

REFERENCES

- [1] Hoang H, Birch A, Callison-Burch C, Federico M, Koehn P, "MOSES: Open Source Toolkit for Statistical Machine Translation.," In Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, 2007, pp. 177-180.
- [2] Bonnie J Dorr, "Machine translation divergences: A formal description and proposed solution," In Computational Linguistics, 1994, pp. 597-633.
- [3] J Cocke, S Della Pietra, Brown P, "A statistical approach to machine translation," In Journal of Computational Linguistics, vol. 16(2), pp. 79-85, 1990.
- [4] Dzmitry, Kyunghyun Cho, Yoshua Bengio, Bahdanau, "Neural machine translation by jointly learning to align and translate," In arXiv preprint arXiv:1409.0473, 2014.
- [5] Weerasinghe R, "A statistical machine translation approach to sinhala-tamil language translation," In Towards an ICT enabled Society, Sri Lanka, 2003, p. 136.
- [6] Sripirakas A, Weerasinghe, Herath S, "Statistical machine translation of systems for Sinhala-Tamil," In International Conference in Advances in ICT for Emerging Regions (ICTer), 2010.
- [7] R Weerasinghe, M Niranjana, R Pushpananda, "Sinhala-Tamil Machine Translation: Towards better Translation Quality," In Australasian Language Technology Association Workshop 2014, Melbourne, 2014.
- [8] S Sheeyam, A Umasuthan, S Rajpirathap, "Real-time direct translation system for Sinhala and Tamil languages," In Computer Science and Information Systems (FedCSIS), 2015.
- [9] R Weerasinghe M Niranjana, R Pushpananda, "Statistical Machine Translation from and into Morphologically Rich and Low Resourced Languages," In International Conference on Intelligent Text Processing and Computational Linguistics, 2015.
- [10] Rosenfeld, Clarkson P R, "Statistical Language Modelling using the CMU-Cambridge Toolkit," In Proceedings ESCA Euro Speech, Greece, 1997.

- [11] S Roukos, T Ward, W J Zhu, K Papineni, "BLEU: a method for automatic evaluation of machine translation.," In Proceedings of the 40th annual meeting on association for computational linguistics, 2002.
- [12] WWSWS. (2018, August) World Socialist Web Site. [Online]. www.wsws.org
- [13] University of Colombo. (2018, August) Language Technology Research Laboratory. [Online]. <http://www.ucsc.cmb.ac.lk/ltrl/>
- [14] Stolcke A et al., "SRILM-An Extensible Language Modeling Toolkit," In INTER-SPEECH, 2002.
- [15] Och FJ, "Minimum Error Rate Training in Statistical Machine Translation," In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, 2003, pp. 160-167.
- [16] Pears R, Koh Y S, Sakthithasan S, "One pass concept change detection for data streams.," In Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, 2013, pp. 461-472.
- [17] Mahendran Jeyakaran, Ruvan Weerasinghe, "A novel kernel regression based machine translation system for sinhala-tamil translation," In Proceedings of the 4th Annual UCSC Research Symposium, 2011.
- [18] Lagus K, Creutz M, "Unsupervised Models for Morpheme Segmentation and Morphology Learning," In ACM Transactions on Speech and Language Processing (TSLP), 2007, p. 4.
- [19] Hoang H, Birch A, Callison-Burch C, Federico M, Koehn P et al., "MOSES: Open Source Toolkit for Statistical Machine Translation.," In Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, 2007, pp. 177-180.
- [20] Ranathunga S, Jayasena S, Dias G, Farhath F, "Integration of Bilingual Lists for Domain-Specific Statistical Machine Translation for Sinhala-Tamil," In Moratuwa Engineering Research Conference (MERCon, Moratuwa, 2018, pp. 538-543.
- [21] Pranavan Theivendiram, Surangika Ranathunga, Sanath Jayasena, Gihan Dias, Fathima Farhath, "Improving Domain-specific SMT for Low-resourced Languages using Data from Different Domains," In 11th Edition of the Language Resources and Evaluation Conference, 2018.

- [22]Prabath Sandaruwan, Malith Thilakarathne, Achini Herath, Surangika Ranathunga, Sanath Jayasena, Gihan Dias, Pasindu Tennage, "Neural Machine Translation for Sinhala and Tamil Languages," In International Conference on Asian Language Processing (IALP), 2017, pp. 189-192.
- [23]Y Kim, Y Deng, J Senellart, A. M. Rush, G Klein, "OpenNMT: Open-source toolkit for neural machine translation," In arxiv preprint arXiv:1701.02810 [cs.CL], 2017.
- [24]Bonnie J, Dorr, "Machine translation divergences: A formal description and proposed solution," In Computational Linguistics, 1994, pp. 597-633.
- [25]Khan M A, Saboor A, "Lexical-semantic divergence In Urdu-to-English example based Machine Translation.," In 6th International Conference In Emerging Technologies (ICET), 2010, October, pp. 316-320.
- [26]Deshmukh P D, Kazi, M M Kale, Kulkarni S B, "Linguistic Divergence Patterns in English to Marathi Translation," In International Journal of Computer Applications, 2014.
- [27]Niladri Sekhar, Dash, "Linguistic Divergences in English to Bengali Translation," In International Journal of English Linguistics, 2013.
- [28]T M Kirby, Ellison, "Measuring language divergence by intra-lexical comparison," In Proceedings of the 21st international conference on computational linguistics and 44th annual meeting of the association for computational linguistics, 2006, pp. 273-280.
- [29]Mourya N, Pandey V, Behera P, "Dealing with Linguistic Divergences in English-Bhojpuri Machine Translation.," In Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP2016), 2016, pp. 103-113.
- [30]Tampoe H.D, "Sinhala and Tamil: A case of Contact-Induced Restructuring," Language and Linguistics, Newcastle University, A PhD Dissertation presented at School of English Literature <http://hdl.handle.net/10443/3552>, 2016.
- [31]R McDonald, "Universal Dependency Annotation for Multilingual Parsing.," In Proceedings of ACL, Sofia, Bulgaria, 2013.
- [32]M Anand Kumar, "Factored Statistical Machine Translation System for English to Tamil Language.," *Pertanika Journal of Social Sciences & Humanities* 22.4, 2014.

- [33] M Anand Kumar, "Morphology based Prototype Statistical Machine Translation System for English to Tamil Language," AMRITA School of Engineering, AMRITA Vishwa Vidyapeetham, Coimbatore, A Thesis Submitted for the Degree of Doctor of Philosophy in the School of Engineering 2013.
- [34] Bureau Indian Standard, "Unified Parts of Speech (POS) Standard in Indian Languages".
- [35] Sudhakar Kumawat, Vinayak Srivastava, Nitish Chandra, "Various Tagsets for Indian Languages and Their Performance In Part Of Speech Tagging.," In Proceedings of 5th IRF International Conference, ISBN: 978-93-82702-67-2, 2014.
- [36] Ramanathan M V, "An Attempt at Multilingual POS Tagging for Tamil."
- [37] IIIT. (2018) IIIT HYDERABAD. [Online]. http://shiva.iiit.ac.in/SPSAL2007/iiit_tagset_guidelines
- [38] Lakshmana Pandian S T, "Morpheme based Language Model for Tamil Part-of-Speech Tagging.," In Polibits 38, 2008, pp. 19-25.
- [39] Natarajan A M, Selvam M, "Improvement of rule based morphological analysis and pos tagging in tamil language via projection and induction techniques.," In International journal of computers, 2009, pp. 357-367.
- [40] Central Institute of Indian Languages. (n.d.). (2018, August) Central Institute of Indian Languages. [Online]. <http://www.ciil.org/>
- [41] Anand Kumar, Shivapratap G, Soman K P, Rajendran S, Dhanalakshmi V, "Tamil POS Tagging using Linear Programming," In International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May, 2009.
- [42] Anand kumar M, Rajendran S, Soman K P, Dhanalakshmi V, "POS Tagger and Chunker for the Tamil Language.," In Proceedings of Tamil Internet Conference., 2009.
- [43] Vasu Renganathan, "Development of Part-of-Speech Tagger for Tamil," In Tamil Internet 2001 Conference August, Kuala Lumpur, 2001, pp. 26-28.
- [44] Ganesan M, "Morph and POS Tagger for Tamil (Software)," Annamalai University , Annamalai Nagar, 2007.
- [45] Rajendran S, "Parsing in Tamil," in LANGUAGE IN INDIA www.languageinindia.com Volume 6: 8 August, 2006.

- [46] AUKBC. (2018, August) AU-KBC RESEARCH CENTRE. [Online]. http://www.au-kbc.org/research_areas/nlp/projects/postagger.html
- [47] Sobha L, Kumara Shanmugam B, Arulmozhi P, "Parts of Speech Tagger for Tamil," In Proceedings of the Symposium on Indian Morphology, Phonology & Language Engineering, Indian Institute of Technology, Kharagpur., 2004.
- [48] Sobha, L Arulmozhi, "A Hybrid POS Tagger for a Relatively Free Word Order Language," In Proceedings of MSPIL-2006, Indian Institute of Technology, Bombay., 2006.
- [49] AUKBC. (2018, August) AUKBC Research Centre. [Online]. <http://www.au-kbc.org/>
- [50] W V Welgama, A R Weerasinghe, Dilmi Gunasekara, "Hybrid Part of Speech Tagger for Sinhala Language.," In International Conference on Advances in ICT for Emerging Regions (ICTer)., 2016, pp. 041 – 048.
- [51] Ranathunga S, Jayasena S, Dias G, Fernando S, "Comprehensive Part-Of-Speech Tag Set and SVM Based POS Tagger for Sinhala.," In WSSANLP 2016, 2016, p. 163.
- [52] N G J Dias, Jayaweera A J P M P, "Hidden Markov Model Based Part of Speech Tagger for Sinhala Language," In arXiv preprint arXiv:1407.2989, 2014.
- [53] A R Weerasinghe, Jayasuriya M, "Learning a stochastic part of speech tagger for sinhala," In 2013 International Conference on Advances in ICT for Emerging Regions (ICTer), 2013.
- [54] Wilson A, Leech G, "Recommendations for the Morphosyntactic Annotation of Corpora.," EAGLES Re-port EAG-TCWG-MAC/R, 1996.
- [55] Kalika Bali, Tanmoy Bhattacharya, Pushpak Bhattacharyya, Girish Nath Jha, Rajendran S, Saravanan K, Sobha L, Sankaran Baskaran, "Designing a Common POS-Tagset Framework for Indian Languages," In The 6th Workshop on Asian Language Resources, 2008.
- [56] Norbert, Suzanne Lenz, Volz, "Multilingual Corpus Tagset Specifications," In MLAP PAROLE 63œ386 WP 4.4, 1996.
- [57] Nancy Ide, Jean, Véronis, "Multext (multilingual tools and corpora)," In Proceedings of the 15th International Conference on Computational Linguistics (COLING -94), Kyoto, Japan, 1994.

- [58] Tomaž, Erjavec, "Multext-east version 3: Multilingual morphosyntactic specifications, lexicons and corpora.," In Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004), Lisboa, Portugal, 2004.
- [59] D Das, Petrov, "Unsupervised part-of-speech tagging with bilingual graph-based projections," In Proceedings of ACL-HLT, 2011.
- [60] B Snyder, J Eisenstein, R Barzilay, T Naseem, "Multilingual part-of-speech tagging: Two unsupervised approaches," In JAIR, 36, 2009.
- [61] Zeman D, "Reusable Tagset Conversion Using Tagset Drivers.," In Proceedings of LREC, Marrakech, Morocco, 2008.
- [62] Tyers, Joakim Nivre, Francis M, "Universal Dependencies for Turkish.," In Proceedings of COLING, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 2016, pp. 3444–3454.
- [63] Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, Christopher D, Marie Catherine De Marneffe, "Universal Stanford Dependencies: a crosslinguistic typology," In Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC), Reykjavík, Iceland, 2014.
- [64] Reut, Tsarfaty, "A unified morpho-syntactic scheme of Stanford Dependencies.," In Proceedings of 51st Annual Meeting of the Association for Computational Linguistics (ACL), 2013, pp. 578-584.
- [65] Alon Y, Halevy, AnHai Doan, "Semantic-Integration Research In the Database Community A Brief Survey.," AI Magazine Volume 26 Number 1, 2005.
- [66] C Lenzerini, M Navathe, Batini, "A Comparative Analysis of Methodologies for Database Schema Integration.," In ACM Computing Survey 18(4), 1986, pp. 323–364.
- [67] Chiang D, "A hierarchical phrase-based model for statistical machine translation," In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005, June, pp. 263-270.
- [68] Aho J D, Ullman A V, "Syntax directed translations and the pushdown assembler," Journal of Computer and System Sciences, pp. 3:37–56., 1969.
- [69] Franz Josef Och, Hermann, Ney, "Improved statistical alignment models," In Proceedings of the 38th Annual Meeting of the ACL, 2000, pp. 440-447.

- [70] Franz Josef Och, Daniel Marcu, Philipp Koehn, "Statistical phrase-based translation," In Proceedings of HLT-NAACL, 2003., pp. 127-133.
- [71] Philipp Koehn, "Pharaoh: a beam search decoder for phrase-based statistical machine translation models," In Proceedings of the Sixth Conference of the Association for Machine Translation in the Americas, 2004, pp. 115-124.
- [72] Mahsa, Abdolhossein Sarrafzadeh, Mohaghegh, "A hierarchical phrase-based model for English-Persian statistical machine translation," In 2012 International Conference on Innovations in Information Technology (IIT), 2012.
- [73] Zeman D, Jawaid B, "Word-order issues in English-to-Urdu statistical machine translation," In The Prague Bulletin of Mathematical Linguistics, 2011, pp. 87-106.
- [74] Nadeem, Khan et al., "English to Urdu hierarchical phrase-based statistical machine translation," In Proceedings of the 4th Workshop on South and Southeast Asian Natural Language Processing, 2013.
- [75] Ulrich, Germann, "Building a statistical machine translation system from scratch: how much bang for the buck can we expect?," In Proceedings of the workshop on Data-driven methods in machine translation-, 2001, p. Volume 14.
- [76] Vasu, Renganathan, "An interactive approach to development of English to Tamil machine translation system on the web," In Proceedings of INFITT, 2002.
- [77] Loganathan R M, "English-Tamil Machine Translation System," Amrita Vishwa Vidyapeetham, Coimbatore, Master of Science by Research Thesis 2010.
- [78] Dhanalakshmi V, Soman K P, Sharmiladevi V Kumar M, "Improving the Performance of English-Tamil Statistical Machine Translation System using Source-Side Pre-Processing.," In arXiv preprint arXiv:1409.8581, 2014.
- [79] Mary Priya, G Santhosh Kumar, Sebastian, "English to malayalam translation: a statistical approach," In Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India. ACM, 2010.
- [80] Daniel Marcu, William, Wong, "A Phrase-Based, Joint Probability Model for Statistical Machine Translation," In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2002), Philadelphia, 2002.

- [81] Kay, Stephan Vogel, Rottmann, "Word reordering in statistical machine translation with a POS-based distortion model," In Proc. of TMI, 2007, pp. 171-180.
- [82] Deepa, Mauro Cettolo, Marcello Federico, Gupta, "POS-based reordering models for statistical machine translation," In Proceedings of the Machine Translation Summit (MT-Summit), 2007.
- [83] Derek F, Wong, Lidia S, Chao, Li Shuo, "Experiments with POS-based restructuring and alignment-based reordering for statistical machine translation," In Proceedings of the Second Workshop on Hybrid Approaches to Translation, 2013.
- [84] Wetzel D, Kaeshammer M, "Enriching phrase-based statistical machine translation with POS information." In Proceedings of the Second Student Research Workshop associated with RANLP, 2011, pp. 33-40.
- [85] N Dhanesh, "Conceptual Framework for Automated English to Tamil Machine Translation System," Scholar, P. G. A.
- [86] Ney H, Ueffing N, "Using pos information for statistical machine translation into morphologically rich languages," In Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 1, 2003, April, pp. 347-354.
- [87] Philipp Koehn, Hieu, Hoang, "Factored translation models," In Proc EMNLP+CoNLL, Prague, 2007, pp. 868–876.
- [88] Sonja Nießen, Hermann, Ney, "Statistical Machine Translation with Scarce Resources Using Morpho-syntactic Information," In Journal of Computational Linguistics, 30(2), 2004, pp. 181–204.
- [89] A Potamianos, I Klasinas, P Karageorgakis, "Towards incorporating language morphology into statistical machine translation systems," In Proceedings Automatic Speech Recogn. and Underst. Workshop (ASRU), 2005.
- [90] Soha, Sultan, "Applying morphology to English-Arabic statistical machine translation.," Diss. Master's Thesis Nr. 11 ETH Zurich In collaboration with Google Inc. 2011.
- [91] Adria de Gispert, Ramis, "Introducing Linguistic knowledge into statistical Machine Translation," TALP Research Center, Speech Processing Group

Department of Signal Theory and Communications, Universitat Politècnica de Catalunya., Ph.D. thesis 2006.

- [92] Sara, Stymne, "Compound Processing for Phrase-Based Statistical Machine Translation," Linköping University, Sweden, Licentiate thesis 2009.
- [93] Rabih M, Zbib, "Using Linguistic Knowledge in Statistical Machine Translation," Massachusetts Institute Of Technology, Ph.D. thesis September, 2010.
- [94] Lee Y S, "Morphological analysis for statistical machine translation," Defense Technical Information Center, 2004.
- [95] S S Arora, K Agrawal, "Pre-processing English-Hindi Corpus for Statistical Machine Translation," In *Computación y Sistemas*, 2017, p. 21(4).
- [96] Chengqing Zong, Bo Xu, Yu Zhou, "Bilingual Chunk Alignment In Statistical Machine Translation," In *IEEE International Conference on Systems, Man and Cybernetics*, 2004.
- [97] Santanu, Sudip Kumar Naskar, Sivaji Bandyopadhyay, Pal, "Word Alignment-Based Reordering of Source Chunks in PB-SM," In *LREC*, 2014.
- [98] Arianna, Daniele Pighin, Marcello Federico, Bisazza, "Chunk-lattices for verb reordering in Arabic-English statistical machine translation," In *Machine translation 26.1-2* (2012), 2012, pp. 85-103.
- [99] Xiaolin, Wang X, Utiyama M, Finch A, Sumita E, Wang, "Refining word segmentation using a manually aligned corpus for statistical machine translation," In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1654-1664.
- [100] Anoop Kunchukuttan, Pushpak Bhattacharyya, Raj Dabre, More Rohit, "Augmenting Pivot based SMT with word segmentation," In *Proceedings of the 12th International Conference on Natural Language Processing*, 2015, pp. 303-307.
- [101] Peyman, Qun Liu, Andy Way, Passban, "Providing Morphological Information for SMT Using Neural Networks," In *The Prague Bulletin of Mathematical Linguistics* 108, 2017, pp. 271-282.
- [102] Wikipedia. (2018, August) Wikipedia. [Online]. http://en.wikipedia.org/wiki/Tamil_language

- [103] Wikipedia. (2018, August) Wikipedia. [Online].
https://en.wikipedia.org/wiki/Sinhalese_language.
- [104] Robert, Caldwell, "A comparative grammar of the Dravidian or South-Indian Family of Languages," London, 1875.
- [105] S Rajendran. (2006, Volume 6 : 8.) Language in India. [Online].
www.languageinindia.com
- [106] Ramakrishnan S, "kiriyAvin thaRkAla thamiz akarAthi," Cre-A, 2006.
- [107] Vilar, D Ney, H Jovicic S, Saric Z, Popovic M, "Augmenting a Small Parallel Text with Morpho-Syntactic Language Resources for Serbian-English Statistical Machine Translation.," In Proceedings of the ACL Workshop on Building and Using Parallel Texts, 2005, pp. 41–48.
- [108] El-Kahlout I D, Oflazer K, "Exploring Different Representational Units in English-to-Turkish Statistical Machine Translation," In: Proceedings of the Second Workshop on Statistical Machine Translation, 2007, pp. 25–32.
- [109] Segalovich I, "A Fast Morphological Algorithm with Unknown Word Guessing Induced by a Dictionary for a Web Search Engine." In In: MLMTA, CiteSeer, 2003, pp. 273–280.
- [110] V ayrynen, J J Creutz, Sadeniemi M, Virpioja S, "Morphology-Aware Statistical Machine Translation based on Morphs Induced in an Unsupervised Manner," In Machine Translation Summit XI, 2007, pp. 491–498.
- [111] Jakob, Elming, Copenhagen Business School, PhD Thesis. 2008.
- [112] George, Doddington, "Automatic evaluation of machine translation quality using n-gram co-occurrence statistics.," In Proceeding of the ARPA Workshop on Human Language Technology., 2002.
- [113] Levenshtein V I, "Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics Doklady," , 1966 February, pp. 707–710.
- [114] Stefan Vogel, Hermann Ney, Alex Zubiaga, Christoph Tillmann, "A DP-based search using monotone alignments in statistical translation.," In Proceedings of the 35th Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics, Somerset, New Jersey, 1997, pp. 289–296.

- [115] B Dorr, R Schwartz, L Micciulla, J Makhoul, Snover M, "A Study of Translation Edit Rate with Targeted Human Annotation" In Proceedings of Association for Machine Translation in the Americas, 2006, pp. 223-231.
- [116] F J Och, "An Efficient method for determining bilingual word classes," In Proceedings of Ninth Conference of the European Chapter of the Association for Computational Linguistics (EACL), 1999.
- [117] Heafield K, "KenLM: Faster and smaller language model queries.," In Proceedings of the Sixth Workshop on Statistical Machine Translation, 2011.
- [118] Huang, Chiang L, "Forest rescoring: Faster decoding with integrated languagemodels," In Annual Meeting-Association For Computational Linguistics., 2007.
- [119] NLTK. (2018, May) NLTK 3.3 documentation. [Online]. <http://www.nltk.org/howto/collocations.html>
- [120] Pavel, Pecina, "An Extensive Empirical Study of Collocation Extraction Methods," In Proceedings of the Association for Computational Linguistics Student Research Workshop, 2005, pp. 13–18.
- [121] P Hanks, W K Church, "Word association norms, mutual information and lexicography," In Proceedings of the 27th meeting of the Association of Computational Linguistics, 1989, pp. 76-83.
- [122] Pathirenehelage N, Ihalapathirana, A Mohamed, M Z Ranathunga, S Jayasena, Dias G, Fernando S, Hameed R A, "Automatic Creation of a Sentence Aligned Sinhala-Tamil Parallel Corpus.," In Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing(WASSANLP), 2016, pp. 124-132.