NAIJARC: A MULTI-CHOICE READING COMPREHENSION DATASET FOR NIGERIAN LANGUAGES

Anuoluwapo Aremu*, **Jesujoba Oluwadara Alabi**[1*], **Daud Abolade**[2*], **Nkechinyere F. Aguobi**[2]
Shamsuddeen Hassan Muhammad[3*], **David Ifeoluwa Adelani**[4*]
*Masakhane NLP [1]Saarland University [2]University of Lagos [3]Imperial College London
[4]University College London
aremuadeola97@gmail.com, d.adelani@ucl.ac.uk

Introduction Reading Comprehension (RC) requires the ability to read a text and demonstrate understanding by answering questions about it. This requires high-level reasoning, an understanding of natural language, and general real-world based knowledge for machines to deliver high-performance results. Developing AI models capable of understanding long paragraphs and answering questions is a major challenge, especially for under-resourced languages. With the quest to improve the machine performance of reading comprehension, there have been several efforts towards curation and acquisition of datasets (Bajgar et al., 2016; Lai et al., 2017; Kočiský et al., 2018; Xuan-Quy et al., 2023). However, many of these works are either in English or do not include any African languages.

With a special lens on African languages, there has recently been a concentrated effort towards building datasets for NLP downstream tasks such as named entity recognition Eiselen (2016); Adelani et al. (2021; 2022b), part-of-speech Nivre et al. (2016); Dione et al. (2023), news topic classification Niyongabo et al. (2020); Hedderich et al. (2020); Alabi et al. (2022); Adelani et al. (2023), and machine translation Reid et al. (2021); NLLB-Team et al. (2022); Adelani et al. (2022a). However, little has been done on question answering and reading comprehension tasks. The closest dataset to ours is the AfriQA (Ogundepo et al., 2023) dataset for open-retrieval question answering, with a primary focus on retrieving correct answers that are answerable on Wikipedia. Recently, Bebebele (Bandarkar et al., 2023a)—a multi-choice RC dataset covering 122 languages was released, however, the questions and passages were translated from English dataset. To the best of our knowledge, there is no *human-annotated* multi-choice reading comprehension dataset for African languages.

In this paper, we create **NaijaRC**—a new multi-choice **N**igerian **R**eading **C**omprehension dataset that is based on high-school RC examination for three Nigerian national languages: Hausa (hau), Igbo (ibo), and Yorùbá (yor). We provide baseline results by performing cross-lingual transfer using the Belebele training data which is majorly from RACE[1] dataset (Lai et al., 2017) based on several pre-trained encoder-only models. Additionally, we provide results by prompting large language models (LLMs) like GPT-4. We provide the code and data on GitHub [2]

Nigerian languages covered The three Nigerian languages covered in the NaijaRC make use of Latin-based script and are tonal. Igbo and Yorùbá additionally makes use of diacritics on their letters like ẹ, ẹ̀, ẹ́, and ụ. Hausa additional makes use of special letters like ɓ, ɗ, ƙ, and ƴ. All the languages have more than 35 million native speakers according to Ethnologue (Eberhard et al., 2020) and are offered in high-school as a compulsory subject in the different regions where they are spoken: Hausa (North), Igbo (South-East), and Yorùbá (South West / North Central).

NaijaRC dataset collection We source past questions from the Hausa, Igbo, and Yorùbá language subjects of the West African Senior School Certificate Examination (WASSCE). The subjects are optionally taken by final-year high school students in south-western Nigeria. Prior to 2017, it used to be compulsory to take this subject. We only found printed versions of the past questions, thus, we had to make use of Optical Character Recognition tools like Google Lens and Apple's OCR to convert the printed format into a digitised version. Due to the diacritics used in the language, the OCR process has some quality issues, which we had to fix manually.

[1]RACE is based on English exams for middle and high school Chinese students, very similar to our dataset.
[2]https://github.com/AremuAdeolaJr/NaijaRC

	Belebele	NaijaRC		
	eng	hau	ibo	yor
TRAIN	67,541	-	-	50
DEV	3,773	-	-	25
TEST	–	88	88	171

(a) Data split for Belebele and NaijaRC

Model	Accuracy				
	eng	hau	ibo	yor	AVG
AfroXLMR-base	67.3	43.2	42.0	34.0	46.6
Serengeti	62.5	50.0	44.3	33.0	47.4
OFA-768	64.5	43.2	48.9	28.3	46.2
GPT-3.5 Turbo	66.3	35.2	30.7	27.7	40.0
GPT-4 Turbo	82.9	48.9	40.9	33.0	51.4

(b) Model accuracy. PLMs fine-tuned on Belebele's train split.

Table 1: **Data split and benchmark results** for Belebele (on DEV set) and NaijaRC (on test set).

The West African Examination Council questions test many aspects of students' competence. However, with regard to the focus of our research, only the comprehension passages and questions were extracted. It is important to note that the number of questions generated for each passage varies from year to year. However, on average, there were five questions for each of the two passages annually.

Our dataset contains comprehension **passages**, **questions**, and **answers** which were carefully validated, verified, and cleaned by native speakers and linguists of the languages. We removed questions relating to the semantics of specific italics or underlined words in the context of usage, as most language models cannot handle them. Table 1(a) provides the data statistics and split for the different language. In Appendix, we also, provide an example of a passage, question and multi-choice answers for Yorùbá and the translation to English.

Experimental setup In this work, we evaluated the performance of fine-tuned pre-trained encoder-only models, specifically AfroXLMR-base Alabi et al. (2022), Serengeti (Adebara et al., 2023) and OFA-768 (Liu et al., 2023), as well as two large language models (LLMs), namely GPT-3.5 and GPT-4. For the encoder based PLMs, we evaluated their performance in a zero-shot cross-lingual setting where we fine-tuned the model on the training split of Belebele [3] (Bandarkar et al., 2023b) dataset for 3 epochs and evaluated the performance of the resulting model on NaijaRC. Furthermore, we prompted GPT-3.5 and GPT-4 providing them with the question, options, and context paragraph. The output of these models are then mapped back to the option labels for evaluation. All these models are evaluated using accuracy as the metric.

Experimental results The experimental results are presented in Table 1(b) including the models' performance on the Belebele's development set and the zero-shot transfer performance on NaijaRC test sets. The result shows that GPT-4 achieved the highest overall accuracy across the four languages, with an accuracy of 51.4%. This score can be attributed to its exceptional performance on the Belebele development set, where it achieved an accuracy of 82.9%. While there is a possibility that the model benefited from pretraining on the exact data in the development, it is crucial to note that this is not a scenario of "winner takes it all". The zero-shot cross-lingual transfer result with PLMs show that different models performed best in specific languages. AfroXLMR-base has highest accuracy on Yorùbá (34.0%), OFA-768 outperformed others on Igbo (48.9%), and Serengeti outperformed others on Hausa (50.0%). Therefore, when excluding English from the average calculation, Serengeti emerged with the highest overall performance. However, GPT-3.5 and GPT-4 showed relatively lower performance compared to multilingual PLMs on the Nigerian languages, with GPT-4 emerging as the superior performer between the two.

Conclusion and Future work In this paper, we presented **NaijaRC**—a new reading comprehension dataset for Nigerian languages. We compared the zero-shot cross-lingual transfer with three PLMs: AfroXLMR-base, Serengeti, and OFA-768 from an English dataset. We also prompted GPT-3.5 and GPT.4, and showed that these models obtained lower performance on NaijaRC dataset when compared to the PLMs.

As future work, we plan to extend our evaluation to few-shot settings. Specifically, we will explore different approaches that can utilize a few examples, like 50 (passage, question, answer)-tuples for the effective adaptation of existing reading comprehension models.

[3] the training data is a combination of several multiple choice datasets and can be retrieved here https://github.com/facebookresearch/belebele

ACKNOWLEDGMENT

David Adelani acknowledges the support of DeepMind Academic Fellowship programme. We are grateful to OpenAI for providing API credits through their Researcher Access API programme to Masakhane for the evaluation of GPT-3.5 and GPT-4 large language models. Jesujoba Alabi was partially supported by the BMBF's (German Federal Ministry of Education and Research) SLIK project under the grant 01IS22015C.

REFERENCES

Ife Adebara, AbdelRahim Elmadany, Muhammad Abdul-Mageed, and Alcides Alcoba Inciarte. SERENGETI: Massively multilingual language models for Africa. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 1498–1537, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-acl.97. URL https://aclanthology.org/2023.findings-acl.97.

David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen Muhammad, Guyo Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ajibade, Tunde Ajayi, Yvonne Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. A few thousand translations go a long way! leveraging pre-trained models for African news translation. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (eds.), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3053–3070, Seattle, United States, July 2022a. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.223. URL https://aclanthology.org/2022.naacl-main.223.

David Adelani, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Constantine Lignos, Jesujoba Alabi, Shamsuddeen Muhammad, Peter Nabende, Cheikh M. Bamba Dione, Andiswa Bukula, Rooweither Mabuya, Bonaventure F. P. Dossou, Blessing Sibanda, Happy Buzaaba, Jonathan Mukiibi, Godson Kalipe, Derguene Mbaye, Amelia Taylor, Fatoumata Kabore, Chris Chinenye Emezue, Anuoluwapo Aremu, Perez Ogayo, Catherine Gitau, Edwin Munkoh-Buabeng, Victoire Memdjokam Koagne, Allahsera Auguste Tapo, Tebogo Macucwa, Vukosi Marivate, Mboning Tchiaze Elvis, Tajuddeen Gwadabe, Tosin Adewumi, Orevaoghene Ahia, Joyce Nakatumba-Nabende, Neo Lerato Mokono, Ignatius Ezeani, Chiamaka Chukwuneke, Mofetoluwa Oluwaseun Adeyemi, Gilles Quentin Hacheme, Idris Abdulmumin, Odunayo Ogundepo, Oreen Yousuf, Tatiana Moteu, and Dietrich Klakow. MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 4488–4508, Abu Dhabi, United Arab Emirates, December 2022b. Association for Computational Linguistics. URL https://aclanthology.org/2022.emnlp-main.298.

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen H. Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba Alabi, Seid Muhie Yimam, Tajuddeen Rabiu Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verrah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwuneke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane MBOUP, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima DIOP, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. MasakhaNER: Named Entity Recognition for African Languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131, 10 2021. ISSN 2307-387X. doi: 10.1162/tacl_a_00416. URL https://doi.org/10.1162/tacl_a_00416.

David Ifeoluwa Adelani, Marek Masiak, Israel Abebe Azime, Jesujoba Oluwadara Alabi, Atnafu Lambebo Tonja, Christine Mwase, Odunayo Ogundepo, Bonaventure F. P. Dossou, Akintunde Oladipo, Doreen Nixdorf, Chris Chinenye Emezue, Sana Al-Azzawi, Blessing K. Sibanda, Davis David, Lolwethu Ndolela, Jonathan Mukiibi, Tunde Oluwaseyi Ajayi, Tatiana Moteu Ngoli, Brian Odhiambo, Abraham Toluwase Owodunni, Nnaemeka C. Obiefuna, Shamsuddeen Hassan Muhammad, Saheed Salahudeen Abdullahi, Mesay Gemeda Yigezu, Tajuddeen Rabiu Gwadabe, Idris Abdulmumin, Mahlet Taye Bame, Oluwabusayo Olufunke Awoyomi, Iyanuoluwa Shode, Tolulope Anu Adelani, Habiba Abdulganiy Kailani, Abdul-Hakeem Omotayo, Adetola Adeeko, Afolabi Abeeb, Anuoluwapo Aremu, Olanrewaju Samuel, Clemencia Siro, Wangari Kimotho, Onyekachi Raphael Ogbu, Chinedu E. Mbonu, Chiamaka I. Chukwuneke, Samuel Fanijo, Jessica Ojo, Oyinkansola F. Awosan, Tadesse Kebede Guge, Sakayo Toadoum Sari, Pamela Nyatsine, Freedmore Sidume, Oreen Yousuf, Mardiyyah Oduwole, Ussen Abre Kimanuka, Kanda Patrick Tshinu, Thina Diko, Siyanda Nxakama, Abdulmejid Tuni Johar, Sinodos Gebre, Muhidin Mohamed, S. A. Mohamed, Fuad Mire Hassan, Moges Ahmed Mehamed, Evrard Ngabire, and Pontus Stenetorp. Masakhanews: News topic classification for african languages. 2023.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. In *Proceedings of the 29th International Conference on Computational Linguistics*, pp. 4336–4349, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics. URL https://aclanthology.org/2022.coling-1.382.

Ondrej Bajgar, Rudolf Kadlec, and Jan Kleindienst. Embracing data abundance: Booktest dataset for reading comprehension. *ArXiv*, abs/1610.00956, 2016.

Lucas Bandarkar, Davis Liang, Benjamin Muller, Mikel Artetxe, Satya Narayan Shukla, Don Husa, Naman Goyal, Abhinandan Krishnan, Luke Zettlemoyer, and Madian Khabsa. The belebele benchmark: a parallel reading comprehension dataset in 122 language variants. *ArXiv*, abs/2308.16884, 2023a. URL https://api.semanticscholar.org/CorpusID:261394577.

Lucas Bandarkar, Davis Liang, Benjamin Muller, Mikel Artetxe, Satya Narayan Shukla, Donald Husa, Naman Goyal, Abhinandan Krishnan, Luke Zettlemoyer, and Madian Khabsa. The belebele benchmark: a parallel reading comprehension dataset in 122 language variants. *arXiv preprint arXiv:2308.16884*, 2023b.

Cheikh M. Bamba Dione, David Adelani, Peter Nabende, Jesujoba Alabi, Thapelo Sindane, Happy Buzaaba, Shamsuddeen Hassan Muhammad, Chris Chinenye Emezue, Perez Ogayo, Anuoluwapo Aremu, Catherine Gitau, Derguene Mbaye, Jonathan Mukiibi, Blessing Sibanda, Bonaventure F. P. Dossou, Andiswa Bukula, Rooweither Mabuya, Allahsera Auguste Tapo, Edwin Munkoh-Buabeng, victoire Memdjokam Koagne, Fatoumata Ouoba Kabore, Amelia Taylor, Godson Kalipe, Tebogo Macucwa, Vukosi Marivate, Tajuddeen Gwadabe, Mboning Tchiaze Elvis, Ikechukwu Onyenwe, Gratien Atindogbe, Tolulope Adelani, Idris Akinade, Olanrewaju Samuel, Marien Nahimana, Th'eogene Musabeyezu, Emile Niyomutabazi, Ester Chimhenga, Kudzai Gotosa, Patrick Mizha, Apelete Agbolo, Seydou Traore, Chinedu Uchechukwu, Aliyu Yusuf, Muhammad Abdullahi, and Dietrich Klakow. Masakhapos: Part-of-speech tagging for typologically diverse african languages. 2023.

David M. Eberhard, Gary F. Simons, and Charles D. Fennig (eds.). Ethnologue: Languages of the world. twenty-third edition., 2020. URL http://www.ethnologue.com.

Roald Eiselen. Government domain named entity recognition for South African languages. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 3344–3348, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA). URL https://aclanthology.org/L16-1533.

Michael A. Hedderich, David Adelani, Dawei Zhu, Jesujoba Alabi, Udia Markus, and Dietrich Klakow. Transfer learning and distant supervision for multilingual transformer models: A study on African languages. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 2580–2591, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.204. URL https://aclanthology.org/2020.emnlp-main.204.

Tomáš Kočiský, Jonathan Schwarz, Phil Blunsom, Chris Dyer, Karl Moritz Hermann, Gábor Melis, and Edward Grefenstette. The NarrativeQA reading comprehension challenge. *Transactions of the Association for Computational Linguistics*, 6:317–328, 2018. doi: 10.1162/tacl_a_00023. URL https://aclanthology.org/Q18-1023.

Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. RACE: Large-scale ReAding comprehension dataset from examinations. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel (eds.), *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 785–794, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/D17-1082. URL https://aclanthology.org/D17-1082.

Yihong Liu, Peiqin Lin, Mingyang Wang, and Hinrich Schütze. Ofa: A framework of initializing unseen subword embeddings for efficient large-scale multilingual continued pretraining, 2023.

Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 1659–1666, 2016.

Rubungo Andre Niyongabo, Qu Hong, Julia Kreutzer, and Li Huang. KINNEWS and KIRNEWS: Benchmarking cross-lingual text classification for Kinyarwanda and Kirundi. In Donia Scott, Nuria Bel, and Chengqing Zong (eds.), *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 5507–5521, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics. doi: 10.18653/v1/2020.coling-main.480. URL https://aclanthology.org/2020.coling-main.480.

NLLB-Team, Marta Ruiz Costa-jussà, James Cross, Onur cCelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Alison Youngblood, Bapi Akula, Loïc Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon L. Spruit, C. Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzm'an, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. No language left behind: Scaling human-centered machine translation. *ArXiv*, abs/2207.04672, 2022.

Odunayo Ogundepo, Tajuddeen R. Gwadabe, Clara E. Rivera, Jonathan H. Clark, Sebastian Ruder, David Ifeoluwa Adelani, Bonaventure F. P. Dossou, Abdou Aziz DIOP, Claytone Sikasote, Gilles Hacheme, Happy Buzaaba, Ignatius Ezeani, Rooweither Mabuya, Salomey Osei, Chris Emezue, Albert Njoroge Kahira, Shamsuddeen H. Muhammad, Akintunde Oladipo, Abraham Toluwase Owodunni, Atnafu Lambebo Tonja, Iyanuoluwa Shode, Akari Asai, Tunde Oluwaseyi Ajayi, Clemencia Siro, Steven Arthur, Mofetoluwa Adeyemi, Orevaoghene Ahia, Aremu Anuoluwapo, Oyinkansola Awosan, Chiamaka Chukwuneke, Bernard Opoku, Awokoya Ayodele, Verrah Otiende, Christine Mwase, Boyd Sinkala, Andre Niyongabo Rubungo, Daniel A. Ajisafe, Emeka Felix Onwuegbuzia, Habib Mbow, Emile Niyomutabazi, Eunice Mukonde, Falalu Ibrahim Lawan, Ibrahim Said Ahmad, Jesujoba O. Alabi, Martin Namukombo, Mbonu Chinedu, Mofya Phiri, Neo Putini, Ndumiso Mngoma, Priscilla A. Amuok, Ruqayya Nasir Iro, and Sonia Adhiambo. Afriqa: Cross-lingual open-retrieval question answering for african languages, 2023.

"Ọmọ bíbí ìlú Béyìíòṣe ni Ìṣọ̀lá. Òun àti àbúrò rẹ̀ Fọláhànmí, nìkan ni Bádéjọ, bàbá wọ́n bí. Àgbẹ̀ oníkòkó aládàáńlá ni Bádéjọ. Bọ́látitó, aya rẹ̀ sì jẹ́ oníṣòwò obì. Bádéjọ kò kàwé ṣùgbọ́n ó pinnu láti tọ́ àwọn ọmọ rẹ̀ méjéèjì débi tí wọ́n bá lè kàwé dé láyé, nítorí pé ìya àìkàwé jẹ ẹ́ púpọ̀ nídi òwò tí ó ń ṣe.

Lẹ́yìn tí Ìṣọ̀lá parí ìwé mẹ́wàá ní ìlú Ìbòdì ni ó gba ìlú Arómisá lọ láti tẹ̀ síwáiú nínú ẹ̀kọ́ rẹ̀ ní fásitì. Ìlú ọbá ni ó sì ti gba oyè ìmọ̀ ìjìnlẹ̀ kejì nínú òfin. Ìṣọ̀lá padà sílúu Béyìíròṣe, ó sì di gbajúgbajà agbẹjọ́rò káàkiri agbègbè náà."

Ilú abínibí Bádéjọ ni

A. Ìbòdì
B. Arómisá
C. Béyìíòṣe
D. Ìlú-ọba

Ìṣọ̀lá is a native of Béyìíòṣe town. He and his younger brother Fọláhànmí, were the only children of their father, Bádéjọ. Bádéjọ is a commercial cocoa farmer. His wife, Bọ́látitó trades kolanuts. Bádéjọ has no formal western education but he was determined to sponsor his two children, as a result of the difficulties he encountered in his business.
After ìṣọ̀lá graduated secondary school at Ìbòdì town, he left for Arómisá to continue his university education. He graduated with a masters in Law degree from abroad. Ìṣọ̀lá returned to Béyìíòṣe, where he became a notable lawyer.

Bádéjọ is a native of which town

A. Ìbòdì
B. Arómisá
C. Béyìíòṣe
D. Ìlú-Ọba

Figure 1: An example of a passage in Yorùbá, a question and corresponding options (A-D). Where C is the correct option. We provide an expert translation in English.

Machel Reid, Junjie Hu, Graham Neubig, and Yutaka Matsuo. AfroMT: Pretraining strategies and reproducible benchmarks for translation of 8 African languages. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 1306–1320, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.99. URL `https://aclanthology.org/2021.emnlp-main.99`.

Dao Xuan-Quy, Le Ngoc-Bich, Vo The-Duy, Phan Xuan-Dung, Ngo Bac-Bien, Nguyen Van-Tien, Nguyen Thi-My-Thanh, and Nguyen Hong-Phuoc. Vnhsge: Vietnamese high school graduation examination dataset for large language models. *arXiv preprint arXiv:2305.12199*, 2023.

A Appendix

Example of the passage in Yorùbá