

EdoQA: A Question-Answering Dataset for Edo

Asher Egerton Idehen¹, Osarugue Erhomonsele², Osaigbovo Evbuomwan³

¹Independent Researcher ²Independent Researcher ³University of Benin

Correspondence: asheregerton1@gmail.com, erhomonseleosarugue@gmail.com, osaigbovo.evbuomwan@uniben.edu

Abstract

We present EdoQA, a question answering dataset for Edo, a Niger-Congo language spoken by about two million people in Edo State, Nigeria. Edo is also called Benin, and has historically been called Bini in English writing; we use Edo throughout this paper, following current guidance from the Benin Traditional Council. The dataset follows the ECLeKTic format adopted by the MRL 2026 Shared Task and contains 260 short-answer questions about Edo and wider Nigerian history, tradition, art, and food, each paired with an evidence link and an English translation. As far as we are aware, EdoQA is the first question answering dataset built specifically for Edo. We release EdoQA to support future evaluation and NLP work on Edo and other Edoid languages.

1 Introduction

The MRL 2026 Shared Task asks contributors to build a community-made, ECLeKTic-style benchmark (Goldman et al., 2025) for as many languages as possible, with a particular focus on low-resource languages and dialects.¹ Each item in this format is a short factual question with a single well-defined answer, an evidence link that supports the answer, and an English translation of both the question and the answer. The goal is to measure how well language models can answer questions that a person from a given language community would know, but that most people outside that community would not.

In this paper we contribute EdoQA, a 260-item dataset of this kind for Edo. We are not aware of any earlier resource of this kind for the language, and very little Edo text of any kind is available online. We describe how we built the dataset, the choices we made about spelling and cultural register, and the results of testing two current language models on it.

¹<https://mrlbenchmarks.github.io/>

2 Background

2.1 The Edo Language

The language in this dataset is Edo, written with a sub-dot under the E. It is a Niger-Congo language of the Edoid branch, spoken by roughly two million people (Eberhard et al., 2019), mostly in Edo State in southern Nigeria and centred on Benin City. It carries the ISO 639-3 code *bin*, based on the language’s older English name. Edo was historically the court and administrative language of the Benin Kingdom, and much of the vocabulary that comes up in everyday questions about Edo culture, titles, ceremonies, and deities still traces back to that history.

Edo is also called Benin, and has historically also been called Bini in English-language writing. Our linguists told us that the Benin Traditional Council has recently asked people to move away from the term Bini, so we use Edo throughout this paper and in the dataset name.

The language has three lexical tones and seven oral vowels, five of which also have nasal counterparts (Elugbe, 1989; Amayo and Elugbe, 1983). Every word ends in a vowel, syllables are simple, and several consonant sequences are written as single digraphs, for example *gb*, *kp*, and *gh*. Tone is phonemic, but in everyday writing it is rarely marked, and we followed that convention in our own dataset (Section 3).

2.2 Edo in NLP

Edo is a low-resource language by any computational measure. It has no Wikipedia edition of its own, and the amount of Edo text available on the open web is small. The only prior computational work on the Edoid language family that we are aware of is Orife (2020), who trained baseline neural machine translation models for four Edoid languages, Èdó, Ésán, Urhobo, and Isoko, using the JW300 corpus. To our knowledge, no question an-

swering or knowledge benchmark dataset has previously been built for Edo specifically.

3 Culturally Informed Design Choices

We wrote every question and answer the way a native speaker would actually say it, rather than in an artificially formal or fully spelled-out register. This section describes three decisions that shaped the dataset.

Vowel elision. Every word in Edo ends in a vowel, and a vowel is usually dropped at the boundary between two words when the second word also begins with a vowel, a process that has been described in detail for the language (Omozuwa, 1989); for example *yo esuku* is usually said and written *y'esuku*. Table 1 shows three ways of writing a phrase with an elision of this kind, between *gbe* and *aro*: spelling every vowel out in full, marking the elision with an apostrophe, or eliding it silently with no mark at all. We chose the apostrophe form throughout the dataset. It reads the way the phrase is actually spoken, and it still lets a reader or a future annotator recover the underlying word, which the fully silent form does not.

Option	Form	Used
Full spacing	gbe aro ghee mwe	
Apostrophe	gb'aro ghee mwe	✓
Silent elision	gbaro ghee mwe	

Table 1: Three ways of writing the same elided phrase, and the option used in EdoQA.

Orthography. Edo does not have one single, universally agreed spelling system, and several orthographies have been proposed over the years (Amayo and Elugbe, 1983). We based our spelling on the widely used system set out in Agheyisi (1986), summarised in Table 2. As noted above, we did not mark tone in the dataset itself, since tone marking is used inconsistently even among fluent Edo readers and adding it would have made every item slower for our reviewing linguist to verify. We recorded, informally, which vowels, consonants, and digraphs each item used, and confirmed that the full inventory below appears at least once across the dataset.

Cultural register. Some direct English phrasing is not appropriate once translated into Edo. The clearest example in our dataset concerns Benin kings: it is not acceptable to say that an *Oba*

Vowels (7 oral; 5 nasalise → 12)	
Oral	i, e, ɛ, a, ɔ, o, u
Nasal	in, ɛn, an, ɔn, un
Consonants (27; digraphs are single units)	
Plain	b, d, f, g, h, k, l, m, n, p, r, s, t, v, w, y, z
Digraphs	gb, kp, kw, gh, kh, vb, mw, rr, rh, ny
Tone	
Marking	á high, à low, a mid (unmarked), ǎ rising, ǎ falling

Table 2: The Edo orthographic inventory we worked from, after Agheyisi (1986). Tone marking exists as a convention but was not applied in EdoQA.

“died,” and the respectful phrasing instead is that he “joined the ancestors.” We applied this kind of substitution wherever a question or answer required it, while keeping the more literal wording in the English translation, so that each language version stays faithful to its own conventions rather than being a word-for-word copy of the other.

4 The EdoQA Dataset

4.1 Construction

Each item started as a question and answer written in English by one team member, together with an evidence link and a country or region tag. A native-speaker linguist then translated the question and answer into Edo, and a second native-speaker linguist reviewed that translation for accuracy and confirmed that the evidence link actually supported the answer.

For 30 items (about 12% of the dataset), the reviewing linguist also recorded an alternate Edo rendering of the question or answer; the English translation stays fixed and only the Edo differs, usually by a word or two where either phrasing is normal in context. We kept both versions in the notes rather than forcing a single choice, so that a difference of this kind is not mistaken later on.

4.2 Statistics

The dataset contains 260 items. Table 3 shows the share of items tagged with each of three regional coverage levels: Edo culture, meaning facts specific to the Edo people or the Benin Kingdom; Nigerian culture, meaning facts about Nigeria more broadly; and a small regional culture category covering related but non-Edo groups in the south of the country. Table 4 shows the same items

broken down by topic, and Table 5 gives basic size statistics for the dataset as a whole. Edo items run slightly longer than their English translations, at a length ratio of about 1.05, and Edo also has a larger vocabulary than English despite a similar total word count, 1,230 unique Edo words against 1,116 unique English words, which likely reflects Edo’s richer verb and pronoun morphology.

Regional coverage	Items	Share
Edo culture	190	73.1%
Nigerian culture	63	24.2%
Regional culture	7	2.7%
Total	260	100%

Table 3: Share of EdoQA items by regional coverage.

Topic	Items	Share
History	90	34.6%
Traditions	71	27.3%
Art	49	18.8%
Food	34	13.1%
Other	16	6.2%
Total	260	100%

Table 4: Share of EdoQA items by topic.

Statistic	Value
Total items	260
English words (total / vocab)	3,611 / 1,116
Edo words (total / vocab)	3,701 / 1,230
Mean English words / item	13.89
Mean Edo words / item	14.23
Sentence length ratio (edo:en)	1.05
Total evidence links	260
Unique evidence links	94

Table 5: Basic statistics for EdoQA, computed over the curated corpus.

Table 6 shows two complete items from the dataset, one about food and one about a yearly festival, to illustrate the format.

4.3 Evidence Sourcing

Nearly all of our evidence links are English-language pages. Edo has no Wikipedia edition, and Edo-language pages on the open web are rare enough that building a 260-item dataset from Edo sources alone was not realistic within our timeline. Our 260 items draw on only 94 unique links, about 2.8 items per link on average, since a handful of pages on Benin Kingdom history and Edo culture

were detailed enough to support several questions each.

This sourcing pattern shapes what kinds of questions we could write. A fact only makes it into the dataset if someone, at some point, wrote it down somewhere findable online, in any language, and in practice that means facts about Benin Kingdom history, well-known festivals, and named historical figures are much easier to source than everyday cultural knowledge that has simply never been written down anywhere. We think this is a major reason history is the largest single topic in Table 4: the Benin Kingdom’s history is comparatively well documented in English, partly because of its connection to the Benin Bronzes and its colonial-era history, while more mundane, niche cultural knowledge was consistently the harder category to find a citable source for.

5 Analysis

5.1 Setup

We tested two instruction-tuned models on the full dataset in a zero-shot, closed-book setting: Gemma 4 31B IT, an open model, and Gemini 3.6 Flash, a closed model added for comparison. Each question was asked once in Edo and once using the English translation, with a single fixed instruction to answer briefly, and a temperature of 0. GPT-4o judged each answer as correct or wrong against the gold answer. We report a lenient accuracy, the share of answers judged correct regardless of language, and a stricter accuracy that also requires the answer to be in the language it was asked in.

5.2 Results

Table 7 gives the overall accuracy for each model in each language. Both models score far lower in Edo than in English: under the stricter measure, by about 41 points for Gemini and 40 points for Gemma.

Table 8 and Table 9 break this down by topic and region, using the stricter accuracy throughout. In English, both models do noticeably better on Nigerian-culture questions than on Edo-specific ones, and the gap is far larger for Gemma (29.5 versus 77.8 percent) than for Gemini (78.4 versus 93.7 percent), consistent with Edo-specific knowledge being less represented in English text than Nigerian culture more broadly, and smaller models being more exposed to that gap.

Èdo question	English translation	Gold answer	Topic
D’uwõnmwẹ ogh’Èdo n’ọ kii vb’a gha rh’ikanmwu yọ, n’a ya rri oghẹde?	Which traditional Èdo soup thickened with potash is served with plantain?	Owo soup	Food
Vb’ẹghẹ n’ẹde, inu ikpẹde a ya ru Igue?	How many days does the Igue festival traditionally span?	Seven days	Traditions

Table 6: Two example items from ÈdoQA.

Model	Èdo		English	Gap
	Strict	Lenient		
Gemini 3.6 Flash	41.5%	46.9%	82.3%	40.8
Gemma 4 31B IT	3.0%	4.2%	42.7%	39.6

Table 7: Overall accuracy by model and language under the lenient and strict measures, and the gap in percentage points between the strict Èdo score and English.

6 Conclusion

We built ÈdoQA, a 260-item question answering dataset for Èdo, following the ECLeKTic format used in the MRL 2026 Shared Task. As far as we can tell, it is the first dataset of this kind for the language. We documented the spelling, elision, and cultural register choices behind it, and tested two current language models on it, finding a large drop in accuracy when the same questions are asked in Èdo rather than English, especially for facts specific to Èdo culture. We hope ÈdoQA is useful both as an evaluation set and as a small, carefully checked source of written Èdo text for future work on the language.

Acknowledgments

We thank the MRL 2026 Shared Task organizers for coordinating this effort.

References

- Rebecca N. Agheyisi. 1986. *An Edo-English Dictionary*. Ethiope Publishing Corporation, Benin City, Nigeria.
- Airen Amayo and Ben Elugbe. 1983. Edo orthography. In Kay Williamson, editor, *Orthographies of Nigerian Languages, Manual II*, pages 1–11. National Language Centre, Federal Ministry of Education, Lagos, Nigeria.
- David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2019. *Ethnologue: Languages of the World*, 22nd edition. SIL International, Dallas, Texas.
- Ben Ohiomamhe Elugbe. 1989. *Comparative Edoid:*

Phonology and Lexicon. Delta Series 6. University of Port Harcourt Press, Port Harcourt, Nigeria.

Omer Goldman, Uri Shaham, Dan Malkin, Sivan Eiger, Avinatan Hassidim, Yossi Matias, Joshua Maynez, Adi Mayrav Gilady, Jason Riesa, Shruti Rijhwani, Laura Rimell, Idan Szpektor, Reut Tsarfaty, and Matan Eyal. 2025. ECLeKTic: a novel challenge set for evaluation of cross-lingual knowledge transfer. *arXiv preprint arXiv:2502.21228*.

Victor E. Omozuwa. 1989. Speech tempo, consonant deletion and tones in Èdo nouns. *Studies in African Linguistics*, 20(3):317–337.

Iroro Orife. 2020. Towards neural machine translation for edoid languages. In *AfricaNLP Workshop, 8th International Conference on Learning Representations (ICLR)*. ArXiv:2003.10704.

A Full Stratified Results

Table 10 collects the topic and regional-coverage breakdowns discussed in Section 5 into a single table.

Topic	n	Gemini 3.6 Flash		Gemma 4 31B IT	
		Èdo	English	Èdo	English
History	90	43.3%	85.6%	1.1%	40.0%
Traditions	71	49.3%	74.6%	5.6%	23.9%
Art	49	38.8%	85.7%	2.0%	57.1%
Food	34	38.2%	79.4%	2.9%	64.7%
Other	16	12.5%	93.8%	6.2%	50.0%

Table 8: Strict accuracy by topic, model, and language.

Region	n	Gemini 3.6 Flash		Gemma 4 31B IT	
		Èdo	English	Èdo	English
Èdo culture	190	42.6%	78.4%	3.2%	29.5%
Nigerian culture	63	38.1%	93.7%	3.2%	77.8%
Regional culture [†]	7	42.9%	85.7%	0.0%	85.7%

Table 9: Strict accuracy by regional coverage, model, and language. [†] fewer than 15 items; too small to read as a stable rate.

Stratum	n	Gemini 3.6 Flash		Gemma 4 31B IT	
		Èdo	English	Èdo	English
<i>By topic</i>					
History	90	43.3%	85.6%	1.1%	40.0%
Traditions	71	49.3%	74.6%	5.6%	23.9%
Art	49	38.8%	85.7%	2.0%	57.1%
Food	34	38.2%	79.4%	2.9%	64.7%
Other	16	12.5%	93.8%	6.2%	50.0%
<i>By regional coverage</i>					
Èdo culture	190	42.6%	78.4%	3.2%	29.5%
Nigerian culture	63	38.1%	93.7%	3.2%	77.8%
Regional culture [†]	7	42.9%	85.7%	0.0%	85.7%

Table 10: Full stratified results: strict accuracy by model, language, and stratum. [†] fewer than 15 items; too small to read as a stable rate.