H PPY BUZ B

Postdoctoral Researcher   Princeton University

Email:hbuzaaba@gmail.com

RESEARCH INTEREST

My current research focuses on multilingual NLP for low-resource languages, especially African languages. In the past, I spent time investigating the use of natural-gradient Bayesian methods to improve uncertainty estimation in pre-trained language models and how to make them robust and adaptable with a keen interest in low-resource setting.

EDUCATION

Ph.D Computer Science, University of Tsukuba   April 2018 - March 2022
Supervisor: Prof. Toshiyuki Amagasa.

Master of Regional Development, Toyo University   Sept 2015 – Sept 2017
Supervisor: Prof. Nakabasami Chieko.

Bachelor of Applied Physics, UR College of Science and Technology (Former KIST)   Jan 2008 - Mar 2012
Thesis Supervisor: Prof. Pogrebnoi Alexander.

WORK EXPERIENCE

Postdoctoral Research Associate   From September 2023
Princeton University   *Princeton NJ, USA*
Postdoctoral Researcher, Approximate Bayesian Inference Team   From June 2022 - August 2023
RIKEN Center for Advanced Intelligence Project   *Tokyo, Japan*
Ph.D Student, Systems and Information Engineering   April 2018 - March 2022
University of Tsukuba   *Tsukuba, Japan*

GRANTS

Culturally Relevant Benchmark Datasets for African Languages   Jan 2025 - June 2026
(Co-PI with Christiane Felbaum, USD 38,300 funded by Princeton AI LAB)

Universal Dependency Treebanks for low-resource African languages   Feb 2024 - Sept 2025
(Co-PI with Christiane Felbaum, USD 41,780 funded by Princeton Language + Intelligence)

African Languages in the Age of AI Speaker Series   Dec 2023
(Co-Organized with Princeton CDH, USD 32,000 funded by Princeton Africa World Initiative)

RESEARCH INTERESTS

Multilingual Natural Language Processing (NLP)
NLP for low-resource languages
Large language model adaptation

TEACHING EXPERIENCE

FRS159 Teaching Computers to Understand African Languages   Fall2024
Course Instructor Princeton University   *Princeton, NJ*

Data Engineering 1 for master's students   Oct – Dec 2019
Teaching Assistant University of Tsukuba   *Tsukuba, Japan*

Database systems for undergraduate students   Oct – Dec 2018
Teaching Assistant University of Tsukuba   *Tsukuba, Japan*

AWARDS   SCHOLARSHIP

Merit Award   2019
Students award for best presenter at the 11th Forum on Data Engineering and Information Management DEIM2019

Mastuo Award   2017
An award for best master thesis of Regional Development Studies, by Toyo University

Abe-Initiative Scholarship   2015
A fully funded masters scholarship awarded to young Africans by the Japanese government

INVITED TALKS

Invited talk: Princeton envision AI conference   Feb 2025
State of Language Technology for low-resource African Languages

Lecture: Library of the Chathams New Jersey Makers day   Mar 2025
State of Technology for African Languages

Invited Talk: MBAZA NLP in preparation for the Global AI summit on Africa   April 2025
Lugha-Llama: Adapting Large Language Models for African Languages

VOLUNTEER SERVICE

Member of MLA Task Force on World Languages and Generative AI	2025
Co-organizing the ACL2025 workshop on AfricaNLP	2025
Co-organizing a panel at ALTA Conference	2025
Co-organizing the ICLR2024 workshop on AfricaNLP	2024
Co-organizing the ICLR2023 workshop on AfricaNLP	2023

Reviewing

AfricaNLP, ACL, ICML	2025
LREC-COLING, AfricaNLP	2024
NeurIPS	2023
NeurIPS, ACML	2022
ICLR	2021

LIST OF PUBLICATIONS

1. **H. Buzaaba**, A. Wettig, D. Adelani C. Fellbaum. Lugha-Llama: Adapting Large Language Models for African Languages *pre-print 2025*.

2. H. Yu, J. Ojo, A. Bukula, J. Zhuang, E. Lee, T. Guge, I. Azime, **H. Buzaaba**, B. Sibanda, G. Kalipe, J. Mukiibi, S. Kabongo, M. Setaka, L. Ndolela, N. Odu, R. Mabuya, J. Zhuang, S. Muhammad, S. Osei, J. Alabi, X. He, M. Ochieng, S. Hooker, S. Samb, J. Murage, D. Adelani. INJONGO: A Multicultural Intent Detection and Slot- lling Dataset for 16 African Languages. *In the 63rd Annual Meeting of the Association for Computational Linguistics ACL2025.*

3. D. Adelani, J. Ojo, I. Azime, J. Zhuang, J. Alabi, X. He, M. Ochieng, S. Hooker, A. Bukula, E. Lee, C. Chukwuneke, **H. Buzaaba**, B. Sibanda, G. Kalipe, J. Mukiibi, S. Kabongo, F. Yuehgoh, M. Setaka, L. Ndolela, N. Odu, R. Mabuya, S. Muhammad, S. Osei, S. Samb, T. Guge, T. Sherman, P. Stenetorp. IrokoBench: A New Benchmark for African Languages in the Age of Large Language Models *In the international Conference of the Nations of Americas Chapter of the Association for Computational Linguistics* **Out Standing Award at NAACL2025**.

4. U. Inyanga, T. Olatoye, **H. Buzaaba**, N. Ermolaev., Computational Approaches to Nigerian Literature: Analyzing Texts in Yoruba and E k. *Alliance of Digital Humanities Organizations international professional/disciplinary conference for 2024 ADHO2024).*

5. O. Ogundepo, T. Gwadabe, C. Rivera, J. Clark, S. Ruder, D. Adelani, B. Dossou, A. DIOP, C. Sikasote, G. Hacheme, **H. Buzaaba**, I. Ezeani, R. Mabuya, S. Osei, C. Emezue, A. Kahira, S. Muhammad, A. Oladipo, A. Owodunni, A. Tonja, I. Shode, A. Asai, T. Ajayi, C. Siro, S. Arthur, M. Adeyemi, O. Ahia, A. Anuoluwapo, O. Awosan, C. Chukwuneke, B. Opoku, A. Ayodele, V. Otiende, C. Mwase, B. Sinkala, A. Rubungo, D. Ajisafe, E. Onwuegbuzia, H. Mbow, E. Niyomutabazi, E. Mukonde, F. Lawan, I. Ahmad, J. Alabi, M. Namukombo, M. Chinedu, M. Phiri, N. Putini, N. Mngoma, P. Amuok, R. Iro, S. Adhiambo. AfriQA: Cross-lingual Open-Retrieval Question Answering for African Languages. *In the International Conference on Empirical Methods in Natural Language Processing EMNLP2023).*

6. C. Bamba, D. Adelani, P. Nabende, J. Alabi, T. Sindane, **H. Buzaaba**, S. Muhammad, C. Emezue, P. Ogayo, A. Aremu, C. Gitau, D. Mbaye, J. Mukiibi, B. Sibanda, B. F. P. Dossou, A. Bukula, R. Mabuya, A. Tapo, E. Munkoh-Buabeng, V. Koagne, F. Kabore, A. Taylor, G. Kalipe, T. Macucwa, V. Marivate, T. Gwadabe, E. Mboning, I. Onyenwe, G. Atindogbe, T. Adelani, I. Akinade, O. Samuel, M. Nahimana, T. Musabeyezu, E. Niyomutabazi, E. Chimhenga, K. Gotosa, P. Mizha, A. Agbolo, S. Traore, C. Uchechukwu, A. M. Abdullahi, D. Klakow. MasakhaPOS: Part-of-Speech Tagging for Typologically Diverse African languages. *In the 61st Annual Meeting of the Association for Computational Linguistics ACL2023).*

7. T. Adewumi, M. Adeyemi, A. Anuoluwapo, B. Peters, **H. Buzaaba**, O. Samuel, A. M. Rufai, B. Ajibade, T. Gwadabe, M. K. Traore, T. Ajayi, S. Muhammad, A. Baruwa, P. Owoicho, T. Ogunremi, P. Ngigi, O. Ahia, R. Nasir, F. Liwicki, M. Liwicki. AfriWOZ: Corpus for Exploiting Cross-Lingual Transferability for Generation of Dialogues in Low-Resource, African Languages. *In the International Joint Conference on Neural Networks IJCNN 2023).*

8. D. Adelani, G. Neubig, S. Ruder, S. Rijhwani, M. Beukman, C. Palen-Michel, C. Lignos, J. Alabi, S. Muhammad, P. Nabende, C. M. B. Dione, A. Bukula, R. Mabuya, B. F. P. Dossou, B. Sibanda, **H. Buzaaba**, J. Mukiibi, G. Kalipe, D. Mbaye, A. Taylor, F. Kabore, C. C. Emezue, A. Aremu, P. Ogayo, C. Gitau, E. Munkoh-Buabeng, V. M. Koagne, A. A. Tapo, T. Macucwa, V. Marivate, M. T. Elvis, T. Gwadabe, T. Adewumi, O. Ahia, J. Nakatumba-Nabende, N. L. Mokono, I. Ezeani, C. Chukwuneke, M. O. Adeyemi, G. Q. Hacheme, I. Abdulmumin, O. Ogundepo, O. Yousuf, T. Moteu, D. 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, EMNLP), 2022.*

9. D. Adelani, J. Alabi, A. Fan, J. Kreutzer, X. Shen, M. Reid, D. Ruiter, D. Klakow, P. Nabende, E. Chang, T. Gwadabe, F. Sackey, B. F. P. Dossou, C. Emezue, C. Leong, M. Beukman, S. Muhammad, G. Jarso, O. Yousuf, A. N. Rubungo, G. Hacheme, E. P. Wairagala, M. U. Nasir, B. Ajibade, T. Ajayi, Y. Gitau, J. Abbott, M. Ahmed, M. Ochieng, A. Aremu, P. Ogayo, J. Mukiibi, F. O. Kabore, G. Kalipe, D. Mbaye, A. A. Tapo, V. M. Koagne, E. Munkoh-Buabeng, V. Wagner, I. Abdulmumin, A. Awokoya, **H. Buzaaba**, B. Sibanda, A. Bukula, S. Manthalu. A Few Thousand Translations Go a Long Way! Leveraging Pre-trained Models for African News Translation. *In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics NAACL), 2022.*

10. **H. Buzaaba**, T. Amagasa. A scheme for efficient question answering with low dimension reconstructed embeddings. *In The 23rd international conference on information integration and web intelligence iiWAS) 2021.*

11. D. I. Adelani, J. Abbott, G. Neubig, D. D'souza, J. Kreutzer, C. Lignos, C. Palen-Michel, **H. Buzaaba**, S. Rijhwani, S. Ruder, S. Mayhew, I. A. Azime, S. H. Muhammad, C. C. Emezue, J. Nakatumba-Nabende, P. Ogayo, A. Anuoluwapo, C. Gitau, D. Mbaye, J. Alabi, S. M. Yimam, T. R. Gwadabe, I. Ezeani, R. A. Niyongabo, J. Mukiibi, V. Otiende, I. Orife, D. David, S. Ngom, T. Adewumi, P. Rayson, M. Adeyemi, G. Muriuki, E. Anebi, C. Chukwuneke, N. Odu, E. P. Wairagala, S. Oyerinde, C. Siro, T. S. Bateesa, T. Oloyede, Y. Wambui, V. Akinode, D. Nabagereka, M. Katusiime, A. Awokoya, M. MBOUP, D. Gebreyohannes, H. Tilaye, K. Nwaike, D. Wolde, A. Faye, B. Sibanda, O. Ahia, B. F. P. Dossou, K. Ogueji, T. I. DIOP, A. Diallo, A. Akinfaderin, T. Marengereke, S. Osei. MasakhaNER: Named Entity Recognition for African Languages. *Transactions of the Association for Computational Linguistics 2021) 9: 1116–1131.*

12. **H. Buzaaba**, T. Amagasa. Question Answering Over Knowledge Base: A Scheme for Integrating Subject and the Identified Relation to Answer Simple Questions. *SN COMPUT. SCI. 2, 25 2021).*

13. **H. Buzaaba**, T. Amagasa. A modular approach for efficient simple question answering over knowledge base.