

References

- Al-Shedivat, M., & Parikh, A. P. (2019). Consistency by agreement in zero-shot neural machine translation. *ArXiv Preprint ArXiv:1904.02338*.
- Arivazhagan, N., Bapna, A., Firat, O., Lepikhin, D., Johnson, M., Krikun, M., Chen, M. X., Cao, Y., Foster, G., & Cherry, C. (2019). Massively multilingual neural machine translation in the wild: Findings and challenges. *ArXiv Preprint ArXiv:1907.05019*.
- Attia, M., Samih, Y., Elkahky, A., Mubarak, H., Abdelali, A., & Darwish, K. (2019). *POS tagging for improving code-switching identification in Arabic*. 18–29.
- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *ArXiv Preprint ArXiv:1409.0473*.
- Bali, K., Sharma, J., Choudhury, M., & Vyas, Y. (2014). “i am borrowing ya mixing?” an analysis of english-hindi code mixing in facebook. *Proceedings of the First Workshop on Computational Approaches to Code Switching*, 116–126.
- Barnett, R., Codó, E., Eppler, E., Forcadell, M., Gardner-Chloros, P., Van Hout, R., Moyer, M., Torras, M. C., Turell, M. T., & Sebba, M. (2000). The LIDES Coding Manual: A Document for Preparing and Analyzing Language Interaction Data Version 1.1—July 1999. *International Journal of Bilingualism*, 4(2), 131–271.
- Carrera, J., Beregovaya, O., & Yanishevsky, A. (2009). Machine translation for cross-language social media. *PROMT Americas Inc*.
- Chanda, A., Das, D., & Mazumdar, C. (2016). Unraveling the English-Bengali code-mixing phenomenon. *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, 80–89.
- Chandu, K. R., Chinnakotla, M., Black, A. W., & Shrivastava, M. (2017). WebShodh: A Code Mixed Factoid Question Answering System for Web. In G. J. F. Jones, S. Lawless, J. Gonzalo, L. Kelly, L. Goeriot, T. Mandl, L. Cappellato, & N. Ferro (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction* (Vol. 10456, pp. 104–111). Springer International Publishing. https://doi.org/10.1007/978-3-319-65813-1_9
- Chen, S. F., & Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, 13(4), 359–394.
- Cheng, Y. (2019). Semi-supervised learning for neural machine translation. In *Joint training for neural machine translation* (pp. 25–40). Springer.

- Chittaranjan, G., Vyas, Y., Bali, K., & Choudhury, M. (2014). *Word-level language identification using crf: Code-switching shared task report of msr india system*. 73–79.
- Choudhury, M., Saraf, R., Jain, V., Mukherjee, A., Sarkar, S., & Basu, A. (2007). Investigation and modeling of the structure of texting language. *International Journal of Document Analysis and Recognition (IJ DAR)*, 10(3), 157–174.
- Cook, P., & Stevenson, S. (2009). *An unsupervised model for text message normalization*. 71–78.
- Davies, E. E., & Bentahila, A. (2007). CAROL MYERS-SCOTTON, Contact linguistics: Bilingual encounters and grammatical outcomes. *Language in Society*, 36(03). <https://doi.org/10.1017/S0047404507070285>
- Dhar, M., Kumar, V., & Shrivastava, M. (2018). Enabling Code-Mixed Translation: Parallel Corpus Creation and MT Augmentation Approach. *Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing*, 131–140. <https://aclanthology.org/W18-3817>
- Dutta, S., Saha, T., Banerjee, S., & Naskar, S. K. (2015). Text normalization in code-mixed social media text. *2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, 378–382. <https://doi.org/10.1109/ReTIS.2015.7232908>
- Firat, O., Sankaran, B., Al-Onaizan, Y., Vural, F. T. Y., & Cho, K. (2016). Zero-resource translation with multi-lingual neural machine translation. *ArXiv Preprint ArXiv:1606.04164*.
- Gambäck, B., & Das, A. (2016). *Comparing the level of code-switching in corpora*. 1850–1855.
- Gu, J., Wang, Y., Cho, K., & Li, V. O. (2019a). Improved zero-shot neural machine translation via ignoring spurious correlations. *ArXiv Preprint ArXiv:1906.01181*.
- Gu, J., Wang, Y., Cho, K., & Li, V. O. (2019b). Improved zero-shot neural machine translation via ignoring spurious correlations. *ArXiv Preprint ArXiv:1906.01181*.
- Gulcehre, C., Firat, O., Xu, K., Cho, K., Barrault, L., Lin, H.-C., Bougares, F., Schwenk, H., & Bengio, Y. (2015). On using monolingual corpora in neural machine translation. *ArXiv Preprint ArXiv:1503.03535*.
- Guzman, G. A., Serigos, J., Bullock, B., & Toribio, A. J. (2016). *Simple tools for exploring variation in code-switching for linguists*. 12–20.
- Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., & Corrado, G. (2017). Google’s multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5, 339–351.

- Kachru, B. B. (1986). The power and politics of English. *World Englishes*, 5(2–3), 121–140. <https://doi.org/10.1111/j.1467-971X.1986.tb00720.x>
- Kalchbrenner, N., & Blunsom, P. (2013). *Recurrent continuous translation models*. 1700–1709.
- Kugathasan, A., & Sumathipala, S. (2020). Standardizing Sinhala Code-Mixed Text using Dictionary based Approach. *2020 International Conference on Image Processing and Robotics (ICIP)*, 1–6. <https://doi.org/10.1109/ICIP48927.2020.9367353>
- Lafferty, J., McCallum, A., & Pereira, F. C. (2001). *Conditional random fields: Probabilistic models for segmenting and labeling sequence data*.
- Lakew, S. M., Cettolo, M., & Federico, M. (2018). A comparison of transformer and recurrent neural networks on multilingual neural machine translation. *ArXiv Preprint ArXiv:1806.06957*.
- Lignos, C., & Marcus, M. (2013). *Toward web-scale analysis of codeswitching*. 90.
- Lui, M., & Baldwin, T. (2012). *Langid. Py: An off-the-shelf language identification tool*. 25–30.
- Mandal, S., Das, S. D., & Das, D. (2018). *Language Identification of Bengali-English Code-Mixed data using Character & Phonetic based LSTM Models* (arXiv:1803.03859). arXiv. <http://arxiv.org/abs/1803.03859>
- Masoud, M., Torregrosa, D., Buitelaar, P., & Arčan, M. (2019). *Back-translation approach for code-switching machine translation: A case study*. 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. M. (1999). *Statistical modeling*. Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.
- Nguyen, D., & Doğruöz, A. S. (2013). *Word level language identification in online multilingual communication*. 857–862.
- Papalexakis, E., Nguyen, D., & Doğruöz, A. S. (2014). *Predicting code-switching in multilingual communication for immigrant communities*. 42–50.
- Rijhwani, S., Sequiera, R., Choudhury, M. C., & Bali, K. (2016). *Translating codemixed tweets: A language detection based system*. 81–82.
- Senaratne, C. D. (2009). *Sinhala-English code-mixing in Sri Lanka: A sociolinguistic study*. LOT.

Sennrich, R., Haddow, B., & Birch, A. (2015). Improving neural machine translation models with monolingual data. *ArXiv Preprint ArXiv:1511.06709*.

Shanmugalingam, K., Sumathipala, S., & Premachandra, C. (2018). *Word level language identification of code mixing text in social media using nlp*. 1–5.

Singh, R., Choudhary, N., & Shrivastava, M. (2018). *Automatic Normalization of Word Variations in Code-Mixed Social Media Text*.

<https://doi.org/10.48550/ARXIV.1804.00804>

Sudsawad, P. (2007). *Knowledge translation: Introduction to models, strategies and measures*. Southwest Educational Development Laboratory

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27.

Tan, X., Chen, J., He, D., Xia, Y., Qin, T., & Liu, T.-Y. (2019). Multilingual neural machine translation with language clustering. *ArXiv Preprint ArXiv:1908.09324*.

Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53(8), 5929–5955. <https://doi.org/10.1007/s10462-020-09838-1>

Veena, P. V., Kumar, M. A., & Soman, K. P. (2017). An effective way of word-level language identification for code-mixed facebook comments using word-embedding via character-embedding. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1552–1556. <https://doi.org/10.1109/ICACCI.2017.8126062>

Volk, M., & Clematide, S. (2014). *Detecting code-switching in a multilingual alpine heritage corpus*. 24–33.

Wong, K.-F., & Xia, Y. (2008). Normalization of Chinese chat language. *Language Resources and Evaluation*, 42(2), 219–242.

Xue, Z., Yin, D., & Davison, B. D. (2011). *Normalizing microtext*. Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence.

Yu, L.-C., He, W.-C., Chien, W.-N., & Tseng, Y.-H. (2013). Identification of code-switched sentences and words using language modeling approaches. *Mathematical Problems in Engineering*, 2013.

Zhang, B., Williams, P., Titov, I., & Sennrich, R. (2020). Improving massively multilingual neural machine translation and zero-shot translation. *ArXiv Preprint ArXiv:2004.11867*.