

CoNLL 2020

**The 24th Conference on Computational  
Natural Language Learning (CoNLL)**

**Proceedings of the Conference**

November 19-20, 2020  
Online

©2020 The Association for Computational Linguistics

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 978-1-952148-63-7

## Introduction

Welcome to the the 24th Conference on Computational Natural Language Learning (CoNLL). This year, CoNLL, like many conferences, is taking place online, due to the impact of the COVID-19 pandemic. In a major shift from previous editions of the conference, CoNLL 2020 focuses on theoretically, cognitively and scientifically motivated approaches to computational linguistics, rather than on work driven by particular engineering applications. This shift was indicated in the call for papers, and was reflected in the instructions given to the reviewers and area chairs.

We received 255 submissions, of which 227 were sent out for review (the remaining submissions were desk rejected or withdrawn by the authors). We accepted 53 of the submissions that were sent out for review (23.3% acceptance rate). CoNLL 2019 features two invited speakers, Emmanuel Dupoux (Ecole des Hautes Etudes en Sciences Sociales and Facebook AI Research, Paris, France) and Kristina Toutanova (Google, Seattle, USA).

We are extremely grateful to all our reviewers (too many to list), and our dedicated area chairs: Aida Nematzadeh, Alvin Grisom II, Andrew Caines, Arianna Bisazza, Barry Devereux, Colin Bannard, Daniel Cer, David Schlangen, Erik Velldal, Greg Durrett, Grzegorz Chrupala, Jacob Andreas, Kevin Duh, Kyle Gorman, Leon Bergen, Marten van Schijndel, Michael Roth, Raffaella Bernardi, Roi Reichert, Sam Bowman, Stella Frank, Tim O'Donnell, Vivek Srikumar, Yevgeni Berzak, Yonatan Belinkov and Yulia Tsvetkov. We are also grateful to Google for generously supporting the conference.

We would also like to thank Julia Hockenmaier and Afra Alishahi and the members of the SIGNLL board for entrusting to us the task of organizing the conference, and Mohit Bansal and Aline Villavicencio, the organizers of CoNLL 2019, for their patient advice.

Raquel Fernández and Tal Linzen  
CoNLL 2020 conference co-chairs



**Organizers:**

Tal Linzen, New York University  
Raquel Fernandez, University of Amsterdam

**Area Chairs:**

Aida Nematzadeh  
Alvin Grisom II  
Andrew Caines  
Arianna Bisazza  
Barry Devereux  
Colin Bannard  
Daniel Cer  
David Schlangen  
Erik Velldal  
Greg Durrett  
Grzegorz Chrupala  
Jacob Andreas  
Kevin Duh  
Kyle Gorman  
Leon Bergen  
Marten van Schijndel  
Michael Roth  
Raffaella Bernardi  
Roi Reichert  
Sam Bowman  
Stella Frank  
Tim O'Donnell  
Vivek Srikumar  
Yevgeni Berzak  
Yonatan Belinkov  
Yulia Tsvetkov

**Invited Speakers:**

Emmanuel Dupoux (Ecole des Hautes Etudes en Sciences Sociales and Facebook AI Research,  
Paris, France)  
Kristina Toutanova (Google, Seattle, USA)



## Table of Contents

<i>Enriching Word Embeddings with Temporal and Spatial Information</i>	
Hongyu Gong, Suma Bhat and Pramod Viswanath .....	1
<i>Interpreting Attention Models with Human Visual Attention in Machine Reading Comprehension</i>	
Ekta Sood, Simon Tannert, Diego Frassinelli, Andreas Bulling and Ngoc Thang Vu .....	12
<i>Neural Proof Nets</i>	
Konstantinos Kogkalidis, Michael Moortgat and Richard Moot .....	26
<i>TaxiNLI: Taking a Ride up the NLU Hill</i>	
Pratik Joshi, Somak Aditya, Aalok Sathe and Monojit Choudhury .....	41
<i>Modeling Subjective Assessments of Guilt in Newspaper Crime Narratives</i>	
Elisa Kreiss, Zijian Wang and Christopher Potts .....	56
<i>On the Frailty of Universal POS Tags for Neural UD Parsers</i>	
Mark Anderson and Carlos Gómez-Rodríguez .....	69
<i>Classifying Syntactic Errors in Learner Language</i>	
Leshem Choshen, Dmitry Nikolaev, Yevgeni Berzak and Omri Abend .....	97
<i>How to Probe Sentence Embeddings in Low-Resource Languages: On Structural Design Choices for Probing Task Evaluation</i>	
Steffen Eger, Johannes Daxenberger and Iryna Gurevych .....	108
<i>Understanding the Source of Semantic Regularities in Word Embeddings</i>	
Hsiao-Yu Chiang, Jose Camacho-Collados and Zachary Pados .....	119
<i>Finding The Right One and Resolving it</i>	
Payal Khullar, Arghya Bhattacharya and Manish Shrivastava .....	132
<i>Bridging Information-Seeking Human Gaze and Machine Reading Comprehension</i>	
Jonathan Malmaud, Roger Levy and Yevgeni Berzak .....	142
<i>A Corpus of Very Short Scientific Summaries</i>	
Yifan Chen, Tamara Polajnar, Colin Batchelor and Simone Teufel .....	153
<i>Recurrent babbling: evaluating the acquisition of grammar from limited input data</i>	
Ludovica Pannitto and Aurélie Herbelot .....	165
<i>Explaining the efficiency of communication: How communicators can reduce their computational burden through interaction</i>	
Jacqueline van Arkel, Marieke Woensdregt, Mark Dingemanse and Mark Blokpoel .....	177
<i>Acquiring language from speech by learning to remember and predict</i>	
Cory Shain and Micha Elsner .....	195
<i>Identifying Incorrect Labels in the CoNLL-2003 Corpus</i>	
Frederick Reiss, Hong Xu, Bryan Cutler, Karthik Muthuraman and Zachary Eichenberger .....	215
<i>When is a bishop not like a rook? When it's like a rabbi! Multi-prototype BERT embeddings for estimating semantic relationships</i>	
Gabriella Chronis and Katrin Erk .....	227

<i>Processing effort is a poor predictor of cross-linguistic word order frequency</i> Brennan Gonerig and Emily Morgan .....	245
<i>Relations between comprehensibility and adequacy errors in machine translation output</i> Maja Popović .....	256
<i>Cross-lingual Embeddings Reveal Universal and Lineage-Specific Patterns in Grammatical Gender Assignment</i> Hartger Veeman, Marc Allasonnière-Tang, Aleksandrs Berdicevskis and Ali Basirat .....	265
<i>Modelling Lexical Ambiguity with Density Matrices</i> Francois Meyer and Martha Lewis .....	276
<i>Catplaying in the snow: Impact of Prior Segmentation on a Model of Visually Grounded Speech</i> William Havard, Laurent Besacier and Jean-Pierre Chevrot .....	291
<i>Learning to ground medical text in a 3D human atlas</i> Dusan Grujicic, Gorjan Radevski, Tinne Tuytelaars and Matthew Blaschko .....	302
<i>Representation Learning for Type-Driven Composition</i> Gijs Wijnholds, Mehrnoosh Sadrzadeh and Stephen Clark .....	313
<i>Word Representations Concentrate and This is Good News!</i> Romain Couillet, Yagmur Gizem Cinar, Eric Gaussier and Muhammad Imran .....	325
<i>“LazImpa”: Lazy and Impatient neural agents learn to communicate efficiently</i> Mathieu Rita, Rahma Chaabouni and Emmanuel Dupoux .....	335
<i>Re-solve it: simulating the acquisition of core semantic competences from small data</i> Aurélie Herbelot .....	344
<i>In Media Res: A Corpus for Evaluating Named Entity Linking with Creative Works</i> Adrian M.P. Brasoveanu, Albert Weichselbraun and Lyndon Nixon .....	355
<i>Analogies minus analogy test: measuring regularities in word embeddings</i> Louis Fournier, Emmanuel Dupoux and Ewan Dunbar .....	365
<i>Word associations and the distance properties of context-aware word embeddings</i> Maria A. Rodriguez and Paola Merlo .....	376
<i>TrClaim-19: The First Collection for Turkish Check-Worthy Claim Detection with Annotator Rationales</i> Yavuz Selim Kartal and Mucahid Kutlu .....	386
<i>Discourse structure interacts with reference but not syntax in neural language models</i> Forrest Davis and Marten van Schijndel .....	396
<i>Continual Adaptation for Efficient Machine Communication</i> Robert Hawkins, Minae Kwon, Dorsa Sadigh and Noah Goodman .....	408
<i>Diverse and Relevant Visual Storytelling with Scene Graph Embeddings</i> Xudong Hong, Rakshith shetty, Asad Sayeed, Khushboo Mehra, Vera Demberg and Bernt Schiele	420
<i>Alleviating Digitization Errors in Named Entity Recognition for Historical Documents</i> Emanuela Boros, Ahmed Hamdi, Elvys Linhares Pontes, Luis Adrián Cabrera-Diego, Jose G. Moreno, Nicolas Sidere and Antoine Doucet .....	431



<i>Analysing Word Representation from the Input and Output Embeddings in Neural Network Language Models</i>	
Steven Derby, Paul Miller and Barry Devereux .....	442
<i>On the Computational Power of Transformers and Its Implications in Sequence Modeling</i>	
Satwik Bhattamishra, Arkil Patel and Navin Goyal .....	455
<i>An Expectation Maximisation Algorithm for Automated Cognate Detection</i>	
Roddy MacSween and Andrew Caines .....	476
<i>Filler-gaps that neural networks fail to generalize</i>	
Debasmita Bhattacharya and Marten van Schijndel .....	486
<i>Don't Parse, Insert: Multilingual Semantic Parsing with Insertion Based Decoding</i>	
Qile Zhu, Haidar Khan, Saleh Soltan, Stephen Rawls and Wael Hamza .....	496
<i>Learning Context-free Languages with Nondeterministic Stack RNNs</i>	
Brian DuSell and David Chiang .....	507
<i>Generating Narrative Text in a Switching Dynamical System</i>	
Noah Weber, Leena Shekhar, Heeyoung Kwon, Niranjan Balasubramanian and Nathanael Chambers .....	520
<i>What Are You Trying to Do? Semantic Typing of Event Processes</i>	
Muhao Chen, Hongming Zhang, Haoyu Wang and Dan Roth .....	531
<i>A Corpus for Outbreak Detection of Diseases Prevalent in Latin America</i>	
Antonella Dellanzo, Viviana Cotik and Jose Ochoa-Luna .....	543
<i>Are Pretrained Language Models Symbolic Reasoners over Knowledge?</i>	
Nora Kassner, Benno Krojer and Hinrich Schütze .....	552
<i>Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues</i>	
Tanmay Parekh, Emily Ahn, Yulia Tsvetkov and Alan W Black .....	565
<i>Identifying robust markers of Parkinson's disease in typing behaviour using a CNN-LSTM network</i>	
Neil Dhir, Mathias Edman, Álvaro Sanchez Ferro, Tom Stafford and Colin Bannard .....	578
<i>An Empirical Study on Model-agnostic Debiasing Strategies for Robust Natural Language Inference</i>	
Tianyu Liu, Zheng Xin, Xiaolan Ding, Baobao Chang and Zhifang Sui .....	596
<i>Cloze Distillation Improves Psychometric Predictive Power</i>	
Tiwalayo Eisape, Noga Zaslavsky and Roger Levy .....	609
<i>Disentangling dialects: a neural approach to Indo-Aryan historical phonology and subgrouping</i>	
Chundra Cathcart and Taraka Rama .....	620
<i>A Dataset for Linguistic Understanding, Visual Evaluation, and Recognition of Sign Languages: The K-RSL</i>	
Alfarabi Imashev, Medet Mukushev, Vadim Kimmelman and Anara Sandygulova .....	631
<i>From Dataset Recycling to Multi-Property Extraction and Beyond</i>	
Tomasz Dwojak, Michał Pietruszka, Łukasz Borchmann, Jakub Chłędowski and Filip Galiński	641
<i>How well does surprisal explain N400 amplitude under different experimental conditions?</i>	
James Michaelov and Benjamin Bergen .....	652



## Conference Program

*Enriching Word Embeddings with Temporal and Spatial Information*

Hongyu Gong, Suma Bhat and Pramod Viswanath

*Interpreting Attention Models with Human Visual Attention in Machine Reading Comprehension*

Ekta Sood, Simon Tannert, Diego Frassinelli, Andreas Bulling and Ngoc Thang Vu

*Neural Proof Nets*

Konstantinos Kogkalidis, Michael Moortgat and Richard Moot

*TaxiNLI: Taking a Ride up the NLU Hill*

Pratik Joshi, Somak Aditya, Aalok Sathe and Monojit Choudhury

*Modeling Subjective Assessments of Guilt in Newspaper Crime Narratives*

Elisa Kreiss, Zijian Wang and Christopher Potts

*On the Frailty of Universal POS Tags for Neural UD Parsers*

Mark Anderson and Carlos Gómez-Rodríguez

*Classifying Syntactic Errors in Learner Language*

Leshem Choshen, Dmitry Nikolaev, Yevgeni Berzak and Omri Abend

*How to Probe Sentence Embeddings in Low-Resource Languages: On Structural Design Choices for Probing Task Evaluation*

Steffen Eger, Johannes Daxenberger and Iryna Gurevych

*Understanding the Source of Semantic Regularities in Word Embeddings*

Hsiao-Yu Chiang, Jose Camacho-Collados and Zachary Pados

*Finding The Right One and Resolving it*

Payal Khullar, Arghya Bhattacharya and Manish Shrivastava

*Bridging Information-Seeking Human Gaze and Machine Reading Comprehension*

Jonathan Malmaud, Roger Levy and Yevgeni Berzak

*A Corpus of Very Short Scientific Summaries*

Yifan Chen, Tamara Polajnar, Colin Batchelor and Simone Teufel

## No Day Set (continued)

*Recurrent babbling: evaluating the acquisition of grammar from limited input data*

Ludovica Pannitto and Aurélie Herbelot

*Explaining the efficiency of communication: How communicators can reduce their computational burden through interaction*

Jacqueline van Arkel, Marieke Woensdregt, Mark Dingemanse and Mark Blokpoel

*Acquiring language from speech by learning to remember and predict*

Cory Shain and Micha Elsner

*Identifying Incorrect Labels in the CoNLL-2003 Corpus*

Frederick Reiss, Hong Xu, Bryan Cutler, Karthik Muthuraman and Zachary Eichenberger

*When is a bishop not like a rook? When it's like a rabbi! Multi-prototype BERT embeddings for estimating semantic relationships*

Gabriella Chronis and Katrin Erk

*Processing effort is a poor predictor of cross-linguistic word order frequency*

Brennan Gonering and Emily Morgan

*Relations between comprehensibility and adequacy errors in machine translation output*

Maja Popović

*Cross-lingual Embeddings Reveal Universal and Lineage-Specific Patterns in Grammatical Gender Assignment*

Hartger Veeman, Marc Allasonnière-Tang, Aleksandrs Berdicevskis and Ali Basirat

*Modelling Lexical Ambiguity with Density Matrices*

Francois Meyer and Martha Lewis

*Catplaying in the snow: Impact of Prior Segmentation on a Model of Visually Grounded Speech*

William Havard, Laurent Besacier and Jean-Pierre Chevrot

*Learning to ground medical text in a 3D human atlas*

Dusan Grujicic, Gorjan Radevski, Tinne Tuytelaars and Matthew Blaschko

*Representation Learning for Type-Driven Composition*

Gijs Wijnholds, Mehrnoosh Sadrzadeh and Stephen Clark

## No Day Set (continued)

*Word Representations Concentrate and This is Good News!*

Romain Couillet, Yagmur Gizem Cinar, Eric Gaussier and Muhammad Imran

*“LazImpa”: Lazy and Impatient neural agents learn to communicate efficiently*

Mathieu Rita, Rahma Chaabouni and Emmanuel Dupoux

*Re-solve it: simulating the acquisition of core semantic competences from small data*

Aurélie Herbelot

*In Media Res: A Corpus for Evaluating Named Entity Linking with Creative Works*

Adrian M.P. Brasoveanu, Albert Weichselbraun and Lyndon Nixon

*Analogies minus analogy test: measuring regularities in word embeddings*

Louis Fournier, Emmanuel Dupoux and Ewan Dunbar

*Word associations and the distance properties of context-aware word embeddings*

Maria A. Rodriguez and Paola Merlo

*TrClaim-19: The First Collection for Turkish Check-Worthy Claim Detection with Annotator Rationales*

Yavuz Selim Kartal and Mucahid Kutlu

*Discourse structure interacts with reference but not syntax in neural language models*

Forrest Davis and Marten van Schijndel

*Continual Adaptation for Efficient Machine Communication*

Robert Hawkins, Minae Kwon, Dorsa Sadigh and Noah Goodman

*Diverse and Relevant Visual Storytelling with Scene Graph Embeddings*

Xudong Hong, Rakshith shetty, Asad Sayeed, Khushboo Mehra, Vera Demberg and Bernt Schiele

*Alleviating Digitization Errors in Named Entity Recognition for Historical Documents*

Emanuela Boros, Ahmed Hamdi, Elvys Linhares Pontes, Luis Adrián Cabrera-Diego, Jose G. Moreno, Nicolas Sidere and Antoine Doucet

*Analysing Word Representation from the Input and Output Embeddings in Neural Network Language Models*

Steven Derby, Paul Miller and Barry Devereux

## No Day Set (continued)

*On the Computational Power of Transformers and Its Implications in Sequence Modeling*

Satwik Bhattamishra, Arkil Patel and Navin Goyal

*An Expectation Maximisation Algorithm for Automated Cognate Detection*

Roddy MacSween and Andrew Caines

*Filler-gaps that neural networks fail to generalize*

Debasmita Bhattacharya and Marten van Schijndel

*Don't Parse, Insert: Multilingual Semantic Parsing with Insertion Based Decoding*

Qile Zhu, Haidar Khan, Saleh Soltan, Stephen Rawls and Wael Hamza

*Learning Context-free Languages with Nondeterministic Stack RNNs*

Brian DuSell and David Chiang

*Generating Narrative Text in a Switching Dynamical System*

Noah Weber, Leena Shekhar, Heeyoung Kwon, Niranjan Balasubramanian and Nathanael Chambers

*What Are You Trying to Do? Semantic Typing of Event Processes*

Muhao Chen, Hongming Zhang, Haoyu Wang and Dan Roth

*A Corpus for Outbreak Detection of Diseases Prevalent in Latin America*

Antonella Dellanzo, Viviana Cotik and Jose Ochoa-Luna

*Are Pretrained Language Models Symbolic Reasoners over Knowledge?*

Nora Kassner, Benno Krojer and Hinrich Schütze

*Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues*

Tanmay Parekh, Emily Ahn, Yulia Tsvetkov and Alan W Black

*Identifying robust markers of Parkinson's disease in typing behaviour using a CNN-LSTM network*

Neil Dhir, Mathias Edman, Álvaro Sanchez Ferro, Tom Stafford and Colin Bannard

*An Empirical Study on Model-agnostic Debiasing Strategies for Robust Natural Language Inference*

Tianyu Liu, Zheng Xin, Xiaoan Ding, Baobao Chang and Zhifang Sui

## No Day Set (continued)

*Cloze Distillation Improves Psychometric Predictive Power*

Tiwalayo Eisape, Noga Zaslavsky and Roger Levy

*Disentangling dialects: a neural approach to Indo-Aryan historical phonology and subgrouping*

Chundra Cathcart and Taraka Rama

*A Dataset for Linguistic Understanding, Visual Evaluation, and Recognition of Sign Languages: The K-RSL*

Alfarabi Imashev, Medet Mukushev, Vadim Kimmelman and Anara Sandygulova

*From Dataset Recycling to Multi-Property Extraction and Beyond*

Tomasz Dwojak, Michał Pietruszka, Łukasz Borchmann, Jakub Chłędowski and Filip Graliński

*How well does surprisal explain N400 amplitude under different experimental conditions?*

James Michaelov and Benjamin Bergen

