PIQA-Yorùbá: A Dataset of Physical Commonsense Reasoning Questions in Yorùbá

Abstract

This paper introduces **PIQA-Yorùbá**, a dataset of 120 question–answer pairs for evaluating physical commonsense reasoning in the Yorùbá language. Each item follows the *Physical Interaction: Question Answering (PIQA)* style, where the system must choose between two candidate solutions to a physically grounded problem, one of which aligns with real-world commonsense knowledge. I translated and culturally adapted existing PIQA-style prompts into Yorùbá. This dataset is intended to support natural language understanding, reasoning, and AI evaluation in one of West Africa's most widely spoken languages.

1. Introduction

Commonsense reasoning is a critical ability for natural language systems. The PIQA benchmark (Bisk et al., 2020) demonstrated the value of testing models on physical commonsense: reasoning about everyday interactions with the world. However, most resources remain in English. To advance inclusivity and linguistic diversity in AI, I present a Yorùbá version of a PIQA-style dataset.

Yorùbá is spoken by more than 40 million people worldwide, primarily in Nigeria, Benin, Togo, and the diaspora. Developing datasets in Yorùbá expands opportunities for research in reasoning, multilingual NLP, and culturally relevant AI systems.

2. Dataset Description

The dataset consists of **120 multiple-choice questions**. Each question describes a physical commonsense scenario with two options. Only one option is correct, reflecting everyday reasoning.

- **Language**: Yorùbá (native orthography, including tone marks where appropriate).
- **Domains**: cooking, clothing, farming, weather, transportation, religion, household practices, festivals.
- **Format**: Each row contains (i) the Yorùbá question prompt, (ii) option A and option B (embedded in the text), and (iii) the correct answer.

Example:

- Question (Yorùbá): “Nígbà tí o bá ń din àkàrà, ṣé epo yẹ **kí ó jin tó kí ó lè bò gbogbo àdùn** tàbí kí ó màà jin ju ko le bo ìsàlẹ àdùn nìkan?”
- Answer: “**kí ó jin tó kí ó lè bò gbogbo àdùn**”

3. Methodology

The dataset was **manually constructed** in two stages:

1. **Seed Questions**: I started with PIQA-style English questions about physical commonsense. These were either directly translated or culturally localized. For instance, “keep soup warm” was adapted to Yorùbá contexts such as *amala*, *ewedu*, *suya*, etc.
2. **Translation to Yorùbá**: Each question and answer was translated into Yorùbá by a fluent speaker, ensuring naturalness and correctness. Care was taken to preserve Yorùbá idiomatic forms. For culturally unique contexts (e.g., *agbada*, *iro and buba*, *emu*), questions were created directly in Yorùbá rather than translated.

No automatic translation was used. All translations and adaptations were **manual**.

4. Cultural Relevance

Unlike a direct translation of English PIQA, this dataset embeds **Yorùbá life and material culture**:

- Food (*amala, ewedu, garri, moin-moin, puff-puff, suya*).
- Clothing (*aso-ebi, agbada, fila, adire*).
- Household practices (*firewood cooking, water storage in clay pots*).
- Farming (*cassava, maize, palm kernels*).
- Festivals and religion (*talking drum, masquerade, kolanut, Yorùbá weddings*).

This grounding ensures that questions are both **commonsense and familiar** to Yorùbá speakers.

5. Applications

- **Multilingual NLP Benchmarks**: Evaluate reasoning in Yorùbá.
- **Commonsense AI**: Measure transfer of commonsense knowledge into low-resource languages.

- **Educational Tools**: Develop Yorùbá learning applications with reasoning tasks.
- **Cultural Preservation**: Capture everyday practices in digital form.

6. Connection to Zabbot

This dataset also supports the mission of Zabbot, a Yorùbá-first language learning and cultural preservation platform. By curating commonsense reasoning questions in Yorùbá, PIQA-Yorùbá contributes to Zabbot's vision of building AI-driven tools that not only teach vocabulary and grammar, but also immerse learners in practical, everyday reasoning rooted in Yorùbá life. In the long term, such datasets can strengthen Zabbot's conversational agents, improve pronunciation feedback, and enrich culturally authentic learning experiences.

7. Limitations

- **Size**: The dataset currently contains 120 phrases, which is larger than the requested minimum of 100 phrases.
- **Manual Scope**: All translations are handcrafted; human bias in phrasing may remain.
- **Orthography**: Tone marks were applied in many cases, but consistency may vary across items.

8. Conclusion

PIQA-Yorùbá provides a foundation for reasoning in a major African language. By aligning physical commonsense tasks with Yorùbá culture and everyday life, it broadens the scope of commonsense AI beyond English and contributes toward linguistic inclusivity.

Reference

- Bisk, Y., Zellers, R., Gao, J., & Choi, Y. (2020). **PIQA: Reasoning about Physical Commonsense in Natural Language.** Proceedings of AAAI. (https://arxiv.org/abs/1911.11641)