

Learning Grammar from Caregiver Feedback



Abdellah Fourtassi
Associate Professor, HDR
Aix Marseille University

The learnability problem

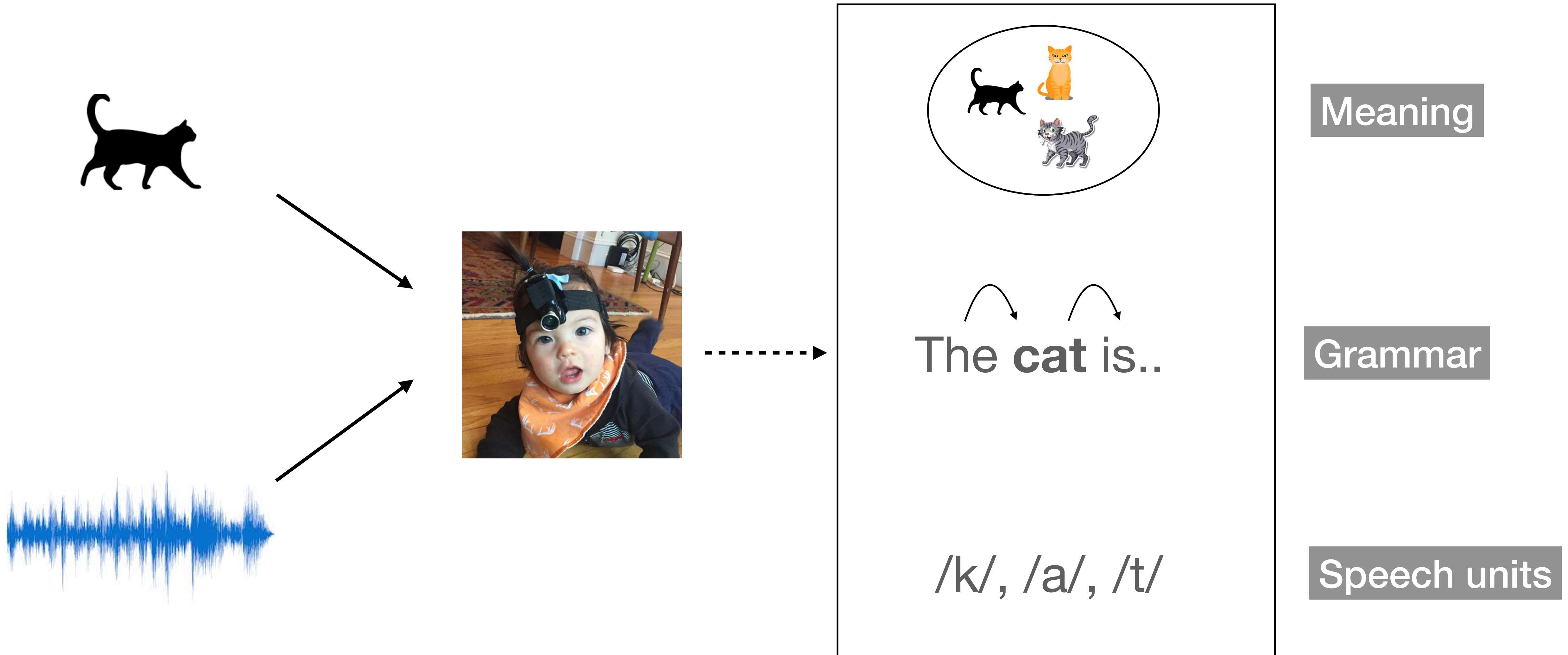
How do humans acquire **rich knowledge** from **poor input**?

Has been debated in philosophy (e.g., empiricism vs. rationalism)

Fundamental to theories of Cognitive Science (e.g., nature vs. nurture)

Language Learning is an important historical and modern case

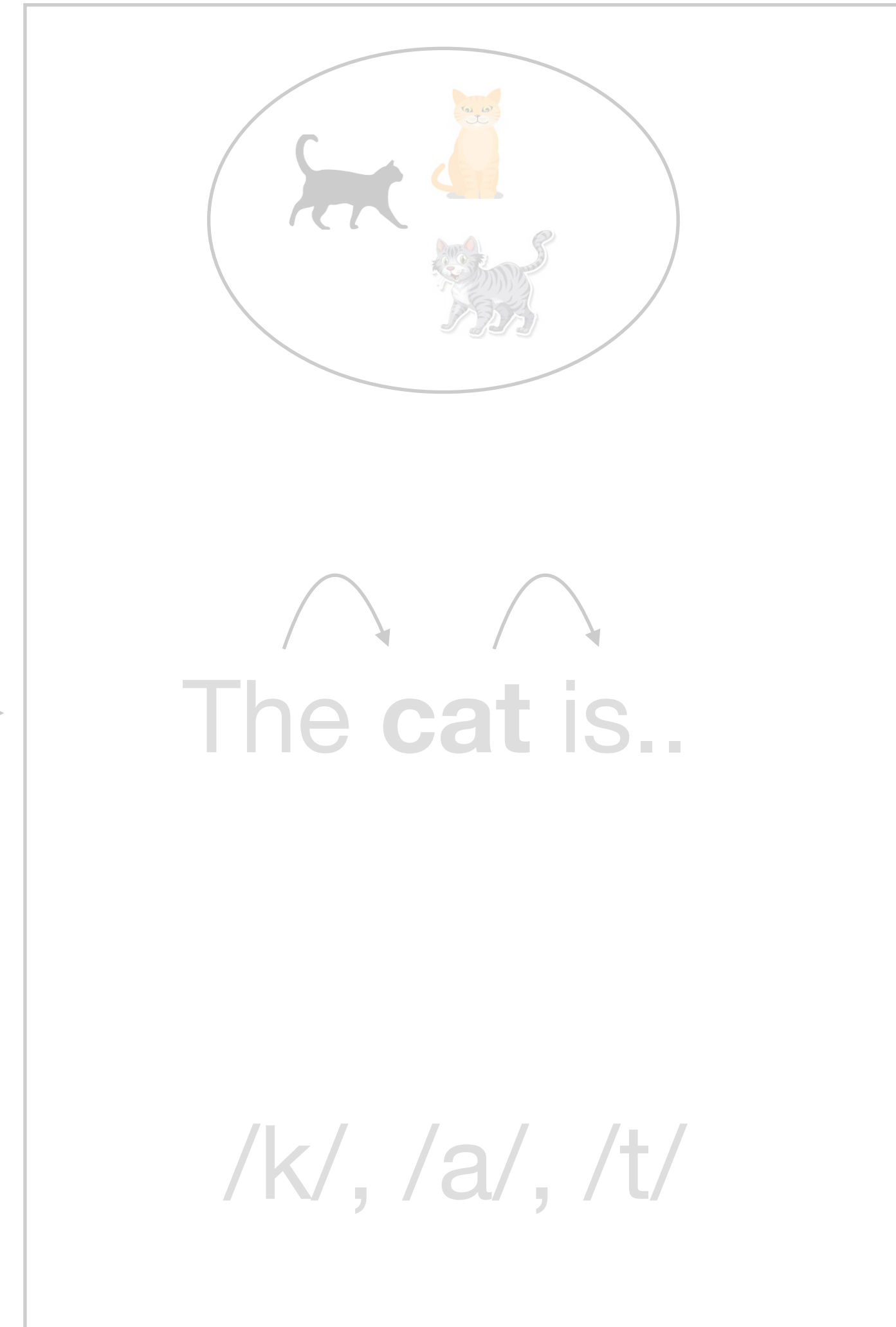
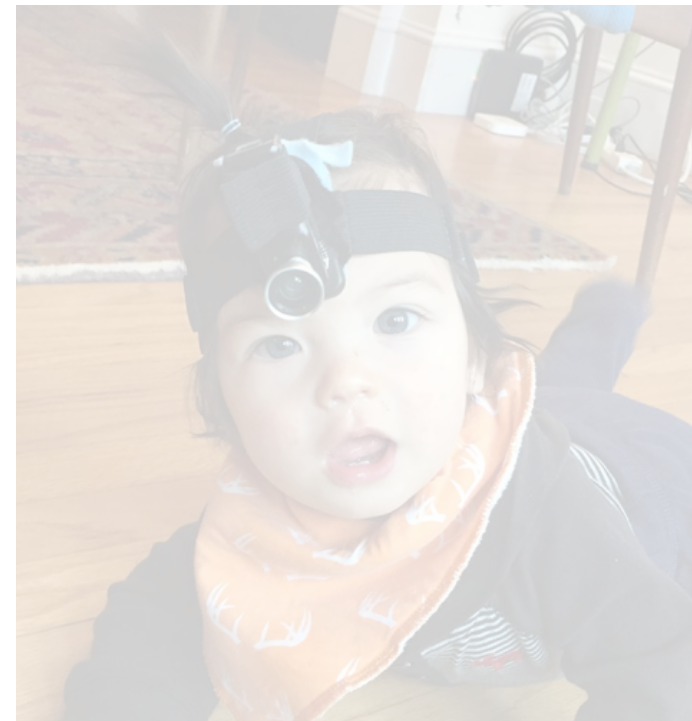
Language Learning



input

?

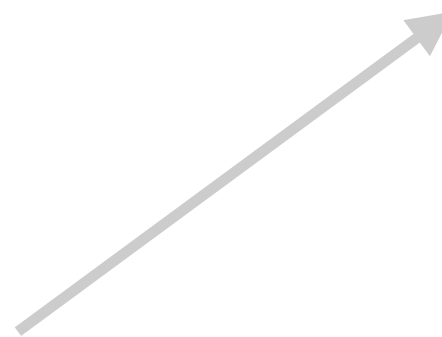
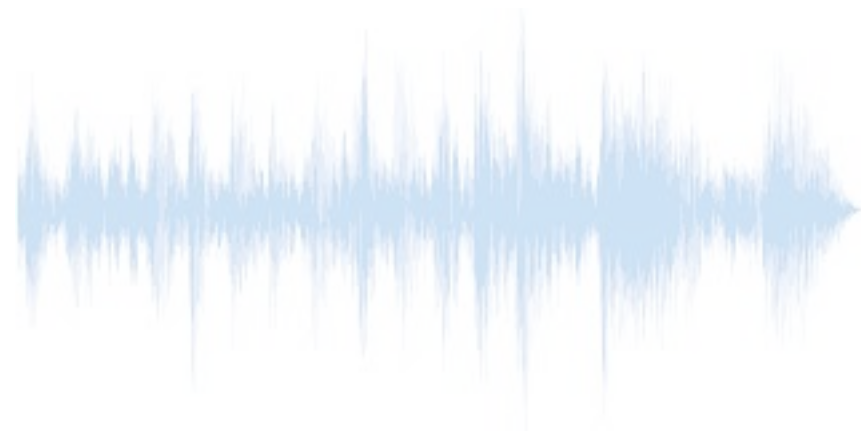
knowledge



Meaning

Grammar

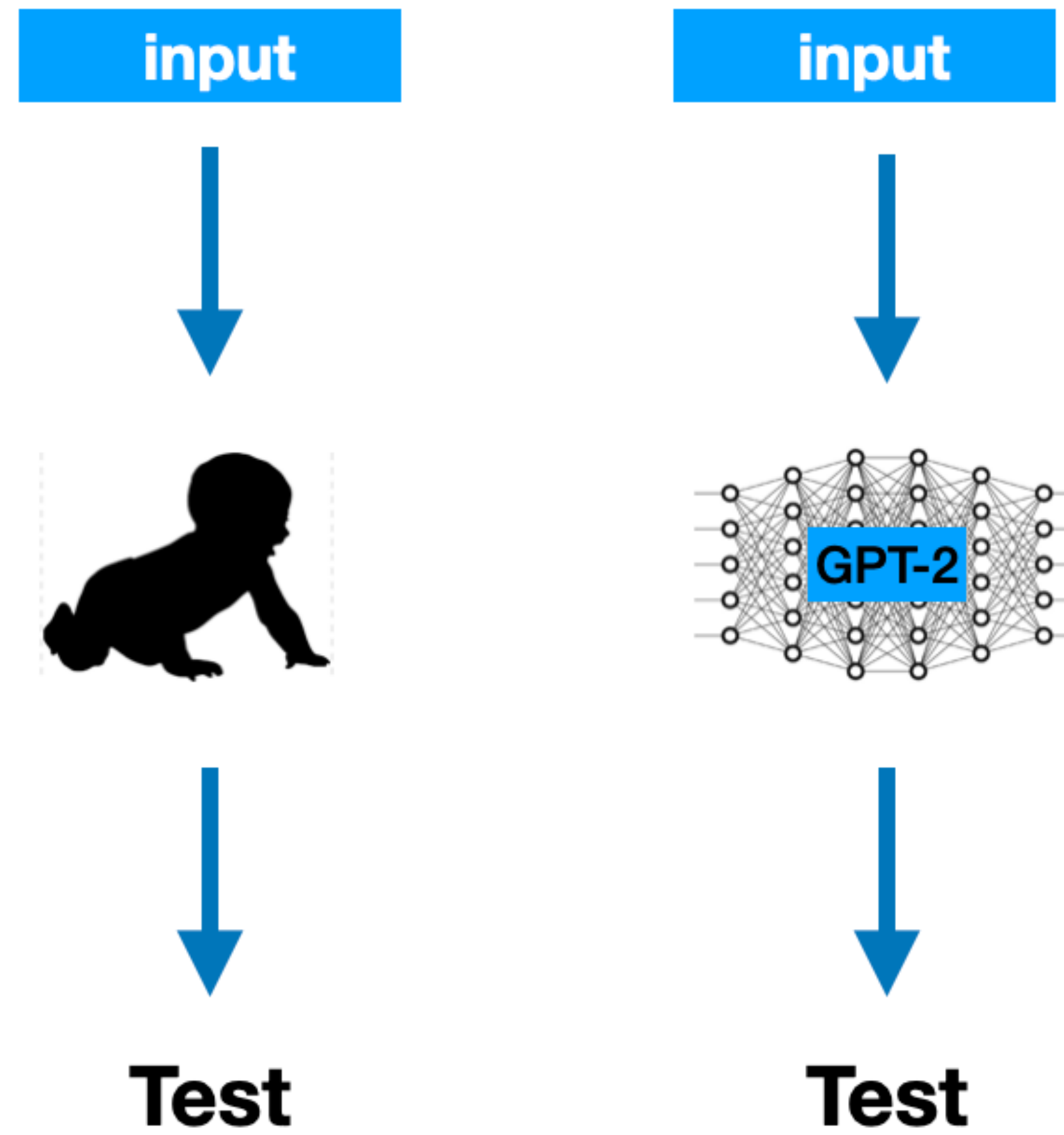
Speech units

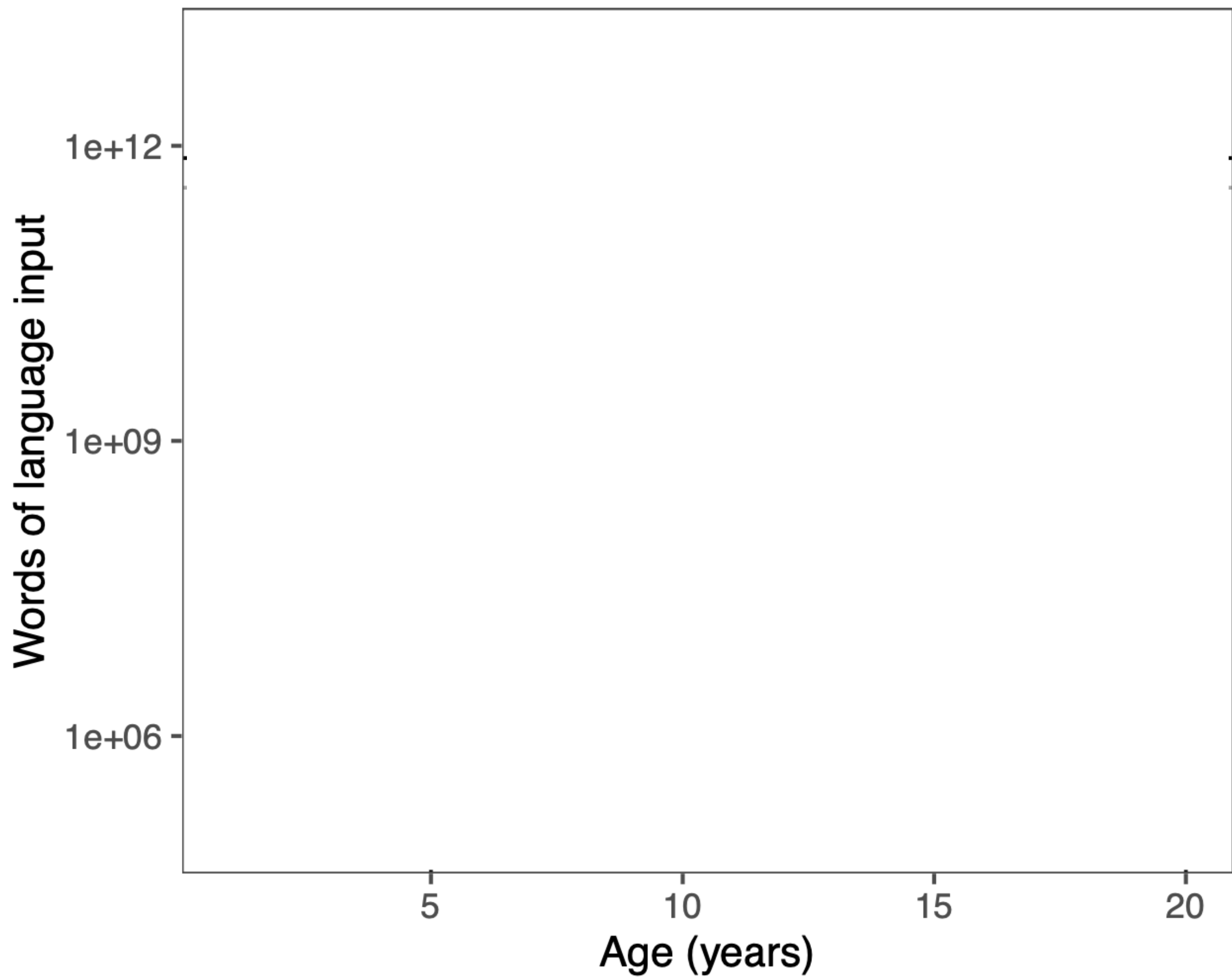


Learnability cannot be tested directly in the lab

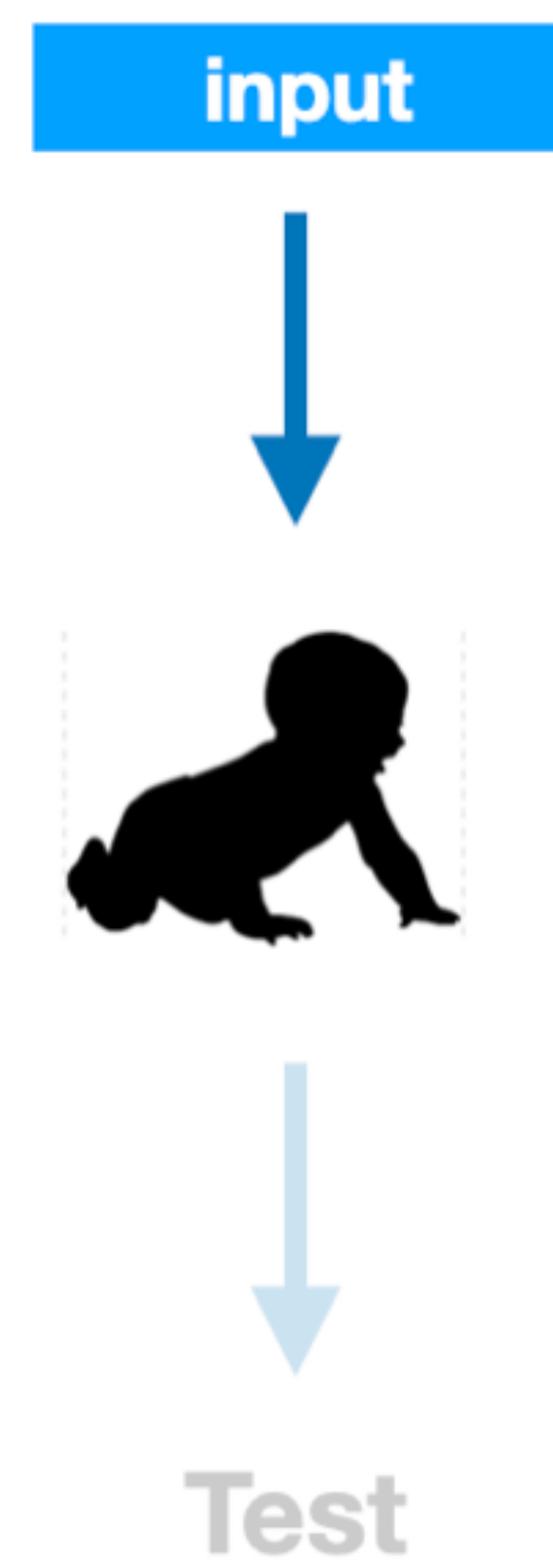
- Language develops over years — we cannot intervene on children's environment
- Animal models can allow manipulation of experience (e.g., perception, memory)
- No “Animal model” for language !

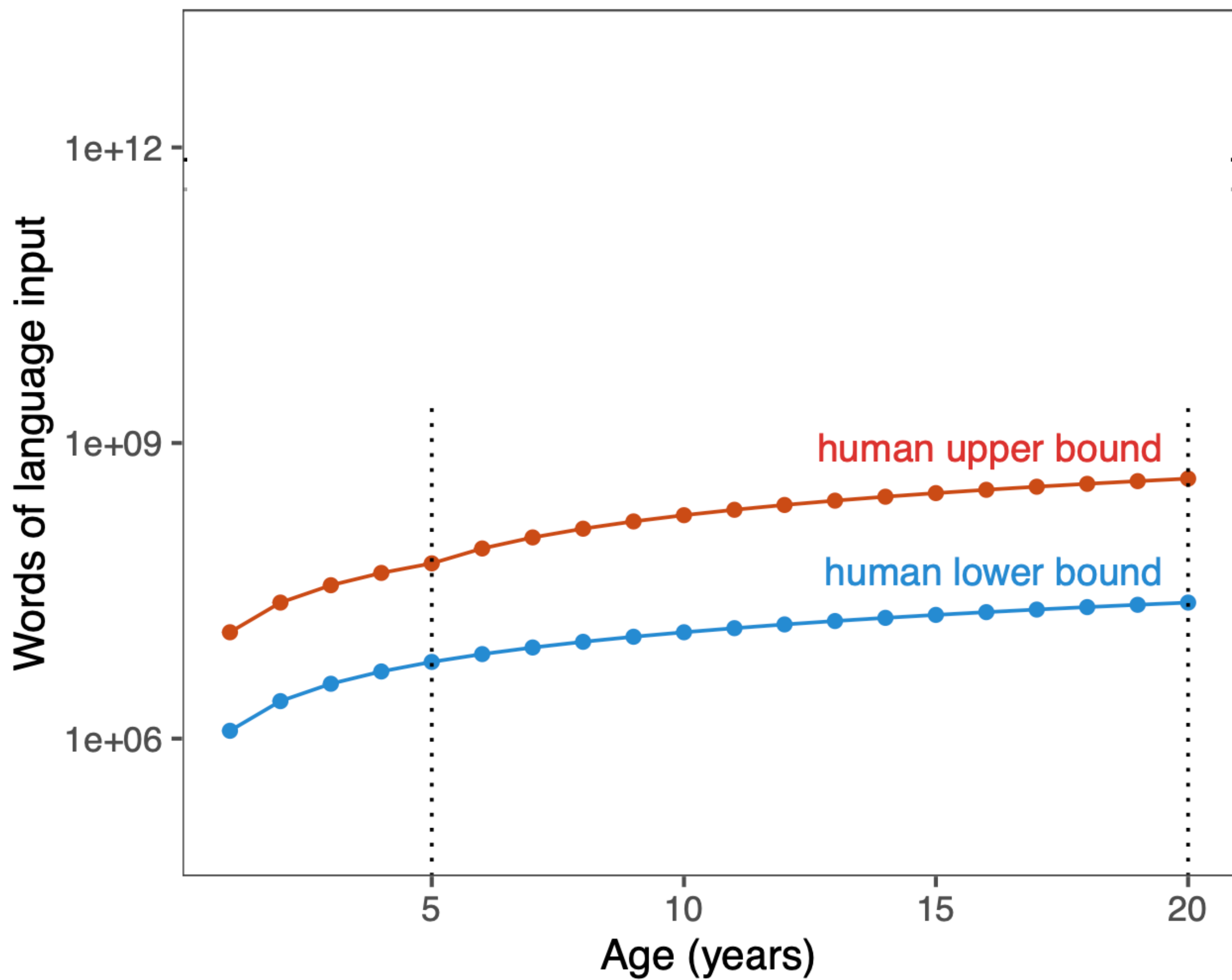
We use an *artificial* model





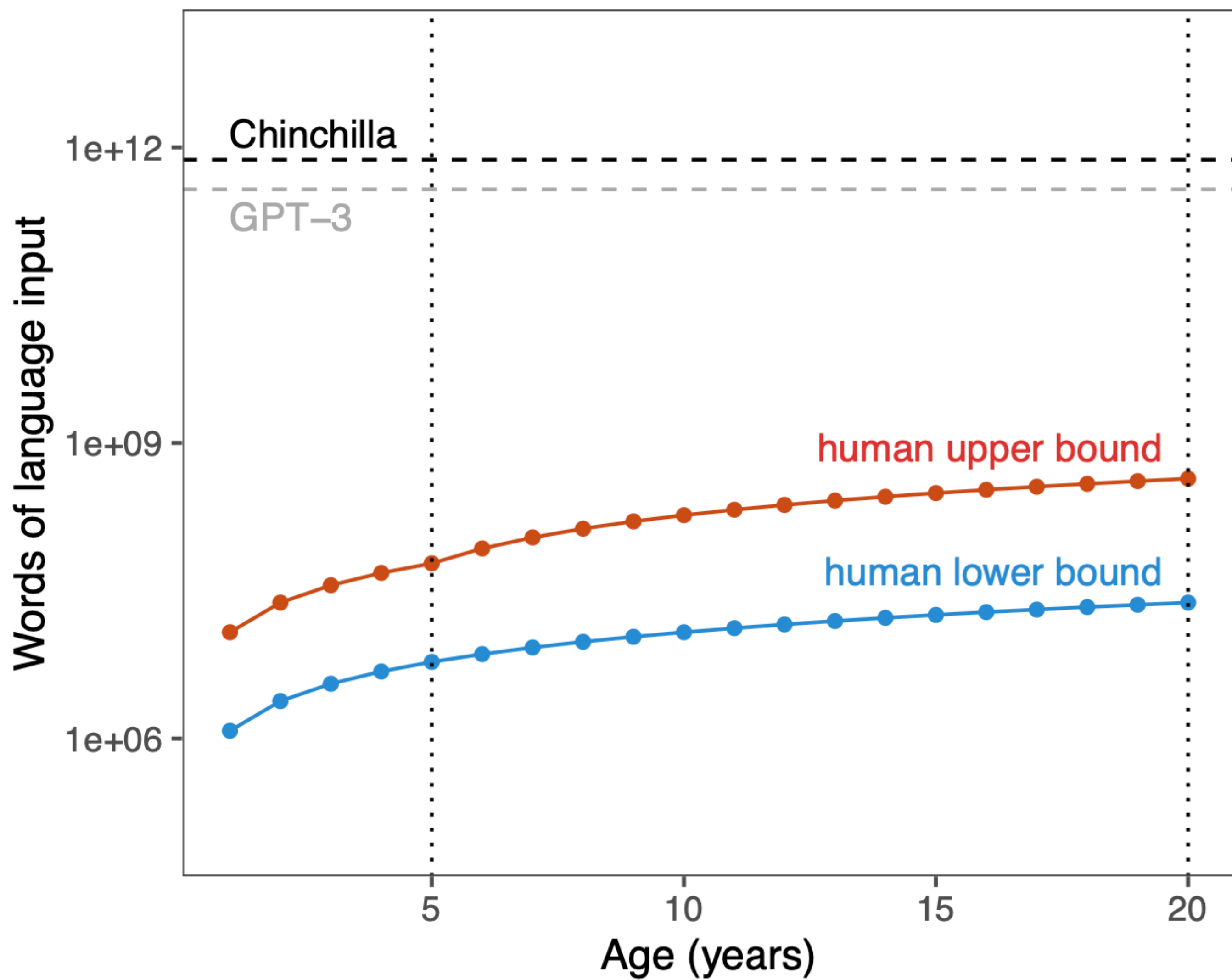
Frank (2023). *Trends In Cognitive Science*





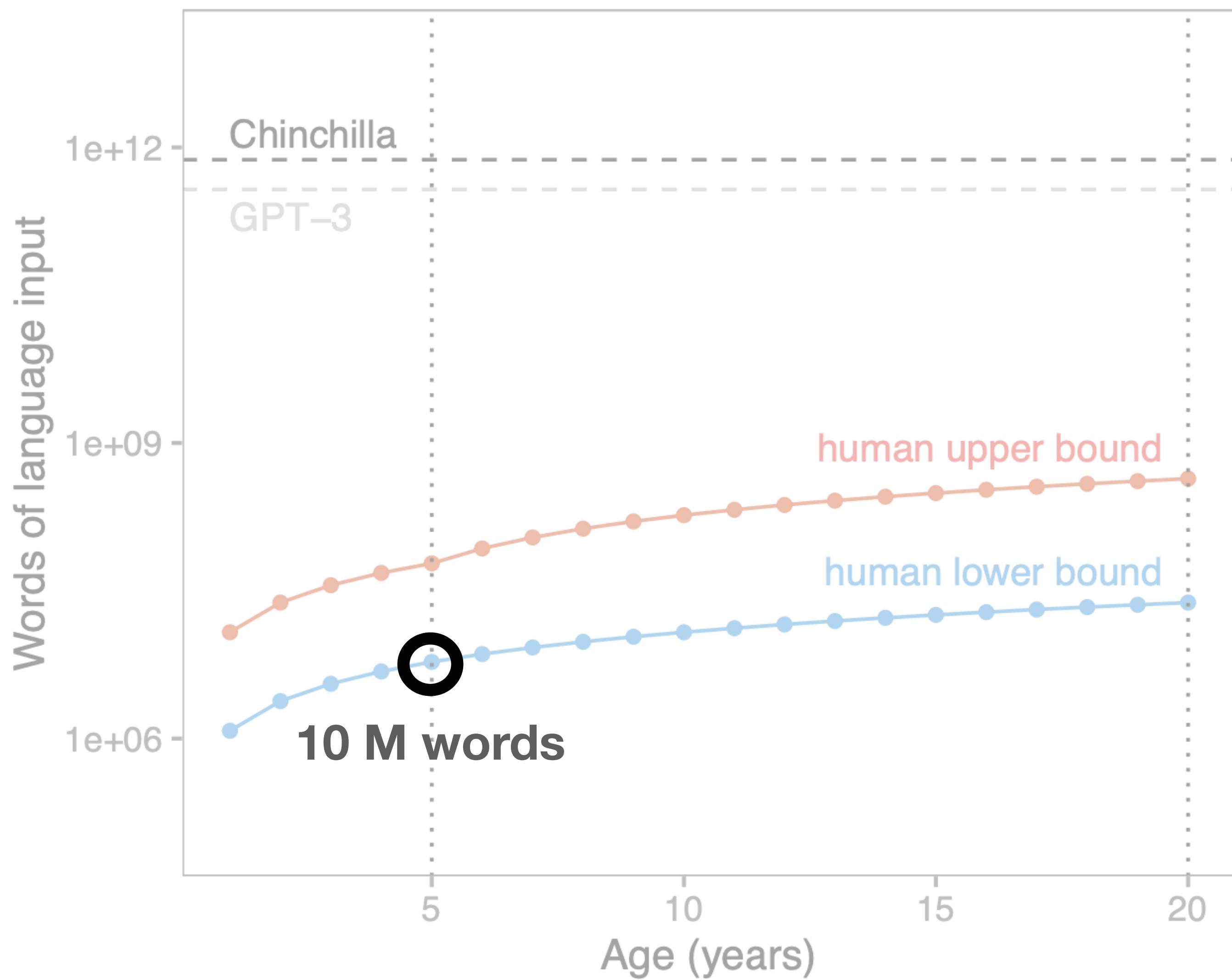
Frank (2023). *Trends In Cognitive Science*

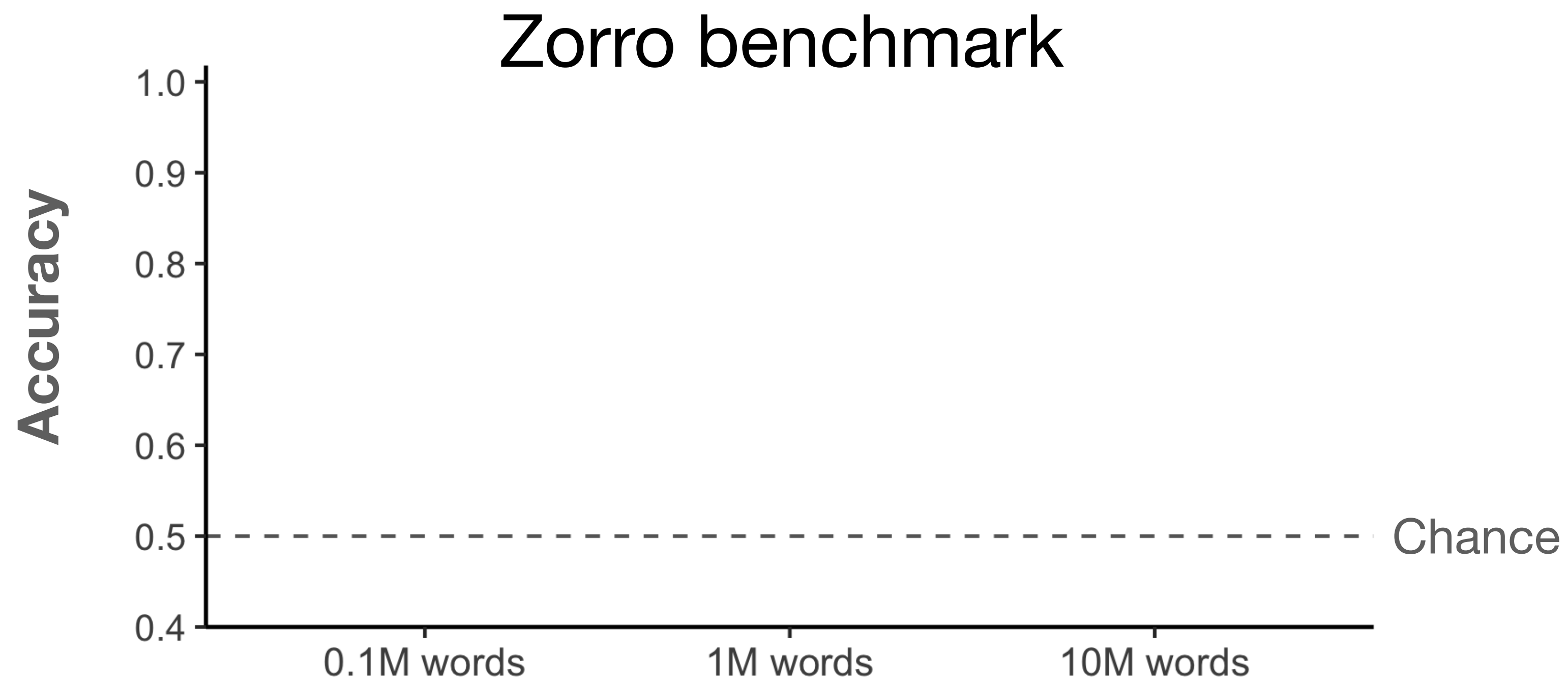


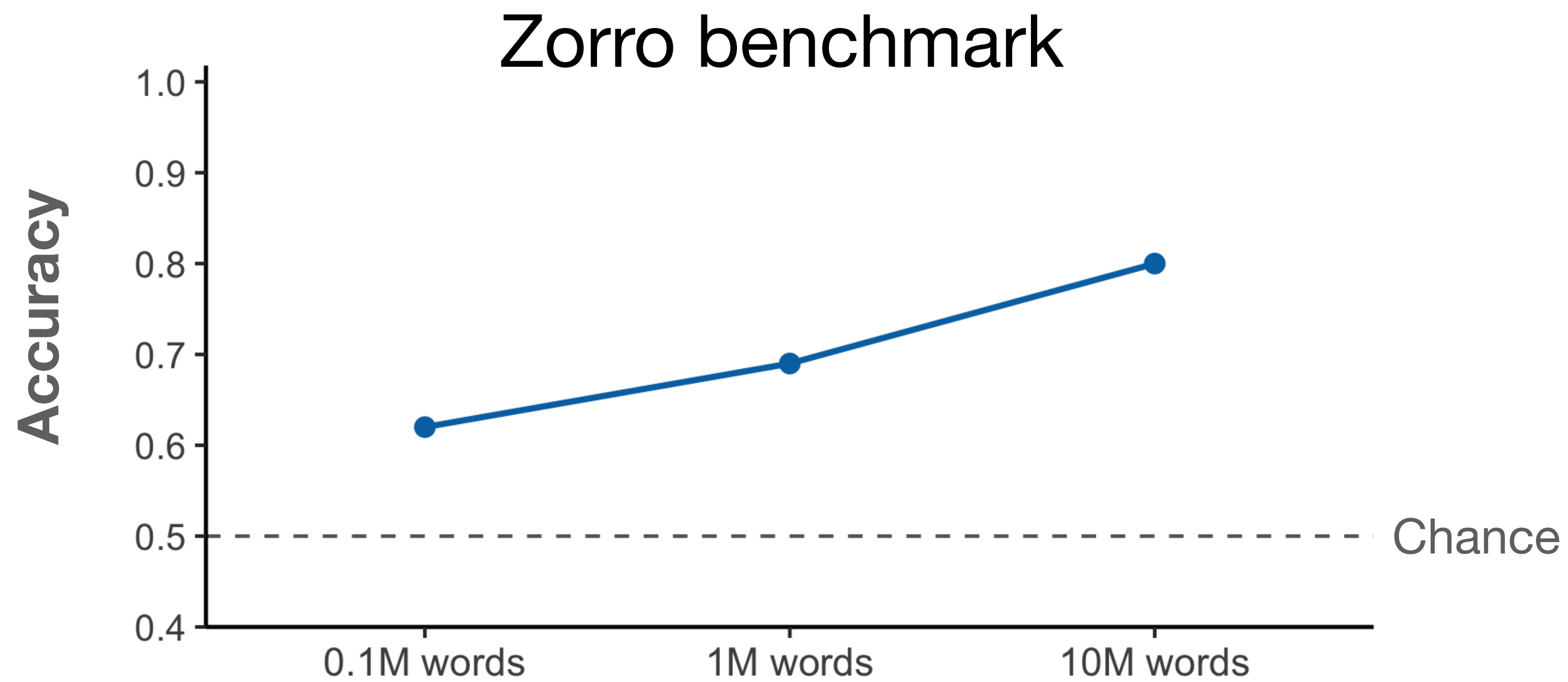


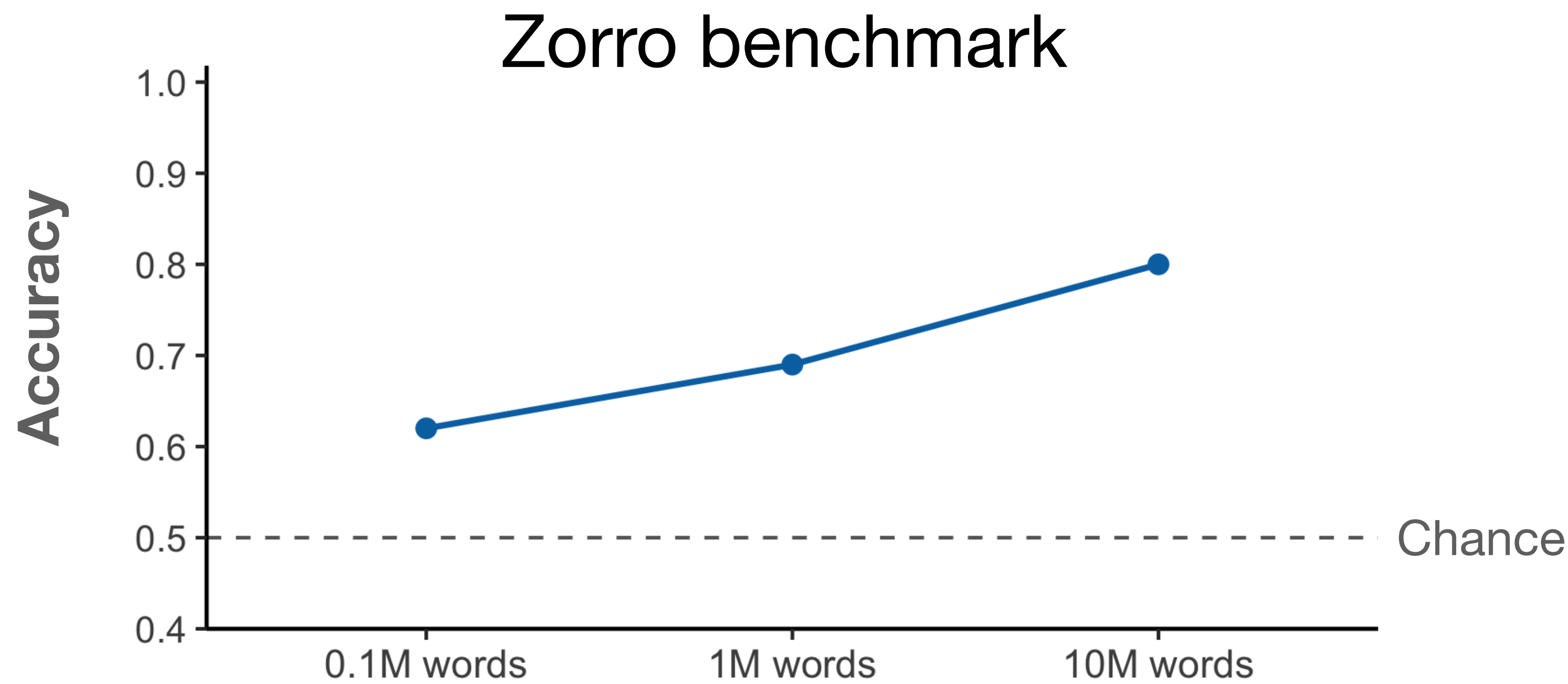
Frank (2023). *Trends In Cognitive Science*











zorro	agr. det.-noun	0.60 ± 0.01	→	0.87 ± 0.01
	agr. subject verb	0.58 ± 0.00		0.70 ± 0.01
	anaphor agr.	0.47 ± 0.00		0.78 ± 0.01
	argument struct.	0.70 ± 0.04		0.80 ± 0.01
	binding	0.49 ± 0.07		0.83 ± 0.02
	case	0.95 ± 0.05		0.99 ± 0.01
	ellipsis	0.36 ± 0.08		0.41 ± 0.02
	filler	0.78 ± 0.09		0.94 ± 0.00
	irregular	0.71 ± 0.00		0.87 ± 0.01
	island	0.51 ± 0.02		0.76 ± 0.04
	local attractor	0.49 ± 0.03		0.84 ± 0.01
	npi licensing	0.67 ± 0.09		0.69 ± 0.07
	quantifiers	0.67 ± 0.03		0.94 ± 0.02

What about children's *production* ?

input



What about children's *production* ?

Children do not just absorb linguistic input

They use their emerging knowledge to communicate

input



What about children's *production* ?

input



"He falled"

Children do not just absorb linguistic input

They use their emerging knowledge to communicate

A window into early produced forms and ***errors***

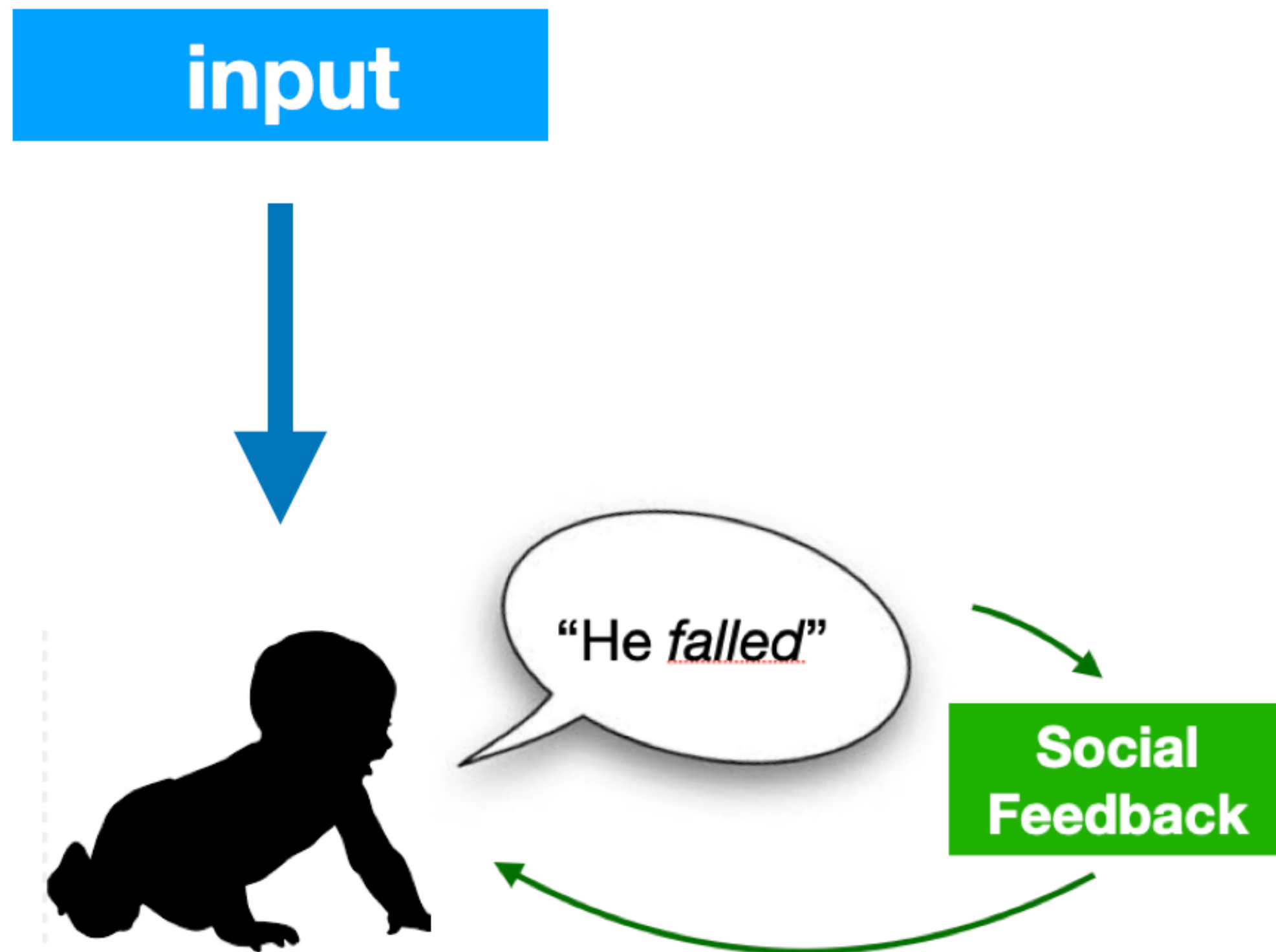
What about children's *production* ?

Children do not just absorb linguistic input

They use their emerging knowledge to communicate

A window into early produced forms and **errors**

We can study grammar learning via communication



Child Conversational Grammaticality (**Child-CoG**)

Child Conversational Grammaticality (**Child-CoG**)

Data

- 20 transcripts of child-caregiver conversation
- Spanning 2 to 5 years old
- Sampled from corpora in CHILDES

Child Conversational Grammaticality (**Child-CoG**)

Data

- 20 transcripts of child-caregiver conversation
- Spanning 2 to 5 years old
- Sampled from corpora in CHILDES

Manual annotation (3 annotators)

Is the utterance grammatical in context?

Child Conversational Grammaticality (**Child-CoG**)

Data

- 20 transcripts of child-caregiver conversation
- Spanning 2 to 5 years old
- Sampled from corpora in CHILDES

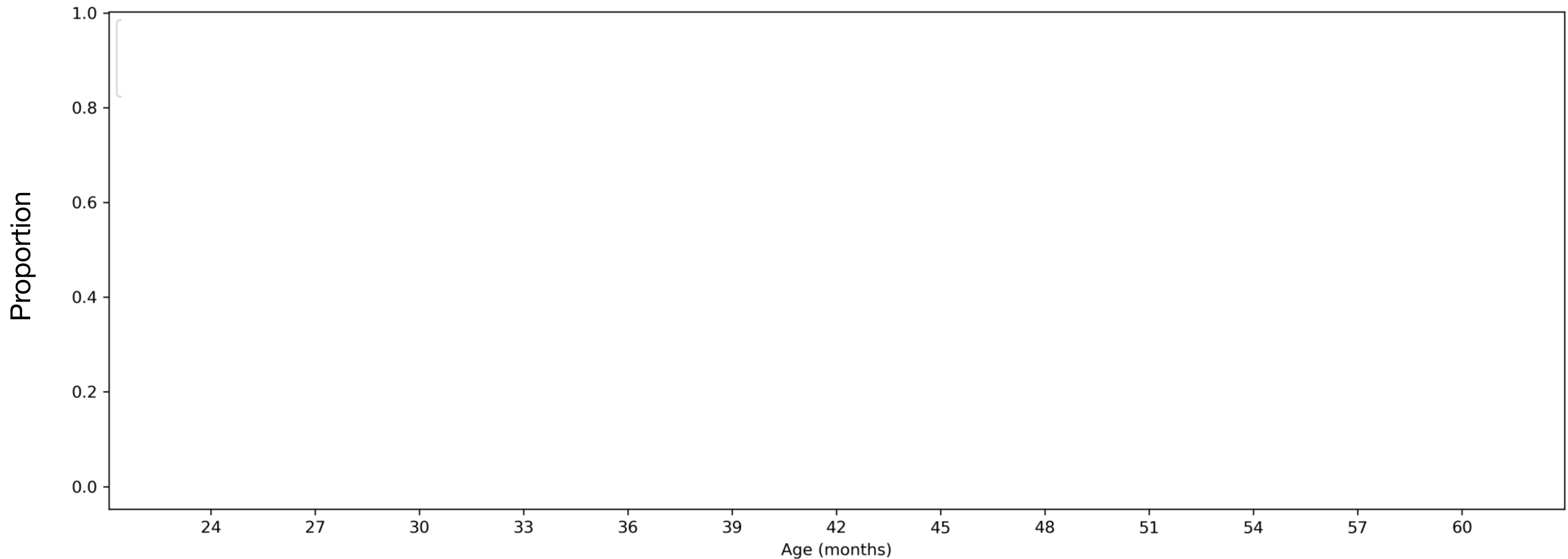
Manual annotation (3 annotators)

Is the utterance grammatical in context?

Supervised LM fine-tuning

Child Conversational Grammaticality (Child-CoG)

Automatic annotation: 276,200 utterances from 321 children and 1900 transcripts



Nikolaus, Agrawal, Kaklamanis, Warstadt, & Fourtassi (2024). *LREC*

Case study: Clarification request as negative evidence

- Child: "*Need some milk.*"
- Caregiver: "*Hm?*"
- Child: "*I need some milk*"

— EllisWeismer corpus, LT/42pc/22175.cha

Case study: Clarification request as negative evidence

- Child: "*Need some milk.*" [subject omission error]
- Caregiver: "*Hm?*" [clarification request]
- Child: "*I need some milk*" [child corrects the error]

— EllisWeismer corpus, LT/42pc/22175.cha

Open questions

How does production-based grammaticality relate to performance on benchmarks such as Zorro and BLiMP?

How should we measure the coverage, frequency, and severity of errors in open-ended generation?

Can production-based measures be standardized into reliable developmental benchmarks?

Thanks