

## REFERENCES

- [1] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [2] Tomas Mikolov, Quoc V Le, and Ilya Sutskever. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*, 2013.
- [3] Anushika Liyanage, Surangika Ranathunga, and Sanath Jayasena. Bilingual lexical induction for sinhala-english using cross lingual embedding spaces. In *2021 Moratuwa Engineering Research Conference (MERCOn)*, pages 579–584. IEEE, 2021.
- [4] Wikipedia contributors. Bilingual lexicon — Wikipedia, the free encyclopedia, 2019. [Online; accessed 22-March-2022].
- [5] Nan Jiang. Lexical representation and development in a second language. *Applied linguistics*, 21(1):47–77, 2000.
- [6] Davide Turcato. Automatically creating bilingual lexicons for machine translation from bilingual text. *arXiv preprint cmp-lg/9807010*, 1998.
- [7] Dhouha Bouamor, Nasredine Semmar, and Pierre Zweigenbaum. Automatic construction of a multiword expressions bilingual lexicon: A statistical machine translation evaluation perspective. In *Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon*, pages 95–108, 2012.
- [8] Ruotian Ma, Minlong Peng, Qi Zhang, and Xuanjing Huang. Simplify the usage of lexicon in chinese ner. *arXiv preprint arXiv:1908.05969*, 2019.
- [9] Ann Irvine and Chris Callison-Burch. A comprehensive analysis of bilingual lexicon induction. *Computational Linguistics*, 43(2):273–310, 2017.

- [10] Surangika Ranathunga, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. Neural machine translation for low-resource languages: A survey. *arXiv preprint arXiv:2106.15115*, 2021.
- [11] Mikel Artetxe, Gorka Labaka, and Eneko Agirre. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 451–462, 2017.
- [12] Yoav Goldberg and Graeme Hirst. Neural network methods in natural language processing. morgan & claypool publishers(2017). 9781627052986 (*zitiert auf Seite 69*).
- [13] Sahar Ghannay, Benoit Favre, Yannick Esteve, and Nathalie Camelin. Word embedding evaluation and combination. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 300–305, 2016.
- [14] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990.
- [15] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [16] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
- [17] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.

- [18] Chao Xing, Dong Wang, Chao Liu, and Yiye Lin. Normalized word embedding and orthogonal transform for bilingual word translation. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1006–1011, 2015.
- [19] Peter H Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.
- [20] Samuel L Smith, David HP Turban, Steven Hamblin, and Nils Y Hammerla. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. *arXiv preprint arXiv:1702.03859*, 2017.
- [21] Mikel Artetxe, Gorka Labaka, Inigo Lopez-Gazpio, and Eneko Agirre. Uncovering divergent linguistic information in word embeddings with lessons for intrinsic and extrinsic evaluation. *arXiv preprint arXiv:1809.02094*, 2018.
- [22] Ivan Vulić, Anna Korhonen, and Goran Glavaš. Improving bilingual lexicon induction with unsupervised post-processing of monolingual word vector spaces. In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 45–54, 2020.
- [23] Vikas Raunak, Vivek Gupta, and Florian Metze. Effective dimensionality reduction for word embeddings. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*, pages 235–243, 2019.
- [24] Christopher M Bishop. *Pattern recognition and machine learning (information science and statistics)*, 2007.
- [25] Peter D Turney and Patrick Pantel. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37:141–188, 2010.
- [26] Enrique Alfonseca, Massimiliano Ciaramita, and Keith Hall. Gazpacho and summer rash: lexical relationships from temporal patterns of web search

- queries. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1046–1055, 2009.
- [27] Taylor Berg-Kirkpatrick and Dan Klein. Simple effective decipherment via combinatorial optimization. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 313–321, 2011.
- [28] Alexandre Klementiev and Dan Roth. Weakly supervised named entity transliteration and discovery from multilingual comparable corpora. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 817–824, 2006.
- [29] Charles Schafer and David Yarowsky. Inducing translation lexicons via diverse similarity measures and bridge languages. In *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*, 2002.
- [30] Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. Inducing crosslingual distributed representations of words. In *Proceedings of COLING 2012*, pages 1459–1474, 2012.
- [31] Stephan Gouws and Anders Søgaard. Simple task-specific bilingual word embeddings. In *HLT-NAACL*, pages 1386–1390, 2015.
- [32] Omer Levy, Anders Søgaard, and Yoav Goldberg. A strong baseline for learning cross-lingual word embeddings from sentence alignments. *arXiv preprint arXiv:1608.05426*, 2016.
- [33] Tianze Shi, Zhiyuan Liu, Yang Liu, and Maosong Sun. Learning cross-lingual word embeddings via matrix co-factorization. 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 567–572, 2015.

- [34] Aitor Ormazabal, Mikel Artetxe, Gorika Labaka, Aitor Soroa, and Eneko Agirre. Analyzing the limitations of cross-lingual word embedding mappings. *arXiv preprint arXiv:1906.05407*, 2019.
- [35] Liangchen Wei Zhi-Hong Deng. A variational autoencoding approach for inducing cross-lingual word embeddings. 2017.
- [36] Barun Patra, Joel Ruben Antony Moniz, Sarthak Garg, Matthew R Gormley, and Graham Neubig. Bilingual lexicon induction with semi-supervision in non-isometric embedding spaces. *arXiv preprint arXiv:1908.06625*, 2019.
- [37] Mikel Artetxe, Gorika Labaka, and Eneko Agirre. Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In *Thirty-second AAAI conference on artificial intelligence*, 2018.
- [38] Alexis Conneau, Guillaume Lample, Marc’Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. Word translation without parallel data. *arXiv preprint arXiv:1710.04087*, 2017.
- [39] Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. Adversarial training for unsupervised bilingual lexicon induction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1959–1970, 2017.
- [40] Mladen Karan, Ivan Vulić, Anna Korhonen, and Goran Glavaš. Classification-based self-learning for weakly supervised bilingual lexicon induction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6915–6922, 2020.
- [41] Manaal Faruqui, Jesse Dodge, Sujay K Jauhar, Chris Dyer, Eduard Hovy, and Noah A Smith. Retrofitting word vectors to semantic lexicons. *arXiv preprint arXiv:1411.4166*, 2014.
- [42] Nikola Mrkšić, Diarmuid O Séaghdha, Blaise Thomson, Milica Gašić, Lina Rojas-Barahona, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen, and Steve

- Young. Counter-fitting word vectors to linguistic constraints. *arXiv preprint arXiv:1603.00892*, 2016.
- [43] Kim Anh Nguyen, Sabine Schulte im Walde, and Ngoc Thang Vu. Neural-based noise filtering from word embeddings. *arXiv preprint arXiv:1610.01874*, 2016.
- [44] Igor Labutov and Hod Lipson. Re-embedding words. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 489–493, 2013.
- [45] Sascha Rothe, Sebastian Ebert, and Hinrich Schütze. Ultradense word embeddings by orthogonal transformation. *arXiv preprint arXiv:1602.07572*, 2016.
- [46] Jiaqi Mu, Suma Bhat, and Pramod Viswanath. All-but-the-top: Simple and effective postprocessing for word representations. *arXiv preprint arXiv:1702.01417*, 2017.
- [47] Takashi Wada, Tomoharu Iwata, Yuji Matsumoto, Timothy Baldwin, and Jey Han Lau. Learning contextualised cross-lingual word embeddings and alignments for extremely low-resource languages using parallel corpora. *arXiv preprint arXiv:2010.14649*, 2020.
- [48] Goran Glavaš and Ivan Vulić. Non-linear instance-based cross-lingual mapping for non-isomorphic embedding spaces. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7548–7555, 2020.
- [49] MD Rajitha, LL Piyarathna, MMD Nayanaajith, and S Surangika. Sinhala and english document alignment using statistical machine translation. In *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pages 29–34. IEEE, 2020.

- [50] Aloka Fernando, Surangika Ranathunga, and Gihan Dias. Data augmentation and terminology integration for domain-specific sinhala-english-tamil statistical machine translation. *arXiv preprint arXiv:2011.02821*, 2020.
- [51] Yanyang Li, Yingfeng Luo, Ye Lin, Quan Du, Huizhen Wang, Shujian Huang, Tong Xiao, and Jingbo Zhu. A simple and effective approach to robust unsupervised bilingual dictionary induction. *arXiv preprint arXiv:2011.14874*, 2020.
- [52] Georgiana Dinu, Angeliki Lazaridou, and Marco Baroni. Improving zero-shot learning by mitigating the hubness problem. *arXiv preprint arXiv:1412.6568*, 2014.
- [53] Tasnim Mohiuddin, M Saiful Bari, and Shafiq Joty. Lnmap: Departures from isomorphic assumption in bilingual lexicon induction through non-linear mapping in latent space. *arXiv preprint arXiv:2004.13889*, 2020.
- [54] Anders Søgaard, Sebastian Ruder, and Ivan Vulić. On the limitations of unsupervised bilingual dictionary induction. *arXiv preprint arXiv:1805.03620*, 2018.
- [55] Long Duong, Hiroshi Kanayama, Tengfei Ma, Steven Bird, and Trevor Cohn. Learning crosslingual word embeddings without bilingual corpora. *arXiv preprint arXiv:1606.09403*, 2016.