

Child Development and Language Model Training: Comparing Psychological Theories of Learning

Luca Capone

Università di Pisa

luca.capone@fileli.unipi.it

Abstract This article presents a comparison between language model training techniques and the phenomenon of child language learning. The concept of learning, while central to human cognition, has struggled to find a stable place within the theoretical debate on language models, largely because training procedures appear to bear little resemblance to human learning and are markedly less efficient. The paper provides an overview of the relationship between the two processes, examining the theoretical challenges raised by dominant approaches in both the language model and developmental psychology literatures. Its aim is to assess whether learning and training are fundamentally incommensurable, or whether a comparison between the two might yield insights into both models' behaviour and child cognitive development. The analysis reveals interesting points of contact, shifting the focus of the discussion away from biological and psychological innate constraints toward environmental and linguistic learning conditions.

Keywords: linguistics, philosophy of language, language models, language acquisition, cognitive development

Received 18/04/2025; accepted 14/10/2025.

«all science would be superfluous if the form of appearance of things directly coincided with their essence» (Marx 1867, vol. 3, part VII, chap. 48, III)

0. Introduction

Language models (LMs) have reshaped debates across linguistic disciplines. Two issues have been of particular concern to scholars: the *grounding* of word embeddings (Bender and Koller 2020) and the question of whether models possess minds (Chalmers 2023, Goldstein and Levinstein 2024). The first issue appears to be nearing resolution. The idea that LMs can provide representations of the meaning of signs of natural-historical languages is gaining traction, albeit within heterogeneous theoretical frameworks (Capone 2021, 2024, Piantadosi and Hill 2022, Lenci 2023, Mandelkern and Linzen 2024, Merrill et al. 2024). The second issue seems likely to remain a matter of principle, one that hinges on how the concept of *mind* is defined. These two debates have partly overlooked what is arguably the most distinctive feature of human cognition that models have managed to

replicate: *learning*. For decades, learning has posed a major challenge in language studies, as exemplified by the *poverty of the stimulus argument* and the *language of thought hypothesis* (Chomsky 1980, Fodor 1975). In the past, nativist positions played a major role in shaping both artificial intelligence research and broader studies of language and cognition. In contrast, learning lies at the origin of the comparatively marginal connectionist tradition. Connectionist theorists argued that, given sufficient training, a model of adequate size could replicate any regular behaviour, including linguistic ones (Smolensky 1988). So far, their claims have been supported by empirical results. However, the question of how LMs training relates to language acquisition struggles to gain a central place in the broader debate. This paper offers an overview of the relationship between models training and child language learning, examining theoretical issues within dominant approaches. The aim is to assess whether learning and training are fundamentally incommensurable phenomena, as claimed by the nativist view, or whether comparing them can yield insights into both models' behaviour and child cognitive development.

1. What it takes to teach a model

A model's performance is a function of the training it undergoes. The primary training method is *language modeling* (the completion of syntagmatic sequences). However, training does not produce a linear progression in the development of a model's capabilities. Wei et al. (2022) argue that many of the skills LMs exhibit are *emergent*. They do not appear immediately, but rather surface after the model has been exposed to a sufficient amount of data (see also Ge et al. 2025, on the evolution of internal features in large LMs). Examples of such emergent capacities include: word unscrambling, basic arithmetic, transliteration into the phonetic alphabet, general knowledge retrieval, truthful question answering, logic-puzzle solving, narrative understanding. While the emergence of these skills mirrors certain aspects of human learning behaviour, the sheer amount of resources (data and computational capacity) required for training underscores a significant inefficiency when compared to child language acquisition and cognitive development (Frank 2023). The amount of linguistic input a child receives by around age 10 (the average age at which children exhibit good language skills) is estimated to be around 100 million words (Hart and Risley 1995, Gilkerson et al. 2017, by contrast, GPT-3's training involved roughly 200 billion words, Warstadt et al. 2023). These findings gave rise to the BabyLM Challenge, a research initiative in which scholars propose training optimization methods aimed at eliciting large LMs abilities in smaller models trained on cognitively plausible datasets, that is, with amounts of data comparable to the amount of speech typically received by a child around the age of 10 (Hu et al. 2024). Strategies explored within this framework include refinements to model architectures, curriculum learning (i.e., organizing training data by increasing linguistic complexity; Soviany et al. 2021), multimodal training, and other methods. Although these proposals are of technical and linguistic interest, none of the approaches has produced overwhelming results, nor have they led to consistent accuracy improvements across all evaluation tasks. In contrast to classical language modeling, *instruction tuning* (IT) appears to yield more promising outcomes (particularly in the form of *reinforcement learning with human feedback*, Ouyang et al. 2022). IT is a fine-tuning technique in which the model is exposed to a dataset of instructions paired with corresponding solutions. The model is required to either generate the correct solution or select among alternatives. This process improves the model's general ability to follow instructions, aligning its outputs more effectively with user expectations. This form of training bears a closer resemblance to human interactive learning and has been shown to accelerate the emergence of sophisticated abilities (Wei

et al. 2022). From this perspective, inefficiency in model learning may be attributable, at least in part, to sub-optimal training methods.

More generally, models trained on datasets ranging from 10 to 100 million words often achieve acceptable results on standard linguistic evaluations (non-generative syntax and semantics benchmarks), performing comparably to much larger models (Capone et al. 2024, Suozzi et al. 2025). However, they do not exhibit the distinctive emergent abilities observed in models trained on substantially larger datasets and with more advanced training techniques.

2. Perspectives on the human-model learning gap

Researchers have proposed three main hypotheses to account for this inefficiency (Frank 2023). The first is the nativist hypothesis, which holds that children possess *core knowledge* or innate psychological structures that explain their early linguistic and cognitive abilities (Pinker 1994, Baron-Cohen 1995, Spelke and Kinzler 2007, Tenenbaum et al. 2011, Leivada et al. 2024, see also Perrine 2023). However, this explanation tends to overlook the developmental dimension and does not offer a productive basis for comparing human and computational learning. Furthermore, nativist positions are difficult to reconcile with findings from developmental psychology, which show that children’s thinking evolves over time, shaped by the material conditions of learning, undergoing qualitative reorganization and passing through critical phases with non-guaranteed outcomes. Finally, the fact that models can acquire certain skills without innate constraints or predefined structures has been taken by some scholars as evidence against the nativist view (Warstadt and Bowman 2022).

The second hypothesis appeals to the grounding argument. Scholars aligned with this perspective argue that, even in early infancy, children formulate *intuitive theories* about the world. Development, in this view, involves the revision of these theories as the infant gathers evidence from experience (Gopnik and Wellman 2012). Accordingly, the absence of grounding (understood as the empirical interaction with the world) would prevent LMs from engaging in genuine learning. Instead, models are often characterized as mere parrots, superficially imitating semantic and reasoning abilities (Bender and Koller 2020, Zečević et al. 2023, Felin and Howleg 2024). However, this explanation also faces challenges, primarily because multimodal models (which incorporate grounded inputs) do not consistently outperform ungrounded models in standard evaluation benchmarks. Multimodal training has yielded poor results as an optimization strategy in the BabyLM Challenges (Warstadt et al. 2023, Hu et al. 2024). A second reason for skepticism toward grounding-based explanations comes from theoretical work in semiotics and linguistics, which suggests that it is not sensorimotor experience itself that gives rise to the kinds of concepts characteristic of logical reasoning. Concepts such as *causality*, *necessity*, or *proportionality* do not emerge directly from perceptual experience. Rather, they function as frameworks that impose a certain order or structure on experience (Eco 1997: §1.11). These concepts are tied to specific linguistic constructions that are acquired through exposure to adult speech (Tomasello 1999: 190). The third hypothesis attributes the difference in learning pace and the quality of skills developed in children and LMs to the type of linguistic input children receive. Most of the studies discussed so far compare biological and artificial learning by focusing mainly on the quantity of words received, a questionable approach (Hart and Risley 1995). A closer look, however, reveals a qualitative gap between language modeling training and the interactive exchanges that children are exposed to from an early age (Nikolaus and Fourtassi 2021). This may help explain why IT (which more closely reflects this type of interactive learning) leads to notable improvements in model performance (see §1). In addition, children’s early efficiency in various cognitive tasks may be overstated due to the way those tasks are administered and interpreted. In many cases, children are supported by external aids, especially visual ones, that make problem-solving easier. These *highly scaffolded* conditions (Frank 2023) contrast sharply with the purely linguistic tasks typically presented to LMs.

The hypotheses proposed to explain the inefficiency of LMs’ training can be grouped into two broad categories. The first includes nativist accounts and those based on the construction of

prelinguistic theories. These perspectives share a common tendency to dismiss the relevance of LMs for the study of human ontogenetic development. The second group is more open to the possibility of a comparison between computational and human learning. In this view, the learning gap is attributed, at least in part, to the interactive (social) nature of child learning, which contrasts with the language modeling training paradigm. With this overview in place, it becomes necessary to take a step back and look through the results to identify the theoretical assumptions underlying the supposed incommensurability between child development and computational training. Should this incommensurability prove unjustified, it would allow for a comparative study of the two phenomena. The aim of the following sections is to assess whether the available evidence on computational training and child learning might, in fact, prove mutually informative, shedding light on both the development of human cognition and the learning processes of artificial systems. The key assumptions supporting this alleged incommensurability can be summarized as follows:

- children learn not only to speak but also to solve complex tasks at a very early age, exhibiting cognitive abilities that are qualitatively not dissimilar to those of adults;
- the learning input can be quantified by the number of words heard, on average, by individuals over the period during which the relevant cognitive skills typically emerge.

The following section examines these two assumptions in detail.

3. Thinking and child speech

Several studies appear to support at least the first of the two assumptions outlined at the end of the previous section. This section presents examples of tests focused on causal reasoning abilities, which are relevant for two main reasons: they have been extensively studied in both psychological and LM research, and they are frequently cited as illustrative cases of sophisticated cognitive performance.

Harris et al. (1996) investigated causal and counterfactual reasoning in children aged 3 to 5. The children were told short stories in which a protagonist had to make a choice. Afterward, they were asked questions such as “What if A had not occurred, then B or not B?”. In a second phase, the children were asked what the character should have done differently to change the outcome. In a related study, McCormack and Hoerl (2005) examined children’s understanding of causal relations using a button-box task. Children (aged 3 to 5) interacted with boxes that produced different visual and acoustic effects depending on which button was pressed. The task required children to infer the relationship between the buttons and the resulting effects. A similar goal motivated the study by Dündar-Coecke et al. (2020), in which children aged 5 to 11 observed simple physical phenomena (such as the absorption of water by a material, the dissolution of salt in liquid, and the rising of a water level when a solid is submerged). After observing, the children were asked to describe what had occurred. In general, these studies conclude that children possess early competence in causal-logical reasoning, sometimes as early as age 3. This view is widely shared. For instance, the developmental psychologist Alison Gopnik argues that children construct intuitive theories of the world, which they revise in response to accumulating evidence (Gopnik 2011, Gopnik and Wellman 2012). This cognitive paradigm, known as *theory-theory*, characterizes the child, even from the earliest stages of life, as a kind of little scientist engaged in constructing an abstract and general model of reality through interaction with the environment. Gopnik and Wellman (2012) specify that an *intuitive theory* refers to a general representation of the *causal structure of the world*.

This paradigm implicitly challenges a strand of developmental psychology rooted in the work of Jean Piaget. A common tendency within this line of research is to naturalize children’s cognitive abilities by tracing their origins to the preverbal sphere, in direct contrast to the theoretical framework introduced by the Swiss psychologist.

3.1. Understanding causal speech

Piaget conducted a series of interviews and experimental tasks with young subjects to illustrate the differences in cognitive operations available to children at various stages of development, compared to those of adults. In two studies in particular (Piaget 1923, 1924), he examines how children use terms that express causal relations (e.g., *because*, *therefore*) and antithetical relations (e.g., *instead of*, *although*). In children's speech, Piaget observed a low frequency of *because* used in contexts of logical justification or reference to material causes. Instead, the majority of uses correspond to what he classifies as psychological causality (for example, "Look, he's laughing! Why? Because he wants to catch the apple."). In contrast, *because* used to explain physical causes (such as "The train can't get past there... because there's too much sand up there.") appears far less often. Demonstrative uses of *because* are similarly rare ("No, it's a boat, because it hasn't any wheels.") (Piaget 1924: 12). These asymmetries are accompanied by persistent difficulties in disambiguating the various meanings of *because*, as well as in correctly completing examiner-provided prompts that include the term (Piaget 1924: 17):

A man fell down in the street... *because he fell ill.*
The man fell from his bicycle... *because he broke his arm.*
I had a bath... *because afterwards I was clean.*

According to Piaget's findings (1924: 19), up to the age of 7 or 8, the word *because* is used by children as an occasional and ambiguous term, with a broad margin of flexibility. It can cover a wide range of relations (including causal, temporal, and teleological connections) without this semantic heterogeneity appearing to trouble the child. In some cases, *because* is even inserted at the beginning of a sentence when it is not functionally required. Piaget notes that, as early as age 3 or 4, expressions denoting relations are used quite appropriately in spontaneous speech. Errors tend to emerge when children are asked to use these terms intentionally. He attributes these difficulties to a failure to differentiate, and to become aware of, the variety of relational meanings that these terms can represent. Piaget characterizes these difficulties as part of a broader lack of order or direction in children's thinking, a form of thought in which juxtaposition often replaces logical relation. In this context, different relational terms appear interchangeable, as they have not yet acquired a stable function in child speech. This situation begins to improve around the age of 8. Piaget's investigations appear to cast doubt, at least, on the first assumption regarding the precocity of child cognitive skills. His results can be summarized as follows: logical and causal reasoning abilities do not develop as early as they might appear. Although spontaneous speech can give the impression of conceptual depth, closer analysis reveals that children do not yet fully grasp the distinctions and functions of causal relations.

Piaget attributes this limited understanding to the child's egocentrism, which he considers a residual form of children's original autism. Vygotsky, while agreeing with Piaget's observations regarding children's incomplete awareness of causal language and the non-precocity of child development, criticizes the Swiss psychologist's theoretical position. He describes the notion of an original autism in the child as an idealistic explanation that does not conform to the realistic character of children's thinking (Vygotsky 1934: 62). Vygotsky's hypothesis is that the child's cognitive development (that is, the emergence of adult, and thus also logical-causal, thought) depends on certain material conditions of learning.

3.2. Material conditions of learning

Vygotsky takes issue with Piaget's methodological approach. He argues that Piaget focused exclusively on children's spontaneous concepts (those derived from everyday experience, such as *riding a bicycle*) while overlooking those that emerge through formal education. Vygotsky refers to these as *scientific* concepts, a term that can be misleading. Mecacci (2017: 99) notes that *scientific* in this context should be understood as *systematic* or *academic*, meaning concepts acquired through structured teaching. This clarifies why Vygotsky's examples of scientific concepts often involve topics such as the history and theory of communism, which were part of the school curriculum in his time.

The police shot the revolutionaries because...
In capitalist countries bread is destroyed although...

Vygotsky replicates Piaget's experiment, using both spontaneous and scientific concepts. The results show that the development of scientific concepts outpaces that of spontaneous ones, with the former reflecting a more advanced level of thinking. Moreover, over time everyday concepts also improve, gradually approaching the level of scientific concepts (around the age of 10, Vygotsky 1934: 214). One possible objection, which Vygotsky addresses directly, is that education merely makes certain topics more familiar to children, thereby making it easier to reason with scientific concepts. However, it is difficult to argue that this difference in performance is due to familiarity alone. Falling off a bicycle or taking a bath can hardly be considered less familiar to children than the class struggle, the exploitation of workers, or the Paris Commune. In short, even in children, grounding alone does not appear to lead to early development of causal-logical reasoning. As in the case of BabyLM challenge models, a good syntactic and semantic competence is not accompanied by early mastery of causal reasoning. It seems that the ability to complete such propositions depends on exposure to specific uses of language and to modes of abstraction and *generalization* that, prior to formal schooling, are not yet fully accessible.

The concept of generalization is the fulcrum of Vygotsky's theory of development. What language introduces into the child's preverbal behaviour is a new way of organizing experience. For Vygotsky, a concept (or meaning) never refers to a single object, but always to a class of objects or phenomena. Every word represents a generalization, and generalization is, in his view, *an extraordinary verbal act of thought, one that reflects reality in a fundamentally different way from immediate perception* (Vygotsky 1934: 47). However, children's thinking does not yet possess the capacity to generalize in the same way adults do. Early meanings are original formations, they reflect experience in a distinctive way, not yet through abstract concepts, but no longer through immediate, preverbal perception. Vygotsky refers to this mode of thought as *complexive thinking*, and infers its properties through a well-known experiment. In the experiment, children are asked to group wooden blocks, differing in shape, size, and colour, using invented words introduced by the examiner. The results show that the perceptual field dominates the children's groupings. The inclusion of a new element into a complex does not follow consistent, unitary criteria (as it does in conceptual thinking), but proceeds in an erratic and unstable fashion. Children construct seemingly open-ended complexes, shifting from one criterion to another (such as proximity or vague perceptual similarity) rather than grounding their generalizations in stable, abstract principles. A similar phenomenon occurs in the early learning of word meanings. Before a word's meaning is stabilized through shared use with adults, its use can expand indefinitely to include heterogeneously related phenomena.

At 251 days, a child uses the word «vau-vau» to designate a porcelain figure of a girl.
At 307 days, this child uses the same word to designate a dog lying in the yard, a portrait of a grandmother and grandfather, a toy horse, and a wall clock. At 331 days, he uses the word to designate a fur boa with a dog's head and another boa without a dog's head (in this context, he paid special attention to the glass eyes). At 334 days, the same name was given to a rubber toy man that squeaked when it was squeezed. At 396 days, it was used to refer to a black button on the child's father's shirt. At 433 days, the child pronounced the same word when he saw pearls on a dress as well as when he noticed a bath thermometer. (Vygotsky 1934, trad eng.: 148).

The type of generalization available to children appears qualitatively different from that of adults, as do their possibilities for using linguistic meanings. This difference becomes especially evident in the early stages of development and in situations that require the deliberate use of concepts. Similar results were found by Luria in his experiments in Uzbekistan, where illiterate adults with no formal education struggled to solve syllogisms, even on familiar topics, and to group objects (whether presented as names or as pictures) according to abstract, non-functional criteria. Both

children and adults showed marked improvement after being exposed to formal schooling and acquiring literacy (Luria 1979: §4).

In both Luria and Vygotsky, subjects gain access to conceptual thinking (and thus to more sophisticated forms of generalization) through exposure to a specific kind of linguistic input. According to the conclusions drawn by both psychologists, the second assumption is also called into question: learning depends not only on the quantity of input, but on its quality. This quality represents one of the key material conditions underlying child development. With both assumptions supporting the incommensurability between computational learning and child language learning now set aside, the next section will outline a theoretical framework within which the two phenomena can be productively compared.

4. The ontogenesis of causal thinking

The results presented so far appear difficult to reconcile. According to cognitive theories aligned with the theory-theory paradigm, capacities for generalization and causal-logical reasoning emerge early and are largely independent of learning, as their experimental findings seem to suggest. In contrast, Piaget and Vygotsky argue that the development of adult thinking is a gradual process, closely tied to language acquisition and education, marked by several intermediate phases and compromise formations. A possible reconciliation is offered by Tomasello who, following Vygotsky, draws a distinction between the types of cognitive abilities involved in the findings of different scholars. He does not deny that, in a given concrete situation, a child will look toward the cause of a phenomenon they are attending to, nor that the same child may correctly identify the referent of the word *dog* when it is uttered by an adult (Tomasello 1999: 118). However, stopping at this level addresses only part of the problem. What matters more is understanding which additional contents the child extends the word to, and which situations or referents are excluded from its use. Only in this way is it possible to assess the properties of children's concepts and the *generalization* operations available to them. Tomasello argues that what is learned during language acquisition is not merely a set of labels for expressing pre-existing representational structures or cognitive capacities. Rather, the child learns a *specific way of constructing their own experience* (Tomasello 1999: 120). According to Tomasello, the immediate sensorimotor modes through which experience is initially organized are profoundly restructured by language acquisition and by the ways in which discourse mediates (or carves out) experience, which can be framed in countless different ways (Tomasello 1999: 126). Words and linguistic constructions embody particular perspectives for representing things, events, and relationships, perspectives that would otherwise remain inaccessible to the child. Access to this adult-like form of linguistic mediation, however, is anything but early. Although the intersubjective (communicative, social) nature of signs becomes evident very early on (as shown by the child's early, superficially appropriate, use of relational terms such as *because*), their perspectival nature only gradually emerges (Tomasello 1999: 123).

Returning to the cases addressed in the previous sections, it is possible to affirm that the concept of causality introduces a specific perspective on events. Piaget emphasized the different senses of *cause* that can be realized through constructions involving the term *because*, grouping together distinct types of phenomena: mechanical causes, psychological causes, logical or formal causes. Beneath a uniformity of form (of the linguistic construction), multiple segments of experience coexist, unified by a particular perspective, by a specific way of generalizing events. Continuing this line of reasoning, one may hypothesize that the mistaken generalization of early word meanings observed by Vygotsky may also extend to linguistic constructions (Tomasello 1999: 135). What emerges from Piaget's and Vygotsky's experiments is a superficial ability to use certain types of linguistic constructions (particularly those involving terms such as *because* and *although*) paired with a limited awareness or understanding of how these constructions organize (differentiate and synthesize) experience. To support this view, Tomasello draws on studies examining the use of connective terms that link two distinct propositions, such as *and*, *because*, and *so*. The psychologist confirms that children begin using these terms quite early, around the ages of 2 to 3. However, an analysis of the simplest among them (*and*) shows that at around age 3, the term is typically used within single propositions that are linked to utterances previously produced

by an adult (Tomasello 2003: 259, Diessel 2004). Between the ages of 4 and 5, instances begin to appear in which *and* is used to connect two propositions, though this remains a secondary and less frequent usage. Peterson and McCabe (1987) investigated the use of conjunctions at age 9. At this stage, *and* remains the most frequently used connective, continuing to be used in an undifferentiated manner to express a wide range of semantic relations, even after more specific connectives have begun to appear in the child's speech. The use of *because* follows a pattern consistent with the observations of Piaget and Vygotsky. Initially, it is employed in a scaffolded manner similar to *and*, and primarily carries a social function, used to justify actions or statements in response to an adult's question or prompt. The autonomous use of *because* to connect two propositions remains at only 40% by age 5 (Tomasello 2003: 260). Similar limitations are observed with all complex constructions: comprehension and voluntary use do not yet align with the apparent fluency of spontaneous speech. Berman and Slobin (1994) document the difficulties children face with narrative discourse, even up to age 9, in tasks requiring them to verbally describe a story represented by a sequence of images. Children consistently struggle to produce coherent narratives that clearly indicate a beginning, progression, and conclusion. They show difficulty tracking referents and ordering events temporally. Tomasello attributes these challenges to the *plurifunctionality* of complex constructions and argues that the mastery of the perspectives embedded in their use develops slowly, over the course of the school years (Tomasello 2003: 274, see also Hickmann 1995). He traces this phenomenon back to the *item-based* nature of early child constructions. That is, the child does not possess formal schemas or abstract propositional functions that can be flexibly applied to various appropriate situations. These constructions are initially mixed formations (if not holophrastic) and are closely tied to specific lexical items, a characteristic that persists for a long time (Tomasello 2003: 264). From this perspective, the child appears less like a scientist and more like a diligent imitator. Rather than uncovering abstract structures behind phenomena and readily generalizing the constructions they are exposed to, the child tends to replicate the specific uses of propositions they have encountered (Tomasello 2003: 261), representing events based on what they have previously heard. This would explain why Vygotsky observed that scientific concepts tend to outpace spontaneous (everyday) ones over the course of development. The precocity often attributed to children may stem from overly hasty attributions of abstract constructions understanding, based on the observation of isolated uses in specific contexts (Tomasello 2003: 274).

Consistent with this line of reasoning, Tomasello believes that causal-logical reasoning skills are the product of specific learning conditions. Over the course of development, children hear particular events (and not others) described in causal terms. This causal structure is not *imposed by the world itself*; it represents a perspective that the child cannot construct alone, since experience can be framed in countless different ways. It must be received from outside, embedded in the forms of the child's native language. This is not to deny that children display some *operative understanding of causality* at a very early age. However, this practical understanding does not automatically translate into genuine capacities for causal reasoning, or into the types of causal concepts discussed so far (Tomasello 1999: 190), just as a child's intuitive grasp of quantities does not equate to sophisticated mathematical abilities¹.

¹ It is worth further clarifying how Tomasello's position differs from earlier nativist accounts. According to Gopnik (2011), children possess an innate capacity to construct causal theories of the world once they are exposed to experience. For Tomasello, however, this is not the case. The ability to construct theories is a *product* of linguistic learning, not a precondition for it. Innate operational capacities, which humans share with other primates, are not nonverbal versions of adult abilities, but rather pre-verbal ways of engaging with experience. Causal reasoning is shaped by the specific language in which it takes form, and languages profoundly reorganize prelinguistic modes of structuring experience. Evidence for this can be found in the fact that the sociogenesis of language could have taken alternative paths, organizing experience according to different linguistic coordinates. An example of this is provided by the contrast between the Roman and Arabic numeral systems and the divergent consequences these had for the development of mathematical science. Although both systems reorganize the same innate operational capacity for dealing with quantities (Tomasello 1999: 185-186), the way each historical system of signs partitions the continuum of content (in this case, quantities) determines the kinds of operations available to speakers. Thus, while all languages may appear capable of expressing what are broadly called causal relations, the way these relations are organized, and the types of logical-discursive operations they enable, can vary greatly. Only specific linguistic training

In short, humans are certainly capable of thinking without the aid of signs, if *thinking* is understood as perceiving, remembering, categorizing, and acting intelligently in the world, much like other primates (Tomasello 1999: 214). Consequently, without prolonged and specific language instruction, the child would not be able to develop sophisticated cognition.

5. Conclusions

It can be argued that computational training and child language learning are not fundamentally incommensurable. Even in children, sophisticated cognitive abilities emerge as a function of exposure to language, with schooling and literacy in particular fostering the development of generalization skills and gradually aligning children's use of words and constructions with that of adults. Although a significant disparity between models and children must still be acknowledged, its magnitude should be reconsidered. A 9 year old child (exposed to approximately 90 million words) is not yet fully in control of the constructions that underlie causal-logical reasoning, nor of complex syntactic structures more broadly. It is therefore unreasonable to expect a model trained on a comparable amount of data to exhibit these abilities. Moreover, children benefit from a form of operational thinking and intersubjective action that underpins linguistic development, an element absent in computational training. Finally, whereas LMs rely primarily on language modeling for training, successful child development appears to require interactive communication, with formal education playing a central role. Given the preceding analysis, among the three hypotheses proposed to explain the disparity between learning and training, the third appears to be the most convincing. Consequently, it cannot be ruled out that further optimization of training methods, combined with improved data curation, may contribute to narrowing the efficiency and comparability gap between model training and child learning. From a broader theoretical perspective, the comparison between the two phenomena leads to a reconsideration of the extent to which human cognition is not only externalized, but constitutively realized through the linguistic medium, rather than being the invariant product of an innate biological endowment. It also becomes less surprising that an LM, understood as a representation of the system of relations operating within a given language state (however approximate), exhibits some of the mechanisms characteristic of verbal thought, even in the absence of its biological substrate.

References

Baron-Cohen, Simon (1995), *Mindblindness: An essay on autism and theory of mind*, Cambridge, MIT Press 1995.

Bender, Emily, Koller, Alexander (2020), «Climbing towards nlu: On meaning, form, and understanding in the age of data». *Proceedings of the 58th Annual Meeting of the ACL*, pp. 5185–5198.

allow individuals to access the form of “logical-causal thought” considered normative in adult cognition. On the basis of these considerations, it can be argued that causal thinking is not an innate feature of human cognition. Rather, it depends on: a) a species-specific predisposition for language learning, b) the development of a natural historical language with specific constructions and modes of organizing experience, and c) the training of individuals in the use of the representational mediation provided by that language.

Berman, Ruth, Slobin, Dan Isaac (1994), *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale, Taylor and Francis 1994.

Brown, Tom B, Mann Benjamin Ryder, Nick, Subbiah, Melanie, Kaplan, Jared, Dhariwal, Prafulla, Neelakantan, Arvind, Shyam, Pranav, Sastry, Girish, Askell, Amanda, Agarwal, Sandhini, Herbert-Voss, Ariel, Krueger, Gretchen, Henighan, Tom, Child, Rewon, Ramesh, Aditya, Ziegler, Daniel M., Wu, Jeffrey, Winter, Clemens, Hesse, Christopher, Chen, Mark, Sigler, Eric, Litwin Mateusz, Gray, Scott, Chess, Benjamin, Clark Jack, Berner, Christopher, McCandlish Sam, Radford Alec, Sutskever Ilya, Amodei Dario (2020), «Language models are few-shot learners», *arXiv*, <https://doi.org/10.48550/arXiv.2005.14165>.

Capone, Luca (2021), «Which theory of language for deep neural networks? Speech and cognition in humans and machines», *Technology and Language*, 2(4), pp. 29-60.

Capone, Luca (2024), «Wittgenstein's remarks on the philosophy of computational linguistics. Philosophical clarification applied to natural language processing». *RIFL*, <https://doi.org/10.4396/SFL202302>

Capone, Luca, Suozzi, Alice, Lebani, Gianluca, Lenci, Alessandro (2024), «Babies: A benchmark for the linguistic evaluation of Italian baby language models», *Proceedings of CLiC-it 2024*, pp. 157–17, <https://aclanthology.org/2024.clicit-1.20/>

Chalmers, David J. (2023), «Could a Large Language Model be Conscious?», *Boston Review*, <https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/>

Chomsky, Noam (1980), *Rules and Representations*, Cambridge, CUP 2005.

Diessel, Holger (2004), *The acquisition of complex sentences*. Cambridge, CUP 2009.

Dündar-Coecke, Selma, Toloime, Andrew, Schlottmann, Anne (2020), «Children's reasoning about continuous causal processes: The role of verbal and non-verbal ability», *BJEP*, 90(2), pp. 364–381.

Eco, Umberto (1997), *Kant e l'ornitorinco*, Milano, La nave di Teseo, 2016.

Felin, Teppo, Howleg, Matthias (2024), «Theory is all you need: Ai, human cognition, and decision making», *SSRN*, <https://doi.org/10.2139/ssrn.4737265>

Fodor, Jerry A. (1975), *The language of thought*, New York, Thomas Crowell 1975.

Frank, Michael C. (2023), «Bridging the data gap between children and large language models», *Trends in Cognitive Sciences*, 27(11), pp. 990–992.

Ge, Xuyang, Shu, wentao, Wu, Jiaying, Zhou, Yunhua, He, Zhengfu, Qiu, Xipeng (2025), «Evolution of Concepts in Language Model Pre-Training», *arXiv preprint*, 10.48550/arXiv.2509.17196

Gilkerson, Jill, Richards, Jeffrey A., Warren, Steven F., Montgomery, Judith K., Greenwood, Charles R., Oller, D. Kimbrough, Hansen John H.L., Terrance, Paul D. (2017), «Mapping the early language environment using all-day recordings and automated analysis», *AJSLP*, 26(2), pp. 248–265.

Goldstein, Simon, Levinstein, Benjamin A. (2024), «Does chatgpt have a mind?», *arXiv*, <https://doi.org/10.48550/arXiv.2407.11015>

Gopnik, Alison (2011), «The theory theory 2. 0: Probabilistic models and cognitive development», *Child Development Perspectives*, 5(3), pp. 161–163.

Gopnik, Alison, Wellman, Henry M. (2012), «Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory», *Psychol. Bull.* 138, pp. 1085–1108.

Harris, Paul L, German, Tim, Mills, Patrick (1996), «Children’s use of counterfactual thinking in causal reasoning», *Cognition*, 61(3), pp. 233–259.

Hart, Betty, Risley, Todd (1995), *Meaningful differences in the everyday experience of young American children*, Baltimore, Brookes 1995.

Hickmann, Maya (1995), «Discourse organization and the development of reference to person, space, and time», in P. Fletcher and B. MacWhinney, eds., *Handbook of child language*, Oxford, Basil Blackwell 1995.

Hu, Michael Y, Mueller, Aaron, Ross, Candace, Williams, Adina, Linzen, Tal, Chengxu, Zhuang, Cotterell Ryan, Choshen, Leshem, Warstadt Alex, Wilcox, Ethan, Gotlieb (2024), «Findings of the second babylm challenge: Sample-efficient pretraining on developmentally plausible corpora», *arXiv*, <https://doi.org/10.48550/arXiv.2412.05149>

Leivada, Evelina, Marcus, Gary, Günther, Fritz, Murphy, Elliot (2024), «A sentence is worth a thousand pictures: Can large language models understand human language and the world behind words?», *arXiv*, <https://doi.org/10.48550/arXiv.2308.00109>

Lenci, Alessandro (2023), «Understanding natural language understanding systems», *Sistemi Intelligenti*, 2, pp. 277–302.

Luria, Aleksandr R. (1979), *The Making of Mind: A Personal Account of Soviet Psychology*, Cambridge, HUP 1979.

Mandelkern, Matthew, Linzen, Tal (2024), «Do language models’ words refer?», *arXiv*, <https://doi.org/10.48550/arXiv.2308.05576>

Marx, Karl (1867), *Das Kapital*, Hamburg, Verlag von Otto Meissner, (*Capital*, transl. by D. Fernbach, London, Penguin Book 1991).

McCormack, Teresa, Hoerl, Christoph (2005), «Children’s reasoning about the causal significance of the temporal order of events», *Developmental Psychology*, 41(1), pp. 54–63.

Mecacci, Luciano (2017), *Lev Vygotskij. Sviluppo, educazione e patologia della mente*, Firenze, Giunti 2021.

Merrill, William, Wu, Zhaofeng, Naka, Norihito, Kim, Yoon, Linzen, Tal (2024), «Can you learn semantics through next-word prediction? The case of entailment», *arXiv*, <https://doi.org/10.48550/arXiv.2402.13956>

Nikolaus, Mitja, Fourtassi, Abdellah (2021), «Modeling the interaction between perception-based and production-based learning in children's early acquisition of semantic knowledge», *Proceedings of the 25th CoNLL*, <https://doi.org/10.18653/v1/2021.conll-1.31>

Ouyang, Long, Wu, Jeff, Almeida, Diogo, Wainwright, Carroll L., Mishkin, Pamela, Zhang, Chong, Agarwal, Sandhini, Slama, Katarina, Ray, Alex, Schulman, John, Hilton, Jacob, Kelton, Fraser, Miller, Luke, Simens, Maddie, Aspell, Amanda, Welinder, Peter, Christiano, Paul, Leike, Jan, Lowe, Ryan (2022), «Training language models to follow instructions with human feedback», *arXiv*, <https://doi.org/10.48550/arXiv.2203.02155>

Perrine, Patric (2023), «Inaccessible neural language models could reinvigorate linguistic nativism», *arXiv*, <https://doi.org/10.48550/arXiv.2301.05272>

Peterson, Carole, McCabe, Allyssa (1987), «The connective 'and': Do older children use it less as they learn other connectives?» *Journal of Child Language*, 14(2), pp. 375–381.

Piaget, Jean (1923), *Le langage et la pensée chez l'enfant*, Delachaux et Niestlé, (*Il linguaggio e il pensiero del fanciullo*, transl. by, C. Musatti Rapuzzi, Firenze, Giunti 1967).

Piaget, Jean (1924), *Le jugement et le raisonnement chez l'enfant*, Delachaux et Niestlé, (*Judgment and reasoning in the child*, transl. by, M. Warden, Routledge 2002).

Piantadosi, Steven T., Hill, Felix (2022), «Meaning without reference in large language models», *arXiv*. <http://arxiv.org/abs/2208.02957>

Pinker, Steven (1994), *The language instinct: How the mind creates language*, New York, Morrow 1994.

Smolensky, Paul (1988), «On the proper treatment of connectionism», *Behavioral and Brain Sciences*, 11(1), pp. 1–23.

Soviany, Petru (2021), «Curriculum learning: A survey», *arXiv*, <https://doi.org/10.48550/arXiv.2101.10382>

Spelke, Elizabeth S., Kinzler, Katherine D. (2007), «Core knowledge», *Dev. Sci.*, 10(1), pp. 89–96.

Suozzi, Alice, Capone, Luca, Lebani, Gianluca E., Lenci, Alessandro (2025), «Bambi: Developing baby language models for Italian», *arXiv*, <https://doi.org/10.48550/arXiv.2503.09481>

Tenenbaum, Joshua B., Kemp, Charles, Griffiths, Thomas L., Goodman, Noah D. (2011), «How to grow a mind: statistics, structure, and abstraction», *Science* 331, pp. 1279–1285

Tomasello, Michael (1999), *The cultural origins of human cognition*, London, HUP 1999.

Tomasello, Michael (2003), *Constructing a language. A usage-based theory of language acquisition*, Cambridge, HUP 2003.

Vygotskij, Lev S. (1934), *Myslenie i reč*, Gosudarstvennoe Sotsial'no-Ekonomicheskoe Izdatel'stvo, (*The collected works of L.S. Vygorsky, Vol. 1*, transl. by. N. Minick, NY, Plenum Press 1987).

Warstadt, Alex, Mueller, Aaron, Choshen, Leshem, Wilcox, Ethan, Zhuang, Chengxu, Ciro Juan, Mosquera Rafael, Paranjabe, Bhargavi, Williams, Adina, Linzen Tal, Cotterell, Ryan (2023), «Findings of the babylm challenge: Sample-efficient pretraining on developmentally plausible corpora», *Proceedings of the BabyLM Challenge at the CoNLL*, <https://doi.org/10.18653/v1/2023.conll-babylm.1>

Warstadt, Alex, Bowman, Samuel R. (2022), «What artificial neural networks can tell us about human language acquisition», *arXiv*, <https://doi.org/10.48550/arXiv.2208.07998>

Wei, Jason, Tay, Yi, Bommasani, Rishi, Raffel, Colin, Zoph, Barret, Borgeaud, Sebastian, Yogatama, Dani, Bosma, Maarten, Zhou, Denny, Metzler, Donald, Chi, Ed H., Hashimoto Tatsunori, Vinyals, Oriol, Liang, Percy, Dean, Jeff, Fedus, William (2022), «Emergent abilities of large language models», *arXiv*, <https://doi.org/10.48550/arXiv.2206.07682>

Zečević, Matej, Willig, Moritz, Dhani, Devendra Singh, Kersting, Kristian (2023), «Causal parrots: Large language models may talk causality but are not causal», *arXiv*, <https://doi.org/10.48550/arXiv.2308.13067>