Appendix A

List of Publications

The work mentioned in this thesis pertains to the following peer-reviewed papers:

1. **Mundotiya, R. K.**, Singh, M. K., Kapur, R., Mishra, S., & Singh, A. K. (2021). Linguistic Resources for Bhojpuri, Magahi, and Maithili: Statistics about Them, Their Similarity Estimates, and Baselines for Three Applications. *Transactions on Asian and Low-Resource Language Information Processing*, 20(6), 1-37.

2. **Mundotiya, R. K.**, Mishra, S., & Singh, A. K. (2021). Hierarchical self attention based sequential labelling model for Bhojpuri, Maithili and Magahi languages. *Journal of King Saud University-Computer and Information Sciences.*

3. **Mundotiya, R. K.**, Mehta, A., Baruah, R., & Singh, A. K. (2021). Integration of Morphological Features and Contextual Weightage using Monotonic Chunk Attention for Part of Speech Tagging. *Journal of King Saud University-Computer and Information Sciences.*

4. **Mundotiya, R. K.**, Mehta, A., & Baruah, R. (2022). Domain Adaptation for POS Tagging with Contrastive Monotonic Chunk-wise Attention. *Neural Processing Letters.*

5. **Mundotiya, R. K.**, Kumar, V., Mehta, A., & Singh, A. K. (2020, October). Attention-based Domain Adaption Using Transfer Learning for Part-of-Speech Tagging: An Experiment on the Hindi Language. *In Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation* (pp. 471-477).

The following papers are in under revision:

1. **Mundotiya, R. K.**, Kumar, S., Chaudhary, U. C., Chauhan, S., Mishra, S., Gatla, P., & Singh, A. K. (2020). Development of a Dataset and a Deep Learning Baseline Named Entity Recognizer for Three Low Resource Languages: Bhojpuri, Maithili and Magahi. *Transactions on Asian and Low-Resource Language Information Processing.*

2. **Mundotiya, R. K.**, Nair, P. A., & Singh, A. K. (2022). Improving Performance for POS Tagging on Low Resource Languages through Meta Learning. *Computational Intelligence.*

Bibliography

[1] Agić, Ž., Hovy, D., and Søgaard, A. (2015). If all you have is a bit of the bible: Learning pos taggers for truly low-resource languages. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 268–272.

[2] Ahmad, M. T., Malik, M. K., Shahzad, K., Aslam, F., Iqbal, A., Nawaz, Z., and Bukhari, F. (2020). Named entity recognition and classification for punjabi shahmukhi. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 19(4).

[3] Akbik, A., Blythe, D., and Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. In *Proceedings of the 27th international conference on computational linguistics*, pages 1638–1649.

[4] Al-Rfou, R., Perozzi, B., and Skiena, S. (2013). Polyglot: Distributed word representations for multilingual nlp. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 183–192.

[5] Alfonseca, E. and Manandhar, S. (2002). An unsupervised method for general named entity recognition and automated concept discovery. In *Proceedings of the 1st international conference on general WordNet, Mysore, India*, pages 34–43.

[6] Ali, W., Lu, J., and Xu, Z. (2020). Siner: A large dataset for sindhi named entity recognition. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 2953–2961.

[7] Altun, Y., Tsochantaridis, I., and Hofmann, T. (2003). Hidden markov support vector machines. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 3–10.

[8] Ando, R. K. and Zhang, T. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6(Nov):1817–1853.

[9] Antony, P., Mohan, S. P., and Soman, K. (2010). Svm based part of speech tagger for malayalam. In *2010 International Conference on Recent Trends in Information, Telecommunication and Computing*, pages 339–341. IEEE.

[10] Antony, P. and Soman, K. (2010). Kernel based part of speech tagger for kannada. In *2010 International Conference on Machine Learning and Cybernetics*, volume 4, pages 2139–2144. IEEE.

[11] Antony, P. and Soman, K. (2011). Parts of speech tagging for indian languages: a literature survey. *International Journal of Computer Applications*, 34(8):0975–8887.

[12] Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*.

[13] Babych, B. and Hartley, A. (2003). Improving machine translation quality with automatic named entity recognition. In *Proceedings of the 7th International EAMT workshop on MT and other language technology tools, Improving MT through other language technology tools, Resource and tools for building MT at EACL 2003*.

[14] Bahdanau, D., Cho, K. H., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*.

[15] Balabantaray, R. C. and Jena, M. K. (2010). Experimenting oriya text chunking with divide-conquer strategy. *International Journal of Advanced Research in Computer Science*, 1(4).

[16] Basili, R., Pazienza, M. T., and Velardi, P. (1996). An empirical symbolic approach to natural language processing. *Artificial Intelligence*, 85(1):59 – 99.

[17] Behera, P. and Muzaffar, S. (2017). Approaches to named entity recognition.

[18] Behera, P., Ojha, A. K., and Jha, G. N. (2015). Issues and challenges in developing statistical pos taggers for sambalpuri. In *Language and Technology Conference*, pages 393–406. Springer.

[19] Bengio, Y., Ducharme, R., and Vincent, P. (2001). A neural probabilistic language model. In *Advances in Neural Information Processing Systems*, pages 932–938.

[20] Bhalla, D., Joshi, N., and Mathur, I. (2013). Improving the quality of mt output using novel name entity translation scheme. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1548–1553. IEEE.

[21] Bharadwaj, A., Mortensen, D. R., Dyer, C., and Carbonell, J. G. (2016). Phonologically aware neural model for named entity recognition in low resource transfer settings. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1462–1472.

[22] Bhat, R. A., Bhat, S. M., and Sharma, D. M. (2014). Towards building a kashmiri treebank: Setting up the annotation pipeline. In *LREC*, pages 748–752.

[23] Bhat, R. A., Bhatt, R., Farudi, A., Klassen, P., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., Vaidya, A., Vishnu, S. R., et al. (2017). The hindi/urdu treebank project. In *Handbook of Linguistic Annotation*, pages 659–697. Springer.

[24] Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., and Xia, F. (2009). A multi-representational and multi-layered treebank for hindi/urdu. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, pages 186–189.

[25] Biemann, C. (2006). Unsupervised part-of-speech tagging employing efficient graph clustering. In *Proceedings of the COLING/ACL 2006 Student Research Workshop*, pages 7–12.

[26] Bisazza, A. and Tump, C. (2018). The lazy encoder: A fine-grained analysis of the role of morphology in neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2871–2876, Brussels, Belgium. Association for Computational Linguistics.

[27] Blitzer, J., McDonald, R., and Pereira, F. (2006). Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 120–128.

[28] Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

[29] Botha, J. and Blunsom, P. (2014). Compositional morphology for word representations and language modelling. In *International Conference on Machine Learning*, pages 1899–1907. PMLR.

[30] Bullinaria, J. A. and Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and svd. *Behavior research methods*, 44(3):890–907.

[31] Cardenas, R., Lin, Y., Ji, H., and May, J. (2019). A grounded unsupervised universal part-of-speech tagger for low-resource languages. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2428–2439, Minneapolis, Minnesota. Association for Computational Linguistics.

[32] Caruana, R., Lawrence, S., and Giles, L. (2001). Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Advances in neural information processing systems*, pages 402–408.

[33] Casas, N., Fonollosa, J. A., Escolano, C., Basta, C., and Costa-jussà, M. R. (2019). The talp-upc machine translation systems for wmt19 news translation task: Pivoting techniques for low resource mt. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 155–162.

[34] Chakrabarty, A., Chaturvedi, A., and Garain, U. (2019). Neumorph: Neural morphological tagging for low-resource languages—an experimental study for indic languages. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(1):1–19.

[35] Chelba, C. and Acero, A. (2006). Adaptation of maximum entropy capitalizer: Little data can help a lot. *Computer Speech & Language*, 20(4):382–399.

[36] Chen, D., Fisch, A., Weston, J., and Bordes, A. (2017). Reading Wikipedia to answer open-domain questions. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1870–1879, Vancouver, Canada. Association for Computational Linguistics.

[37] Chen, D. and Manning, C. D. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 740–750.

[38] Chen, L. and Moschitti, A. (2019). Transfer learning for sequence labeling using source model and target data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6260–6267.

[39] Chiu, C. and Raffel, C. (2018). Monotonic chunkwise attention. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

[40] Chiu, J. P. and Nichols, E. (2016). Named entity recognition with bidirectional lstm-cnns. *Transactions of the association for computational linguistics*, 4:357–370.

[41] Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014a). On the properties of neural machine translation: Encoder–decoder approaches. *Syntax, Semantics and Structure in Statistical Translation*, page 103.

[42] Cho, K., van Merrienboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014b). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Moschitti, A., Pang, B., and Daelemans, W., editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1724–1734. ACL.

[43] Chomsky, N. (2002). *Syntactic structures*. Walter de Gruyter.

[44] Christianson, C., Duncan, J., and Onyshkevych, B. (2018). Overview of the darpa lorelei program. *Machine Translation*, 32(1):3–9.

[45] Collobert, R. and Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM.

[46] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *J. Mach. Learn. Res.*, 12(null):2493–2537.

[47] Conforti, C., Huck, M., and Fraser, A. (2018). Neural morphological tagging of lemma sequences for machine translation. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 39–53.

[48] Crystal, D. (2000). *Language death.* Ernst Klett Sprachen.

[49] C.S, M., RK Rao, P., and Lalitha Devi, S. (2012). Tamil NER - coping with real time challenges. In *Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages*, pages 23–38, Mumbai, India. The COLING 2012 Organizing Committee.

[50] Cucchiarelli, A. and Velardi, P. (2001). Unsupervised named entity recognition using syntactic and semantic contextual evidence. *Computational Linguistics*, 27(1):123–131.

[51] Cucerzan, S. and Yarowsky, D. (1999). Language independent named entity recognition combining morphological and contextual evidence. In *1999 joint SIGDAT conference on empirical methods in natural language processing and very large corpora.*

[52] Cui, L. and Zhang, Y. (2019). Hierarchically-refined label attention network for sequence labeling. In Inui, K., Jiang, J., Ng, V., and Wan, X., editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4113–4126. Association for Computational Linguistics.

[53] Daimary, S. K., Goyal, V., Barbora, M., and Singh, U. (2018). Development of part of speech tagger for assamese using hmm. *International Journal of Synthetic Emotions (IJSE)*, 9(1):23–32.

[54] Dalal, A., Nagaraj, K., Sawant, U., and Shelke, S. (2006). Hindi part-of-speech tagging and chunking: A maximum entropy approach. *Proceedings of the NLPAI Machine Learning Contest*, 6.

[55] Dalal, A., Nagaraj, K., Swant, U., Shelke, S., and Bhattacharyya, P. (2007). Building feature rich pos tagger for morphologically rich languages: Experience in hindi. *ICON*.

[56] Dalvi, F., Durrani, N., Sajjad, H., Belinkov, Y., and Vogel, S. (2017). Understanding and improving morphological learning in the neural machine translation decoder. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 142–151.

[57] Daumé III, H. (2007). Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263, Prague, Czech Republic. Association for Computational Linguistics.

[58] Daumé III, H., Kumar, A., and Saha, A. (2010). Frustratingly easy semi-supervised domain adaptation. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, pages 53–59. Association for Computational Linguistics.

[59] Daumé III, H. and Marcu, D. (2005). Learning as search optimization: Approximate large margin methods for structured prediction. In *Proceedings of the 22nd international conference on Machine learning*, pages 169–176.

[60] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407.

[61] Deng, S., Zhang, N., Sun, Z., Chen, J., and Chen, H. (2020). When low resource nlp meets unsupervised language model: Meta-pretraining then meta-learning for few-shot text classification (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13773–13774.

[62] Dernoncourt, F., Lee, J. Y., and Szolovits, P. (2017). Neuroner: an easy-to-use program for named-entity recognition based on neural networks. *arXiv preprint arXiv:1705.05487*.

[63] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

[64] Dhanalakshmi, V., Anandkumar, M., Vijaya, M., Loganathan, R., Soman, K., and Rajendran, S. (2008). Tamil part-of-speech tagger based on svmtool. In *Proceedings of the COLIPS International Conference on natural language processing (IALP), Chiang Mai, Thailand*, volume 2008, pages 59–64.

[65] Dou, Z., Yu, K., and Anastasopoulos, A. (2019). Investigating meta-learning algorithms for low-resource natural language understanding tasks. In Inui, K., Jiang, J., Ng, V., and Wan, X., editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 1192–1197. Association for Computational Linguistics.

[66] Dozat, T., Qi, P., and Manning, C. D. (2017). Stanford's graph-based neural dependency parser at the CoNLL 2017 shared task. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 20–30, Vancouver, Canada. Association for Computational Linguistics.

[67] Dryer, M. S. and Haspelmath, M., editors (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

[68] Du, C., Sun, H., Wang, J., Qi, Q., and Liao, J. (2020). Adversarial and domain-aware bert for cross-domain sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4019–4028.

[69] Ekbal, A. and Bandyopadhyay, S. (2008a). Development of bengali named entity tagged corpus and its use in ner systems. In *Proceedings of the 6th Workshop on Asian Language Resources*.

[70] Ekbal, A. and Bandyopadhyay, S. (2008b). Part of speech tagging in bengali using support vector machine. In *2008 International Conference on Information Technology*, pages 106–111. IEEE.

[71] Ekbal, A., Haque, R., and Bandyopadhyay, S. (2007). Bengali part of speech tagging using conditional random field. In *Proceedings of Seventh International Symposium on Natural Language Processing (SNLP2007)*, pages 131–136.

[72] Elman, J. L. (1990). Finding structure in time. *Cognitive science*, 14(2):179–211.

[73] Etzioni, O., Cafarella, M., Downey, D., Popescu, A.-M., Shaked, T., Soderland, S., Weld, D. S., and Yates, A. (2005). Unsupervised named-entity extraction from the web: An experimental study. *Artificial intelligence*, 165(1):91–134.

[74] Farkas, R. and Szarvas, G. (2006). Statistical named entity recognition for hungarian–analysis of the impact of feature space characteristics.

[75] Faruqui, M., Dodge, J., Jauhar, S. K., Dyer, C., Hovy, E., and Smith, N. A. (2015). Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1606–1615, Denver, Colorado. Association for Computational Linguistics.

[76] Feng, X., Feng, Z., Zhao, W., Qin, B., and Liu, T. (2020). Enhanced neural machine translation by joint decoding with word and pos-tagging sequences. *Mobile Networks and Applications*, 25(5):1722–1728.

[77] Ferraro, J. P., Daumé III, H., DuVall, S. L., Chapman, W. W., Harkema, H., and Haug, P. J. (2013). Improving performance of natural language processing part-of-speech tagging on clinical narratives through domain adaptation. *Journal of the American Medical Informatics Association*, 20(5):931–939.

[78] Finn, C., Abbeel, P., and Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR.

[79] Fonseca, E. R., Rosa, J. L. G., and Aluísio, S. M. (2015). Evaluating word embeddings and a revised corpus for part-of-speech tagging in portuguese. *Journal of the Brazilian Computer Society*, 21(1):1–14.

[80] Gers, F. A., Schmidhuber, J. A., and Cummins, F. A. (2000). Learning to forget: Continual prediction with lstm. *Neural Computation*, 12(10):2451–2471.

[81] Ghaddar, A. and Langlais, P. (2018). Robust lexical features for improved neural network named-entity recognition. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1896–1907.

[82] Globerson, A. and Roweis, S. (2006). Nightmare at test time: robust learning by feature deletion. In *Proceedings of the 23rd international conference on Machine learning*, pages 353–360.

[83] Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings.

[84] Goldberg, Y. and Hirst, G. (2017). *Neural Network Methods in Natural Language Processing*. Morgan & Claypool Publishers.

[85] Greenberg, J. H. et al. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, 2:73–113.

[86] Grishman, R. and Sundheim, B. M. (1996). Message understanding conference-6: A brief history. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*.

[87] Grönroos, S.-A., Virpioja, S., and Kurimo, M. (2017). Extending hybrid word-character neural machine translation with multi-task learning of morphological analysis. In *Proceedings of the Second Conference on Machine Translation*, pages 296–302, Copenhagen, Denmark. Association for Computational Linguistics.

[88] Gu, J., Hassan, H., Devlin, J., and Li, V. O. K. (2018). Universal neural machine translation for extremely low resource languages. In Walker, M. A., Ji, H., and Stent, A., editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 344–354. Association for Computational Linguistics.

[89] Gui, T., Zhang, Q., Huang, H., Peng, M., and Huang, X. (2017). Part-of-speech tagging for Twitter with adversarial neural networks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2411–2420, Copenhagen, Denmark. Association for Computational Linguistics.

[90] Guo, J., Shah, D. J., and Barzilay, R. (2018). Multi-source domain adaptation with mixture of experts. *arXiv preprint arXiv:1809.02256*.

[91] Gupta, A., Krishna, A., Goyal, P., and Hellwig, O. (2020). Evaluating neural morphological taggers for sanskrit. *SIGMORPHON 2020*, page 198.

[92] Gutmann, M. U. and Hyvärinen, A. (2012). Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *Journal of Machine Learning Research*, 13(2).

[93] Harbert, W. (2006). *The Germanic Languages*. Cambridge University Press.

[94] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

[95] Heigold, G., Neumann, G., and van Genabith, J. (2017). An extensive empirical evaluation of character-based morphological tagging for 14 languages. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 505–513.

[96] Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554.

[97] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

[98] Houdt, G. V., Mosquera, C., and Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53(8):5929–5955.

[99] Howard, J. and Ruder, S. (2018). Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 328–339.

[100] Huang, F. and Yates, A. (2009). Distributional representations for handling sparsity in supervised sequence-labeling. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 495–503. Association for Computational Linguistics.

[101] Huang, F. and Yates, A. (2010). Exploring representation-learning approaches to domain adaptation. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, pages 23–30. Association for Computational Linguistics.

[102] Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

[103] Indurthi, S. R., Chung, I., and Kim, S. (2019). Look harder: A neural machine translation model with hard attention. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3037–3043.

[104] Jagadeesh, M., Kumar, M. A., and Soman, K. (2016). Deep belief network based part-of-speech tagger for telugu language. In *Proceedings of the Second International Conference on Computer and Communication Technologies*, pages 75–84. Springer.

[105] Jain, A., Tayal, D. K., Yadav, D., and Arora, A. (2020). Research trends for named entity recognition in hindi language. In *Data Visualization and Knowledge Engineering*, pages 223–248. Springer.

[106] Ji, S., Yun, H., Yanardag, P., Matsushima, S., and Vishwanathan, S. V. N. (2016). Wordrank: Learning word embeddings via robust ranking. In Su, J., Carreras, X., and Duh, K., editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 658–668. The Association for Computational Linguistics.

[107] Joshi, S. (2019). Similar yet different languages sanskrit and german.

[108] Kann, K., Bjerva, J., Augenstein, I., Plank, B., and Søgaard, A. (2018). Character-level supervision for low-resource pos tagging. In *Proceedings of the Workshop on Deep Learning Approaches for Low-Resource NLP*, pages 1–11.

[109] Kann, K., Lacroix, O., and Søgaard, A. (2020). Weakly supervised pos taggers perform poorly on truly low-resource languages. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8066–8073.

[110] Kaur, A., Josan, G., and Kaur, J. (2009). Named entity recognition for punjabi: A conditional random field approach. In *Proceedings of 7th international conference on Natural Language ProcessingICON-09. Macmillan Publishers, India.*

[111] Kaur, A. and Josan, G. S. (2015). Evaluation of named entity features for punjabi language. *Procedia Computer Science*, 46:159 – 166. Proceedings of the International Conference on Information and Communication Technologies, ICICT 2014, 3-5 December 2014 at Bolgatty Palace & Island Resort, Kochi, India.

[112] Kaur, P., Goyal, V., Shah, K. S., and Singh, U. (2018). Hybrid chunker for gujarati language. In *Networking Communication and Data Knowledge Engineering*, pages 217–226. Springer.

[113] Kim, J.-K., Kim, Y.-B., Sarikaya, R., and Fosler-Lussier, E. (2017). Cross-lingual transfer learning for pos tagging without cross-lingual resources. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2832–2838.

[114] Kim, Y. (2014). Convolutional neural networks for sentence classification. In Moschitti, A., Pang, B., and Daelemans, W., editors, *Proceedings of the 2014 Conference*

on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1746–1751. ACL.

[115] Kim, Y., Jernite, Y., Sontag, D., and Rush, A. M. (2016). Character-aware neural language models. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[116] Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

[117] Kruengkrai, C., Uchimoto, K., Kazama, J., Wang, Y., Torisawa, K., and Isahara, H. (2009). An error-driven word-character hybrid model for joint chinese word segmentation and pos tagging. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 513–521. Association for Computational Linguistics.

[118] Kudo, T. and Matsumoto, Y. (2001). Chunking with support vector machines. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

[119] Kumar, D. and Josan, G. S. (2010). Part of speech taggers for morphologically rich indian languages: a survey. *International Journal of Computer Applications*, 6(5):32–41.

[120] Kumar, R., Lahiri, B., and Alok, D. (2012). Developing a POS tagger for Magahi: A comparative study. In *Proceedings of the 10th Workshop on Asian Language Resources*, pages 105–114, Mumbai, India. The COLING 2012 Organizing Committee.

[121] Kumar, S., Kumar, M. A., and Soman, K. (2016). Experimental analysis of malayalam pos tagger using epic framework in scala. *ARPN J. Eng. Appl. Sci*, 11.

[122] Kumar, S., Kumar, M. A., and Soman, K. (2019). Deep learning based part-of-speech tagging for malayalam twitter data (special issue: deep learning techniques for natural language processing). *Journal of Intelligent Systems*, 28(3):423–435.

[123] Kupiec, J. (1992). Robust part-of-speech tagging using a hidden markov model. *Computer speech & language*, 6(3):225–242.

[124] Lafferty, J. D., McCallum, A., and Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, page 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[125] Lake, B., Salakhutdinov, R., Gross, J., and Tenenbaum, J. (2011). One shot learning of simple visual concepts. In *Proceedings of the annual meeting of the cognitive science society*, volume 33.

[126] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California. Association for Computational Linguistics.

[127] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., and Soricut, R. (2020). ALBERT: A lite BERT for self-supervised learning of language representations. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

[128] Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR.

[129] Levy, O. and Goldberg, Y. (2014a). Linguistic regularities in sparse and explicit word representations. In *Proceedings of the eighteenth conference on computational natural language learning*, pages 171–180.

[130] Levy, O. and Goldberg, Y. (2014b). Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems*, pages 2177–2185.

[131] Lin, J. C.-W., Shao, Y., Djenouri, Y., and Yun, U. (2021). Asrnn: a recurrent neural network with an attention model for sequence labeling. *Knowledge-Based Systems*, 212:106548.

[132] Ling, W., Dyer, C., Black, A. W., Trancoso, I., Fermandez, R., Amir, S., Marujo, L., and Luís, T. (2015). Finding function in form: Compositional character models for open vocabulary word representation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1520–1530, Lisbon, Portugal. Association for Computational Linguistics.

[133] Liu, K., Chapman, W., Hwa, R., and Crowley, R. S. (2007). Heuristic sample selection to minimize reference standard training set for a part-of-speech tagger. *Journal of the American Medical Informatics Association*, 14(5):641–650.

[134] Liu, L., Shang, J., Ren, X., Xu, F., Gui, H., Peng, J., and Han, J. (2018a). Empower sequence labeling with task-aware neural language model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

[135] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

[136] Liu, Y., Zhu, Y., Che, W., Qin, B., Schneider, N., and Smith, N. A. (2018b). Parsing tweets into universal dependencies. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 965–975.

[137] Lu, B.-L., Ma, Q., Ichikawa, M., and Isahara, H. (2003). Efficient part-of-speech tagging with a min-max modular neural-network model. *Applied Intelligence*, 19(1):65–81.

[138] Luong, M.-T., Socher, R., and Manning, C. D. (2013). Better word representations with recursive neural networks for morphology. In *Proceedings of the seventeenth conference on computational natural language learning*, pages 104–113.

[139] Ma, Q., Yu, L., Tian, S., Chen, E., and Ng, W. W. (2019). Global-local mutual attention model for text classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(12):2127–2139.

[140] Ma, X. and Hovy, E. (2016). End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1064–1074, Berlin, Germany. Association for Computational Linguistics.

[141] Mahar, J. A. and Memon, G. Q. (2010a). Rule based part of speech tagging of sindhi language. In *2010 International Conference on Signal Acquisition and Processing*, pages 101–106. IEEE.

[142] Mahar, J. A. and Memon, G. Q. (2010b). Sindhi part of speech tagging system using wordnet. *International Journal of Computer Theory and Engineering*, 2(4):538.

[143] Malaviya, C., Gormley, M. R., and Neubig, G. (2018). Neural factor graph models for cross-lingual morphological tagging. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2653–2663.

[144] Manju, K., Soumya, S., and Idicula, S. M. (2009). Development of a pos tagger for malayalam-an experience. In *2009 International Conference on Advances in Recent Technologies in Communication and Computing*, pages 709–713. IEEE.

[145] Manning, C. and Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

[146] Marcus, M. P., Marcinkiewicz, M. A., and Santorini, B. (1993). Building a large annotated corpus of english: the penn treebank. *Computational Linguistics*, 19(2):313–330.

[147] März, L., Trautmann, D., and Roth, B. (2019). Domain adaptation for part-of-speech tagging of noisy user-generated text. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3415–3420.

[148] Matthews, A., Neubig, G., and Dyer, C. (2018). Using morphological knowledge in open-vocabulary neural language models. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1435–1445.

[149] McCallum, A., Freitag, D., and Pereira, F. C. (2000). Maximum entropy markov models for information extraction and segmentation. In *Icml*, volume 17, pages 591–598.

[150] Meftah, S. and Semmar, N. (2018). A neural network model for part-of-speech tagging of social media texts. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

[151] Mikolov, T. (2007). Language modeling for speech recognition in czech. *Ph. D. dissertation, Masters thesis.*

[152] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. In Bengio, Y. and LeCun, Y., editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings.*

[153] Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., and Joulin, A. (2018). Advances in pre-training distributed word representations. In Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., and Tokunaga, T., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018.* European Language Resources Association (ELRA).

[154] Mikolov, T., Kopecky, J., Burget, L., Glembek, O., et al. (2009). Neural network based language models for highly inflective languages. In *2009 IEEE international conference on acoustics, speech and signal processing*, pages 4725–4728. IEEE.

[155] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

[156] Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

[157] Miller, J., Torii, M., and Vijay-Shanker, K. (2007). Adaptation of pos tagging for multiple biomedical domains. In *Biological, translational, and clinical language processing*, pages 179–180.

[158] Miller, J. E., Bloodgood, M., Torii, M., and Vijay-Shanker, K. (2006). Rapid adaptation of pos tagging for domain specific uses. In *Proceedings of the HLT-NAACL BioNLP Workshop on Linking Natural Language and Biology*, pages 118–119. Association for Computational Linguistics.

[159] Mishra, N. and Jain, S. (2017). Pos tagging of hindi language using hybrid approach. In *International Conference on Next Generation Computing Technologies*, pages 723–735. Springer.

[160] Mishra, N. and Mishra, A. (2011). Part of speech tagging for hindi corpus. In *2011 International Conference on Communication Systems and Network Technologies*, pages 554–558. IEEE.

[161] Mishra, P., Mujadia, V., and Sharma, D. M. (2017). Pos tagging for resource poor languages through feature projection. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*, pages 50–55.

[162] Miyato, T., Dai, A. M., and Goodfellow, I. (2017). Adversarial training methods for semi-supervised text classification. *ICLR*.

[163] Morwal, S., Jahan, N., and Chopra, D. (2012). Named entity recognition using hidden markov model (hmm). *International Journal on Natural Language Computing (IJNLC)*, 1(4):15–23.

[164] Moryossef, A., Aharoni, R., and Goldberg, Y. (2019). Filling gender & number gaps in neural machine translation with black-box context injection. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 49–54, Florence, Italy. Association for Computational Linguistics.

[165] Motlani, R., Lalwani, H., Shrivastava, M., and Sharma, D. M. (2015). Developing part-of-speech tagger for a resource poor language: Sindhi. In *Proceedings of the 7th Language and Technology Conference (LTC 2015), Poznan, Poland.*

[166] Mundotiya, R. K., Kumar, V., Mehta, A., and Singh, A. K. (2020a). Attention-based domain adaption using transfer learning for part-of-speech tagging: An experiment on the hindi language. In *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, pages 471–477.

[167] Mundotiya, R. K., Singh, M. K., Kapur, R., Mishra, S., and Singh, A. K. (2020b). Basic linguistic resources and baselines for bhojpuri, magahi and maithili for natural language processing. *CoRR*, abs/2004.13945.

[168] Mundotiya, R. K., Singh, M. K., Kapur, R., Mishra, S., and Singh, A. K. (2021). Linguistic resources for bhojpuri, magahi and maithili: Statistics about them, their similarity estimates, and baselines for three applications. *Transactions on Asian and Low-Resource Language Information Processing*, (in press).

[169] Murthy, R., Khapra, M. M., and Bhattacharyya, P. (2018). Improving ner tagging performance in low-resource languages via multilingual learning. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 18(2):9.

[170] Narayan, R., Chakraverty, S., and Singh, V. (2014). Neural network based parts of speech tagger for hindi. *IFAC Proceedings Volumes*, 47(1):519–524.

[171] Nguyen, V. Q., Anh, T. N., and Yang, H.-J. (2019). Real-time event detection using recurrent neural network in social sensors. *International Journal of Distributed Sensor Networks*, 15(6):1550147719856492.

[172] Nichol, A., Achiam, J., and Schulman, J. (2018). On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*.

[173] Niehues, J. and Cho, E. (2017). Exploiting linguistic resources for neural machine translation using multi-task learning. In *Proceedings of the Second Conference on Machine Translation*, pages 80–89.

[174] Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., et al. (2016). Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666.

[175] Nongmeikapam, K. and Bandyopadhyay, S. (2012). A transliteration of crf based manipuri pos tagging. *Procedia Technology*, 6:582–589.

[176] Ojha, A. K., Behera, P., Singh, S., and Jha, G. N. (2015). Training & evaluation of pos taggers in indo-aryan languages: a case of hindi, odia and bhojpuri. In *the*

proceedings of 7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, pages 524–529.

[177] Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., and Smith, N. A. (2013). Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 380–390.

[178] Padró, M. and Padró, L. (2005). A named entity recognition system based on a finite automata acquisition algorithm. *Procesamiento del Lenguaje Natural*, 35:319–326.

[179] Pagliardini, M., Gupta, P., and Jaggi, M. (2017). Unsupervised learning of sentence embeddings using compositional n-gram features. *arXiv preprint arXiv:1703.02507*.

[180] Pakoci, E., Popović, B., and Pekar, D. (2019). Using morphological data in language modeling for serbian large vocabulary speech recognition. *Computational intelligence and neuroscience*, 2019.

[181] Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

[182] Pasca, M., Lin, D., Bigham, J., Lifchits, A., and Jain, A. (2006). Organizing and searching the world wide web of facts-step one: the one-million fact extraction challenge. In *AAAI*, volume 6, pages 1400–1405.

[183] Passban, P., Liu, Q., and Way, A. (2016). Boosting neural pos tagger for farsi using morphological information. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 16(1):1–15.

[184] Passban, P., Liu, Q., and Way, A. (2018). Improving character-based decoding using target-side morphological information for neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 58–68, New Orleans, Louisiana. Association for Computational Linguistics.

[185] Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

[186] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.

[187] Plank, B. and Agić, Ž. (2018). Distant supervision from disparate sources for low-resource part-of-speech tagging. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 614–620, Brussels, Belgium. Association for Computational Linguistics.

[188] Plank, B. and Klerke, S. (2019). Lexical resources for low-resource pos tagging in neural times. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 25–34.

[189] Plank, B., Søgaard, A., and Goldberg, Y. (2016). Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 412–418, Berlin, Germany. Association for Computational Linguistics.

[190] Pourdamghani, N., Ghazvininejad, M., and Knight, K. (2018). Using word vectors to improve word alignments for low resource machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 524–528.

[191] Prabha, G., Jyothsna, P., Shahina, K., Premjith, B., and Soman, K. (2018). A deep learning approach for part-of-speech tagging in nepali language. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1132–1136. IEEE.

[192] Priyadarshi, A. and Saha, S. K. (2020). Towards the first maithili part of speech tagger: Resource creation and system development. *Computer Speech & Language*, 62:101054.

[193] PVS, A. and Karthik, G. (2007). Part-of-speech tagging and chunking using conditional random fields and transformation based learning. *Shallow Parsing for South Asian Languages*, 21.

[194] R, V. and Lalitha Devi, S. (2008). Domain focused named entity recognizer for Tamil using conditional random fields. In *Proceedings of the IJCNLP-08 Workshop on Named Entity Recognition for South and South East Asian Languages*.

[195] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

[196] Rao, P. R., Malarkodi, C., Ram, R. V. S., and Devi, S. L. (2015). Esm-il: Entity extraction from social media text for indian languages@ fire 2015-an overview. In *FIRE workshops*, pages 74–80.

[197] Ratnaparkhi, A. (1996). A maximum entropy model for part-of-speech tagging. In *Conference on empirical methods in natural language processing*.

[198] Rau, L. F. (1991). Extracting company names from text. In *[1991] Proceedings. The Seventh IEEE Conference on Artificial Intelligence Application*, volume 1, pages 29–32. IEEE.

[199] Ravi, S. and Larochelle, H. (2016). Optimization as a model for few-shot learning.

[200] Rei, M., Crichton, G., and Pyysalo, S. (2016). Attending to characters in neural sequence labeling models. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 309–318.

[201] Rijhwani, S., Zhou, S., Neubig, G., and Carbonell, J. G. (2020). Soft gazetteers for low-resource named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8118–8123.

[202] Riloff, E., Jones, R., et al. (1999). Learning dictionaries for information extraction by multi-level bootstrapping. In *AAAI/IAAI*, pages 474–479.

[203] Rishikesh (2018). Parts of speech tagger for maithili language using hmm. *International Journal of Innovations & Advancement in Computer Science*, 7:206.

[204] Rozenfeld, B., Feldman, R., and Fresko, M. (2006). A systematic cross-comparison of sequence classifiers. In *Proceedings of the 2006 SIAM International Conference on Data Mining*, pages 564–568. SIAM.

[205] Ruder, S., Peters, M. E., Swayamdipta, S., and Wolf, T. (2019). Transfer learning in natural language processing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, pages 15–18.

[206] Rumelhart, D., Hinton, G., and Williams, R. (1988). Learning internal representations by error propagation. In *Neurocomputing: foundations of research*, pages 673–695.

[207] Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. (1986). Learning internal representations by back-propagating errors. *Nature*, 323(99):533–536.

[208] Sagot, B. and Alonso, H. M. (2017). Improving neural tagging with lexical information. In *Proceedings of the 15th International Conference on Parsing Technologies*, pages 25–31.

[209] Saha, S. K., Sarkar, S., and Mitra, P. (2008). A hybrid feature set based maximum entropy hindi named entity recognition. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*.

[210] Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

[211] Santos, C. D. and Zadrozny, B. (2014). Learning character-level representations for part-of-speech tagging. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1818–1826.

[212] Santos, C. N. d. and Guimaraes, V. (2015). Boosting named entity recognition with neural character embeddings. *arXiv preprint arXiv:1505.05008*.

[213] Sarawagi, S. and Cohen, W. W. (2004). Semi-markov conditional random fields for information extraction. *Advances in neural information processing systems*, 17:1185–1192.

[214] Scherrer, Y. and Rabus, A. (2019). Neural morphosyntactic tagging for rusyn. *Natural Language Engineering*, 25(5):633–650.

[215] Schnabel, T. and Schütze, H. (2013). Towards robust cross-domain domain adaptation for part-of-speech tagging. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 198–206, Nagoya, Japan. Asian Federation of Natural Language Processing.

[216] Schnabel, T. and Schütze, H. (2014). Flors: Fast and simple domain adaptation for part-of-speech tagging. *Transactions of the Association for Computational Linguistics*, 2:15–26.

[217] Schuster, M. and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681.

[218] Schwenk, H. (2007). Continuous space language models. *Computer Speech & Language*, 21(3):492–518.

[219] Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell system technical journal*, 30(1):50–64.

[220] Shao, Y., Lin, J. C.-W., Srivastava, G., Jolfaei, A., Guo, D., and Hu, Y. (2021). Self-attention-based conditional random fields latent variables model for sequence labeling. *Pattern Recognition Letters*, 145:157–164.

[221] Shrivastava, M. and Bhattacharyya, P. (2008). Hindi pos tagger using naive stemming: harnessing morphological information without extensive linguistic knowledge. In *International Conference on NLP (ICON08), Pune, India*.

[222] Singh, A. K. (2008). Named entity recognition for south and south east asian languages: taking stock. In *Proceedings of the IJCNLP-08 Workshop on Named Entity Recognition for South and South East Asian Languages*.

[223] Singh, J., Joshi, N., and Mathur, I. (2013). Development of marathi part of speech tagger using statistical approach. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1554–1559. IEEE.

[224] Singh, S., Gupta, K., Shrivastava, M., and Bhattacharyya, P. (2006). Morphological richness offsets resource demand-experiences in constructing a pos tagger for hindi. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 779–786. Association for Computational Linguistics.

[225] Singh, S. and Jha, G. N. (2015). Statistical tagger for bhojpuri (employing support vector machine). In *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1524–1529. IEEE.

[226] Singh, T. D. and Bandyopadhyay, S. (2008). Morphology driven manipuri pos tagger. In *Proceedings of the IJCNLP-08 Workshop on NLP for less privileged languages*.

[227] Singh, T. D., Ekbal, A., and Bandyopadhyay, S. (2008). Manipuri pos tagging using crf and svm: A language independent approach. In *proceeding of 6th International conference on Natural Language Processing (ICON-2008)*, pages 240–245.

[228] Singha, K. R., Purkayastha, B. S., and Singha, K. D. (2012). Part of speech tagging in manipuri with hidden markov model. *International Journal of Computer Science Issues (IJCSI)*, 9(6):146.

[229] Sinha, K., Dong, Y., Cheung, J. C. K., and Ruths, D. (2018). A hierarchical neural attention-based text classifier. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 817–823, Brussels, Belgium. Association for Computational Linguistics.

[230] Socher, R., Bauer, J., Manning, C. D., and Ng, A. Y. (2013). Parsing with compositional vector grammars. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 455–465.

[231] Søgaard, A. (2013). Part-of-speech tagging with antagonistic adversaries. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 640–644.

[232] Song, Y., Klassen, P., Xia, F., and Kit, C. (2012). Entropy-based training data selection for domain adaptation. In Kay, M. and Boitet, C., editors, *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Posters, 8-15 December 2012, Mumbai, India*, pages 1191–1200. Indian Institute of Technology Bombay.

[233] Subbārāo, K. V. (2012). *South Asian languages: A syntactic typology.* Cambridge University Press.

[234] Sun, X. and Lu, W. (2020). Understanding attention for text classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3418–3428.

[235] Täckström, O., Das, D., Petrov, S., McDonald, R., and Nivre, J. (2013). Token and type constraints for cross-lingual part-of-speech tagging. *Transactions of the Association for Computational Linguistics*, 1:1–12.

[236] Tandon, J., Chaudhry, H., Bhat, R. A., and Sharma, D. (2016). Conversion from paninian karakas to universal dependencies for Hindi dependency treebank. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pages 141–150, Berlin, Germany. Association for Computational Linguistics.

[237] Tawfik, A., Emam, M., Essam, K., Nabil, R., and Hassan, H. (2019). Morphology-aware word-segmentation in dialectal Arabic adaptation of neural machine translation. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 11–17, Florence, Italy. Association for Computational Linguistics.

[238] Tham, M. J. (2018). Challenges and issues in developing an annotated corpus and hmm pos tagger for khasi. In *The 15th International Conference on Natural Language Processing.*

[239] Tkachenko, A. and Sirts, K. (2018). Modeling composite labels for neural morphological tagging. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 368–379.

[240] Todi, K. K., Mishra, P., and Sharma, D. M. (2018). Building a kannada pos tagger using machine learning and neural network models. *arXiv preprint arXiv:1808.03175*.

[241] Turian, J., Ratinov, L., and Bengio, Y. (2010). Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 384–394. Association for Computational Linguistics.

[242] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

[243] Vu, T.-T., Phung, D., and Haffari, G. (2020a). Effective unsupervised domain adaptation with adversarially trained language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6163–6173, Online. Association for Computational Linguistics.

[244] Vu, V.-H., Nguyen, Q.-P., Nguyen, K.-H., Shin, J.-C., and Ock, C.-Y. (2020b). Korean-vietnamese neural machine translation with named entity recognition and part-of-speech tags. *IEICE Transactions on Information and Systems*, 103(4):866–873.

[245] Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. R. (2019). GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

[246] Wang, C., Chen, W., and Xu, B. (2017). Named entity recognition with gated convolutional neural networks. In *Chinese computational linguistics and natural language processing based on naturally annotated big data*, pages 110–121. Springer.

[247] Wei, W., Wang, Z., Mao, X., Zhou, G., Zhou, P., and Jiang, S. (2021). Position-aware self-attention based neural sequence labeling. *Pattern Recognition*, 110:107636.

[248] Winograd, T. (1972). Understanding natural language. *Cognitive psychology*, 3(1):1–191.

[249] Wright, D. and Augenstein, I. (2020). Transformer based multi-source domain adaptation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7963–7974, Online. Association for Computational Linguistics.

[250] Wu, M., Liu, F., and Cohn, T. (2018). Evaluating the utility of hand-crafted features in sequence labelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2850–2856.

[251] Xiao, M. and Guo, Y. (2013). Domain adaptation for sequence labeling tasks with a probabilistic language adaptation model. In *International Conference on Machine Learning*, pages 293–301.

[252] Xin, Y., Hart, E., Mahajan, V., and Ruvini, J. D. (2018). Learning better internal structure of words for sequence labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2584–2593.

[253] Yadav, V., Sharp, R., and Bethard, S. (2018). Deep affix features improve neural named entity recognizers. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 167–172.

[254] Yajnik, A. (2017). Part of speech tagging using statistical approach for nepali text. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 11(1):76–79.

[255] Yang, G. and Xu, H. (2020). A residual bilstm model for named entity recognition. *IEEE Access*, 8:227710–227718.

[256] Yang, J., Liang, S., and Zhang, Y. (2018). Design challenges and misconceptions in neural sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3879–3889, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

[257] Yang, X., Liu, Y., Xie, D., Wang, X., and Balasubramanian, N. (2019a). Latent part-of-speech sequences for neural machine translation. In *Proceedings of the 2019*

Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 780–790, Hong Kong, China. Association for Computational Linguistics.

[258] Yang, Z., Dai, Z., Yang, Y., Carbonell, J. G., Salakhutdinov, R., and Le, Q. V. (2019b). Xlnet: Generalized autoregressive pretraining for language understanding. In Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. B., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 5754–5764.

[259] Yang, Z., Salakhutdinov, R., and Cohen, W. W. (2017). Transfer learning for sequence tagging with hierarchical recurrent networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

[260] Yasunaga, M., Kasai, J., and Radev, D. (2018). Robust multilingual part-of-speech tagging via adversarial training. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 976–986, New Orleans, Louisiana. Association for Computational Linguistics.

[261] Yin, Y., Su, J., Wen, H., Zeng, J., Liu, Y., and Chen, Y. (2019). Pos tag-enhanced coarse-to-fine attention for neural machine translation. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 18(4):1–14.

[262] Yoonus, M. M. and Sinha, S. (2011). A hybrid pos tagger for indian languages. *Language in India*, 11(9).

[263] Zennaki, O., Semmar, N., and Besacier, L. (2019). A neural approach for inducing multilingual resources and natural language processing tools for low-resource languages. *Natural Language Engineering*, 25(1):43–67.

[264] Zhang, M., Zhang, Y., Che, W., and Liu, T. (2014). Type-supervised domain adaptation for joint segmentation and pos-tagging. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 588–597.

[265] Zhang, S. and Elhadad, N. (2013). Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts. *Journal of biomedical informatics*, 46(6):1088–1098.

[266] Zhang, X., Zhao, J. J., and LeCun, Y. (2015). Character-level convolutional networks for text classification. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 649–657.

[267] Zhang, Y., Chen, H., Zhao, Y., Liu, Q., and Yin, D. (2018a). Learning tag dependencies for sequence tagging. In *IJCAI*, pages 4581–4587.

[268] Zhang, Y., Sun, X., Ma, S., Yang, Y., and Ren, X. (2018b). Does higher order lstm have better accuracy for segmenting and labeling sequence data? In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 723–733.

[269] Zhang, Y. and Wallace, B. C. (2017). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 253–263.

[270] Zhao, Z., Liu, T., Li, S., Li, B., and Du, X. (2017). Ngram2vec: Learning improved word representations from ngram co-occurrence statistics. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 244–253.

[271] Zheng, H., Fu, J., Mei, T., and Luo, J. (2017). Learning multi-attention convolutional neural network for fine-grained image recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 5209–5217.