

RepL4NLP 2021

**The 6th Workshop on Representation Learning for NLP**

**Proceedings of the Workshop**

August 6, 2021  
Bangkok, Thailand (online)

©2021 The Association for Computational Linguistics  
and The Asian Federation of Natural Language Processing

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-954085-72-5

## Introduction

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  
Roe 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

Eleonora Giunchiglia  
Hongyu Gong  
Kartik Goyal  
Robin Jia  
Aishwarya Kamath  
Santosh Kesiraju  
Ekaterina Kochmar  
Vid Kocijan  
Shankar Kumar  
Andrey Kutuzov  
Matthieu Labeau  
John P. Lalor  
Jey Han Lau  
Carolin Lawrence  
Xiang Lorraine Li  
Tao Li  
Yitong Li  
Bill Yuchen Lin  
Peng Liu  
Olga Majewska  
Sabrina Mielke  
Victor Milewski  
Jeff Mitchell  
Daichi Mochihashi  
Ashutosh Modi  
Nicholas Monath  
Lili Mou  
Maximilian Mozes  
Khalil Mrini  
Nikita Nangia  
Jason Naradowsky  
Thien Huu Nguyen  
Truc-Vien T. Nguyen  
Jekaterina Novikova  
Tsuyoshi Okita  
Ankur Padia  
Letitia Parcalabescu  
Matthew Peters  
Tiago Pimentel  
Vipul Raheja  
Surangika Ranathunga  
Abhilasha Ravichander  
Subendhu Rongali  
Frank Rudzicz  
Ehsan Shareghi  
Hao Tang  
Shuai Tang  
Dung Thai  
Jörg Tiedemann  
Nadi Tomeh  
Marcos Treviso

Adam Trischler  
Lifu Tu  
Shikhar Vashishth  
Hai Wang  
Hua Wang  
Rodrigo Wilkens  
Adina Williams  
Yuexin Wu  
Yuxiang Wu  
Wenpeng Yin  
Yordan Yordanov  
Hong Yu  
Wei Zhang  
Xiangyang Zhou  
Dong Zhou  
Yi Zhu  
Imed Zitouni  
Robert Östling





## Table of Contents

<i>Improving Cross-lingual Text Classification with Zero-shot Instance-Weighting</i> Irene Li, Prithviraj Sen, Huaiyu Zhu, Yunyao Li and Dragomir Radev .....	1
<i>Probing Multilingual Language Models for Discourse</i> Murathan Kurfalı and Robert Östling .....	8
<i>Comprehension Based Question Answering using Bloom’s Taxonomy</i> Pritish Sahu, Michael Cogswell, Ajay Divakaran and Sara Rutherford-Quach .....	20
<i>Larger-Scale Transformers for Multilingual Masked Language Modeling</i> Naman Goyal, Jingfei Du, Myle Ott, Giri Anantharaman and Alexis Conneau .....	29
<i>Learning Sparse Sentence Encoding without Supervision: An Exploration of Sparsity in Variational Autoencoders</i> Victor Prokhorov, Yingzhen Li, Ehsan Shareghi and Nigel Collier .....	34
<i>Temporal-aware Language Representation Learning From Crowdsourced Labels</i> Yang Hao, Xiao Zhai, Wenbiao Ding and Zitao Liu .....	47
<i>Structure-aware Sentence Encoder in Bert-Based Siamese Network</i> Qiwei Peng, David Weir and Julie Weeds .....	57
<i>Preserving Cross-Linguality of Pre-trained Models via Continual Learning</i> Zihan Liu, Genta Indra Winata, Andrea Madotto and Pascale Fung .....	64
<i>Text Style Transfer: Leveraging a Style Classifier on Entangled Latent Representations</i> Xiaoyan Li, Sun Sun and Yunli Wang .....	72
<i>Inductively Representing Out-of-Knowledge-Graph Entities by Optimal Estimation Under Translational Assumptions</i> Damai Dai, Hua Zheng, Fuli Luo, Pengcheng Yang, Tianyu Liu, Zhifang Sui and Baobao Chang .....	83
<i>Revisiting Pretraining with Adapters</i> Seungwon Kim, Alex Shum, Nathan Susanj and Jonathan Hilgart .....	90
<i>Knodle: Modular Weakly Supervised Learning with PyTorch</i> Anastasiia Sedova, Andreas Stephan, Marina Speranskaya and Benjamin Roth .....	100
<i>X2Parser: Cross-Lingual and Cross-Domain Framework for Task-Oriented Compositional Semantic Parsing</i> Zihan Liu, Genta Indra Winata, Peng Xu and Pascale Fung .....	112
<i>Unsupervised Representation Disentanglement of Text: An Evaluation on Synthetic Datasets</i> Lan Zhang, Victor Prokhorov and Ehsan Shareghi .....	128
<i>Learn The Big Picture: Representation Learning for Clustering</i> Sumanta Kashyapi and Laura Dietz .....	141
<i>Probing Cross-Modal Representations in Multi-Step Relational Reasoning</i> Iuliia Parfenova, Desmond Elliott, Raquel Fernández and Sandro Pezzelle .....	152
<i>In-Batch Negatives for Knowledge Distillation with Tightly-Coupled Teachers for Dense Retrieval</i> Sheng-Chieh Lin, Jheng-Hong Yang and Jimmy Lin .....	163

<i>NPVec1: Word Embeddings for Nepali - Construction and Evaluation</i> Pravesh Koirala and Nopal B. Niraula .....	174
<i>Deriving Word Vectors from Contextualized Language Models using Topic-Aware Mention Selection</i> Yixiao Wang, Zied Bouraoui, Luis Espinosa Anke and Steven Schockaert .....	185
<i>Zero-shot Sequence Labeling for Transformer-based Sentence Classifiers</i> Kamil Bujel, Helen Yannakoudakis and Marek Rei .....	195
<i>Predicting the Success of Domain Adaptation in Text Similarity</i> Nick Pogrebnjakov and Shohreh shaghighian .....	206
<i>Syntagmatic Word Embeddings for Unsupervised Learning of Selectional Preferences</i> Renjith P. Ravindran, Akshay Badola and Narayana Kavi Murthy .....	213
<i>Bayesian Model-Agnostic Meta-Learning with Matrix-Valued Kernels for Quality Estimation</i> Abiola Obamuyide, Marina Fomicheva and Lucia Specia .....	223
<i>Knowledge Informed Semantic Parsing for Conversational Question Answering</i> Raghuveer Thirukovalluru, Mukund Sridhar, Dung Thai, Shruti Chanumolu, Nicholas Monath, Sankaranarayanan Ananthakrishnan and Andrew McCallum .....	231
<i>Simultaneously Self-Attending to Text and Entities for Knowledge-Informed Text Representations</i> Dung Thai, Raghuveer Thirukovalluru, Trapit Bansal and Andrew McCallum .....	241
<i>Deriving Contextualised Semantic Features from BERT (and Other Transformer Model) Embeddings</i> Jacob Turton, Robert Elliott Smith and David Vinson .....	248
<i>Syntactic Perturbations Reveal Representational Correlates of Hierarchical Phrase Structure in Pre-trained Language Models</i> Matteo Alleman, Jonathan Mamou, Miguel A Del Rio, Hanlin Tang, Yoon Kim and SueYeon Chung	263
<i>Box-To-Box Transformations for Modeling Joint Hierarchies</i> Shib Sankar Dasgupta, Xiang Lorraine Li, Michael Boratko, Dongxu Zhang and Andrew McCallum	277
<i>An Overview of Uncertainty Calibration for Text Classification and the Role of Distillation</i> Han Guo, Ramakanth Pasunuru and Mohit Bansal .....	289
<i>Entity and Evidence Guided Document-Level Relation Extraction</i> Kevin Huang, Peng Qi, Guangtao Wang, Tengyu Ma and Jing Huang .....	307
<i>Scaling Deep Contrastive Learning Batch Size under Memory Limited Setup</i> Luyu Gao, Yunyi Zhang, Jiawei Han and Jamie Callan .....	316
<i>Direction is what you need: Improving Word Embedding Compression in Large Language Models</i> Klaudia Bałazy, Mohammadreza Banaei, Rémi Lebret, Jacek Tabor and Karl Aberer .....	322

## 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

**August 6, 2021 (continued)**

*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 shaghighian

*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*

Raghuveer Thirukovalluru, Mukund Sridhar, Dung Thai, Shruti Chanumolu, Nicholas Monath, Sankaranarayanan Ananthakrishnan and Andrew McCallum

**August 6, 2021 (continued)**

*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 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

*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 Bałazy, Mohammadreza Banaei, Rémi Lebret, Jacek Tabor and Karl Aberer

