

richness and diversity of the Myanmar language. Based on the results presented in Table 13, it can be concluded that utilizing character or sub-syllable units in Myanmar is preferable to using syllables for transliteration. The reason is that Myanmar syllables have limited processing capability when dealing with exceptional structures, whereas sub-syllables and characters provide more flexibility.

8. Conclusion and Future Works

Based on our findings, an internal dictionary of Myanmar-English NE terminology that incorporates transliterations for both local and western languages have been developed. The experiments involved utilizing Transformer-based NN techniques to assess their efficacy on the data we prepared. Despite the limited scope of the NE corpus utilized, our neural network models were able to achieve satisfactory outcomes for transliteration duties. It is anticipated that as more data is gathered and further testing is conducted, neural network-based transliteration models will become increasingly relevant in this domain. This investigation represents the initial exploration of utilizing neural networks for Myanmar NE transliteration. In the future, developing an interactive and user-friendly transliteration tool or application that incorporates the Transformer model can facilitate real-world applications and assist users in seamlessly transcribing Myanmar named entities into English, potentially aiding in fields like language translation, information retrieval, and natural language processing.

Acknowledgments

I would like to express my appreciation to my dedicated supervisor for their continuous guidance, motivation, and unwavering support throughout the course of this study. Their extensive knowledge and innovative thinking have been indispensable pillars of support throughout my work.

Conflicts of Interest

The authors state no conflict of interest

References

- [1] Bradley D. "Burmic languages in Myanmar". Continuum of richness of languages and dialects in Myanmar. 2015:167-90.
- [2] Chang, C.B., 2009. English loanword adaptation in Burmese. *Journal of the Southeast Asian Linguistics Society*, 1, pp.77-94.
- [3] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
- [4] Ding, C., Pa, W.P., Utiyama, M. and Sumita, E., 2018. Burmese (Myanmar) name romanization: A sub-syllabic segmentation scheme for statistical solutions. In *Computational Linguistics: 15th International Conference of the Pacific Association for Computational Linguistics, PACLING 2017, Yangon, Myanmar, August 16–18, 2017, Revised Selected Papers 15* (pp. 191-202). Springer Singapore.
- [5] Ding C. Transliteration of Foreign Words in Burmese. arXiv preprint arXiv:2110.03163. 2021 Oct 7.
- [6] Ding, C., Utiyama, M. and Sumita, E., 2018. NOVA: A feasible and flexible annotation system for joint tokenization and part-of-speech tagging. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 18(2), pp.1-18.
- [7] Ding, C., Aye, H.T.Z., Pa, W.P., Nwet, K.T., Soe, K.M., Utiyama, M. and Sumita, E., 2019. Towards Burmese (Myanmar) morphological analysis: Syllable-based tokenization and part-of-speech tagging. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(1), pp.1-34.
- [8] Ding, C., Yee, S.S.S., Pa, W.P., Soe, K.M., Utiyama, M. and Sumita, E., 2020. A Burmese (Myanmar) treebank: Guideline and analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(3), pp.1-13.
- [9] Hosken M, TunTunLwin M. "Representing Myanmar in Unicode". Unicode Technical Note. 2012;13:1-67.
- [10] Koehn, P., Och, F.J. and Marcu, D., 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 127-133).
- [11] Koehn, P., Federico, M., Shen, W., Bertoldi, N., Bojar, O., Callison-Burch, C., Cowan, B., Dyer, C., Hoang, H., Zens, R. and Constantin, A., 2007, August. Open source toolkit for statistical machine translation: Factored translation models and confusion network decoding. In *CLSP Summer Workshop Final Report WS-2006*, Johns Hopkins University.
- [12] Klein G, Hernandez F, Nguyen V, Senellart J. The OpenNMT neural machine translation toolkit: 2020 edition. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track) 2020 Oct* (pp. 102-109).
- [13] Klein G, Kim Y, Deng Y, Senellart J, Rush AM. "Openmt: Open-source toolkit for neural machine translation". arXiv preprint arXiv:1701.02810. 2017 Jan 10.
- [14] Kumai H, Sagawa H, Morimoto Y. "NTCIR-7 Patent Translation Experiments at Hitachi". In *NTCIR 2008*.
- [15] Klakow D, Peters J. "Testing the correlation of word error rate and perplexity". *Speech Communication*. 2002 Sep 1;38(1-2):19-28.
- [16] Li, H., Kumaran, A., Pervouchine, V. and Zhang, M., 2009, August. Report of NEWS 2009 machine transliteration shared task. In *Proceedings of the 2009 Named Entities Workshop: Shared Task on Transliteration (NEWS 2009)* (pp. 1-18).
- [17] Liu, L., Utiyama, M., Finch, A. and Sumita, E., 2016, June. Agreement on target-bidirectional neural machine translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 411-416).
- [18] Merhav Y, Ash S. Design challenges in named entity transliteration. arXiv preprint arXiv:1808.02563. 2018 Aug 7.
- [19] Mon, A.M. and Soe, K.M., 2020, November. Phrase-based named entity transliteration on Myanmar-English terminology dictionary. In *2020 23rd conference of the oriental COCOSDA international committee for the co-ordination and standardisation of speech databases and assessment techniques (O-COCOSDA)* (pp. 38-43). IEEE.

- [20] Naing, H.M.S., Thu, Y.K., Pa, W.P., Kato, H., Finch, A., Sumita, E. and Hori, C., 2015. Rule Based Katakana to Myanmar Transliteration for Post-editing Machine Translation. In Proceedings of the Annual Conference of the Language Processing Society of Japan (pp. 257-260).
- [21] Och, F.J. and Ney, H., 2003. A systematic comparison of various statistical alignment models. Computational linguistics, 29(1), pp.19-51.
- [22] Och, F.J., 2003, July. Minimum error rate training in statistical machine translation. In Proceedings of the 41st annual meeting of the Association for Computational Linguistics (pp. 160-167).
- [23] ShweSin, Y.M., Soe, K.M. and Htwe, K.Y., 2018, October. Large scale Myanmar to English neural machine translation system. In 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE) (pp. 464-465). IEEE.
- [24] Su, Yuanhang, and C-C. Jay Kuo. "On extended long short-term memory and dependent bidirectional recurrent neural network." *Neurocomputing* 356 (2019): 151-161.
- [25] Su, Yuanhang, and C-C. Jay Kuo. "Recurrent neural networks and their memory behavior: a survey." *APSIPA Transactions on Signal and Information Processing* 11, no. 1 (2022).
- [26] Su, Yuanhang, Yuzhong Huang, and C-C. Jay Kuo. "Dependent bidirectional RNN with extended-long short-term memory." (2018).
- [27] Su, Yuanhang, Kai Fan, Nguyen Bach, C-C. Jay Kuo, and Fei Huang. "Unsupervised multi-modal neural machine translation." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10482-10491. 2019.
- [28] Thu, Y. K., Finch, A., Sagisaka, Y., & Sumita, E. (2013). "A study of myanmar word segmentation schemes for statistical machine translation" (Doctoral dissertation, MERAL Portal).
- [29] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Authors Profile



In 2012, **Aye Myat Mon** earned her Master of Computer Science (M.C.Sc.) degree with credits from the University of Computer Studies, Patheingyi, Myanmar. She is currently pursuing her Ph.D. at the University of Computer Studies, Yangon, Myanmar, and is an active member of the Natural Language Processing and Speech Processing Lab at UCSY. Her research interests encompass Natural Language Processing, Machine Learning, and Deep Learning. Aye Myat Mon has also been involved in the Asia Language Treebank (ALT) projects, a collaborative endeavor between the National Institute of Information and Communications Technology (NICT), Japan, and UCSY. In addition, she completed an internship at the Advanced Translation Technology Laboratory (ASTREC), Universal Communication Research Institute, NICT, Kyoto, Japan, from April 2019 to March 2020.



Dr. Khin Mar Soe completed her Ph.D. in Information Technology in the year 2005. Currently, she holds the position of a professor and serves as the Head of the Natural Language Processing Lab within the Faculty of Computer Science at the University of Computer Studies, Yangon. She actively supervises Master's theses and Ph.D. research in the field of Natural Language Processing. Additionally, Dr. Khin Mar Soe has contributed to projects such as ASEAN MT, a machine translation initiative for South East Asian languages, and the Asia Language Treebank (ALT) project, which involved research collaboration between NICT, Japan, and UCSY.