Modeling Overregularization in Children with Small Language Models

Akari Haga[1] Saku Sugawara[2] Akiyo Fukatsu[3] Miyu Oba[1]
Hiroki Ouchi[1] Taro Watanabe[1] Yohei Oseki[3]

[1]Nara Institute of Science and Technology [2]National Institute of Informatics [3]The University of Tokyo

Background

Humans can learn language with less training data than current language models (LMs).

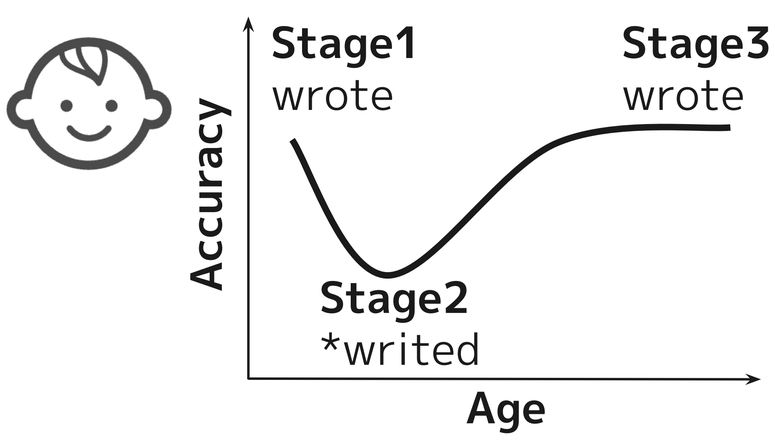
Research in the field of NLP

Huebner+ (2021) and Eldan+ (2023) suggested that emulating human learning processes can improve the training efficiency of language models.

Research in the field of first language acquisition

Children tend to add -ed (overgeneralize) to the past tense of all verbs at middle stage of learning (e.g., write→writed, take→taked).

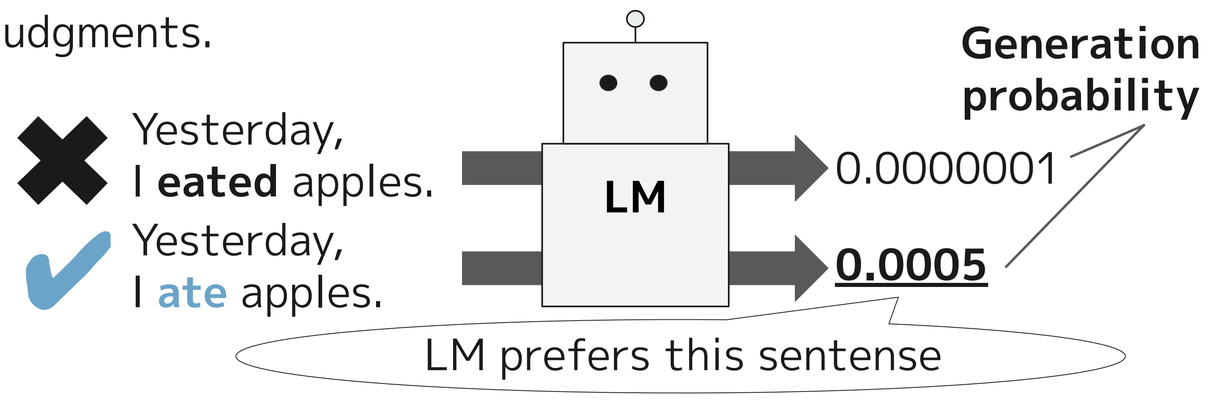
Overgeneralization errors and the U-shaped learning curve are crucial for efficient language acquisition (Bowerman+ 1982; Carlucci+ 2013).

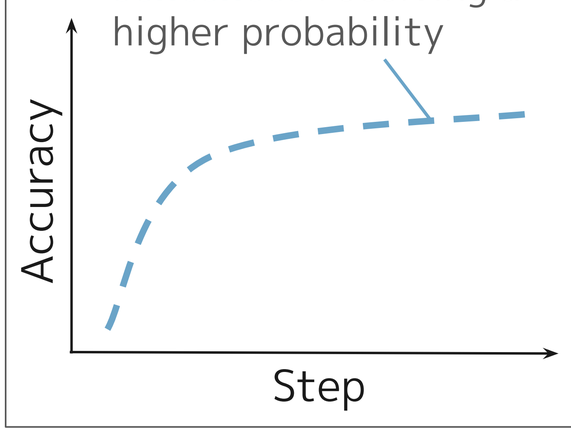
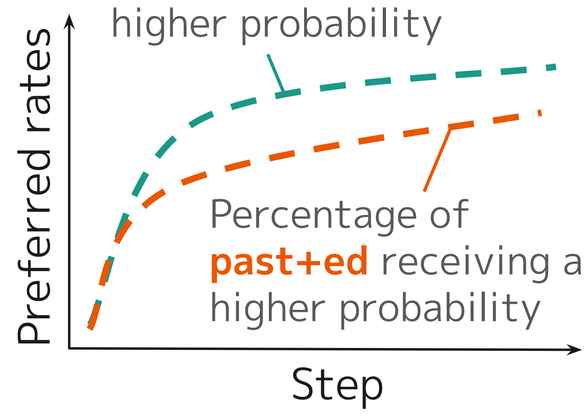


Our Approach

Our hypothesis：
Simulating children's errors allows LMs to achieve **human-like and efficient learning**

We analyzed the learning curve and error preferences of verb inflections in small-scale LMs using acceptability judgments.



How to analyze LM's learning curves	How to analyze LM's error preferences
Percentage of **correct sentences** receiving a higher probability	Percentage of **base+ed** receiving a higher probability
	

Experimental Settings

Training Data

(i) AO-CHILDES (Huebner and Willits, 2021)
(ii) Wikipedia (Huebner+ 2021)
500,000 sentences randomly collected from the English Wikipedia corpus
(iii) AO-CHILDES+Wikipedia

Minimal Pair Data

We created 1,000 sets of sentences containing the following forms:

- Correct form (e.g., John **wrote** this article.)
- Base+ed form (e.g., John **writed** this article.)
- Past+ed form (e.g., John **wroted** this article.)

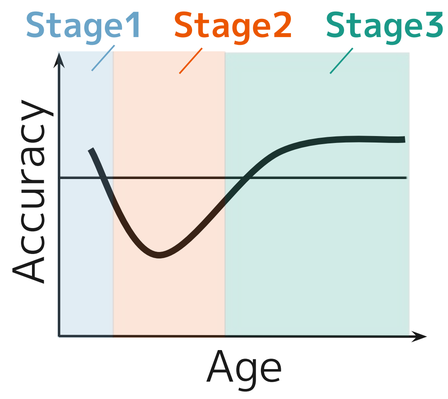
Models

Model	Layers	Heads	Enbeddings	Parameters
nanoGPT	6	6	384	13.77M
nanoGPT (2.90M)	3	3	192	2.90M
BabyBERTa	8	8	256	8.52M

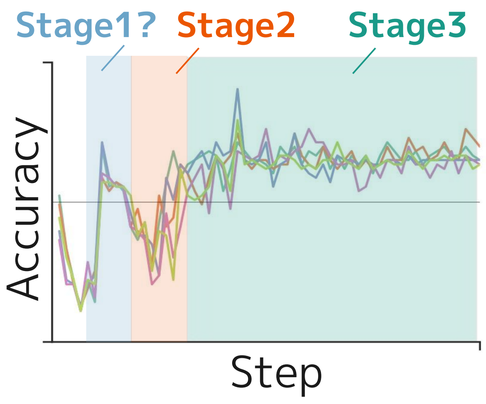
Results

Correct Form vs. Overregularized Form

Children (Marcus+ 1992)

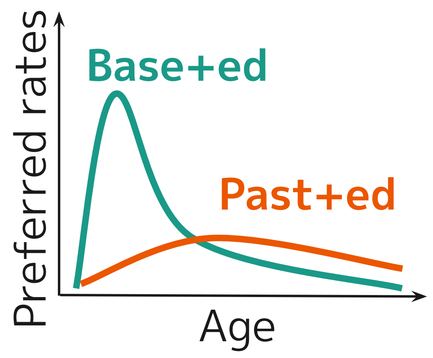


nanoGPT (2.90M)

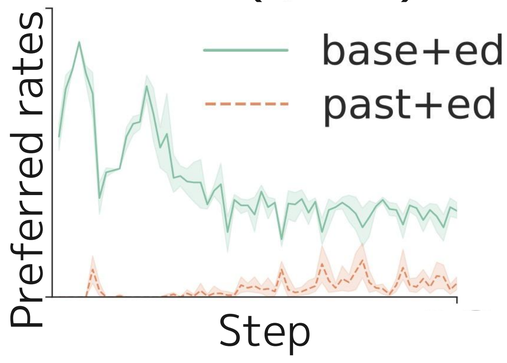


Error Preferences

Children (Kuczaj '77)



naoGPT (2.90M)



- **3 stages was observed only for certain verb types**
- **The preference for errors differed from that of children**