Welcome to the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)! The workshop was co-located with the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), and was held on August 6, 2021 as an online workshop. The workshop was organised by Anna Rogers, Iacer Calixto, Ivan Vulić, Naomi Saphra, Nora Kassner, Oana-Maria Camburu, Trapit Bansal, and Vered Shwartz; and advised by Chris Dyer, Ed Grefenstette, Isabelle Augenstein, Karl Moritz Hermann, Kyunghyun Cho, and Laura Rimell. The workshop is annually organised by the ACL Special Interest Group on Representation Learning (SIGREP).

The 6th Workshop on Representation Learning for NLP aims to continue the success of the previous editions, and remains a strong and established venue for representation learning in NLP, attracting more than 60 submissions this year. This edition continued to invite papers of a theoretical or experimental nature describing recent advances in vector space models of meaning, compositionality, and the application of deep neural networks and spectral methods to NLP. A strong focus was put on topics of developing new representations, evaluating existing representations, efficient and sustainable learning and inference, and representation learning beyond the English language and text only (e.g., multi-modal, cross-lingual, knowledge-informed learning).

We take this opportunity to thank the RepL4NLP-2021 program committee for their help and thorough reviews. We also thank the authors who presented their work at the workshop, and the workshop participants for the valuable feedback and discussions. Finally, we are deeply honored to have four excellent talks from our invited speakers Sameer Singh, Karen Livescu, Noah Smith, and Lena Voita.

The RepL4NLP-2021 Workshop organizers
Organizers:
Anna Rogers, University of Copenhagen
Iacer Calixto, New York University & University of Amsterdam
Ivan Vulić, University of Cambridge & PolyAI
Naomi Saphra, New York University
Nora Kassner, LMU Munich
Oana-Maria Camburu, University of Oxford
Trapit Bansal, UMass Amherst
Vered Shwartz, Allen Institute for AI & University of Washington

Senior Advisers:
Chris Dyer, DeepMind
Edward Grefenstette, Facebook AI Research & University College London
Isabelle Augenstein, University of Copenhagen
Karl Moritz Hermann, Saiga
Kyunghyun Cho, New York University
Laura Rimell, DeepMind

Keynote Speakers:
Sameer Singh, University of California Irvine
Karen Livescu, Toyota Technological Institute at Chicago
Noah Smith, Allen Institute for AI & University of Washington
Lena Voita, University of Edinburgh & University of Amsterdam

Program Committee:
Muhammad Abdul-Mageed
Guy Aglionby
Roee Aharoni
Nicholas Andrews
Mikel Artetxe
Federico Bianchi
Andrew Caines
Helena Caseli
Yue Chen
Mingda Chen
Alexandra Chronopoulou
Manuel Ciosici
Paula Czarnowska
Giuseppe Antonio Di Luna
Philipp Dufter
Javid Ebrahimi
Vladimir Eidelman
Yanai Elazar
Sergey Feldman
Eraldo Fernandes
Elisabetta Fersini
Orhan Firat
Thibault Févry
Rainer Gemulla
# Table of Contents

*Improving Cross-lingual Text Classification with Zero-shot Instance-Weighting*
   Irene Li, Prithviraj Sen, Huaiyu Zhu, Yunyao Li and Dragomir Radev .......................... 1

*Probing Multilingual Language Models for Discourse*
   Murathan Kurfali and Robert Östling ........................................... 8

*Comprehension Based Question Answering using Bloom’s Taxonomy*
   Pritish Sahu, Michael Cogswell, Ajay Divakaran and Sara Rutherford-Quach .................. 20

*Larger-Scale Transformers for Multilingual Masked Language Modeling*
   Naman Goyal, Jingfei Du, Myle Ott, Giri Anantharaman and Alexis Conneau ........................ 29

*Learning Sparse Sentence Encoding without Supervision: An Exploration of Sparsity in Variational Autoencoders*
   Victor Prokhorov, Yingzhen Li, Ehsan Shareghi and Nigel Collier ............................. 34

*Temporal-aware Language Representation Learning From Crowdsourced Labels*
   Yang Hao, Xiao Zhai, Wenbiao Ding and Zitao Liu ........................................... 47

*Structure-aware Sentence Encoder in Bert-Based Siamese Network*
   Qiwei Peng, David Weir and Julie Weeds ....................................................... 57

*Preserving Cross-Linguality of Pre-trained Models via Continual Learning*
   Zihan Liu, Genta Indra Winata, Andrea Madotto and Pascale Fung ............................. 64

*Text Style Transfer: Leveraging a Style Classifier on Entangled Latent Representations*
   Xiaoyan Li, Sun Sun and Yunli Wang ..................................................... 72

*Inductively Representing Out-of-Knowledge-Graph Entities by Optimal Estimation Under Translational Assumptions*
   Damai Dai, Hua Zheng, Fuli Luo, Pengcheng Yang, Tianyu Liu, Zhifang Sui and Baobao Chang 83

*Revisiting Pretraining with Adapters*
   Seungwon Kim, Alex Shum, Nathan Susanj and Jonathan Hilgart ............................... 90

*Knodle: Modular Weakly Supervised Learning with PyTorch*
   Anastasiia Sedova, Andreas Stephan, Marina Speranskaya and Benjamin Roth .................. 100

*X2Parser: Cross-lingual and Cross-Domain Framework for Task-Oriented Compositional Semantic Parsing*
   Zihan Liu, Genta Indra Winata, Peng Xu and Pascale Fung .......................................... 112

*Unsupervised Representation Disentanglement of Text: An Evaluation on Synthetic Datasets*
   Lan Zhang, Victor Prokhorov and Ehsan Shareghi .............................................. 128

*Learn The Big Picture: Representation Learning for Clustering*
   Sumanta Kashyapi and Laura Dietz ......................................................... 141

*Probing Cross-Modal Representations in Multi-Step Relational Reasoning*
   Iuliia Parfenova, Desmond Elliott, Raquel Fernández and Sandro Pezzelle .................... 152

*In-Batch Negatives for Knowledge Distillation with Tightly-Coupled Teachers for Dense Retrieval*
   Sheng-Chieh Lin, Jheng-Hong Yang and Jimmy Lin .............................................. 163
NPVec1: Word Embeddings for Nepali - Construction and Evaluation
Pravesh Koirala and Nobal B. Niraula .......................................................... 174

Deriving Word Vectors from Contextualized Language Models using Topic-Aware Mention Selection
Yixiao Wang, Zied Bouraoui, Luis Espinosa Anke and Steven Schockaert .................. 185

Zero-shot Sequence Labeling for Transformer-based Sentence Classifiers
Kamil Bujel, Helen Yannakoudakis and Marek Rei ........................................... 195

Predicting the Success of Domain Adaptation in Text Similarity
Nick Pogrebnyakov and Shohreh shaghaghan .......................... 206

Syntagmatic Word Embeddings for Unsupervised Learning of Selectional Preferences
Renjith P. Ravindran, Akshay Badola and Narayana Kavi Murthy ....................... 213

Bayesian Model-Agnostic Meta-Learning with Matrix-Valued Kernels for Quality Estimation
Abiola Obamuyide, Marina Fomicheva and Lucia Specia                              223

Knowledge Informed Semantic Parsing for Conversational Question Answering
Raghuveer Thirukovalluru, Mukund Sridhar, Dung Thai, Shruti Chanumolu, Nicholas Monath,
Sankaranarayanan Ananthakrishnan and Andrew McCallum .................................. 231

Simultaneously Self-Attending to Text and Entities for Knowledge-Informed Text Representations
Dung Thai, Raghuveer Thirukovalluru, Trapit Bansal and Andrew McCallum .................. 241

Deriving Contextualised Semantic Features from BERT (and Other Transformer Model) Embeddings
Jacob Turton, Robert Elliott Smith and David Vinson .................................... 248

Syntactic Perturbations Reveal Representational Correlates of Hierarchical Phrase Structure in Pre-trained Language Models
Matteo Alleman, Jonathan Mamou, Miguel A Del Rio, Hanlin Tang, Yoon Kim and SueYeon Chung

Box-To-Box Transformations for Modeling Joint Hierarchies
Shib Sankar Dasgupta, Xiang Lorraine Li, Michael Boratko, Dongxu Zhang and Andrew McCallum

An Overview of Uncertainty Calibration for Text Classification and the Role of Distillation
Han Guo, Ramakanth Pasunuru and Mohit Bansal .................................. 289

Entity and Evidence Guided Document-Level Relation Extraction
Kevin Huang, Peng Qi, Guangtao Wang, Tengyu Ma and Jing Huang ...................... 307

Scaling Deep Contrastive Learning Batch Size under Memory Limited Setup
Luyu Gao, Yunyi Zhang, Jiawei Han and Jamie Callan ................................ 316

Direction is what you need: Improving Word Embedding Compression in Large Language Models
Klaudia Bałazy, Mohammadreza Banaei, Rémi Lebret, Jacek Tabor and Karl Aberer .......... 322

x
Conference Program

*Improving Cross-lingual Text Classification with Zero-shot Instance-Weighting*
Irene Li, Prithviraj Sen, Huaiyu Zhu, Yunyao Li and Dragomir Radev

*Probing Multilingual Language Models for Discourse*
Murathan Kurfalı and Robert Östling

*Comprehension Based Question Answering using Bloom’s Taxonomy*
Pritish Sahu, Michael Cogswell, Ajay Divakaran and Sara Rutherford-Quach

*Larger-Scale Transformers for Multilingual Masked Language Modeling*
Naman Goyal, Jingfei Du, Myle Ott, Giri Anantharaman and Alexis Conneau

*Learning Sparse Sentence Encoding without Supervision: An Exploration of Sparsity in Variational Autoencoders*
Victor Prokhorov, Yingzhen Li, Ehsan Shareghi and Nigel Collier

*Temporal-aware Language Representation Learning From Crowdsourced Labels*
Yang Hao, Xiao Zhai, Wenbiao Ding and Zitao Liu

*Structure-aware Sentence Encoder in Bert-Based Siamese Network*
Qiwei Peng, David Weir and Julie Weeds

*Preserving Cross-Linguality of Pre-trained Models via Continual Learning*
Zihan Liu, Genta Indra Winata, Andrea Madotto and Pascale Fung

*Text Style Transfer: Leveraging a Style Classifier on Entangled Latent Representations*
Xiaoyan Li, Sun Sun and Yunli Wang

*Inductively Representing Out-of-Knowledge-Graph Entities by Optimal Estimation Under Translational Assumptions*
Damai Dai, Hua Zheng, Fuli Luo, Pengcheng Yang, Tianyu Liu, Zhifang Sui and Baobao Chang

*Revisiting Pretraining with Adapters*
Seungwon Kim, Alex Shum, Nathan Susanj and Jonathan Hilgart

*Knodle: Modular Weakly Supervised Learning with PyTorch*
Anastasiia Sedova, Andreas Stephan, Marina Speranskaya and Benjamin Roth
X2Parser: Cross-Lingual and Cross-Domain Framework for Task-Oriented Compositional Semantic Parsing
Zihan Liu, Genta Indra Winata, Peng Xu and Pascale Fung

Unsupervised Representation Disentanglement of Text: An Evaluation on Synthetic Datasets
Lan Zhang, Victor Prokhorov and Ehsan Shareghi

Learn The Big Picture: Representation Learning for Clustering
Sumanta Kashyapi and Laura Dietz

Probing Cross-Modal Representations in Multi-Step Relational Reasoning
Iuliia Parfenova, Desmond Elliott, Raquel Fernández and Sandro Pezzelle

In-Batch Negatives for Knowledge Distillation with Tightly-Coupled Teachers for Dense Retrieval
Sheng-Chieh Lin, Jheng-Hong Yang and Jimmy Lin

NPVec1: Word Embeddings for Nepali - Construction and Evaluation
Pravesh Koirala and Nobal B. Niraula

Deriving Word Vectors from Contextualized Language Models using Topic-Aware Mention Selection
Yixiao Wang, Zied Bouraoui, Luis Espinosa Anke and Steven Schockaert

Zero-shot Sequence Labeling for Transformer-based Sentence Classifiers
Kamil Bujel, Helen Yannakoudakis and Marek Rei

Predicting the Success of Domain Adaptation in Text Similarity
Nick Pogrebnyakov and Shohreh shaghaghian

Syntagmatic Word Embeddings for Unsupervised Learning of Selectional Preferences
Renjith P. Ravindran, Akshay Badola and Narayana Kavi Murthy

Bayesian Model-Agnostic Meta-Learning with Matrix-Valued Kernels for Quality Estimation
Abiola Obamuyide, Marina Fomicheva and Lucia Specia

Knowledge Informed Semantic Parsing for Conversational Question Answering
Raghuvir Thirukovalluru, Mukund Sridhar, Dung Thai, Shruti Chandumolu, Nicholas Monath, Sankaranarayanan Ananthakrishnan and Andrew McCallum
Simultaneously Self-Attending to Text and Entities for Knowledge-Informed Text Representations
Dung Thai, Raghuveer Thirukovalluru, Trapit Bansal and Andrew McCallum

Deriving Contextualised Semantic Features from BERT (and Other Transformer Model) Embeddings
Jacob Turton, Robert Elliott Smith and David Vinson

Syntactic Perturbations Reveal Representational Correlates of Hierarchical Phrase Structure in Pretrained Language Models
Matteo Alleman, Jonathan Mamou, Miguel A Del Rio, Hanlin Tang, Yoon Kim and Sue Yoon Chung

Box-To-Box Transformations for Modeling Joint Hierarchies
Shib Sankar Dasgupta, Xiang Lorraine Li, Michael Boratko, Dongxu Zhang and Andrew McCallum

An Overview of Uncertainty Calibration for Text Classification and the Role of Distillation
Han Guo, Ramakanth Pasunuru and Mohit Bansal

Entity and Evidence Guided Document-Level Relation Extraction
Kevin Huang, Peng Qi, Guangtao Wang, Tengyu Ma and Jing Huang

Scaling Deep Contrastive Learning Batch Size under Memory Limited Setup
Luyu Gao, Yunyi Zhang, Jiawei Han and Jamie Callan

Direction is what you need: Improving Word Embedding Compression in Large Language Models
Klaudia Balazy, Mohammadreza Banaei, Rémi Lebret, Jacek Tabor and Karl Aberer