

Less is More: Pre-Training Cross-Lingual Small-Scale Language Models with Cognitively-Plausible Curriculum Learning Strategies

Suchir Salhan^{1,2†*}, Richard Diehl Martinez¹, Zébulon Goriely¹ & Paula Buttery^{1,2†}

¹ Department of Computer Science & Technology, University of Cambridge, U.K. ² ALTA Institute, Cambridge, U.K.

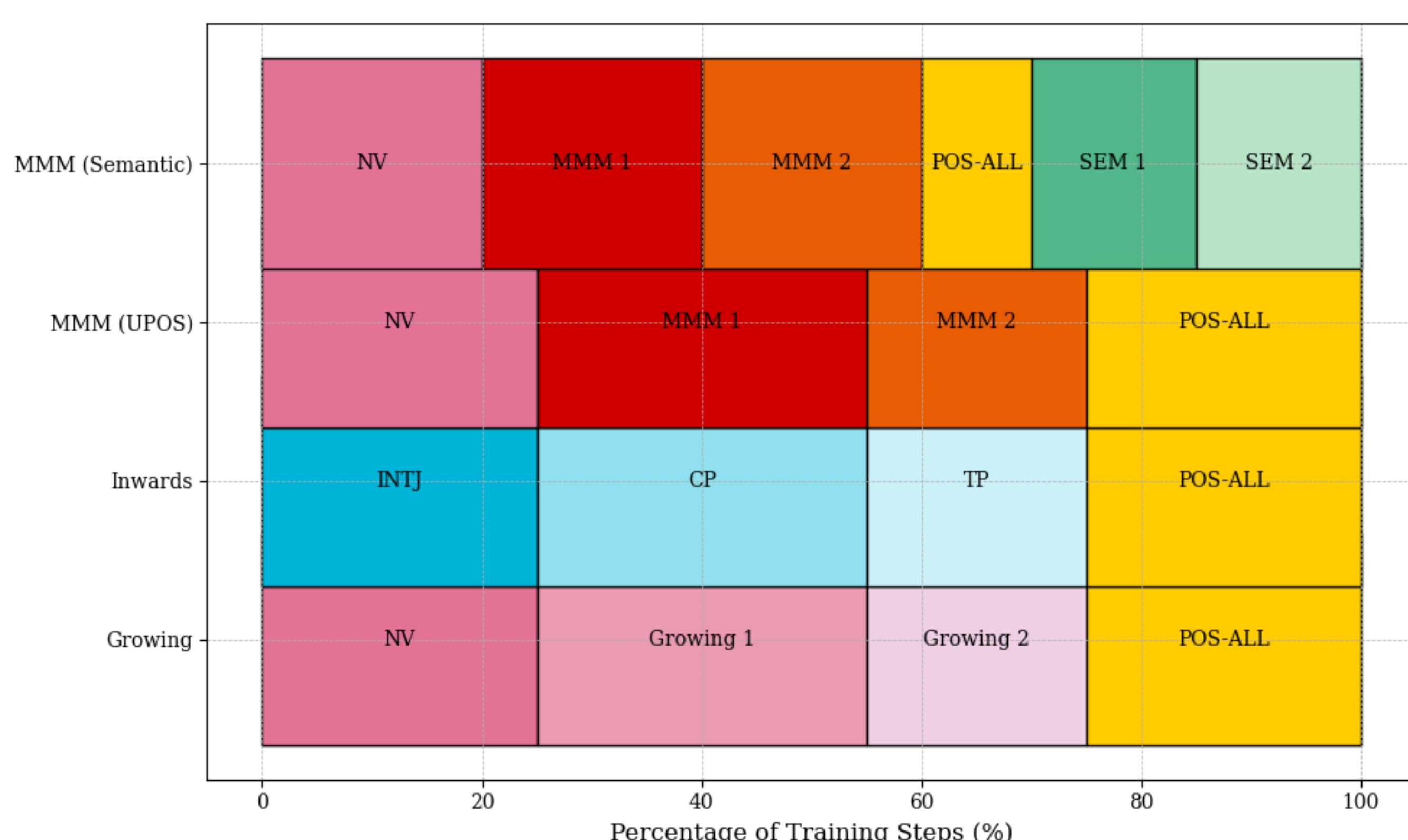
[†] supported by Cambridge University Press & Assessment * Corresponding Author: sas245@cam.ac.uk

Curriculum Learning (CL) strategies, used to improve the cognitive plausibility of small language models, have not led to considerable improvements over non-curriculum models in the BabyLM Challenge. We find fine-grained curricula, precisely replicating the developmental sequences of contemporary language acquisition theories, can outperform non-curriculum baselines, with benefits from fine-grained language-specific curricula rather than universal curricula cross-lingually.

SSLMs beyond English

Curriculum Learning has been a popular strategy to improve the cognitive plausibility of Small-Scale Language Models (SSLMs), trained on corpora that approximate the volume and nature of input a six-year-old learner is expected to receive to support data-efficient training comparably to human learners. As a **Cross-Lingual Extension** of the BabyLM Challenge, we release **MAO-CHILDES**, a universal POS-tag annotated **age-ordered training corpora** using **Child-Directed Speech (CDS)** from Child Language Data Exchange System for five languages (English, Japanese, Chinese, French, German). We assess whether **theoretical linguistic acquisition theories** can be used to specify **more fine-grained acquisition-inspired curricula** cross-lingually.

Cognitively-Plausible Pre-Training: Growing, Inwards & MMM



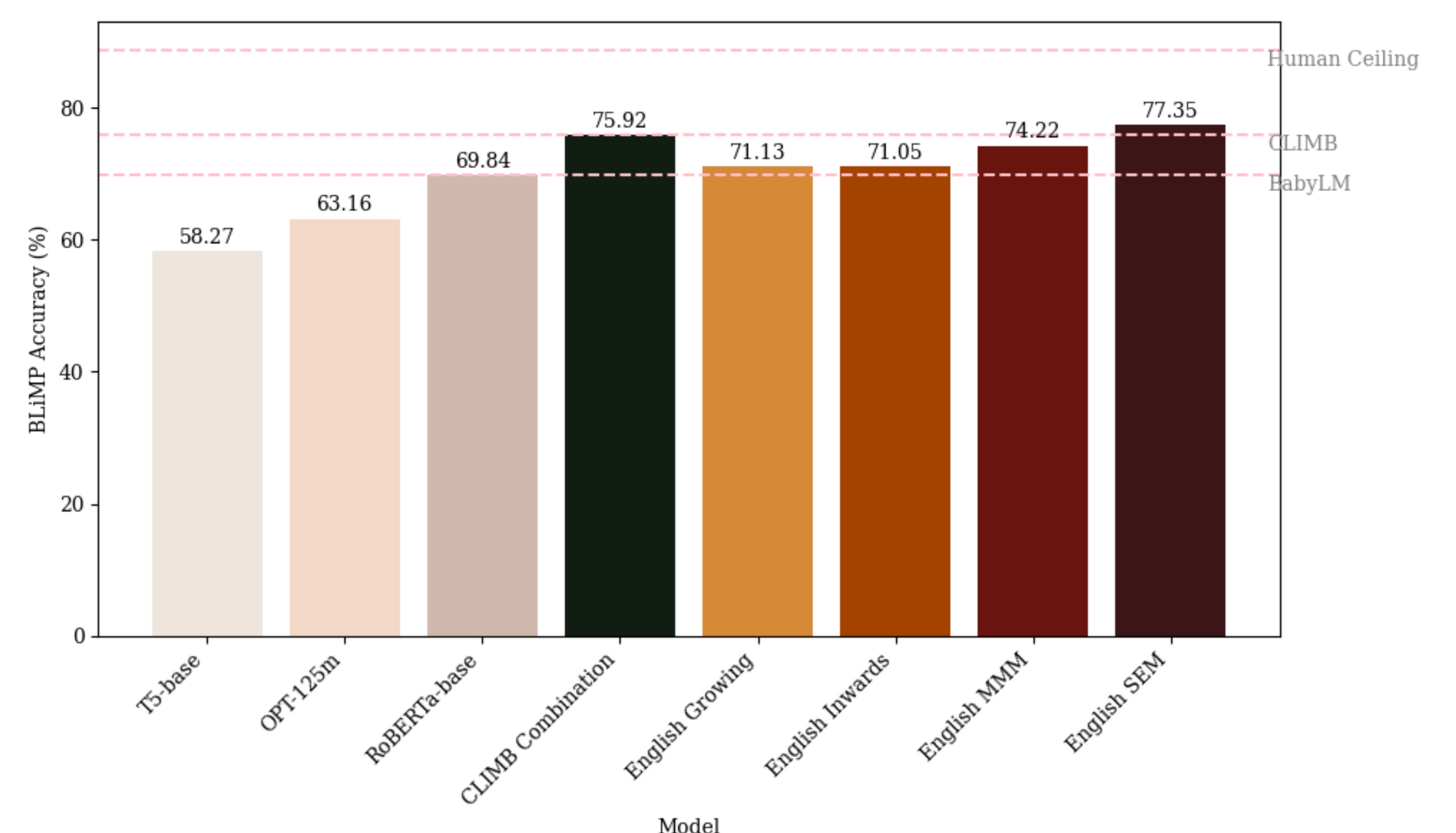
We compare the performance of three objective curricula (Growing, Inwards and MMM) that precisely replicate the predictions of acquisition theories on a standard BabyBERTa architecture. The objective curricula perform multi-task masked language modelling over words belonging to a specific UPOS tag subset. This corresponding to the acquisition model’s predicted developmental sequence in first language acquisition. Growing and Inwards predict **universal developmental sequences cross-lingually**, while MMM predicts learners follow a **language-specific** ordering. We implement a coarse universal MMM (UPOS) prediction and a language-specific hypothesis (MMM SEM).

Results: Less is More

Language-specific curricula (MMM SEM) achieve the best performance (Table 1). SSLMs with CL can **perform competitively against LLMs** (Table 2). Our Japanese model with MMM (UPOS) **outperforms GPT-2** despite $\times 10^3 - 10^4$ less training data.

Baselines: Mao-BabyBERTa (“vanilla” SSLM architecture with no objective curricula) and the three Objective Curricula (Growing, Inwards, and MMM) are evaluated on **language-specific syntactic minimal pairs datasets**: BLIMP (English), JBLIMP (Japanese), SLING (Chinese), CLAMS (French and German). Performance is compared to SSLM trained on non-CDS corpora of the same size from Wikipedia (Wiki). The graph shows MMM (SEM) **outperforms less fine-grained CL strategies** implemented in CLIMB (Diehl Martinez et al, 2023) in English.

	Model	en	ja	zh	fr	de
No CL	SSLM (wiki)	64.60%	55.42%	48.01%	70.68%	59.63%
	Mao-BabyBERTa	75.48% *	61.21%	51.32%	80.00%	68.78%
CL	Growing	71.13%	79.30%	56.22%	76.21%	71.13%
	Inwards	71.05%	81.32%	54.26%	79.01%	69.34%
	MMM (UPOS)	74.22%	87.31%	58.79%	75.93%	73.25%
	(SEM)	77.35%		55.01%		



	en	ja	zh	fr	de
Best CL	77.35	87.31	58.79	79.01	73.25
LLM	80.10	77.95	83.41	83.00	92.16

Our trained models and datasets can be found on Cambridge CLIMB’s HuggingFace: <https://huggingface.co/climb-mao>. With thanks to **Andrew Caines** for his support and supervision, the other co-authors of CLIMB, **Hope McGovern**, **Chris Davis** and **Lisa Beinborn**, and **Mila Marcheva** and **Núria Bosch-Masip Masip**.