ACL 2020

Figurative Language Processing

Proceedings of the Second Workshop

July 9, 2020

©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

ISBN 978-1-952148-12-5

Introduction

Figurative language processing is a rapidly growing area in Natural Language Processing (NLP), including processing of metaphors, idioms, puns, irony, sarcasm, as well as other figures. Characteristic to all areas of human activity (from poetic to ordinary to scientific) and, thus, to all types of discourse, figurative language becomes an important problem for NLP systems. Its ubiquity in language has been established in a number of corpus studies, and the role it plays in human reasoning has been confirmed in psychological experiments. This makes figurative language an important research area for computational and cognitive linguistics, and its automatic identification and interpretation indispensable for any semantics-oriented NLP application.

This workshop is the second in a series of biannual workshops on Figurative Language Processing. This new workshop series builds upon the successful start of the Metaphor in NLP workshop series (at NAACL– HLT 2013, ACL 2014, NAACL–HLT 2015, NAACL–HLT 2016), expanding its scope to incorporate the rapidly growing body of research on various types of figurative language such as sarcasm, irony and puns, with the aim of maintaining and nourishing a community of NLP researchers interested in this topic. The workshop features both regular research papers and two shared tasks on metaphor and sarcasm detection. In the regular research track, we received 20 research paper submissions and accepted 9 (3 oral presentations and 6 posters). The featured papers cover a range of aspects of figurative language processing such as metaphor identification (Dankers et al.; Zayed, McCrae and Buitelaar), metaphor in the visual modality (Bizzoni and Dobnik), annotation of oxymorons (La Pietra and Massini), satirical and humorous headline generation (Weller et al.; Horvitz et al.) and recognising euphemisms and dysphemisms (Felt and Riloff). The workshop program also features a keynote talk by Marilyn Walker, Department of Computer Science, University of California Santa Cruz, on the topic of "Generating Expressive Language by Mining User Reviews".

The two shared tasks on metaphor and sarcasm detection serve to benchmark various computational approaches to metaphor and sarcasm, clarifying the state of this steadily growing field and facilitating further research.

For the metaphor shared task, we used the VU Amsterdam Metaphor Corpus (VUA) corpus as one of the corpora for the shared tasks. New to this year's benchmarking tasks, we added a corpus of TOEFL essays written by non-native speakers of English annotated for metaphor (a subset from the publicly available ETS Corpus of Non-Native Written English), allowing us to broaden the genres covered in the task and in accordance with findings in the literature demonstrating the potential of information on metaphor usage for assessing English proficiency of students.

The shared task was organized into four tracks: a Verbs track and an All Content Part-of-Speech track for both VUA and TOEFL. Overall, there were 1,224 submissions from 71 teams. There were 805 submissions from the 14 teams who submitted system papers; one paper was withdrawn before publication. In terms of performance, the current published state-of-art on VUA corpus has been matched by the best participating system, while a new state-of-art was established for the TOEFL corpus. We observed the following general trends: (1) Transformer architectures were highly popular and resulted in competitive performance; (2) New sources of information were explored by participants, such as fine-grained POS, spell-corrected variants of words (for TOEFL data), sub-word level information (e.g., character embeddings), idioms, sensorimotor and embodiment-related information; (3) The relative performance rankings of teams were largely consistent between VUA and TOEFL datasets; (4) Performance of participating systems was generally better on Verbs than on the All POS tracks, across both corpora.

The shared task on sarcasm detection was designed to benchmark the usefulness of modeling conversation context (i.e., all the prior dialogue turns) for sarcasm detection. Two types of social media

content are used as training data for the two tracks – microblogging platforms such as Twitter and online discussion forums such as Reddit. Overall, we received an overwhelming number of submissions: 655 for the Reddit track and 1070 for the Twitter track. The CodaLab leaderboard showcases results from 39 systems for the Reddit track and 38 systems for the Twitter track, respectively. Out of all submissions, 14 shared task system papers were submitted. Almost all the submitted systems have used the transformer-architecture that seems to perform better than RNN-architectures, even without any task-specific fine-tuning. The best system has shown the usefulness of augmenting "other" dataset(s) during training. In terms of context, novel approaches include: CNN-LSTM based summarization of the prior dialogue turns, time-series fusion with proxy labels, an ensemble of a variety of transformers with different depth of context, aspect-based sentiment classification for the immediate context, etc. When explicitly modeling the number of turns, systems have shown better accuracy with a depth of a maximum of three prior turns. In the future, we plan to continuously grow the training corpus, collecting data from a variety of subreddits, in case of Reddit, and different topics from Twitter.

We wish to thank everyone who showed interest and submitted a paper, all of the authors for their contributions, the members of the Program Committee for their thoughtful reviews, the invited speaker for sharing her perspective on the topic, and all the attendees of the workshop. All of these factors contribute to a truly enriching event!

Workshop co–chairs:
Beata Beigman Klebanov, Educational Testing Service (USA)
Ekaterina Shutova, University of Amsterdam (The Netherlands)
Patricia Lichtenstein, University of California, Merced (USA)
Smaranda Muresan, Columbia University (USA)
Chee Wee (Ben) Leong, Educational Testing Service (USA)
Anna Feldman, Montclair State University (USA)
Debanjan Ghosh, Educational Testing Service (USA)

Organizers:

Beata Beigman Klebanov, Educational Testing Service (USA)
Ekaterina Shutova, University of Amsterdam (The Netherlands)
Patricia Lichtenstein, University of California, Merced (USA)
Smaranda Muresan, Columbia University (USA)
Chee Wee (Ben) Leong, Educational Testing Service (USA)
Anna Feldman, Montclair State University (USA)
Debanjan Ghosh, Educational Testing Service (USA)

Program Committee:

Khalid Alnajjar, University of Helsinki (Finland)
Yulia Badryzlova, National Research University Higher School of Economics (Russia)
Suma Bhat, University of Illinois at Urbana-Champaign (USA)
Paul Cook, University of New Brunswick (Canada)
Ellen Dodge, Google (USA)
Jonathan Dunn, University of Canterbury (New Zealand)
Anjalie Field, Carnegie Mellon University (USA)
Michael Flor, Educational Testing Service (USA)
Mark Granroth-Wilding, University of Helsinki (Finland)
Mika Hämäläinen, University of Helsinki (Finland)
Veronique Hoste, Ghent University (Belgium)
Eduard Hovy, Carnegie Mellon University (USA)
Rebecca Hwa, University of Pittsburgh (USA)
Hyeju Jang, University of British Columbia (Canada)
Valia Kordoni, Humboldt-Universitaet zu Berlin (Germany)
Mark Last, Ben-Gurion University of the Negev (Israel)
Mark Lee, University of Birmingham (UK)
Els Lefever, LT3, Ghent University (Belgium)
Jean Maillard, Facebook AI (USA)
Katja Markert, Heidelberg University (Germany)
James Martin, University of Colorado Boulder (USA)
Michael Mohler, Language Computer Corporation (USA)
Elena Musi, University of Liverpool (UK)
Preslav Nakov, Qatar Computing Research Institute, HBKU (Qatar)
Srini Narayanan, Google Research (USA)
Thierry Poibeau, CNRS (France)
Vinodkumar Prabhakaran, Google Research (USA)
Marek Rei, University of Cambridge (UK)
Ellen Riloff, University of Utah (USA)
Sarah Rosenthal, IBM Research (USA)
Paolo Rosso, University of Politècnica de València (Spain)
Giovanni Da San Martino, Qatar Computing Research Institute (Qatar)
Eyal Sagi, University of St. Francis (USA)
Agata Savary, University of Tours (France)
Sabine Schulte im Walde, University of Stuttgart (Germany)
Samira Shaikh, University of North Carolina, Charlotte (USA)
Thamar Solorio Martinez, University of Houston (USA)

Kevin Stowe, TU Darmstadt, UKP Lab (Germany)
Tomek Strzalkowski, Rensselaer Polytechnic Institute (USA)
Stan Szpakowicz, University of Ottawa (Canada)
Marc Tomlinson, Language Computer Corporation (USA)
Yulia Tsvetkov, Carnegie Mellon University (USA)
Aline Villavicencio, University of Sheffield (UK)

Invited Speaker:

Marilyn Walker, University of California Santa Cruz (USA)

Table of Contents

A Report on the 2020 Sarcasm Detection Shared Task
Debanjan Ghosh, Avijit Vajpayee and Smaranda Muresan 1

Augmenting Data for Sarcasm Detection with Unlabeled Conversation Context
Hankyol Lee, Youngjae Yu and Gunhee Kim 12

A Report on the 2020 VUA and TOEFL Metaphor Detection Shared Task
Chee Wee (Ben) Leong, Beata Beigman Klebanov, Chris Hamill, Egon Stemle, Rutuja Ubale and Xianyang Chen 18

DeepMet: A Reading Comprehension Paradigm for Token-level Metaphor Detection
Chuandong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiyi Li, Rongbo Wang and Zhiqun Chen 30

Context-Driven Satirical News Generation
Zachary Horvitz, Nam Do and Michael L. Littman 40

Sarcasm Detection using Context Separators in Online Discourse
TANVI DADU and Kartikey Pant 51

Sarcasm Detection in Tweets with BERT and GloVe Embeddings
Akshay Khatri and Pranav P 56

C-Net: Contextual Network for Sarcasm Detection
Amit Kumar Jena, Aman Sinha and Rohit Agarwal 61

Applying Transformers and Aspect-based Sentiment Analysis approaches on Sarcasm Detection
Taha Shangipour ataei, Soroush Javdan and Behrouz Minaei-Bidgoli 67

Sarcasm Identification and Detection in Conversion Context using BERT
kalaivani A and Thenmozhi D 72

Neural Sarcasm Detection using Conversation Context
Nikhil Jaiswal 77

Context-Aware Sarcasm Detection Using BERT
Arup Baruah, Kaushik Das, Ferdous Barbhuiya and Kuntal Dey 83

Transformers on Sarcasm Detection with Context
Amardeep Kumar and Vivek Anand 88

A Novel Hierarchical BERT Architecture for Sarcasm Detection
Himani Srivastava, Vaibhav Varshney, Surabhi Kumari and Saurabh Srivastava 93

Detecting Sarcasm in Conversation Context Using Transformer-Based Models
Adithya Avvaru, Sanath Vobilisetty and Radhika Mamidi 98

Using Conceptual Norms for Metaphor Detection
Mingyu WAN, Kathleen Ahrens, Emmanuele Chersoni, Menghan Jiang, Qi Su, Rong Xiang and Chu-Ren Huang 104

ALBERT-BiLSTM for Sequential Metaphor Detection
Shuqun Li, Jingjie Zeng, Jinhui Zhang, Tao Peng, Liang Yang and Hongfei Lin 110

Character aware models with similarity learning for metaphor detection
Tarun Kumar and Yashvardhan Sharma 116

Sky + Fire = Sunset. Exploring Parallels between Visually Grounded Metaphors and Image Classifiers
Yuri Bizzoni and Simon Dobnik 126

Recognizing Euphemisms and Dysphemisms Using Sentiment Analysis
Christian Felt and Ellen Riloff 136

IlliniMet: Illinois System for Metaphor Detection with Contextual and Linguistic Information
Hongyu Gong, Kshitij Gupta, Akriti Jain and Suma Bhat 146

Adaptation of Word-Level Benchmark Datasets for Relation-Level Metaphor Identification
Omnia Zayed, John Philip McCrae and Paul Buitelaar 154

Generating Ethnographic Models from Communities' Online Data
Tomek Strzalkowski, Anna Newheiser, Nathan Kemper, Ning Sa, Bharvee Acharya and Gregorios Katsios 165

Oxymorons: a preliminary corpus investigation
Marta La Pietra and Francesca Masini 176

Can Humor Prediction Datasets be used for Humor Generation? Humorous Headline Generation via Style Transfer
Orion Weller, Nancy Fulda and Kevin Seppi 186

Evaluating a Bi-LSTM Model for Metaphor Detection in TOEFL Essays
Kevin Kuo and Marine Carpuat 192

Neural Metaphor Detection with a Residual biLSTM-CRF Model
Andrés Torres Rivera, Antoni Oliver, Salvador Climent and Marta Coll-Florit 197

Augmenting Neural Metaphor Detection with Concreteness
Ghadi Alnafesah, Harish Tayyar Madabushi and Mark Lee 204

Supervised Disambiguation of German Verbal Idioms with a BiLSTM Architecture
Rafael Ehren, Timm Lichte, Laura Kallmeyer and Jakub Waszczuk 211

Metaphor Detection using Context and Concreteness
Rowan Hall Maudslay, Tiago Pimentel, Ryan Cotterell and Simone Teufel 221

Being neighbourly: Neural metaphor identification in discourse
Verna Dankers, Karan Malhotra, Gaurav Kudva, Volodymyr Medentsiy and Ekaterina Shutova 227

Go Figure! Multi-task transformer-based architecture for metaphor detection using idioms: ETS team in 2020 metaphor shared task
Xianyang Chen, Chee Wee (Ben) Leong, Michael Flor and Beata Beigman Klebanov 235

Metaphor Detection using Ensembles of Bidirectional Recurrent Neural Networks
Jennifer Brooks and Abdou Youssef 244

Metaphor Detection Using Contextual Word Embeddings From Transformers
Jerry Liu, Nathan O'Hara, Alexander Rubin, Rachel Draelos and Cynthia Rudin 250

Testing the role of metadata in metaphor identification
Egon Stemle and Alexander Onysko 256

Sarcasm Detection Using an Ensemble Approach
Jens Lemmens, Ben Burtenshaw, Ehsan Lotfi, Ilia Markov and Walter Daelemans 264

A Transformer Approach to Contextual Sarcasm Detection in Twitter
Hunter Gregory, Steven Li, Pouya Mohammadi, Natalie Tarn, Rachel Draelos and Cynthia Rudin 270

Transformer-based Context-aware Sarcasm Detection in Conversation Threads from Social Media
Xiangjue Dong, Changmao Li and Jinho D. Choi 276

Conference Program

Thursday July 9, 2020

06:05 *A Report on the 2020 Sarcasm Detection Shared Task*
Debanjan Ghosh, Avijit Vajpayee and Smaranda Muresan

06:20 *Augmenting Data for Sarcasm Detection with Unlabeled Conversation Context*
Hankyol Lee, Youngjae Yu and Gunhee Kim

06:35 *A Report on the 2020 VUA and TOEFL Metaphor Detection Shared Task*
Chee Wee (Ben) Leong, Beata Beigman Klebanov, Chris Hamill, Egon Stemle, Rutuja Ubale and Xianyang Chen

06:50 *DeepMet: A Reading Comprehension Paradigm for Token-level Metaphor Detection*
Chuandong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiyi Li, Rongbo Wang and Zhiqun Chen

07:05 *Context-Driven Satirical News Generation*
Zachary Horvitz, Nam Do and Michael L. Littman

07:40 *Sarcasm Detection using Context Separators in Online Discourse*
TANVI DADU and Kartikey Pant

07:45 *Sarcasm Detection in Tweets with BERT and GloVe Embeddings*
Akshay Khatri and Pranav P

07:50 *C-Net: Contextual Network for Sarcasm Detection*
Amit Kumar Jena, Aman Sinha and Rohit Agarwal

08:00 *Applying Transformers and Aspect-based Sentiment Analysis approaches on Sarcasm Detection*
Taha Shangipour ataei, Soroush Javdan and Behrouz Minaei-Bidgoli

08:05 *Sarcasm Identification and Detection in Conversion Context using BERT*
kalaivani A and Thenmozhi D

08:10 *Neural Sarcasm Detection using Conversation Context*
Nikhil Jaiswal

Thursday July 9, 2020 (continued)

08:40 *Context-Aware Sarcasm Detection Using BERT*
Arup Baruah, Kaushik Das, Ferdous Barbhuiya and Kuntal Dey

08:45 *Transformers on Sarcasm Detection with Context*
Amardeep Kumar and Vivek Anand

08:50 *A Novel Hierarchical BERT Architecture for Sarcasm Detection*
Himani Srivastava, Vaibhav Varshney, Surabhi Kumari and Saurabh Srivastava

09:00 *Detecting Sarcasm in Conversation Context Using Transformer-Based Models*
Adithya Avvaru, Sanath Vobilisetty and Radhika Mamidi

09:05 *Using Conceptual Norms for Metaphor Detection*
Mingyu WAN, Kathleen Ahrens, Emmanuele Chersoni, Menghan Jiang, Qi Su, Rong Xiang and Chu-Ren Huang

09:10 *ALBERT-BiLSTM for Sequential Metaphor Detection*
Shuqun Li, Jingjie Zeng, Jinhui Zhang, Tao Peng, Liang Yang and Hongfei Lin

09:15 *Character aware models with similarity learning for metaphor detection*
Tarun Kumar and Yashvardhan Sharma

10:00 *Sky + Fire = Sunset. Exploring Parallels between Visually Grounded Metaphors and Image Classifiers*
Yuri Bizzoni and Simon Dobnik

10:15 *Recognizing Euphemisms and Dysphemisms Using Sentiment Analysis*
Christian Felt and Ellen Riloff

10:30 *IlliniMet: Illinois System for Metaphor Detection with Contextual and Linguistic Information*
Hongyu Gong, Kshitij Gupta, Akriti Jain and Suma Bhat

10:35 *Adaptation of Word-Level Benchmark Datasets for Relation-Level Metaphor Identification*
Omnia Zayed, John Philip McCrae and Paul Buitelaar

10:40 *Generating Ethnographic Models from Communities' Online Data*
Tomek Strzalkowski, Anna Newheiser, Nathan Kemper, Ning Sa, Bharvee Acharya and Gregorios Katsios

Thursday July 9, 2020 (continued)

10:50 *Oxymorons: a preliminary corpus investigation*
Marta La Pietra and Francesca Masini

10:55 *Can Humor Prediction Datasets be used for Humor Generation? Humorous Headline Generation via Style Transfer*
Orion Weller, Nancy Fulda and Kevin Seppi

11:00 *Evaluating a Bi-LSTM Model for Metaphor Detection in TOEFL Essays*
Kevin Kuo and Marine Carpuat

11:30 *Neural Metaphor Detection with a Residual biLSTM-CRF Model*
Andrés Torres Rivera, Antoni Oliver, Salvador Climent and Marta Coll-Florit

11:35 *Augmenting Neural Metaphor Detection with Concreteness*
Ghadi Alnafesah, Harish Tayyar Madabushi and Mark Lee

11:40 *Supervised Disambiguation of German Verbal Idioms with a BiLSTM Architecture*
Rafael Ehren, Timm Lichte, Laura Kallmeyer and Jakub Waszczuk

11:50 *Metaphor Detection using Context and Concreteness*
Rowan Hall Maudslay, Tiago Pimentel, Ryan Cotterell and Simone Teufel

11:55 *Being neighbourly: Neural metaphor identification in discourse*
Verna Dankers, Karan Malhotra, Gaurav Kudva, Volodymyr Medentsiy and Ekaterina Shutova

12:00 *Go Figure! Multi-task transformer-based architecture for metaphor detection using idioms: ETS team in 2020 metaphor shared task*
Xianyang Chen, Chee Wee (Ben) Leong, Michael Flor and Beata Beigman Klebanov

12:10 *Metaphor Detection using Ensembles of Bidirectional Recurrent Neural Networks*
Jennifer Brooks and Abdou Youssef

12:15 *Metaphor Detection Using Contextual Word Embeddings From Transformers*
Jerry Liu, Nathan O'Hara, Alexander Rubin, Rachel Draelos and Cynthia Rudin

12:20 *Testing the role of metadata in metaphor identification*
Egon Stemle and Alexander Onysko

Thursday July 9, 2020 (continued)

12:50 *Sarcasm Detection Using an Ensemble Approach*
Jens Lemmens, Ben Burtenshaw, Ehsan Lotfi, Ilia Markov and Walter Daelemans

12:55 *A Transformer Approach to Contextual Sarcasm Detection in Twitter*
Hunter Gregory, Steven Li, Pouya Mohammadi, Natalie Tarn, Rachel Draelos and Cynthia Rudin

13:00 *Transformer-based Context-aware Sarcasm Detection in Conversation Threads from Social Media*
Xiangjue Dong, Changmao Li and Jinho D. Choi