Unsupervised Cross-Lingual Representation Learning

Sebastian Ruder[1], Anders Søgaard[2], Ivan Vulić[3]
[1] DeepMind, London
[2] Department of Computer Science, University of Copenhagen
[3] Language Technology Lab, University of Cambridge
sebastian@ruder.io   soegaard@di.ku.dk   iv250@cam.ac.uk

1 Motivation and Objectives

Cross-lingual word representations offer an elegant and language-pair independent way to represent content across different languages. They enable us to reason about *word meaning in multilingual contexts* and serve as an integral source of knowledge for multilingual applications such as machine translation (Artetxe et al., 2018d; Qi et al., 2018; Lample et al., 2018b) or multilingual search and question answering (Vulić and Moens, 2015). In addition, they are a key facilitator of *cross-lingual transfer* and *joint multilingual training*, offering support to NLP applications in a large spectrum of languages (Søgaard et al., 2015; Ammar et al., 2016a). While NLP is increasingly more embedded into a variety of products related to, e.g., translation, conversational or search tasks, resources such as annotated training data are still lacking or insufficient to induce satisfying models for many resource-poor languages. There are often no trained linguistic annotators for these languages, and markets may be too small or premature to invest in such training. This is a major challenge, but cross-lingual modelling and transfer can help by exploiting observable correlations between major languages and low-resource languages.

Recent work has already verified the usefulness of cross-lingual word representations in a wide variety of downstream tasks, and has provided extensive model classifications in several survey papers (Upadhyay et al., 2016; Ruder et al., 2018b). They cluster *supervised* cross-lingual word representation models according to the *bilingual supervision* required to induce such shared cross-lingual semantic spaces, covering models based on word alignments and readily available bilingual dictionaries (Mikolov et al., 2013; Smith et al., 2017), sentence-aligned parallel data (Gouws et al., 2015), document-aligned data (Søgaard et al., 2015; Vulić

and Moens, 2016), or even image tags and captions (Rotman et al., 2018). The current trend (or rather 'obsession') in cross-lingual word embedding learning, however, concerns models that require a tiny amount of supervision (i.e., *weakly-supervised alignment* models that require only dozens of word translation pairs) or no supervision at all (*fully unsupervised* models).[1] Such resource-light unsupervised methods are based on the assumption that monolingual word vector spaces are approximately isomorphic (Conneau et al., 2018a). Therefore, they require only monolingual data and hold promise to enable cross-lingual NLP modeling in the absence of any bilingual resources. As a consequence, they offer support to a wider array of language pairs than supervised models, and promise to deliver language technology to truly resource-poor languages and dialects. However, due to the strong assumption on the similarity of space topology, these models often diverge to non-optimal solutions, and their robustness is one of the crucial research questions at present (Søgaard et al., 2018).

In this tutorial, we provide a comprehensive survey of the exciting recent work on cutting-edge weakly-supervised and unsupervised cross-lingual word representations. After providing a brief history of supervised cross-lingual word representations, we focus on: **1)** how to induce weakly-supervised and unsupervised cross-lingual word representations in truly resource-poor settings where bilingual supervision cannot be guaranteed; **2)** critical examinations of different training conditions and requirements under which unsupervised algorithms can and cannot work effectively; **3)** more robust methods for distant language pairs that

---

[1] Learning unsupervised cross-lingual models has indeed taken the field by storm: there are 10+ papers on this very topic published in EMNLP 2018 proceedings alone, with even more papers available on arXiv.

can mitigate instability issues and low performance for distant language pairs; **4)** how to comprehensively evaluate such representations; and **5)** diverse applications that benefit from cross-lingual word representations (e.g., MT, dialogue, cross-lingual sequence labeling and structured prediction applications, cross-lingual IR).

We will introduce researchers to state-of-the-art methods for constructing resource-light cross-lingual word representations and discuss their applicability in a broad range of downstream NLP applications, covering bilingual lexicon induction, machine translation (both neural and phrase-based), dialogue, and information retrieval tasks. We will deliver a detailed survey of the current cutting-edge methods, discuss best training and evaluation practices and use-cases, and provide links to publicly available implementations, datasets, and pretrained models and word embedding collections.[2]

2 Tutorial Overview

Part I: Introduction We first present an overview of cross-lingual NLP research, situating the current work on unsupervised cross-lingual representation learning, and motivating the need for multilingual training and cross-lingual transfer for resource-poor languages with weak supervision or no bilingual supervision at all. We also present key downstream applications for cross-lingual word representations, such as bilingual lexicon induction and unsupervised MT (Lample et al., 2018b). These tasks will be used throughout the tutorial to analyze the performance of different methods.

Almost all of the work on unsupervised cross-lingual representation learning fall into the category of *mapping-based approaches* (Ruder et al., 2018b). Such approaches to cross-lingual learning learn mapping functions between pretrained monolingual word embedding spaces; this is in contrast with approaches based on joint learning, data augmentation, or grounding. We show that such approaches to cross-lingual learning, while so far unexplored, can also be unsupervised. We will put focus on a standardized two-step mapping-based framework (Artetxe et al., 2018a) that generalizes all mapping-based approaches, and analyze the importance of each component of the framework. The two-step framework decomposes unsupervised cross-lingual representation learning into

initial seed induction and iterative supervised bootstrapping.

Part II: Unsupervised and Weakly Supervised Alignment as Initial Seed Induction + Iterative Supervised Alignment We will analyze the impact of seed bilingual lexicon size and quality (e.g., cognates, named entities, or shared numerals) on the quality of weakly supervised cross-lingual word representations. Unsupervised and weakly supervised approaches can be directly compared by compared the quality of the learned dictionary seeds (Parts III and IV) to using cognates, named entities, etc.

Part III: Adversarial Seed Induction The underlying modus operandi of all adversarial methods will be demonstrated on the example of the MUSE architecture (Conneau et al., 2018a); this is by far the most cited adversarial seed induction method. We will then present similar adversarial methods and discuss their modeling choices, implementation tricks, and various trade-offs. We will also present our own direct comparisons of various GAN algorithms (e.g., WGAN, GP-WGAN, and CT-GAN) within the MUSE framework.

Part IV: Non-Adversarial Seed Induction In the next part, we will present several non-adversarial alternatives for unsupervised seed induction based on convex relaxations, point set registration methods, and evolutionary strategies. We will again dissect all components of the unsupervised methods and point to minor, but important implementation tricks and hyper-parameters that often slip under the radar (e.g., vocabulary size, post-mapping refinements, preprocessing steps such as mean centering and unit length normalisation, selected semantic similarity measures, hubness reduction mechanisms). We will also introduce the newest research that extends these methods from bilingual settings to multilingual settings (with more than 2 languages represented in the same shared space).

Part V: Stochastic Dictionary Induction improves Iterative Alignment We will then discuss stochastic approaches to improve the iterative refinement of the dictionary. Stochastic dictionary induction was introduced in Artetxe et al. (2018b), and we show that this bootstrapping technique improves performance and robustness, and is the main reason Artetxe et al. (2018b) achieves state-of-the-

art performance for many language pairs. This part of our tutorial explores variation of stochastic dictionary induction.

Part VI: Robustness and (In)stability Unsupervised methods rely on the assumption that monolingual word vector spaces are approximately isomorphic and there exists a linear mapping between the two spaces. This assumption is not true for many cases, which leads to degenerate or suboptimal solutions. The efficacy and stability of unsupervised methods relies on multiple factors such as: monolingual representation models, domain (dis)similarity, language pair proximity and other typological properties, chosen hyper-parameters, etc. In this part, we will analyze the current problems with robustness and stability of weakly-supervised and unsupervised alignment methods in relation to all these factors, and introduce latest solutions to alleviate these problems. We will provide advice on how to approach weakly-supervised and unsupervised training based on a series of empirical observations available in recent literature (Søgaard et al., 2018; Hartmann et al., 2018). We will also discuss the (im)possibility of learning non-linear mappings using either non-linear generators or locally linear maps (Nakashole, 2018).

We will **conclude** by providing publicly available software packages and implementations, as well as available training datasets and evaluation protocols and systems. We will also list current state-of-the-art results on standard evaluation datasets, and sketch future research paths.

3 Outline

Part I: Introduction: Motivating and situating cross-lingual word representation learning; presentation of mapping-based approaches *(30 minutes)*

- Current challenges in cross-lingual NLP. NLP for resource-poor languages.

- Bilingual data and cross-lingual supervision. Why do we need weakly supervised and unsupervised cross-lingual representation learning?

- Bilingual supervision and typology of supervised cross-lingual representation models.

- Learning with word-level supervision: mapping-based approaches.

Part II: Unsupervised and Weakly Supervised Alignment as Initial Seed Induction + Iterative Supervised Alignment *(30 minutes)*

- A general framework for mapping-based approaches.

- Importance of seed bilingual lexicons.

- Learning alignment with weak supervision: small seed lexicons, shared words, numerals.

Part III: Adversarial Seed Induction *(30 minutes)*

- Fully unsupervised models using adversarial training; MUSE and related approaches.

Part IV: Non-Adversarial Seed Induction *(25 minutes)*

- Fully unsupervised models using optimal transport, Wasserstein distance, Sinkhorn distance, and other alternatives.

- Importance of minor technical "tricks": pre-mapping and post-mapping steps: length normalisation, mean centering, whitening and de-whitening, making the methods more robust

Part V: Stochastic Dictionary Induction improves Iterative Alignment *(15 minutes)*

- An overview of methods to improve iterative refinement of the dictionary.

Part VI: Robustness and (In)stability *(35 minutes)*

- Impact of language similarity and typological properties.

- Impact of chosen monolingual models, domain similarity, and hyper-parameters.

- Convergence criteria, possible and impossible setups for unsupervised methods.

- How to build more robust and more stable unsupervised methods?

Discussion and Final Remarks *(15 minutes)*

- Towards cross-lingual contextualised word embeddings.

- Publicly available software and training data.

- Publicly available evaluation systems.

- Concluding remarks, remaining challenges, future work, a short discussion.

4 Tutorial Breadth

Based on the representative set of papers listed in the selected bibliography, we anticipate that the 75%-80% of the tutorial will cover other people's work, while the rest concerns the work where at least one of the three presenters has been actively involved in. Note that the three presenters have been the main authors of the recent book on cross-lingual word representations which aimed at making a systematic overview of the field.

5 Prerequisites

- *Machine Learning*: Basic knowledge of common neural network components like word embeddings, RNNs, CNNs, denoising autoencoders, and encoder-decoder models.

- *Computational Linguistics*: Familiarity with standard NLP tasks such as machine translation.

6 Presenters

Ivan Vulić, PhD, Senior Research Associate, Language Technology Lab, University of Cambridge. 9 West Road, CB3 9DP, Cambridge, UK; Senior Scientist, PolyAI, London, UK. iv250@cam.ac.uk. Ivan is interested in representation learning, distributional, lexical, and multi-modal semantics in monolingual and multilingual contexts, and transfer learning for enabling cross-lingual NLP applications. His work has been published at top-tier *ACL and *IR conferences. Ivan co-lectured a tutorial on multilingual topic models and applications at ECIR 2013 and WSDM 2014, a tutorial on cross-lingual word representations at EMNLP 2017, and a tutorial on language understanding for conversational AI at NAACL 2018. He also co-organised a workshop on Vision and Language at EMNLP 2015 and co-organises the ACL 2019 workshop on linguistic typology for cross-lingual NLP. He serves as an area chair for the multilinguality track at NAACL 2019 and word-level semantics at ACL 2019.

Sebastian Ruder is a research scientist at DeepMind. His research focuses on transfer learning in NLP and transferring models to low-resource languages. He has published widely read reviews of related areas, co-organised the NLP Session at the Deep Learning Indaba 2018, and co-organises the ACL 2019 workshop on representation learning and the European NLP Summit 2019 (EurNLP-2019).

Anders Søgaard, PhD, Dr.Phil, Full Professor in NLP and Machine Learning, Department of Computer Science, University of Copenhagen. soegaard@di.ku.dk. Anders is interested in machine learning for NLP. He currently holds a Google Focused Research Award and a Facebook Research Award and has won best paper awards at NAACL, EACL, CoNLL and *SEM. He has previously given tutorials at COLING and EMNLP, as well as an ESSLLI course. He has been an area chair of many top NLP/AI conferences.

The presenters have recently published a handbook for Morgan & Claypool on *cross-lingual word embeddings*.

7 Other Important Information

Previous Tutorial Editions The EMNLP 2017 tutorial on cross-lingual word embeddings presented much of the earlier work from 2013-2016 that require large amounts of parallel data (i.e., supervised cross-lingual representations). In contrast, this tutorial focuses on cutting-edge unsupervised and weakly supervised approaches from the period of 2016-2018, which will be highly relevant to the audience, and will provide a complete overview of the current cutting-edge research in the field.

Acknowledgments

Selected Bibliography

Oliver Adams, Adam Makarucha, Graham Neubig, Steven Bird, and Trevor Cohn. 2017. Cross-lingual word embeddings for low-resource language modeling. In *Proceedings of EACL*, pages 937–947.

Jean Alaux, Edouard Grave, Marco Cuturi, and Armand Joulin. 2018. Unsupervised hyperalignment for multilingual word embeddings. *CoRR*, abs/1811.01124.

Hanan Aldarmaki, Mahesh Mohan, and Mona Diab. 2018. Unsupervised word mapping using structural similarities in monolingual embeddings. *Transactions of the ACL*, 6:185–196.

David Alvarez-Melis and Tommi Jaakkola. 2018. Gromov-Wasserstein alignment of word embedding spaces. In *Proceedings of EMNLP*, pages 1881–1890.

David Alvarez-Melis, Tommi S. Jaakkola, and Stefanie Jegelka. 2018. Structured optimal transport. In *Proceedings of AISTATS*, pages 1771–1780.

Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah Smith. 2016a. Many languages, one parser. *Transactions of the ACL*, 4:431–444.

Waleed Ammar, George Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A. Smith. 2016b. Massively multilingual word embeddings. *CoRR*, abs/1602.01925.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2016. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In *Proceedings of EMNLP*, pages 2289–2294.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of ACL*, pages 451–462.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018a. Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In *Proceedings of AAAI*, pages 5012–5019.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018b. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of ACL*, pages 789–798.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018c. Unsupervised statistical machine translation. In *Proceedings of EMNLP*, pages 3632–3642.

Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2018d. Unsupervised neural machine translation. In *Proceedings of ICLR*.

Antonio Valerio Miceli Barone. 2016. Towards cross-lingual distributed representations without parallel text trained with adversarial autoencoders. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 121–126.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the ACL*, 5:135–146.

Chloé Braud, Ophélie Lacroix, and Anders Søgaard. 2017. Cross-lingual and cross-domain discourse segmentation of entire documents. In *Proceedings of ACL*, pages 237–243.

Jose Camacho-Collados, Mohammad Taher Pilehvar, Nigel Collier, and Roberto Navigli. 2017. SemEval-2017 Task 2: Multilingual and cross-lingual semantic word similarity. In *Proceedings of SEMEVAL*, pages 15–26.

Xilun Chen, Ahmed Hassan Awadallah, Hany Hassan, Wei Wang, and Claire Cardie. 2018. Zero-resource multilingual model transfer: Learning what to share.

Xilun Chen and Claire Cardie. 2018. Unsupervised multilingual word embeddings. In *Proceedings of EMNLP*, pages 261–270.

Ta Chung Chi and Yun-Nung Chen. 2018. CLUSE: Cross-lingual unsupervised sense embeddings. In *Proceedings of EMNLP*, pages 271–281.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018a. Word translation without parallel data. In *Proceedings of ICLR (Conference Track)*.

Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel Bowman, Holger Schwenk, and Veselin Stoyanov. 2018b. XNLI: Evaluating cross-lingual sentence representations. In *Proceedings of EMNLP*, pages 2475–2485.

Georgiana Dinu, Angeliki Lazaridou, and Marco Baroni. 2015. Improving zero-shot learning by mitigating the hubness problem. In *Proceedings of ICLR (Workshop Papers)*, volume abs/1412.6568.

Zi-Yi Dou, Zhi-Hao Zhou, and Shujian Huang. 2018. Unsupervised bilingual lexicon induction via latent variable models. In *Proceedings of EMNLP*, pages 621–626.

Yerai Doval, Jose Camacho-Collados, Luis Espinosa Anke, and Steven Schockaert. 2018. Improving cross-lingual word embeddings by meeting in the middle. In *Proceedings of EMNLP*, pages 294–304.

Philipp Dufter and Hinrich Schütze. 2018. A stronger baseline for multilingual word embeddings. *CoRR*, abs/1811.00586.

Philipp Dufter, Mengjie Zhao, Martin Schmitt, Alexander Fraser, and Hinrich Schütze. 2018. Embedding learning through multilingual concept induction. In *Proceedings of ACL*, pages 1520–1530.

Long Duong, Trevor Cohn, Steven Bird, and Paul Cook. 2015. Cross-lingual transfer for unsupervised dependency parsing without parallel data. In *Proceedings of CoNLL*, pages 113–122.

Long Duong, Hiroshi Kanayama, Tengfei Ma, Steven Bird, and Trevor Cohn. 2016. Learning crosslingual word embeddings without bilingual corpora. In *Proceedings of EMNLP*, pages 1285–1295.

Greg Durrett, Adam Pauls, and Dan Klein. 2012. Syntactic transfer using a bilingual lexicon. In *Proceedings of EMNLP*, pages 1–11.

Goran Glavaš and Ivan Vulić. 2018a. Discriminating between lexico-semantic relations with the specialization tensor model. In *Proceedings of NAACL-HLT*, pages 181–187.

Goran Glavaš and Ivan Vulić. 2018b. Explicit retrofitting of distributional word vectors. In *Proceedings of ACL*, pages 34–45.

Goran Glavaš, Robert Litschko, Sebastian Ruder, and Ivan Vulić. 2019. How to (properly) evaluate cross-lingual word embeddings: On strong baselines, comparative analyses, and some misconceptions. In *Proceedings of ACL*.

Denis Gordeev, Alexey Rey, and Dmitry Shagarov. 2018. Unsupervised cross-lingual matching of product classifications. In *Proceedings of the 23rd Conference of Open Innovations Association (FRUCT)*, pages 62:459–62:464.

Stephan Gouws, Yoshua Bengio, and Greg Corrado. 2015. BilBOWA: Fast bilingual distributed representations without word alignments. In *Proceedings of ICML*, pages 748–756.

Edouard Grave, Armand Joulin, and Quentin Berthet. 2018. Unsupervised alignment of embeddings with Wasserstein Procrustes. *CoRR*, abs/1805.11222.

Jiang Guo, Wanxiang Che, David Yarowsky, Haifeng Wang, and Ting Liu. 2015. Cross-lingual dependency parsing based on distributed representations. In *Proceedings of ACL*, pages 1234–1244.

Mareike Hartmann, Yova Kementchedjhieva, and Anders Søgaard. 2018. Why is unsupervised alignment of English embeddings from different algorithms so hard? In *Proceedings of EMNLP*, pages 582–586.

Tatsunori Hashimoto, David Alvarez-Melis, and Tommi Jaakkola. 2016. Word embeddings as metric recovery in semantic spaces. *Transactions of the ACL*, 4:273–286.

John Hewitt, Daphne Ippolito, Brendan Callahan, Reno Kriz, Derry Tanti Wijaya, and Chris Callison-Burch. 2018. Learning translations via images with a massively multilingual image dataset. In *Proceedings of ACL*, pages 2566–2576.

Geert Heyman, Bregt Verreet, Ivan Vulić, and Marie-Francine Moens. 2019. Learning unsupervised multilingual word embeddings with incremental multilingual hubs. In *Proceedings of NAACL-HLT*, pages 1890–1902.

Geert Heyman, Ivan Vulić, and Marie-Francine Moens. 2017. Bilingual lexicon induction by learning to combine word-level and character-level representations. In *Proceedings of EACL*, pages 1085–1095.

Yedid Hoshen and Lior Wolf. 2018. Non-adversarial unsupervised word translation. In *Proceedings of EMNLP*, pages 469–478.

Kejun Huang, Matt Gardner, Evangelos Papalexakis, Christos Faloutsos, Nikos Sidiropoulos, Tom Mitchell, Partha P. Talukdar, and Xiao Fu. 2015. Translation invariant word embeddings. In *Proceedings of EMNLP*, pages 1084–1088.

Lifu Huang, Kyunghyun Cho, Boliang Zhang, Heng Ji, and Kevin Knight. 2018. Multi-lingual common semantic space construction via cluster-consistent word embedding. In *Proceedings of EMNLP*, pages 250–260.

Pratik Jawanpuria, Arjun Balgovind, Anoop Kunchukuttan, and Bamdev Mishra. 2019. Learning multilingual word embeddings in latent metric space: A geometric approach. *Transactions of the ACL*.

Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. 2018. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *Proceedings of EMNLP*, pages 2979–2984.

Yova Kementchedjhieva, Sebastian Ruder, Ryan Cotterell, and Anders Søgaard. 2018. Generalizing Procrustes analysis for better bilingual dictionary induction. In *Proceedings of CoNLL*, pages 211–220.

Yunsu Kim, Jiahui Geng, and Hermann Ney. 2018. Improving unsupervised word-by-word translation with language model and denoising autoencoder. In *Proceedings of EMNLP*, pages 862–868.

Guillaume Lample, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018a. Unsupervised machine translation using monolingual corpora only. In *Proceedings of ICLR*.

Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018b. Phrase-based & neural unsupervised machine translation. In *Proceedings of EMNLP*, pages 5039–5049.

Angeliki Lazaridou, Georgiana Dinu, and Marco Baroni. 2015. Hubness and pollution: Delving into cross-space mapping for zero-shot learning. In *Proceedings of ACL*, pages 270–280.

Omer Levy, Anders Søgaard, and Yoav Goldberg. 2017. A strong baseline for learning cross-lingual word embeddings from sentence alignments. In *Proceedings of EACL*, pages 765–774.

Robert Litschko, Goran Glavaš, Simone Paolo Ponzetto, and Ivan Vulić. 2018. Unsupervised cross-lingual information retrieval using monolingual data only. In *Proceedings of SIGIR*, pages 1253–1256.

Stephen Mayhew, Chen-Tse Tsai, and Dan Roth. 2017. Cheap translation for cross-lingual named entity recognition. In *Proceedings of EMNLP*, pages 2536–2545.

Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of EMNLP*, pages 62–72.

Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. 2013. Exploiting similarities among languages for machine translation. *CoRR*, abs/1309.4168.

Tasnim Mohiuddin and Shafiq Joty. 2019. Revisiting adversarial autoencoder for unsupervised word translation with cycle consistency and improved training. In *Proceedings of NAACL-HLT*, pages 3857–3867.

Nikola Mrkšić, Ivan Vulić, Diarmuid Ó Séaghdha, Ira Leviant, Roi Reichart, Milica Gašić, Anna Korhonen, and Steve Young. 2017. Semantic specialization of distributional word vector spaces using monolingual and cross-lingual constraints. *Transactions of the ACL*, 5:309–324.

Tanmoy Mukherjee, Makoto Yamada, and Timothy Hospedales. 2018. Learning unsupervised word translations without adversaries. In *Proceedings of EMNLP*, pages 627–632.

Ndapa Nakashole. 2018. NORMA: Neighborhood sensitive maps for multilingual word embeddings. In *Proceedings of EMNLP*, pages 512–522.

Ndapa Nakashole and Raphael Flauger. 2018. Characterizing departures from linearity in word translation. In *Proceedings of ACL*, pages 221–227.

Ndapandula Nakashole and Raphael Flauger. 2017. Knowledge distillation for bilingual dictionary induction. In *Proceedings of EMNLP*, pages 2497–2506.

Yves Peirsman and Sebastian Padó. 2010. Cross-lingual induction of selectional preferences with bilingual vector spaces. In *Proceedings of NAACL-HLT*, pages 921–929.

Edoardo Maria Ponti, Ivan Vulić, Goran Glavaš, Nikola Mrkšić, and Anna Korhonen. 2018. Adversarial propagation and zero-shot cross-lingual transfer of word vector specialization. In *Proceedings of EMNLP*, pages 282–293.

Ye Qi, Devendra Sachan, Matthieu Felix, Sarguna Padmanabhan, and Graham Neubig. 2018. When and why are pre-trained word embeddings useful for neural machine translation? In *Proceedings of NAACL-HLT*, pages 529–535.

Guy Rotman, Ivan Vulić, and Roi Reichart. 2018. Bridging languages through images with Deep Partial Canonical Correlation Analysis. In *Proceedings of ACL*, pages 910–921.

Sebastian Ruder, Ryan Cotterell, Yova Kementchedjhieva, and Anders Søgaard. 2018a. A discriminative latent-variable model for bilingual lexicon induction. In *Proceedings of EMNLP*, pages 458–468.

Sebastian Ruder, Ivan Vulić, and Anders Søgaard. 2018b. A survey of cross-lingual embedding models. *Journal of Artificial Intelligence Research*.

Tal Schuster, Ori Ram, Regina Barzilay, and Amir Globerson. 2019. Cross-lingual alignment of contextual word embeddings, with applications to zero-shot dependency parsing. In *Proceedings of NAACL-HLT*, pages 1599–1613.

Sidak Pal Singh, Andreas Hug, Aymeric Dieuleveut, and Martin Jaggi. 2018. Wasserstein is all you need. *CoRR*, abs/1808.09663.

Karan Singla, Dogan Can, and Shrikanth Narayanan. 2018. A multi-task approach to learning multilingual representations. In *Proceedings of ACL*, pages 214–220.

Samuel L. Smith, David H.P. Turban, Steven Hamblin, and Nils Y. Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. In *Proceedings of ICLR (Conference Track)*.

Anders Søgaard, Željko Agić, Héctor Martínez Alonso, Barbara Plank, Bernd Bohnet, and Anders Johannsen. 2015. Inverted indexing for cross-lingual NLP. In *Proceedings of ACL*, pages 1713–1722.

Anders Søgaard, Sebastian Ruder, and Ivan Vulić. 2018. On the limitations of unsupervised bilingual dictionary induction. In *Proceedings of ACL*, pages 778–788.

Chen-Tse Tsai and Dan Roth. 2016. Cross-lingual wikification using multilingual embeddings. In *Proceedings of NAACL-HLT*, pages 589–598.

Shyam Upadhyay, Manaal Faruqui, Chris Dyer, and Dan Roth. 2016. Cross-lingual models of word embeddings: An empirical comparison. In *Proceedings of ACL*, pages 1661–1670.

Shyam Upadhyay, Manaal Faruqui, Gökhan Tür, Dilek Z. Hakkani-Tür, and Larry P. Heck. 2018a. (almost) zero-shot cross-lingual spoken language understanding. In *Proceedings of ICASSP*, pages 6034–6038.

Shyam Upadhyay, Nitish Gupta, and Dan Roth. 2018b. Joint multilingual supervision for cross-lingual entity linking. In *Proceedings of EMNLP*, pages 2486–2495.

Shyam Upadhyay, Yogarshi Vyas, Marine Carpuat, and Dan Roth. 2018c. Robust cross-lingual hypernymy detection using dependency context. In *Proceedings of NAACL-HLT*, pages 607–618.

Ivan Vulić and Anna Korhonen. 2016. On the role of seed lexicons in learning bilingual word embeddings. In *Proceedings of ACL*, pages 247–257.

Ivan Vulic and Marie-Francine Moens. 2013. A study on bootstrapping bilingual vector spaces from non-parallel data (and nothing else). In *Proceedings of EMNLP*, pages 1613–1624.

Ivan Vulić and Marie-Francine Moens. 2015. Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. In *Proceedings of SIGIR*, pages 363–372.

Ivan Vulić and Marie-Francine Moens. 2016. Bilingual distributed word representations from document-aligned comparable data. *Journal of Artificial Intelligence Research*, 55:953–994.

Ivan Vulić, Nikola Mrkšić, and Anna Korhonen. 2017. Cross-lingual induction and transfer of verb classes based on word vector space specialisation. In *Proceedings of EMNLP*, pages 2536–2548.

Takashi Wada and Tomoharu Iwata. 2018. Unsupervised cross-lingual word embedding by multilingual neural language models. *CoRR*, abs/1809.02306.

Ruochen Xu, Yiming Yang, Naoki Otani, and Yuexin Wu. 2018. Unsupervised cross-lingual transfer of word embedding spaces. In *Proceedings of EMNLP*, pages 2465–2474.

Pengcheng Yang, Fuli Luo, Shuangzhi Wu, Jingjing Xu, Dongdong Zhang, and Xu Sun. 2018a. Learning unsupervised word mapping by maximizing mean discrepancy. *CoRR*, abs/1811.00275.

Zhen Yang, Wei Chen, Feng Wang, and Bo Xu. 2018b. Unsupervised neural machine translation with weight sharing. In *Proceedings of ACL*, pages 46–55.

Meng Zhang, Yang Liu, Huanbo Luan, Yiqun Liu, and Maosong Sun. 2016a. Inducing bilingual lexica from non-parallel data with earth mover's distance regularization. In *Proceedings of COLING*, pages 3188–3198.

Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. 2017a. Adversarial training for unsupervised bilingual lexicon induction. In *Proceedings of ACL*, pages 1959–1970.

Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. 2017b. Earth Mover's distance minimization for unsupervised bilingual lexicon induction. In *Proceedings of EMNLP*, pages 1934–1945.

Mozhi Zhang, Keyulu Xu, Ken-ichi Kawarabayashi, Stefanie Jegelka, and Jordan Boyd-Graber. 2019. Are girls neko or shojo? Cross-lingual alignment of non-isomorphic embeddings with iterative normalization.

Yuan Zhang, David Gaddy, Regina Barzilay, and Tommi Jaakkola. 2016b. Ten pairs to tag – Multilingual POS tagging via coarse mapping between embeddings. In *Proceedings of NAACL-HLT*, pages 1307–1317.

Yftah Ziser and Roi Reichart. 2018. Deep pivot-based modeling for cross-language cross-domain transfer with minimal guidance. In *Proceedings of EMNLP*, pages 238–249.