

FLP 2022

3rd Workshop on Figurative Language Processing

Proceedings of the Workshop

December 8, 2022

©2022 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 978-1-959429-11-1

Introduction

Welcome to the 3rd Workshop on Figurative Language Processing (FigLang 2022), to be held on December 8, 2022 as part of EMNLP in Abu Dhabi.

The use of figurative language enriches human communication by allowing us to express complex ideas and emotions. Consequently, it is not surprising that figurative language processing has become a rapidly growing area in Natural Language Processing (NLP), including metaphors, idioms, puns, irony, sarcasm, among others. 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 several 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. Recent advent of large language model-based NLP has led to novel techniques for understanding, interpreting, and creating figurative language.

This workshop is the third in a series of biannual workshops on Figurative Language Processing (following ACL 2018 and ACL 2020 installments). 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 euphemism detection and understanding of a variety (e.g., metaphor, simile, idiom, and sarcasm) of figurative language through textual explanations. The workshop is privileged to present two invited talks this year. Penny Pexman and Aline Villavicencio will be presenting talks at this year’s workshop.

In the regular research track, we received sixteen research paper submissions and accepted twelve. The featured papers cover a range of aspects of figurative language processing such as metaphor prediction and understanding (Berger; Li et al.; Wachowiak et al.; Dankin et al.; Sengupta et al.), translation of idiomatic expressions (Santing et al.), metaphor-rich translation in fictitious language (Jansen and Boyd-Graber), measure of surprise in humor and metaphor (Bunescu and Uduehi), multimodal metaphor detection in videos (Alnajjar et al.), identifying figurative content in drug lexicon (Reyes and Saldivar), and answering questions from figurative contexts (Rakshit and Flanigan).

The two shared tasks on euphemism detection and understanding of figurative language via textual explanations serve to benchmark various computational approaches to euphemism and different types of figurative language, clarifying the state of this steadily growing field and facilitating further research.

The Shared Task on Euphemism Detection invited teams to submit systems for the following task: given a text containing a potentially euphemistic term (PET), determine whether the PET is being used euphemistically or literally. The dataset used consisted of texts from the GloWbE corpus, human-annotated to be euphemistic (1) or literal (0). The goal of this task was to investigate the performance of current NLP methods on a euphemism-related task, establish a baseline from which to launch future work on euphemisms, and analyze additional enhancements attempted by participants. 46 participants spanning 13 teams attempted the task, and 9 system descriptions were submitted. Teams tested approaches such as transformer models, data balancing, linguistically motivated methods, etc., with the highest F1-scores being around 0.88.

The second shared task on understanding figurative language is designed to challenge the participants to build models to not only identify the type of figurative language but also to explain the decision via natural language. The task is based on the recently developed FLUTE dataset, which is based on four types of figurative language – idiom, sarcasm, metaphor, and simile. Out of all the models submitted, four system papers were submitted to the shared task. Although all the submitted models were based on the transformer architecture, participants did attempt different approaches – such as using elaboration of the situation first as additional contexts, sequential training on a variety of NLI datasets, and conducting multi sequence2sequence tasks. Two participants attained the highest accuracy (accuracy@60) scores of 63.33.

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 speakers for sharing their perspective on the topic, and all the attendees of the workshop. All of these factors contribute to a truly enriching event!

Debanjan Ghosh, Beata Beigman Klebanov, Smaranda Muresan, Anna Feldman, Soujanya Poria, Tuhin Chakrabarty, Workshop Co-Chairs

Organizing Committee

Workshop Organizers

Debanjan Ghosh, Educational Testing Service, USA
Beata Beigman Klebanov, Educational Testing Service, USA
Smaranda Muresan, Columbia University, USA
Anna Feldman, Montclair State University, USA
Soujanya Poria, Singapore University of Technology and Design, Singapore
Tuhin Chakrabarty, Columbia University, USA

Figurative Language Understanding Shared Task Organizers

Tuhin Chakrabarty, Columbia University, USA
Arkadiy Saakyan, Columbia University, USA
Debanjan Ghosh, Educational Testing Service, USA
Smaranda Muresan, Columbia University, USA

Euphemism Detection Shared Task Organizers

Patrick Lee, Montclair State University, USA
Anna Feldman, Montclair State University, USA
Jing Peng, Montclair State University, USA

Program Committee

Program Committee

Tariq Alhindi, Khalid Alnajjar

Yulia Badryzlova, Yuri Bizzoni, Susan Brown

Verna Dankers, Jonathan Dunn

Michael Flor

Katy Gutierrez

Mika Hämäläinen

Hyeju Jang

Valia Kordoni

Mark Lee, Els Lu

Michael Mohler, Elena Musi

Preslav Nakov

Thierry Poibeau

Paul Rayson, Ellen Riloff, Andres Torres Rivera

Farig Sadeque, Eyal Sagi, Vered Shwartz, Oprea Silviu, Carlo Stapparava, Egon Stemle, Kevin Stowe, Tomek Strzalkowski, Stan Szpakowicz

Simone Teufel, Yufei Tian

Tony Veale

Sabine Schulte im Walde

Invited Speakers

Penny Pexman, University of Calgary, Canada

Aline Villavicencio, University of Sheffield, UK

Keynote Talk: Irony Acquisition: How Children Learn to Detect Sarcasm

Penny M. Pexman
University of Calgary

Abstract: One of the challenges children face in learning to navigate the social world is created by the fact that people often speak indirectly, for example, with sarcasm or verbal irony. Research has shown that typically developing children don't usually begin to convey and appreciate ironic intent until the early school years. Children's use and appreciation of ironic language develop over a fairly long developmental window, and are related to their cognitive development and social experiences. Most of these insights have come from research that is focused on the product of interpretation: the understanding that children convey through verbal descriptions, ratings, or yes/no decisions. In a series of studies, we developed methodology that allows us to explore the process of children's irony interpretation. Using a variant of the visual world paradigm, we track children's eye gaze and reaching behavior as they judge speaker intent for ironic language that unfolds in real time. We have used this paradigm to identify factors that make irony particularly challenging for children. Most recently, those studies have helped us to devise a training paradigm to teach children to detect sarcastic speech. I'll discuss what our findings tell us about what it takes to develop a sense of sarcasm.

Bio: Penny Pexman is currently Professor of Psychology and Associate Vice-President (Research) at the University of Calgary. Penny earned her PhD in Psychology at the University of Western Ontario in 1998 and joined the University of Calgary the same year. Her research expertise is in psycholinguistics, cognitive neuroscience, and social-cognitive development. In broad terms, she is interested in how we derive meaning from language, and how those processes are changed by context or experience. Her research investigates several aspects of language understanding, ranging from lexical-semantic processes to figurative language. Penny has published over 150 journal articles and book chapters on those topics. An award-winning mentor and researcher, Penny is an elected Fellow of both the Canadian Psychological Association and the Association for Psychological Science.

Keynote Talk: Modelling Multiword Expressions and Idiomatcity: an Acid Test for Understanding

Aline Villavicencio
University of Sheffield

Abstract: Advances in large-scale word representation models have been successful in capturing distinct (and very specific) word usages in context. However, these models still face a serious challenge when dealing with non-literal or non-compositional language, like that involved in Multiword Expressions (MWEs) such as noun compounds (grandfather clock), light verb constructions (give a talk), and verb particle constructions (give up). MWEs are an integral part of the mental lexicon of native speakers often used to express complex ideas in a conventionalised way accepted by a given linguistic community, but often displaying a wealth of idiosyncrasies, from lexical, syntactic and semantic to statistical which means that they represent a real challenge for current NLP techniques. However, their accurate integration has the potential for improving the precision, naturalness and fluency of downstream tasks like machine translation and text simplification. In this talk, I will present an overview of how advances in word representations have made an impact for the identification and modelling of idiomatcity and MWEs. I will concentrate on what models seem to incorporate of idiomatcity, as idiomatic interpretation may require knowledge that goes beyond what can be gathered from the individual words of an expression (e.g. “dark horse” as an unknown candidate who unexpectedly succeeds).

Bio: Aline Villavicencio is the Chair in Natural Language Processing at the Department of Computer Science, University of Sheffield (UK). Prior to that she was affiliated as a Reader to the Institute of Informatics, Federal University of Rio Grande do Sul (Brazil), and as a Lecturer at the University of Essex (UK). She received her PhD from the University of Cambridge (UK) in 2001, and held postdoc positions at the University of Cambridge and University of Essex (UK). She was a Visiting Scholar at the Massachusetts Institute of Technology (USA, 2011-2012 and 2014-2015), at the École Normale Supérieure (France, 2014), an Erasmus-Mundus Visting Scholar at Saarland University (Germany in 2012/2013) and at the University of Bath (UK, 2006-2009). She held a Research Fellowship from the Brazilian National Council for Scientific and Technological Development (Brazil, 2009-2017). She is a member of the editorial board of Computational Linguistics, TACL and of JNLE. She was a PC Co-Chair of the 60th Meeting of the Association for Computational Linguistics (ACL 2022), and was a PC Co-Chair of CoNLL-2019, Senior Area Chair for ACL-2020 and ACL-2019 among others and General co-chair for the 2018 International Conference on Computational Processing of Portuguese. She is also a member of the NAACL board, SIGLEX board and of the program committees of various ACL and AI conferences, and has co-chaired several ACL workshops on Cognitive Aspects of Computational Language Acquisition and on Multiword Expressions. Her research interests include lexical semantics, multilinguality, multiword expressions and cognitively motivated NLP, and has co-edited special issues and books dedicated to these topics.

Table of Contents

<i>TEDB System Description to a Shared Task on Euphemism Detection 2022</i>	
Peratham Wiriyathamabhum	1
<i>A Prompt Based Approach for Euphemism Detection</i>	
Abulimiti Maimaitituoheti, Yang Yong and Fan Xiaochao	8
<i>Transfer Learning Parallel Metaphor using Bilingual Embeddings</i>	
Maria Berger	13
<i>Ring That Bell: A Corpus and Method for Multimodal Metaphor Detection in Videos</i>	
Khalid Alnajjar, Mika Hämmäläinen and Shuo Zhang	24
<i>Picard understanding Darmok: A Dataset and Model for Metaphor-Rich Translation in a Constructed Language</i>	
Peter A. Jansen and Jordan Boyd-Graber	34
<i>The Secret of Metaphor on Expressing Stronger Emotion</i>	
Yucheng Li, Frank Guerin and Chenghua Lin	39
<i>Drum Up SUPPORT: Systematic Analysis of Image-Schematic Conceptual Metaphors</i>	
Lennart Wachowiak, Dagmar Gromann and Chao Xu	44
<i>Effective Cross-Task Transfer Learning for Explainable Natural Language Inference with T5</i>	
Irina Bigoulaeva, Rachneet Singh Sachdeva, Harish Tayyar Madabushi, Aline Villavicencio and Iryna Gurevych	54
<i>Detecting Euphemisms with Literal Descriptions and Visual Imagery</i>	
Ilker Kesen, Aykut Erdem, Erkut Erdem and Iacer Calixto	61
<i>Distribution-Based Measures of Surprise for Creative Language: Experiments with Humor and Metaphor</i>	
Razvan C. Bunescu and Oseremen O. Uduehi	68
<i>Euphemism Detection by Transformers and Relational Graph Attention Network</i>	
Yuting Wang, Yiyi Liu, Ruqing Zhang, Yixing Fan and Jiafeng Guo	79
<i>Just-DREAM-about-it: Figurative Language Understanding with DREAM-FLUTE</i>	
Yuling Gu, Yao Fu, Valentina Pyatkin, Ian Magnusson, Bhavana Dalvi Mishra and Peter Clark	84
<i>Bayes at FigLang 2022 Euphemism Detection shared task: Cost-Sensitive Bayesian Fine-tuning and Venn-Abers Predictors for Robust Training under Class Skewed Distributions</i>	
Paul Trust, Kadusabe Provia and Kizito Omala	94
<i>Food for Thought: How can we exploit contextual embeddings in the translation of idiomatic expressions?</i>	
Lukas Santing, Ryan Jean-Luc Sijstermans, Giacomo Anerdi, Pedro Jeuris, Marijn ten Thij and Riza Batista-Navarro	100
<i>EUREKA: EUPhemism Recognition Enhanced through Knn-based methods and Augmentation</i>	
Sedrick Scott Keh, Rohit Bharadwaj, Emmy Liu, Simone Tedeschi, Varun Gangal and Roberto Navigli	111
<i>An insulin pump? Identifying figurative links in the construction of the drug lexicon</i>	
Antonio Reyes and Rafael Saldivar	118

<i>Can Yes-No Question-Answering Models be Useful for Few-Shot Metaphor Detection?</i>	
Lena Dankin, Kfir Bar and Nachum Dershowitz	125
<i>An Exploration of Linguistically-Driven and Transfer Learning Methods for Euphemism Detection</i>	
Devika Tiwari and Natalie Parde	131
<i>Back to the Roots: Predicting the Source Domain of Metaphors using Contrastive Learning</i>	
Meghdut Sengupta, Milad Alshomary and Henning Wachsmuth	137
<i>SBU Figures It Out: Models Explain Figurative Language</i>	
Yash Kumar Lal and Mohaddeseh Bastan	143
<i>NLP@UIT at FigLang-EMNLP 2022: A Divide-and-Conquer System For Shared Task On Understanding Figurative Language</i>	
Khoa Thi-Kim Phan, Duc-Vu Nguyen and Ngan Luu-Thuy Nguyen	150
<i>Adversarial Perturbations Augmented Language Models for Euphemism Identification</i>	
Guneet Kohli, Prabsimran Kaur and Jatin Bedi	154
<i>FigurativeQA: A Test Benchmark for Figurativeness Comprehension for Question Answering</i>	
Geetanjali Rakshit and Jeffrey Flanigan	160
<i>Exploring Euphemism Detection in Few-Shot and Zero-Shot Settings</i>	
Sedrick Scott Keh	167
<i>On the Cusp of Comprehensibility: Can Language Models Distinguish Between Metaphors and Non-sense?</i>	
Bernadeta Griciūtė, Marc Tanti and Lucia Donatelli	173
<i>A Report on the FigLang 2022 Shared Task on Understanding Figurative Language</i>	
Arkadiy Saakyan, Tuhin Chakrabarty, Debanjan Ghosh and Smaranda Muresan	178
<i>A Report on the Euphemisms Detection Shared Task</i>	
Patrick Lee, Anna Feldman and Jing Peng	184

Program

Thursday, December 8, 2022

08:50 - 09:00 *Opening Remarks*

09:00 - 10:30 *Research Track*

Ring That Bell: A Corpus and Method for Multimodal Metaphor Detection in Videos

Khalid Alnajjar, Mika Hämmäläinen and Shuo Zhang

Food for Thought: How can we exploit contextual embeddings in the translation of idiomatic expressions?

Lukas Santing, Ryan Jean-Luc Sijstermans, Giacomo Anerdi, Pedro Jeuris, Marijn ten Thij and Riza Batista-Navarro

Distribution-Based Measures of Surprise for Creative Language: Experiments with Humor and Metaphor

Razvan C. Bunescu and Oseremen O. Uduehi

The Secret of Metaphor on Expressing Stronger Emotion

Yucheng Li, Frank Guerin and Chenghua Lin

Back to the Roots: Predicting the Source Domain of Metaphors using Contrastive Learning

Meghdut Sengupta, Milad Alshomary and Henning Wachsmuth

Can Yes-No Question-Answering Models be Useful for Few-Shot Metaphor Detection?

Lena Dankin, Kfir Bar and Nachum Dershowitz

On the Cusp of Comprehensibility: Can Language Models Distinguish Between Metaphors and Nonsense?

Bernadeta Griciūtė, Marc Tanti and Lucia Donatelli

10:30 - 11:00 *Coffee Break*

11:00 - 12:30 *Research Track + Shared Tasks*

A Report on the Euphemisms Detection Shared Task

Patrick Lee, Anna Feldman and Jing Peng

Thursday, December 8, 2022 (continued)

EUREKA: EUPhemism Recognition Enhanced through Knn-based methods and Augmentation

Sedrick Scott Keh, Rohit Bharadwaj, Emmy Liu, Simone Tedeschi, Varun Gangal and Roberto Navigli

Detecting Euphemisms with Literal Descriptions and Visual Imagery

Ilker Kesen, Aykut Erdem, Erkut Erdem and Iacer Calixto

A Report on the FigLang 2022 Shared Task on Understanding Figurative Language

Arkadiy Saakyan, Tuhin Chakrabarty, Debanjan Ghosh and Smaranda Muresan

Just-DREAM-about-it: Figurative Language Understanding with DREAM-FLUTE

Yuling Gu, Yao Fu, Valentina Pyatkin, Ian Magnusson, Bhavana Dalvi Mishra and Peter Clark

Effective Cross-Task Transfer Learning for Explainable Natural Language Inference with T5

Irina Bigoulaeva, Rachneet Singh Sachdeva, Harish Tayyar Madabushi, Aline Villavicencio and Iryna Gurevych

Drum Up SUPPORT: Systematic Analysis of Image-Schematic Conceptual Metaphors

Lennart Wachowiak, Dagmar Gromann and Chao Xu

Transfer Learning Parallel Metaphor using Bilingual Embeddings

Maria Berger

12:30 - 14:00 *Lunch Break*

14:00 - 15:00 *Keynote Talk 1: Aline Villavicencio: Modelling Multiword Expressions and Idiomaticity: an Acid Test for Understanding*

15:00 - 15:30 *Research Track*

An insulin pump? Identifying figurative links in the construction of the drug lexicon

Antonio Reyes and Rafael Saldivar

Picard understanding Darmok: A Dataset and Model for Metaphor-Rich Translation in a Constructed Language

Peter A. Jansen and Jordan Boyd-Graber

Thursday, December 8, 2022 (continued)

FigurativeQA: A Test Benchmark for Figurativeness Comprehension for Question Answering

Geetanjali Rakshit and Jeffrey Flanigan

15:30 - 16:00 *Coffee Break*

16:00 - 17:30 *Poster Session (Shared Tasks + Findings)*

TEDB System Description to a Shared Task on Euphemism Detection 2022

Peratham Wiriathamabhum

A Prompt Based Approach for Euphemism Detection

Abulimiti Maimaitituoheti, Yang Yong and Fan Xiaochao

Euphemism Detection by Transformers and Relational Graph Attention Network

Yuting Wang, Yiyi Liu, Ruqing Zhang, Yixing Fan and Jiafeng Guo

Bayes at FigLang 2022 Euphemism Detection shared task: Cost-Sensitive Bayesian Fine-tuning and Venn-Abers Predictors for Robust Training under Class Skewed Distributions

Paul Trust, Kadusabe Provia and Kizito Omala

An Exploration of Linguistically-Driven and Transfer Learning Methods for Euphemism Detection

Devika Tiwari and Natalie Parde

Adversarial Perturbations Augmented Language Models for Euphemism Identification

Guneet Kohli, Prabsimran Kaur and Jatin Bedi

Exploring Euphemism Detection in Few-Shot and Zero-Shot Settings

Sedrick Scott Keh

SBU Figures It Out: Models Explain Figurative Language

Yash Kumar Lal and Mohaddeseh Bastan

NLP@UIT at FigLang-EMNLP 2022: A Divide-and-Conquer System For Shared Task On Understanding Figurative Language

Khoa Thi-Kim Phan, Duc-Vu Nguyen and Ngan Luu-Thuy Nguyen

Thursday, December 8, 2022 (continued)

Visualizing the Obvious: A Concreteness-based Ensemble Model for Noun Property Prediction

Yue Yang, Artemis Panagopoulou, Marianna Apidianaki, Mark Yatskar and Chris Callison-Burch

Sarcasm Detection is Way Too Easy! An Empirical Comparison of Human and Machine Sarcasm Detection

Ibrahim Abu Farha, Steven Wilson, Silviu Oprea and Walid Magdy

A Unified Framework for Pun Generation with Humor Principles

Yufei Tian, Divyanshu Sheth and Nanyun Peng

It's Better to Teach Fishing than Giving a Fish: An Auto-Augmented Structure-aware Generative Model for Metaphor Detection

Huawen Feng and Qianli Ma

Systematicity in GPT-3's Interpretation of Novel English Noun Compounds

Siyang Li, Riley Carlson and Christopher Potts

PoeLM: A Meter- and Rhyme-Controllable Language Model for Unsupervised Poetry Generation

Aitor Ormazabal, Mikel Artetxe, Manex Agirrezabal, Aitor Soroa and Eneko Agirre

Scientific and Creative Analogies in Pretrained Language Models

Tamara Czinczoll, Helen Yannakoudakis, Pushkar Mishra and Ekaterina Shutova

Cards Against AI: Predicting Humor in a Fill-in-the-blank Party Game

Dan Ofer and Dafna Shahaf

17:30 - 17:55 *Break*

17:55 - 19:00 *Keynote Talk 2: Penny M. Pexman: Irony Acquisition: How Children Learn to Detect Sarcasm*