Beyond the Data Gap:

Children (and not LLMs) create languages, violate their input statistics, and exhibit critical periods

Annika McDermott-Hinman & Roman Feiman

Forthcoming as commentary on BBS target article by Futrell & Mahowald (2026): "How Linguistics learned to stop worrying and love the language models

Large Language Models as Models of Human Language Competence

Nativist Theories

- There must be a rich set of innate content that supports language learning
- It is impossible for children to learn what they do with the input they get without innate content

e.g., Chomsky (1959; 1965); Lenneberg (1967)

Constructivist Theories

- Language structure is built from experience
- Domain-general learning mechanisms support language learning. There is no innate language-specific content

e.g., Bates et al (1996); Tomasello (2003)

Language Model Success: in-principle learnability

Large language models successfully learn to produce nearly error-free language using only domain-general, connectionist architecture

This seems to undermine the poverty of the stimulus argument – seems to support constructivist theories

The Data Gap

The nativist response

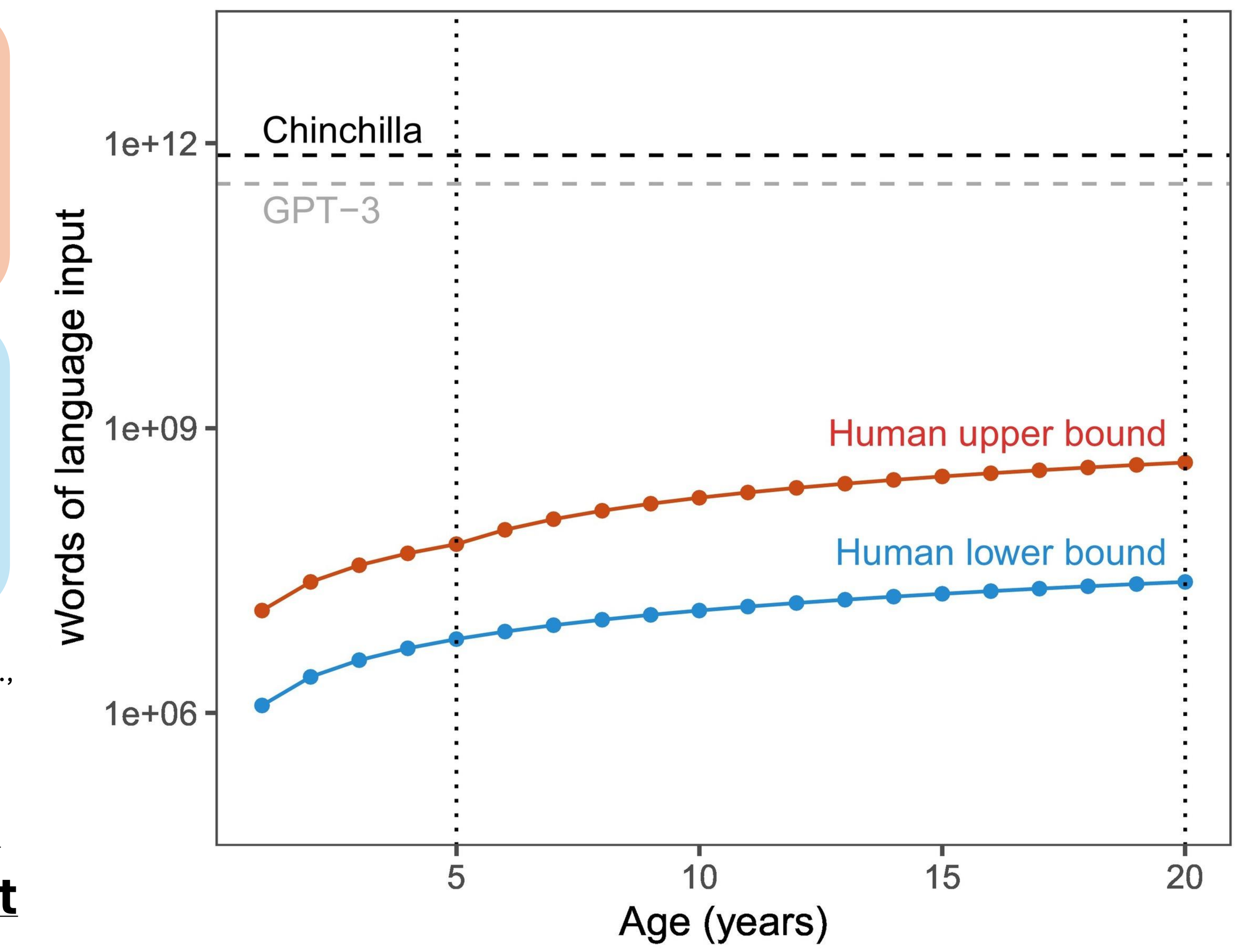
Language models seem to learn the structure of language, but they do it with up to **5 orders of magnitude** more input than children do

The current constructivist project

Build language models that can learn language from human-scale quantities of data (Futrell & Mahowald, 2025; Frank, 2023; Hu et al., 2024)

If we successfully close the data gap (see, e.g., Hu et al., 2024), have the constructivists won the day?

We argue no, because more separates children and language models than just the data gap.



Frank (2023) Trends in Cognitive Sciences

Children violate their input to create languages

Gestural communication systems developed by isolated deaf children to communicate with their hearing families

Communicative gestures of hearing adults

Nicaraguan Sign Language: developed de novo by a community of deaf children with no external signed input

	Gestures	Home-sign	1st cohort	2nd cohort
Morphologically distinguished parts of speech, syntactic roles	X	✓	✓	✓
Marked grammatical arguments	X	✓	✓	✓
Syntactic embedding	unknown	X	✓	✓
Discrete, compositional morphology	X	unknown	~	✓
Spatial modulation	unknown	unknown	X	✓

Abner et al. (2019); Coppola & Newport (2005); Goldin-Meadow et al. (1994); Goldin-Meadow & Mylander (1983); Kocab et al. (2023); Senghas et al. (1997); Senghas & Coppola (2001); Senghas (2003); Senghas et al. (2004)

Language models do not violate their input

By design, language models recapitulate the statistical properties of their training data

- To create a language from sparse input is to *violate* those statistical properties

How might language models emulate children?

- Simple neural agents can create rudimentary compositional systems in:
  - Cooperative reference games (Lazaridou & Baroni, 2020; Steinert-Threlkeld, 2020; Boldt & Mortensen, 2024)
  - Iterative learning scenarios (Jarvis et al., 2026)
- Unclear whether (or under which conditions) language models could match children in spontaneously generated complexity
  - May require external pressure to communicate

Language models fail to model children's language competence

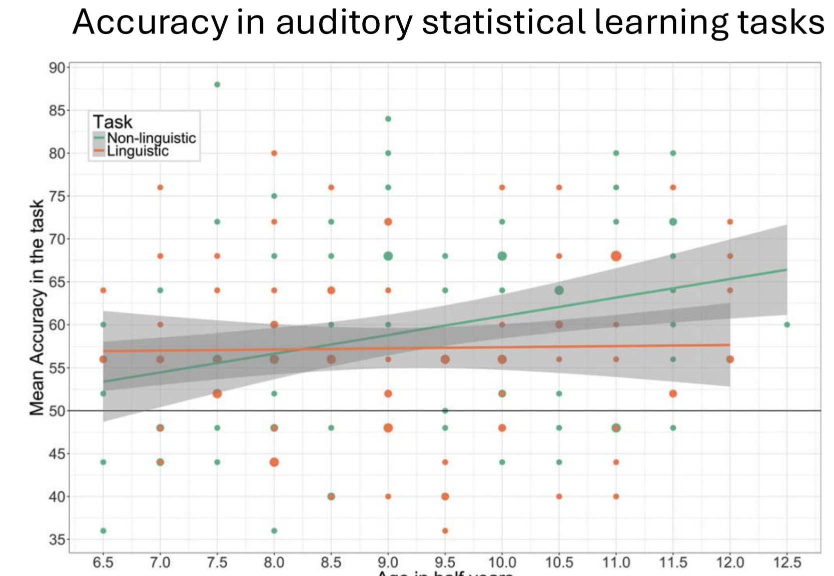
To undermine linguistic nativism, LLMs must account for:

- children's ability to mismatch and outstrip their input
- adults' inability to learn a new native language

Critical periods: language learning dissociates from statistical learning

General statistical learning abilities improve throughout development (Shufaniya & Arnon, 2018)

Accuracy in auditory statistical learning tasks



Shufaniya & Arnon (2018)

Children's ability to learn their language's phonology declines in infancy (Werker, 2024), and to learn its grammar declines around puberty (Hartshorne et al., 2018; Johnson & Newport, 1989; Lenneberg, 1967)

Language models show no critical period

In language models, learning a language *accelerates* the rate of subsequently learning a second language (Oba et al., 2023; cf. Hernandez et al., 2005)

How might language models emulate children?

- A critical period can be induced in a language model – by imposing a regularizer partway through training (Constantinescu et al., 2025)
  - More analogous to maturational decrease in language-specific plasticity
- Non-language neural nets can spontaneously show critical periods: In deep perceptual networks, improperly correlated data can permanently impair the model's learning (Kleinman et al., 2023)

Children have thoughts before words

Children

13 months
"[jʌwʌblublublujujæwæ]"

15 months
"Mommy"

22 months
"It's a kite"

26 months
"I made my own toast , okay?"

Bernstein-Ratner (1984); Braunwald (1971); MacWhinney (2000)

LLMs

3.3M training tokens
"This is,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,, the the the the,,,,,,,,,,......"

33M training tokens
"This is a few of the first of the same of the world's the most of the first of the the same of the first of the world."

330M training tokens
"This is a great way to make a difference in your life."

Chang et al. (2024)

Language models learn only the structure of language

- These differences stem from differences in the function of language for LMs and children
  - LMs learn patterns of co-occurrence – first words are the most common
  - Children think about the world – first words express those thoughts

How might language models emulate children?

- Multimodal LMs might bridge the gap
  - Early attempts show severe limits in even basic referential abilities – and lack of shape bias (Vong et al., 2024)
- Children's preverbal conceptual and perceptual capacities likely structure their thought (Spelke & Kinzler, 2007; Carey, 2009)
  - Children have access to abstract concepts before they know the corresponding words (see, e.g., McDermott-Hinman et al., 2026)
  - LMs may need to model humans' capacities for thought to fully model their language

References

Abner, N., Flaherty, M., Stangl, K., Coppola, M., Brentari, D., & Goldin-Meadow, S. (2019). The noun-verb distinction in established and emergent sign systems. *Language*, *95*(2), 230–267. https://doi.org/10.1353/lan.2019.0030

Bates, E., Elman, J., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. The MIT Press. https://doi.org/10.7551/mitpress/5929.001.0001

Bernstein-Ratner, N. (1984). Phonological rule usage in mother-child speech. *Journal of Phonetics*, *12*(3), 245–254. https://doi.org/10.1016/S0095-4470(19)30881-2

Boldt, B., & Mortensen, D. (2024). *A Review of the Applications of Deep Learning-Based Emergent Communication* (arXiv:2407.03302). arXiv. https://doi.org/10.48550/arXiv.2407.03302

Braunwald, S. R. (1971). Mother-Child Communication: The Function of Maternal-Language Input. WORD, *27*(1–3), 28–50. https://doi.org/10.1080/00437956.1971.11435613

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Chang, T. A., Tu, Z., & Bergen, B. K. (2024). Characterizing Learning Curves During Language Model Pre-Training: Learning, Forgetting, and Stability. *Transactions of the Association for Computational Linguistics*, *12*, 1346–1362. https://doi.org/10.1162/tacl_a_00708

Chomsky, N. (1959). Review of Verbal Behavior, by B.F. Skinner. *Language*, *35*(1), 26–58.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Constantinescu, I., Pimentel, T., Cotterell, R., & Warstadt, A. (2025). Investigating Critical Period Effects in Language Acquisition through Neural Language Models. *Transactions of the Association for Computational Linguistics*, *13*, 96–120. https://doi.org/10.1162/tacl_a_00725

Coppola, M., & Newport, E. L. (2005). Grammatical Subjects in home sign: Abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences*, *102*(52), 19249–19253. https://doi.org/10.1073/pnas.0509306102

Frank, M. C. (2023). Bridging the data gap between children and large language models. *Trends in Cognitive Sciences*, *27*(11), 990–992. https://doi.org/10.1016/j.tics.2023.08.007

Futrell, R., & Mahowald, K. (2025). How Linguistics Learned to Stop Worrying and Love the Language Models. *Behavioral and Brain Sciences*, 1–98. https://doi.org/10.1017/S0140525X2510112X

Goldin-Meadow, S., Butcher, C., Mylander, C., & Dodge, M. (1994). Nouns and Verbs in a Self-Styled Gesture System: What's in a Name? *Cognitive Psychology*, *27*, 259–319.

Goldin-Meadow, S., & Mylander, C. (1983). Gestural Communication in Deaf Children: Noneffect of Parental Input on Language Development. *Science*, *221*(4608), 372–374. https://doi.org/10.1126/science.6867713

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, *9*(5), 220–225. https://doi.org/10.1016/j.tics.2005.03.003

Hu, M. Y., Mueller, A., Ross, C., Williams, A., Linzen, T., Zhuang, C., Cotterell, R., Choshen, L., Warstadt, A., & Wilcox, E. G. (2024). *Findings of the Second BabyLM Challenge: Sample-Efficient Pretraining on Developmentally Plausible Corpora*. 1–21.

Jarvis, D., Klein, R., Rosman, B., & Saxe, A. M. (2026). Compositionality and systematicity emerge from iterated learning in deep linear networks. *Proceedings of the National Academy of Sciences*, *123*(19), e2509739123. https://doi.org/10.1073/pnas.2509739123

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Kleinman, M., Achille, A., & Soatto, S. (2023). Critical Learning Periods for Multisensory Integration in Deep Networks. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 24296–24305. https://doi.org/10.1109/CVPR52729.2023.02327

Kocab, A., Senghas, A., Coppola, M., & Snedeker, J. (2023). Potentially recursive structures emerge quickly when a new language community forms. *Cognition*, *232*, 105261. https://doi.org/10.1016/j.cognition.2022.105261

Lazaridou, A., & Baroni, M. (2020). *Emergent Multi-Agent Communication in the Deep Learning Era* (arXiv:2006.02419). arXiv. https://doi.org/10.48550/arXiv.2006.02419

Lenneberg, E. H. (1967). *Biological foundations of language*. John Wiley & Sons.

MacWhinney, B. (2000). *The CHILDES Project: The database* (Vol. 2). Lawrence Erlbaum Associates, Inc.

McDermott-Hinman, A., Zimmerman, S., Snedeker, J., & Feiman, R. (2026). Separating cognitive development from language development in the acquisition of negation using international adoption. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001938

Oba, M., Kuribayashi, T., Ouchi, H., & Watanabe, T. (2023). *Second Language Acquisition of Neural Language Models* (arXiv:2306.02920). arXiv. https://doi.org/10.48550/arXiv.2306.02920

Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cognitive Development*, *18*(4), 511–531. https://doi.org/10.1016/j.cogdev.2003.09.006

Senghas, A., & Coppola, M. (2001). Children Creating Language: How Nicaraguan Sign Language Acquired a Spatial Grammar. *Psychological Science*, *12*(4), 323–328. https://doi.org/10.1111/1467-9280.00359

Senghas, A., Coppola, M., Newport, E., & Supalla, T. (1997). Argument Structure in Nicaraguan Sign Language: The Emergence of Grammatical Devices. *Proceedings of the Boston University Conference on Language Development*, *21*, 550–561.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua. *Science*, *305*(5691), 1779–1782. https://doi.org/10.1126/science.1100199

Shufaniya, A., & Arnon, I. (2018). Statistical Learning Is Not Age-Invariant During Childhood: Performance Improves With Age Across Modality. *Cognitive Science*, *42*(8), 3100–3115. https://doi.org/10.1111/cogs.12692

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Steinert-Threlkeld, S. (2020). Toward the Emergence of Nontrivial Compositionality. *Philosophy of Science*, *87*(5), 897–909. https://doi.org/10.1086/710628

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge University Press.

Vong, W. K., Wang, W., Orhan, A. E., & Lake, B. M. (2024). Grounded language acquisition through the eyes and ears of a single child. *Science*, *383*(6682), 504–511. https://doi.org/10.1126/science.adi1374

Werker, J. F. (2024). Phonetic perceptual reorganization across the first year of life: Looking back. *Infant Behavior and Development*, *75*, 101935. https://doi.org/10.1016/j.infbeh.2024.101935