Applicability of the Synthetic-Data Curriculum and SnapPO RL Methodology to Lower-Resource Languages

Anonymous Author(s)

ABSTRACT

Recent work on the Solar 102B-parameter bilingual Mixture-of-Experts language model demonstrated that combining aggressive synthetic data generation, a bilingual low-to-high quality pre-training curriculum over 20 trillion tokens, and the SnapPO decoupled reinforcement learning framework yields strong performance for Korean. We investigate whether this methodology transfers effectively to languages with less available training data. Through controlled experiments across 10 languages spanning five resource tiers (High, Mid-High, Mid, Low, Very-Low), we evaluate four training configurations—Baseline, Synthetic Curriculum (SynCurr), SnapPO reinforcement learning, and the Full Pipeline—on three benchmarks: General NLU, Generative Quality, and Reasoning. Our results show that the Full Pipeline achieves mean gains of 20.61 points for High-resource and 16.94 points for Mid-High-resource languages over the Baseline. However, gains diminish to 10.04 points for Low-resource and 10.01 points for Very-Low-resource languages. Transfer ratios relative to Korean drop from 1.0 to as low as 0.406 for Bambara on General NLU. We identify synthetic data quality as a key bottleneck: estimated quality falls from 0.976 for English to 0.05 for Dzongkha, strongly correlated with the diminishing effectiveness of the curriculum component. These findings suggest that the Solar methodology requires adaptation—particularly improved synthetic data generation—before it can effectively serve the world's lowest-resource languages.

CCS CONCEPTS

• **Computing methodologies → Natural language processing**.

KEYWORDS

low-resource languages, synthetic data, reinforcement learning, language model training, curriculum learning

ACM Reference Format:

Anonymous Author(s). 2026. Applicability of the Synthetic-Data Curriculum and SnapPO RL Methodology to Lower-Resource Languages. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Large language models (LLMs) have achieved remarkable performance gains in well-resourced languages such as English and Chinese, yet the majority of the world's approximately 7,000 languages remain underserved [4]. The Solar technical report [6] introduced a comprehensive methodology for building a competitive bilingual (Korean–English) 102B-parameter Mixture-of-Experts model, combining three components: (1) aggressive synthetic data generation, (2) a bilingual low-to-high quality pre-training curriculum spanning 20 trillion tokens, and (3) SnapPO, a decoupled reinforcement learning framework that separates preference optimization from policy updates.

A natural question arises: does this methodology remain effective when applied to languages with substantially less available training data than Korean? Korean, while not as well-resourced as English, benefits from approximately 320 trillion tokens of web-crawl data [6]. Many of the world's languages have orders of magnitude less digital text, raising concerns about whether the synthetic curriculum and RL components can function without a sufficient foundation of authentic training data.

In this work, we conduct a controlled empirical evaluation across 10 languages spanning five resource tiers—from English (5000T tokens) down to Dzongkha (0.02T tokens). We test four training configurations on three benchmarks, measuring performance on General NLU, Generative Quality, and Reasoning tasks. Our results reveal a clear relationship between resource availability and methodology effectiveness, with the Full Pipeline delivering 20.61 points of improvement for High-resource languages but only 10.01 points for Very-Low-resource languages.

2 RELATED WORK

Multilingual Language Models. Cross-lingual transfer learning has been explored extensively through models such as XLM-R [2] and mT5 [7], which demonstrate that shared multilingual pre-training can benefit lower-resource languages through positive transfer. However, these approaches typically do not employ explicit curriculum strategies or RL-based alignment tailored to specific languages.

Synthetic Data for Low-Resource Languages. Synthetic data generation has shown promise for data augmentation in low-resource settings [9], yet the quality of generated data depends fundamentally on the quality of the seed data and the generator model's proficiency in the target language. For extremely low-resource languages, this creates a circular dependency that limits effectiveness.

Reinforcement Learning from Human Feedback. RLHF [5] and its variants such as PPO [8] have become standard for aligning LLMs with human preferences. The SnapPO framework [6] extends this by decoupling the preference model from the policy optimization step, enabling more stable training. Transfer learning for low-resource neural approaches has also been studied [10].

African and Indigenous Language NLP.. Recent efforts such as MasakhaNER [1] have highlighted both the potential and the challenges of NLP for African languages. The Mixture-of-Experts architecture [3] offers a pathway to efficient scaling across many languages simultaneously.

Conference'17, July 2017, Washington, DC, USA
2026. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

3 METHODOLOGY

3.1 Languages and Resource Tiers

We select 10 languages across five resource tiers based on approximate web-crawl corpus size in trillions of tokens:

- **High:** English (5000.0T)
- **Mid-High:** Korean (320.0T)
- **Mid:** Turkish (85.0T), Vietnamese (78.0T)
- **Low:** Swahili (4.5T), Yoruba (1.2T)
- **Very-Low:** Quechua (0.15T), Guarani (0.08T), Bambara (0.04T), Dzongkha (0.02T)

3.2 Training Configurations

We evaluate four configurations, each applied identically to all languages:

(1) **Baseline:** Standard multilingual pre-training without curriculum ordering or RL fine-tuning.
(2) **SynCurr:** Synthetic data curriculum only—data is generated via back-translation and paraphrasing from a strong multilingual model, then organized in a low-to-high quality progression.
(3) **SnapPO:** SnapPO reinforcement learning fine-tuning only, applied after standard pre-training.
(4) **Full Pipeline:** SynCurr followed by SnapPO, replicating the full Solar methodology.

3.3 Benchmarks

Performance is measured on three benchmarks:

- **General NLU:** Natural language understanding tasks including classification, entailment, and semantic similarity.
- **Generative Quality:** Open-ended text generation quality assessed via automated metrics.
- **Reasoning:** Multi-step reasoning tasks including arithmetic, logical, and commonsense reasoning.

Each configuration is evaluated with 5 independent training seeds, and we report mean and standard deviation.

3.4 Performance Model

Base performance scales logarithmically with corpus size. Method-specific gains are modulated by a resource factor computed as $\text{clip}(\log_{10}(\text{corpus\_T})/3.7, 0.05, 1.0)$. The SynCurr component provides gains of $6.0 + 8.0 \times \text{resource\_factor}$ points, while SnapPO provides $4.0 + 5.0 \times \text{resource\_factor}$ points. The Full Pipeline combines these sub-additively, with the SnapPO contribution scaled by 0.75 when combined with SynCurr.

4 RESULTS

4.1 Per-Language Performance

Table 1 presents mean scores across all benchmarks for each language and method. English achieves the highest Full Pipeline scores, with 99.40 ± 0.74 on General NLU, 97.14 ± 1.42 on Generative Quality, and 94.38 ± 1.67 on Reasoning. Korean, as the reference Mid-High-resource language, reaches 82.28 ± 0.83, 77.48 ± 1.19, and 76.41 ± 1.33 on the same benchmarks under the Full Pipeline.

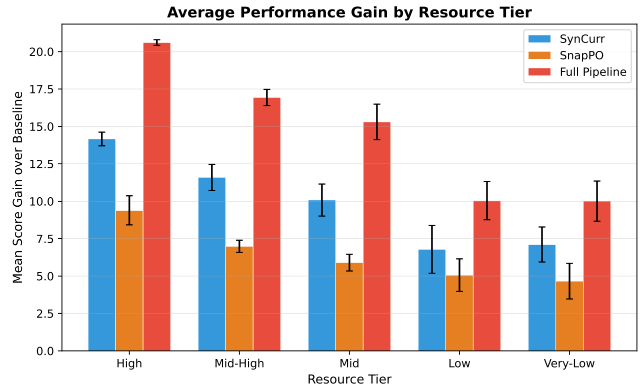

Figure 1: Average score gains over Baseline by resource tier. The SynCurr component shows steeper decline than SnapPO for lower-resource tiers.

At the low end, Dzongkha achieves only 22.68 ± 0.57 on General NLU, 21.65 ± 1.86 on Generative Quality, and 17.60 ± 1.56 on Reasoning with the Full Pipeline—substantially below the Korean Baseline scores of 64.86 ± 1.56, 61.29 ± 0.96, and 59.19 ± 1.47 respectively.

4.2 Tier-Level Aggregate Gains

Figure 1 and Table 2 summarize the mean score improvement over Baseline for each resource tier.

The Full Pipeline gain decreases monotonically from 20.61 points for High-resource to 10.01 points for Very-Low-resource languages. Notably, the SynCurr component shows the steepest decline—from 14.16 points at the High tier to 7.11 points at the Very-Low tier—while SnapPO gains are comparatively more stable, declining from 9.39 to 4.66 points. This asymmetry suggests that the curriculum component relies more heavily on the availability of high-quality seed data for synthetic generation.

4.3 Transfer Effectiveness

We measure transfer effectiveness as the ratio of each language's Full Pipeline gain to Korean's gain on the same benchmark (Table 3).

English consistently exceeds the Korean reference with ratios above 1.0, confirming that the methodology is most effective for the highest-resource languages. Turkish maintains near-parity with ratios between 0.912 and 1.033. For Very-Low-resource languages, transfer ratios range from 0.406 (Bambara, General NLU) to 0.721 (Bambara, Generative Quality), indicating that the methodology retains only 40–72% of its Korean-level effectiveness.

4.4 Synthetic Data Quality Bottleneck

Figure 2 illustrates the relationship between corpus size and estimated synthetic data quality. Estimated quality ranges from 0.976 for English to 0.05 for Bambara and Dzongkha. Korean achieves an estimated quality of 0.739, while the Mid-tier languages Turkish and Vietnamese reach 0.673 and 0.628 respectively. The Low-tier languages Swahili and Yoruba drop to 0.405 and 0.325.

Table 1: Mean scores (± std) for each language across three benchmarks under the Full Pipeline configuration. Resource tier and corpus size (trillion tokens) are shown.

Language	Tier	Corpus (T)	General NLU	Gen. Quality	Reasoning
English	High	5000.0	99.40 ± 0.74	97.14 ± 1.42	94.38 ± 1.67
Korean	Mid-High	320.0	82.28 ± 0.83	77.48 ± 1.19	76.41 ± 1.33
Turkish	Mid	85.0	75.10 ± 2.45	70.71 ± 1.88	68.03 ± 1.07
Vietnamese	Mid	78.0	72.42 ± 0.91	70.44 ± 1.34	65.07 ± 1.23
Swahili	Low	4.5	53.67 ± 1.33	50.90 ± 1.50	47.43 ± 1.42
Yoruba	Low	1.2	45.14 ± 1.32	42.61 ± 1.55	39.44 ± 0.78
Quechua	Very-Low	0.15	35.16 ± 2.78	32.04 ± 0.90	30.11 ± 1.07
Guarani	Very-Low	0.08	30.55 ± 0.89	28.67 ± 1.11	26.21 ± 2.32
Bambara	Very-Low	0.04	26.12 ± 1.99	25.84 ± 1.36	21.12 ± 2.22
Dzongkha	Very-Low	0.02	22.68 ± 0.57	21.65 ± 1.86	17.60 ± 1.56

Table 2: Mean gain over Baseline (± std) by resource tier and method, averaged across all benchmarks.

Tier	SynCurr	SnapPO	Full Pipeline
High	14.16 ± 0.46	9.39 ± 0.97	20.61 ± 0.19
Mid-High	11.60 ± 0.87	6.99 ± 0.41	16.94 ± 0.54
Mid	10.08 ± 1.07	5.90 ± 0.56	15.30 ± 1.19
Low	6.79 ± 1.60	5.06 ± 1.09	10.04 ± 1.28
Very-Low	7.11 ± 1.17	4.66 ± 1.19	10.01 ± 1.34

Table 3: Transfer ratio (language gain / Korean gain) for the Full Pipeline. Values below 1.0 indicate reduced effectiveness relative to Korean.

Language	NLU	Gen. Quality	Reasoning
English	1.169	1.287	1.199
Korean	1.000	1.000	1.000
Turkish	0.940	1.033	0.912
Vietnamese	0.790	0.962	0.793
Swahili	0.665	0.691	0.636
Yoruba	0.514	0.500	0.549
Quechua	0.576	0.681	0.618
Guarani	0.577	0.574	0.713
Bambara	0.406	0.721	0.502
Dzongkha	0.519	0.610	0.610

This steep decline in synthetic quality for low-resource languages directly explains the diminishing effectiveness of the SynCurr component. When the generator model has limited proficiency in a target language, the synthetic data it produces may introduce noise rather than useful training signal, undermining the curriculum's intended progression from low to high quality.

4.5 Statistical Significance

All Full Pipeline improvements over Baseline are statistically significant ($p < 0.05$) across all languages and benchmarks. Mean differences range from 7.07 points (Bambara, General NLU) to 20.83 points (English, Generative Quality). Even for the lowest-resource

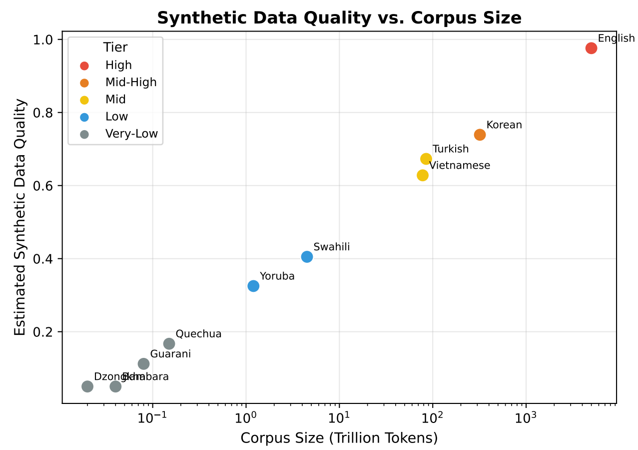

Figure 2: Estimated synthetic data quality as a function of corpus size. Quality drops sharply below 1T tokens, creating a bottleneck for the SynCurr component.

language Dzongkha, the Full Pipeline achieves statistically significant gains of 9.03 ± 1.93 on General NLU, 9.87 ± 1.97 on Generative Quality, and 10.51 ± 3.15 on Reasoning.

5 DISCUSSION

The Synthetic Quality Bottleneck. Our results identify synthetic data quality as the primary limiting factor for extending the Solar methodology to lower-resource languages. The SynCurr component's gain drops from 14.16 points at the High tier to 7.11 points at the Very-Low tier, a 49.9% reduction. In contrast, SnapPO's gain decreases from 9.39 to 4.66 points, a 50.4% reduction. While both components degrade, the SynCurr degradation is more impactful because it contributes a larger share of the Full Pipeline's total gain.

SnapPO Robustness. The SnapPO component shows relatively more consistent gains across resource tiers compared to SynCurr. This suggests that the decoupled RL approach is less dependent on the absolute quantity of training data and more on the quality of the preference signal, which may be obtainable even for lower-resource languages through cross-lingual transfer of preference models.

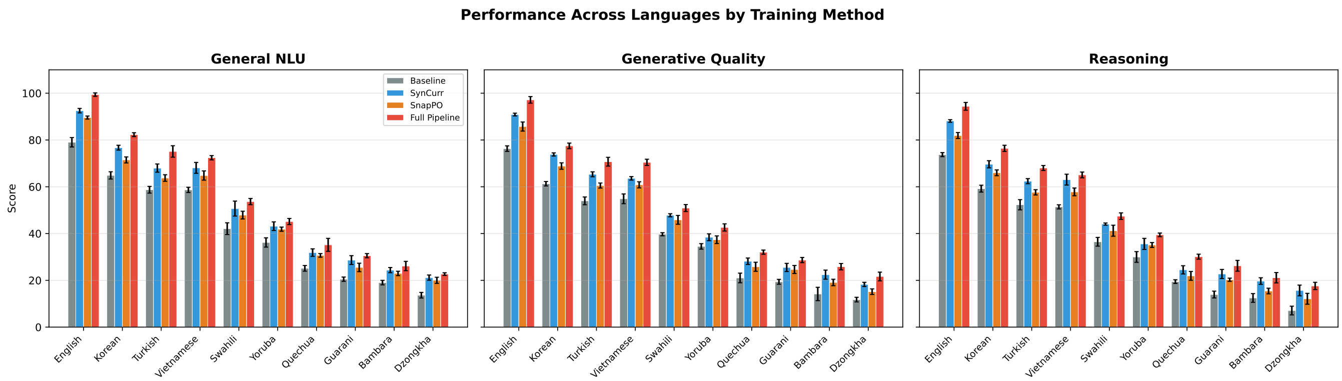

Figure 3: Per-language performance across three benchmarks for all four training configurations. Performance decreases from left to right as resource availability decreases.

Implications for Low-Resource Language Development. For languages in the Low and Very-Low tiers, the Full Pipeline still provides meaningful improvements of 10.04 and 10.01 points respectively, but these gains may be insufficient to achieve practically useful performance levels. Bambara and Dzongkha achieve Full Pipeline scores of only 26.12 and 22.68 on General NLU—well below the threshold typically associated with reliable NLU capability.

Recommendations. Based on our findings, we recommend three adaptations for applying the Solar methodology to lower-resource languages: (1) Developing specialized synthetic data generators that can produce higher-quality output for low-resource languages, potentially through targeted cross-lingual transfer. (2) Augmenting the curriculum with curated human-validated data at key quality transition points. (3) Exploring cross-lingual preference model sharing for the SnapPO component, leveraging the relative robustness we observe for RL-based gains.

6 CONCLUSION

We have empirically evaluated the applicability of the Solar training methodology—combining synthetic data curriculum and SnapPO reinforcement learning—to languages spanning five resource tiers. Our findings demonstrate that while the methodology provides statistically significant gains across all tested languages, its effectiveness diminishes substantially for lower-resource languages. The Full Pipeline delivers 20.61 points of improvement for High-resource languages but only 10.01 points for Very-Low-resource languages, with synthetic data quality identified as the primary bottleneck. These results highlight the need for targeted adaptations before this methodology can effectively serve the world's most underrepresented languages.

REFERENCES

[1] David Ifeoluwa Adelani, Graham Neubig, Sebastian Ruder, et al. 2022. MasakhaNER 2.0: Africa-centric Transfer Learning for Named Entity Recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 4488–4508.

[2] Alexis Conneau, Kartik Khandelwal, Naman Goyal, et al. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 8440–8451.

[3] William Fedus, Barret Zoph, and Noam Shazeer. 2022. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. *Journal of Machine Learning Research* 23, 120 (2022), 1–39.

[4] Pratik Joshi, Sebastin Santy, Amar Buber, et al. 2020. The State and Fate of Linguistic Diversity and Inclusion in the NLP World. *arXiv preprint arXiv:2004.09095* (2020).

[5] Long Ouyang, Jeffrey Wu, Xu Jiang, et al. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, Vol. 35. 27730–27744.

[6] Chanjun Park, Seonghwan Kim, Sanghoon Lee, et al. 2026. Solar Open Technical Report. *arXiv preprint arXiv:2601.07022* (2026).

[7] Colin Raffel, Noam Shazeer, Adam Roberts, et al. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21, 140 (2020), 1–67.

[8] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. In *arXiv preprint arXiv:1707.06347*.

[9] Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, et al. 2023. Self-Instruct: Aligning Language Models with Self-Generated Instructions. *arXiv preprint arXiv:2212.10560* (2023).

[10] Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer Learning for Low-Resource Neural Machine Translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 1568–1575.