wachtel 1988) present in young children learning new words is an innate bias or a learnt one and why. Lewis et al. (2020) offer a review of current proposals as to the origins of this bias which, beyond including the theory that this bias is innate or early emerging (Markman, Wasow, and Hansen 2003), also include different proposals about what may lead children to learn this bias, such as it following from pragmatic reasoning about a speakers intentions (E. V. Clark 1988; Diesendruck and Markson 2001; M. C. Frank, Goodman, and Tenenbaum 2009) or from a more basic tracking of statistical patterns in language mappings (Regier 2005; McMurray, Horst, and Samuelson 2012). These last two studies in fact used neural networks to support their claims about the mutual exclusivity bias following from statistical regularities. They proposed that observing the predominance of one-to-one mappings in the input can lead to the generalisation that one-to-one mappings should also extend to novel data. However, the tests that both these studies used did not truly show that models had a mutual exclusivity bias when presented with unseen input, only that they showed a preference for one-to-one mappings in seen input. As Gandhi and Lake (2020) demonstrate using multiple neural network architectures, these models do not show a mutual exclusivity bias for out-of-sample generalisation—or unseen data. Models must have some prior heuristic bias or reasoning ability to truly develop a mutual exclusivity bias. Ohmer, König, and Franke (2020) show that by integrating the ability to reason about speakers intentions (M. C. Frank et al. 2009) into neural networks, models will develop over time a true mutual exclusivity bias when presented with novel stimuli. Importantly, they also show using an ablation study that in the absence of the ability for pragmatic reasoning, models do not develop a mutual exclusivity bias. In so doing, they have not only shown that the bias is learnable, but have demonstrated what leads their models to learn it in the first place. Using a proof of concept approach, these researchers were able to offer evidence in support of a specific proposal for the

origins of the bias, namely that the bias is learnable and follows from pragmatic reasoning.

In language acquisition research, the burden of proof often falls on hypotheses proposing that some linguistic behaviour is learnt, while nativist hypotheses most often rely on the absence of evidence, such as *poverty of the stimulus*-style arguments (A. Clark and Lappin 2011; Lasnik and Lidz 2016). The proof-of-concept approach can model how in practice some linguistic outcome can be learnt, offering support for these types of hypotheses. This approach has been one of the most fruitful in recent years and has been used to support multiple hypotheses about the learnability of: the shape bias (Ohmer et al. 2021), basic syntactic dependencies (Huebner et al. 2021), syntactic and semantic categories (Wang et al. 2023), associative word learning strategies (Vong and Lake 2022), word meaning acquisition (Vong et al. 2024), words from sustained attention (Tsutsui et al. 2020), and logical reasoning words (Portelance, Frank, and Jurafsky 2024). These examples all favour the effective reproduction of children’s learning outcomes, but offer very little insight into the processes that lead to them.

Within the realist-instrumentalist space depicted in Figure 1, proof-of-concept models in language acquisition should be placed very low on the y-axis since they do not commit to the reality of the components, structures, or mechanisms of the proposed models. They help us see what kind of learning or processing of language is possible. They may focus on very specific linguistic phenomena (Ohmer, König, and Franke 2020) or capture a wide range of linguistic behaviour (Huebner et al. 2021), without necessarily committing to the internal structure or representations of the model. Which aspects of such models are closer to the reality of human acquisition and processing will be determined by further commitments within the cognitive modelling approach discussed earlier or with empirical tests within the hypothesis generation approach which we will discuss next.

2.3.1 | Key Insight

Though proofs-of-concept are important and can help us better define our hypotheses, they cannot offer any form of hypothesis confirmation or testing, nor do they make predictions which could be leveraged in experimental settings. Without explicitly linking our models’ representations and learning mechanisms to cognitive processes, they cannot lead to conclusions about human language acquisition; they can only help us formalise and define our existing hypotheses further.

2.4 | Models for Hypothesis Generation

The fourth and final approach may remedy some of the shortcomings of the proof of concept models. Much like the previous approach, it involves the study of neural network models as independent language learners. However, instead of offering support for existing proposals, it uses these models to propose new hypotheses about language learning in children, which can

themselves make testable predictions. For these reasons, we called them models for hypothesis generation.

Beyond the trend of evaluating existing large language models' ability to learn natural language, there has also been a push to design new models that are once again inspired by human learners (Lake et al. 2017; Bisk et al. 2020; McClelland et al. 2020; Warstadt et al. 2023; Mahowald et al. 2024). One of the main proposals in this direction is the use of multimodal input data, where models learn language from grounded contexts, not only using linguistic or symbolic input, but paired linguistic input with situational information like auditory, visual, social or embodied input, and beyond (Lake et al. 2017; Chevalier-Boisvert et al. 2019; Bender and Koller 2020; Bisk et al. 2020; McClelland et al. 2020; Hill et al. 2020).

When we design these new models for a learning task, we must formally define their architecture and learning procedures, laying out all the structural and functional assumptions we have about how best to have our model accomplish this learning task. In so doing computational models can be thought of as formal theories. If these models are theories, then they can be used to generate and test novel hypotheses. Models can be especially useful for testing hypotheses about causal relations between learning outcomes and either input data or structural biases, since we can easily design experiments that vary the presence or absence of hypothesised causes and study the effects of our manipulations—something that is not easily done with children.

Now, in considering these neural networks as language learners, we are not going so far as to say that our models are theories of human language processing. Instead, the hypotheses we may test and generate from them are a priori about how neural networks learn. In order to propose that our hypotheses may extend to human learners, we must establish that there are similarities between the model's learning tasks and human language learning and furthermore, that the same possible causes and effects are present in children's language learning environment. Then, we can propose that our hypothesis extends to human learners, though additional evidence will still be needed to test and prove it. Thus, using this hypothesis generation approach, we can help expand our understanding of language learning in children not by providing evidence in favour of existing proposals, but as a tool to inspire new ones.

Examples of this approach have used a variety of multimodal neural networks architectures to propose novel hypotheses about language learning, including about: defining new models of the interdependence between phonetic and phonological acquisition (Beguš 2020), the origins of the shape bias (Portelance et al. 2021), the learning principles behind systematicity (Lake and Baroni 2023), and the synergistic interactions between production-based learning (Nikolaus and Fourtassi 2021) as well as active-based learning (Gelderloos, Kamelabad, and Alishahi 2020) with perceptual cross-situational word learning.

As an example of how this approach may be applied, let us consider the case of Portelance et al. (2021). The shape bias is a word learning bias whereby when children are presented with a new word referring to a novel category of referents, they assume

that these objects are more likely to share shape properties over other visual features like colour or texture. It is a learnt bias, appearing around 2 years old (Heibeck and Markman 1987; Landau, Smith, and Jones 1988), but why we should have it in our languages in the first place is an open question. Portelance et al. use neural emergent language agents that learn to communicate about raw pixelated images as a way to model both language acquisition and evolution. Much like humans, agents develop efficient communication systems (Kirby et al. 2015; Kemp, Xu, and Regier 2018; Gibson et al. 2019). The authors find that the bias emerges and persists between generations of 'parent' and 'child' agents as a function of communicative need to distinguish word classes based on shape. They then suggest that this same cause or pressure may be present for human learners since shape is often an affordance of objects, especially man-made objects, and communication is often about function. They make the prediction that the shape bias should be stronger in linguistic societies with a stronger association between shape and objects' use. In fact, this prediction was later corroborated in concurrent experimental work (Jara-Ettinger et al. 2022). Thus, neural networks were used to propose a novel hypothesis about a word learning bias in children, which generated a testable prediction. This prediction was then substantiated by experimental work in humans, strengthening the proposed hypothesis.

It is common to design models based on assumptions and knowledge regarding human cognition. In other words, the modelling uses what is known about human cognition to design models that reflect and explain it. Models used for hypothesis generation do the opposite. They use what is known about models to generate hypotheses regarding how humans learn and process language. As such, they are tools for experimental refutation or corroboration of hypotheses. They help researchers add to their commitments, moving models up the y-axis in Figure 1. By testing different internal components of models against empirical data, we can gather more evidence for specific model architectures and get closer to models that are weakly or strongly cognitive. Therefore, using models to generate hypothesis that are empirically testable is extremely useful in increasing verisimilitude of linguistic models.

2.4.1 | Key Insight

Models for hypothesis generation a priori define novel hypotheses about language learning in neural networks. They can however be leveraged to make testable predictions. If these predictions are found to hold in human learners, then these novel hypotheses may also be extended to inform our theories about language acquisition in children. As such, these models may encourage the exploration of new theories and ideas in our field.

3 | Final Discussion

In this paper we outlined four ways in which neural networks may be useful for studying questions about language acquisition. Beyond the traditional cognitive modelling approach,

newer neural networks which are not explicitly designed as models of human processing have drawn researchers to think of novel ways to use these models in the field. We define three such methods: cognitive models with weaker assumptions, proof of concept models, and models for hypothesis generation.

Each of these approaches can be useful in different ways. They allow us to test and validate hypotheses, define novel language learning metrics, instantiate examples of learnability hypotheses, and finally generate novel testable prediction for new language learning theories. All of these approaches play an important role and depend on each other for their development. For example, before a researcher commits to a model as cognitive, they may first try a proof-of-concept version of the model and subsequently use an aspect of the model to generate a testable prediction for an empirical study. Empirical studies corroborating different aspects of models in turn allow researchers to commit to the cognitive reality of their models. For this reason, in an ideal research ecosystem we should expect to see examples of all of these approaches. We connected these four approaches to the discussion of realism and instrumentalism within philosophy of science to show that recent developments are in line with foundational goals within science and linguistics, and are a healthy part of scientific development in the field.

In recent years, proof-of-concept models have seen a significant rise in popularity, while cognitive neural network models have all but disappeared. This may be because making stronger theoretical commitments, like those required for cognitive models, has become a more difficult endeavour. Those willing to design new models and test new ideas have instead shifted towards approaches with fewer theoretical commitments like models for hypothesis generation or cognitive models with weaker assumptions (we, the authors, are also guilty of this). We would like to encourage researchers to be more willing to commit to their models and theories, even if this may eventually mean that they were wrong. Science progresses fastest when there are theories to refute or corroborate. As members of the community, we should also be more accepting of cognitive modelling proposals which make strong assumptions about the processes behind word learning, even if evidence for them is still lacking, as these models will only encourage further exploration and development in the field. There is still so much to learn about how we learn language!

In writing this paper, we took the opportunity to highlight the impressive and creative work of up and coming researchers in the field. We hope that this paper may serve as a guide for future research and encourage others to use these models while knowing what can and cannot be concluded about human cognition from these contemporary neural network models.

Author Contributions

This paper is inspired by the first chapter of Eva Portelance's dissertation which covered the neural network model taxonomy discussed here. Masoud Jasbi contributed the discussion about how each of these approaches map to the realism versus instrumentalism debates in philosophy of science.

Acknowledgements

The authors would like to acknowledge the following people for their helpful comments and discussion at different stages of the development process for this piece: Dan Jurafsky, Michael C. Frank, Christopher Potts, Alexander Warstadt, Adam Sennet, and Patrick Skeel.

Endnotes

¹ Backchannelling refers to the subtle cues that listeners make to signal understanding or attentive listening to a speaker during conversation (e.g., head nods or 'uh-huh').

² A similar approach is also used in Portelance et al. (2023).

References

- Baroni, M. 2022. "On the Proper Role of Linguistically Oriented Deep Net Analysis in Linguistic Theorising." In *Algebraic Structures in Natural Language*, 1–16. Milton Park: Taylor & Francis Group.
- Beguš, G. 2020. "Generative Adversarial Phonology: Modeling Unsupervised Phonetic and Phonological Learning With Neural Networks." *Frontiers in Artificial Intelligence* 3: 44. <https://doi.org/10.3389/frai.2020.00044>.
- Bender, E. M., and A. Koller. 2020. "Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198.
- Bisk, Y., A. Holtzman, J. Thomason, et al. 2020. "Experience Grounds Language." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 8718–8735.
- Bodur, K., M. Nikolaus, L. Prévot, and A. Fourtassi. 2023. "Using Video Calls to Study Children's Conversational Development: The Case of Backchannel Signaling." *Frontiers of Computer Science* 5: 1088752. <https://doi.org/10.3389/fcomp.2023.1088752>.
- Bommasani, R., D. A. Hudson, E. Adeli, et al. 2022. "On the Opportunities and Risks of Foundation Models." *arXiv Preprint arXiv:2108.07258*.
- Brown, T., B. Mann, N. Ryder, et al. 2020. "Language Models Are Few-Shot Learners." In *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Vol. 33, 1877–1901. New York: Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc94967418bfb8ac142f64a-Paper.pdf.
- Cairns, P., R. Shillcock, N. Chater, and J. Levy. 1997. "Bootstrapping Word Boundaries: A Bottom-Up Corpus-Based Approach to Speech Segmentation." *Cognitive Psychology* 33, no. 2: 111–153. <https://doi.org/10.1006/cogp.1997.0649>.
- Chakravartty, A. 2017. "Scientific Realism." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Stanford, CA: Stanford University. <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>.
- Chang, F., G. S. Dell, and K. Bock. 2006. "Becoming Syntactic." *Psychological Review* 113, no. 2: 234–272. <https://doi.org/10.1037/0033-295x.113.2.234>.
- Chang, T. A., and B. K. Bergen. 2022. "Word Acquisition in Neural Language Models." *Transactions of the Association for Computational Linguistics* 10: 1–16. https://doi.org/10.1162/tacl_a_00444.
- Chevalier-Boisvert, M., D. Bahdanau, S. Lahlou, et al. 2019. "BabyAI: First Steps Towards Grounded Language Learning With a Human in the Loop." In *Proceedings of the International Conference on Learning Representations*, Vol. 105.
- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton & Co.
- Chomsky, N. 1964. "Current Issues in Linguistic Theory." In *The Structure of Language: Readings in the Philosophy of Language*, edited by J. Fodor and J. Katz. Hoboken, NJ: Prentice-Hall.

- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Christiansen, M. H., J. Allen, and M. S. Seidenberg. 1998. "Learning to Segment Speech Using Multiple Cues: A Connectionist Model." *Language & Cognitive Processes* 13, no. 2–3: 221–268. <https://doi.org/10.1080/016909698386528>.
- Christiansen, M. H., R. Dale, and F. Real. 2010. "Connectionist Explorations of Multiple-Cue Integration in Syntax Acquisition." In *Neocognitivism: The New Science of Cognitive Development*, edited by S. P. Johnson, 87–108. Oxford: Oxford University Press.
- Clair, M. C. S., P. Monaghan, and M. H. Christiansen. 2010. "Learning Grammatical Categories From Distributional Cues: Flexible Frames for Language Acquisition." *Cognition* 116, no. 3: 341–360. <https://doi.org/10.1016/j.cognition.2010.05.012>.
- Clark, A., and S. Lappin. 2011. *Linguistic Nativism and the Poverty of the Stimulus*. Hoboken, NJ: John Wiley & Sons.
- Clark, E. V. 1988. "On the Logic of Contrast." *Journal of Child Language* 15, no. 2: 317–335. <https://doi.org/10.1017/s0305000900012393>.
- Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2019, June. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, edited by J. Burstein, C. Doran, and T. Solorio, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>.
- Diesendruck, G., and L. Markson. 2001. "Children's Avoidance of Lexical Overlap: A Pragmatic Account." *Developmental Psychology* 37, no. 5: 630–641. <https://doi.org/10.1037/0012-1649.37.5.630>.
- Dittmann, A. T. 1972. "Developmental Factors in Conversational Behavior." *Journal of Communication* 22, no. 4: 404–423. <https://doi.org/10.1111/j.1460-2466.1972.tb00165.x>.
- Elman, J. L. 1990. "Finding Structure in Time." *Cognitive Science* 14, no. 2: 179–211. https://doi.org/10.1207/s15516709cog1402_1.
- Elman, J. L. 1991. "Distributed Representations, Simple Recurrent Networks, and Grammatical Structure." *Machine Learning* 7, no. 2–3: 195–225. <https://doi.org/10.1007/bf00114844>.
- Elman, J. L., E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT press.
- Frank, M. C., N. D. Goodman, and J. B. Tenenbaum. 2009. "Using Speakers' Referential Intentions to Model Early Cross-Situational Word Learning." *Psychological Science* 20, no. 5: 578–585. <https://doi.org/10.1111/j.1467-9280.2009.02335.x>.
- Frank, S. L., P. Monaghan, and C. Tsoukala. 2019. "Neural Network Models of Language Acquisition and Processing." In *Human Language: From Genes and Brain to Behavior*, 277–293. Cambridge, MA: MIT Press.
- Frank, S. L., and R. M. Willems. 2017. "Word Predictability and Semantic Similarity Show Distinct Patterns of Brain Activity During Language Comprehension." *Language, Cognition and Neuroscience* 32, no. 9: 1192–1203. <https://doi.org/10.1080/23273798.2017.1323109>.
- Futrell, R., E. Wilcox, T. Morita, P. Qian, M. Ballesteros, and R. Levy. 2019. "Neural Language Models as Psycholinguistic Subjects: Representations of Syntactic State." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 32–42.
- Gandhi, K., and B. M. Lake. 2020. "Mutual Exclusivity as a Challenge for Deep Neural Networks." *Advances in Neural Information Processing Systems* 33: 14182–14192.
- Gelderloos, L., A. M. Kamelabad, and A. Alishahi. 2020. "Active Word Learning Through Self-Supervision." In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society (CogSci)*. <https://cognitivesciencesociety.org/cogsci20/papers/0183/0183.pdf>.
- Gibson, E., R. Futrell, S. P. Piantadosi, et al. 2019. "How Efficiency Shapes Human Language." *Trends in Cognitive Sciences* 23, no. 12: 389–407. <https://doi.org/10.1016/j.tics.2019.09.005>.
- Guest, O., and A. E. Martin. 2021. "How Computational Modeling Can Force Theory Building in Psychological Science." *Perspectives on Psychological Science* 16, no. 4: 789–802. <https://doi.org/10.1177/1745691620970585>.
- Haley, C. 2020. "This Is a BERT. Now There Are Several of Them. Can They Generalize to Novel Words?" In *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, 333–341.
- Heibeck, T. H., and E. M. Markman. 1987. "Word Learning in Children: An Examination of Fast Mapping." *Child Development* 58, no. 4: 1021–1034. <https://doi.org/10.1111/j.1467-8624.1987.tb01438.x>.
- Hess, L. J., and J. R. Johnston. 1988. "Acquisition of Back Channel Listener Responses to Adequate Messages." *Discourse Processes* 11, no. 3: 319–335. <https://doi.org/10.1080/01638538809544706>.
- Hill, F., A. Lampinen, R. Schneider, et al. 2020. "Environmental Drivers of Systematicity and Generalization in a Situated Agent." In *Proceedings of the International Conference on Learning Representations*.
- Hollenstein, N., A. de la Torre, N. Langer, and C. Zhang. 2019. "CogniVal: A Framework for Cognitive Word Embedding Evaluation." In *Proceedings of the 23rd Conference on Computational Natural Language Learning*, 538–549.
- Hu, J., J. Gauthier, P. Qian, E. Wilcox, and R. Levy. 2020. "A Systematic Assessment of Syntactic Generalization in Neural Language Models." In *Proceedings of the Association for Computational Linguistics*.
- Huebner, P. A., E. Sulem, F. Cynthia, and D. Roth. 2021. "BabyBERTa: Learning More Grammar With Small-Scale Child-Directed Language." In *Proceedings of the 25th Conference on Computational Natural Language Learning*, 624–646.
- Huebner, P. A., and J. A. Willits. 2018. "Structured Semantic Knowledge Can Emerge Automatically From Predicting Word Sequences in Child-Directed Speech." *Frontiers in Psychology* 9: 133. <https://doi.org/10.3389/fpsyg.2018.00133>.
- Jara-Ettinger, J., R. P. Levy, J. Sakel, T. Huanca, and E. Gibson. 2022. "The Origins of the Shape Bias: Evidence From the Tsimane." *Journal of Experimental Psychology* 151, no. 10: 2437–2447. <https://doi.org/10.1037/xge0001195>.
- Johns, B. T., R. K. Jamieson, and M. N. Jones. 2023. "Scalable Cognitive Modelling: Putting Simon's (1969) Ant Back on the Beach." *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 77, no. 3: 185–201. <https://doi.org/10.1037/cep0000306>.
- Kemp, C., Y. Xu, and T. Regier. 2018. "Semantic Typology and Efficient Communication." *Annual Review of Linguistics* 4, no. 1: 109–128. <https://doi.org/10.1146/annurev-linguistics-011817-045406>.
- Kirby, S., M. Tamariz, H. Cornish, and K. Smith. 2015. "Compression and Communication in the Cultural Evolution of Linguistic Structure." *Cognition* 141: 87–102. <https://doi.org/10.1016/j.cognition.2015.03.016>.
- Lakatos, I. 1978 In *The Methodology of Scientific Research Programmes: Philosophical Papers*, edited by J. Worrall and G. Currie. Cambridge: Cambridge University Press.
- Lake, B. M., and M. Baroni. 2018. "Generalization Without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks." In *Proceedings of the International Conference on Machine Learning*, 2873–2882.

- Lake, B. M., and M. Baroni. 2023. "Human-Like Systematic Generalization Through a Meta-Learning Neural Network." *Nature* 623, no. 7985: 1–7. <https://doi.org/10.1038/s41586-023-06668-3>.
- Lake, B. M., T. Linzen, and M. Baroni. 2019. "Human Few-Shot Learning of Compositional Instructions." In *Proceedings of CogSci*.
- Lake, B. M., T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. 2017. "Building Machines That Learn and Think Like People." *Behavioral and Brain Sciences* 40: e253. <https://doi.org/10.1017/s0140525x16001837>.
- Landau, B., L. B. Smith, and S. S. Jones. 1988. "The Importance of Shape in Early Lexical Learning." *Cognitive Development* 3: 299–321. [https://doi.org/10.1016/0885-2014\(88\)90014-7](https://doi.org/10.1016/0885-2014(88)90014-7).
- Lappin, S. 2021. *Deep Learning and Linguistic Representation*. Boca Raton, FL: CRC Press.
- Lasnik, H., and J. Lidz. 2016. "The Argument From the Poverty of the Stimulus." In *The Oxford Handbook of Universal Grammar*, 221–248. Oxford: Oxford University Press.
- Lau, J. H., A. Clark, and S. Lappin. 2017. "Grammaticality, Acceptability, and Probability: A Probabilistic View of Linguistic Knowledge." *Cognitive Science* 41, no. 5: 1202–1241. <https://doi.org/10.1111/cogs.12414>.
- Lavechin, M., Y. Sy, H. Titeux, et al. 2023. "Babyslm: Language-Acquisition-Friendly Benchmark of Self-Supervised Spoken Language Models." In *Interspeech*.
- Lewis, M., V. Cristiano, B. M. Lake, T. Kwan, and M. C. Frank. 2020. "The Role of Developmental Change and Linguistic Experience in the Mutual Exclusivity Effect." *Cognition* 198: 104191. <https://doi.org/10.1016/j.cognition.2020.104191>.
- Li, B., W. Wu, Z. Tang, et al. 2024. "Devbench: A Comprehensive Benchmark for Software Development." *arXiv Preprint arXiv:2403.08604*.
- Linzen, T. 2019. "What Can Linguistics and Deep Learning Contribute to Each Other? Response to Pater." *Language* 95, no. 1: e99–e108. <https://doi.org/10.1353/lan.2019.0015>.
- Linzen, T., E. Dupoux, and Y. Goldberg. 2016. "Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies." *Transactions of the Association for Computational Linguistics* 4: 521–535. https://doi.org/10.1162/tacl_a_00115.
- Liu, J., M. Nikolaus, K. Bodur, and A. Fourtassi. 2022. "Predicting Backchannel Signaling in Child-Caregiver Multimodal Conversations." In *Companion Publication of the 2022 International Conference on Multimodal Interaction*, 196–200.
- Mahowald, K., A. A. Ivanova, I. A. Blank, N. Kanwisher, J. B. Tenenbaum, and E. Fedorenko. 2024. "Dissociating Language and Thought in Large Language Models." *Trends in Cognitive Sciences* 28, no. 6: 517–540. <https://doi.org/10.1016/j.tics.2024.01.011>.
- Manning, C. D., K. Clark, J. Hewitt, U. Khandelwal, and O. Levy. 2020. "Emergent Linguistic Structure in Artificial Neural Networks Trained by Self-Supervision." *Proceedings of the National Academy of Sciences* 117, no. 48: 30046–30054. <https://doi.org/10.1073/pnas.1907367117>.
- Markman, E. M., and G. F. Wachtel. 1988. "Children's Use of Mutual Exclusivity to Constrain the Meanings of Words." *Cognitive Psychology* 20, no. 2: 121–157. [https://doi.org/10.1016/0010-0285\(88\)90017-5](https://doi.org/10.1016/0010-0285(88)90017-5).
- Markman, E. M., J. L. Wasow, and M. B. Hansen. 2003. "Use of the Mutual Exclusivity Assumption by Young Word Learners." *Cognitive Psychology* 47, no. 3: 241–275. [https://doi.org/10.1016/s0010-0285\(03\)00034-3](https://doi.org/10.1016/s0010-0285(03)00034-3).
- Mayor, J., and K. Plunkett. 2010. "A Neurocomputational Account of Taxonomic Responding and Fast Mapping in Early Word Learning." *Psychological Review* 117, no. 1: 1–31. <https://doi.org/10.1037/a0018130>.
- McClelland, J. L., F. Hill, M. Rudolph, J. Baldridge, and H. Schütze. 2020. "Placing Language in an Integrated Understanding System: Next Steps toward Human-Level Performance in Neural Language Models." *Proceedings of the National Academy of Sciences* 117, no. 42: 25966–25974. <https://doi.org/10.1073/pnas.1910416117>.
- McClelland, J. L., D. E. Rumelhart, and PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT press.
- McMurray, B., J. S. Horst, and L. K. Samuelson. 2012. "Word Learning Emerges From the Interaction of Online Referent Selection and Slow Associative Learning." *Psychological Review* 119, no. 4: 831–877. <https://doi.org/10.1037/a0029872>.
- Michaelov, J. A., M. D. Bardolph, S. Coulson, and B. K. Bergen. 2021. "Different Kinds of Cognitive Plausibility: Why Are Transformers Better Than RNNs at Predicting N400 Amplitude?" In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society (CogSci)*.
- Nikolaus, M., and A. Fourtassi. 2021. "Modeling the Interaction Between Perception-Based and Production-Based Learning in Children's Early Acquisition of Semantic Knowledge." In *Proceedings of the 25th Conference on Computational Natural Language Learning*, 391–407.
- Ohmer, X., P. König, and M. Franke. 2020. "Reinforcement of Semantic Representations in Pragmatic Agents Leads to the Emergence of a Mutual Exclusivity Bias." In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society (CogSci)*, Vol. 42, 1779–1785.
- Ohmer, X., M. Marino, P. König, and M. Franke. 2021. "Why and How to Study the Impact of Perception on Language Emergence in Artificial Agents." In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society (CogSci)*.
- OpenAI. 2023. "GPT-4 Technical Report." *arXiv Preprint arXiv:2303.08774*.
- Pearl, L. 2023. "Computational Cognitive Modeling for Syntactic Acquisition: Approaches That Integrate Information From Multiple Places." *Journal of Child Language* 50, no. 6: 1–21. <https://doi.org/10.1017/s0305000923000247>.
- Phillips, L., and L. Pearl. 2015. "The Utility of Cognitive Plausibility in Language Acquisition Modeling: Evidence From Word Segmentation." *Cognitive Science* 39, no. 8: 1824–1854. <https://doi.org/10.1111/cogs.12217>.
- Piantadosi, S. 2023. "Modern Language Models Refute Chomsky's Approach to Language." *Lingbuzz Preprint: 007180*.
- Popper, K. 1962. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.
- Portelance, E., J. Degen, and M. C. Frank. 2020. "Predicting Age of Acquisition in Early Word Learning Using Recurrent Neural Networks." In *Proceedings of 42nd Annual Meeting of the Cognitive Science Society (CogSci)*, 1057–1063.
- Portelance, E., Y. Duan, M. C. Frank, and G. Lupyan. 2023. "Predicting Age of Acquisition for Children's Early Vocabulary in Five Languages Using Language Model Surprisal." *Cognitive Science* 47, no. 9: e13334. <https://doi.org/10.1111/cogs.13334>.
- Portelance, E., M. C. Frank, and D. Jurafsky. 2024. "Learning the Meanings of Function Words From Grounded Language Using a Visual Question Answering Model." *Cognitive Science* 48, no. 5: e13448. <https://doi.org/10.1111/cogs.13448>.
- Portelance, E., M. C. Frank, D. Jurafsky, A. Sordoni, and R. Laroché. 2021. "The Emergence of the Shape Bias Results From Communicative Efficiency." In *Proceedings of the 25th Conference on Computational Natural Language Learning*, 607–623.
- Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.
- Psillos, S. 2000. "The Present State of the Scientific Realism Debate." *British Journal for the Philosophy of Science* 51, no. 4: 705–728. <https://doi.org/10.1093/bjps/51.4.705>.

- Putnam, H. 1982. "Three Kinds of Scientific Realism." *Philosophical Quarterly* 32, no. 128: 195–200. <https://doi.org/10.2307/2219323>.
- Regier, T. 2005. "The Emergence of Words: Attentional Learning in Form and Meaning." *Cognitive Science* 29, no. 6: 819–865. https://doi.org/10.1207/s15516709cog0000_31.
- Rogers, T. T., and J. L. McClelland. 2004. *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT press.
- Rowbottom, D. P. 2019. *The Instrument of Science: Scientific Anti-Realism Revitalised*. London: Routledge.
- Rumelhart, D. E., and J. L. McClelland. 1986. "On Learning the Past Tenses of English Verbs." In *Parallel Distributed Processing*, edited by J. McClelland, D. Rumelhart, and the PDP Research Group. Cambridge, MA: MIT Press.
- Sagae, K. 2021. "Tracking Child Language Development With Neural Network Language Models." *Frontiers in Psychology* 12. <https://doi.org/10.3389/fpsyg.2021.674402>.
- Samuelson, L. K., S. C. Kucker, and J. P. Spencer. 2017. "Moving Word Learning to a Novel Space: A Dynamic Systems View of Referent Selection and Retention." *Cognitive Science* 41, no. S1: 52–72. <https://doi.org/10.1111/cogs.12369>.
- Sprouse, J., B. Yankama, S. Indurkha, S. Fong, and R. C. Berwick. 2018. "Colorless Green Ideas Do Sleep Furiously: Gradient Acceptability and the Nature of the Grammar." *Linguistic Review* 35, no. 3: 575–599. <https://doi.org/10.1515/tlr-2018-0005>.
- Touvron, H., L. Martin, K. Stone, et al. 2023. "Llama 2: Open Foundation and Fine-Tuned Chat Models." *arXiv Preprint arXiv:2307.09288*.
- Tsuji, S., A. Cristia, and E. Dupoux. 2021. "SCALa: A Blueprint for Computational Models of Language Acquisition in Social Context." *Cognition* 213: 104779. <https://doi.org/10.1016/j.cognition.2021.104779>.
- Tsutsui, S., A. Chandrasekaran, M. A. Reza, D. Crandall, and C. Yu. 2020. "A Computational Model of Early Word Learning From the Infant's Point of View." In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society (CogSci)*. <https://cognitivesciencesociety.org/cogsci20/papers/0180/0180.pdf>.
- Van Fraassen, B. C. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Vong, W. K., and B. M. Lake. 2022. "Cross-Situational Word Learning With Multimodal Neural Networks." *Cognitive Science* 46, no. 4: e13122. <https://doi.org/10.1111/cogs.13122>.
- Vong, W. K., W. Wang, A. E. Orhan, and B. M. Lake. 2024. "Grounded Language Acquisition Through the Eyes and Ears of a Single Child." *Science* 383, no. 6682: 504–511. <https://doi.org/10.1126/science.ad11374>.
- Wang, W., W. K. Vong, N. Kim, and B. M. Lake. 2023. "Finding Structure in One Child's Linguistic Experience." *Cognitive Science* 47, no. 6: e13305. <https://doi.org/10.1111/cogs.13305>.
- Warstadt, A., and S. R. Bowman. 2022. "What Artificial Neural Networks Can Tell Us About Human Language Acquisition." In *Algebraic Structures in Natural Language*, edited by S. Lappin and J.-P. Bernardy. Milton Park: Taylor & Francis Group.
- Warstadt, A., A. Mueller, L. Choshen, et al. 2023. "Findings of the BabyLM Challenge: Sample-Efficient Pretraining on Developmentally Plausible Corpora." In *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*, edited by A. Warstadt, et al., 1–34. Singapore: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.conll-babylm.1>.
- Warstadt, A., A. Singh, and S. R. Bowman. 2019. "Neural Network Acceptability Judgments." *Transactions of the Association for Computational Linguistics* 7: 625–641. https://doi.org/10.1162/tac1_a_00290.