FigLang 2024

4th Workshop on Figurative Language Processing

Proceedings of the Workshop

June 21, 2024

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-110-0

Introduction

Welcome to the 4th Workshop on Figurative Language Processing (FigLang 2024), to be held on June 21, 2024 as part of NAACL in Mexico City, Mexico.

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 fourth in a series of biannual workshops on Figurative Language Processing (following ACL 2018, ACL 2020, and EMNLP 2022 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 Multilingual Euphemism Detection and Multimodal Figurative Language. The workshop is privileged to present one invited talk this year. Dr. Vered Shwartz will be presenting talks at this year's workshop on whether LLMs have solved figurative language.

In the regular research track, we received twenty two research paper submissions and accepted nine. The featured papers cover a range of aspects of figurative language processing such as disagreement in sarcasm detection (Jang et al.), multimodal generation such as images (Khaliq et al.), metaphor detection in cross-lingual setting (Hulsing et al.) annotation guidelines for identifying metaphors (Dippet et al.), metaphor annotation in Mexican Spanish popular science tweets (Montero et al.), expectation-realization model for metaphor detection (Uduehi and Bunescu), idiom detection (Fornaciari et al.), distribution of personification in Hungarian (Simon), and a summary paper on challenges of rhetorical figures detection (Kuhn and Mitrović).

The two shared tasks on Multilingual Euphemism Detection and Multimodal Figurative Language 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.

In the Multilingual Euphemism Detection Shared Task, participants were invited to develop models to classify texts in various languages as either euphemistic or not. The previous iteration used only an English dataset. This time, we included data in American English (EN), Spanish (ES), Yorùbá (YO), and Mandarin Chinese (ZH) to broaden the insights across languages and facilitate transfer learning for identifying cross-lingual patterns. The datasets consisted of texts from diverse sources including online articles, webpages, transcribed texts, and social media posts. Each text, containing up to three sentences with a potentially euphemistic term (PET), was annotated by humans to indicate euphemistic (1) or non-euphemistic (0) usage. During the development phase, participants were provided with datasets in all four languages. During the test phase, participants were provided a test set for each language and had the option of submitting predictions for one to four of them for scoring. However, all teams ultimately chose to submit predictions for all four. Submissions were evaluated based on the Macro-F1 score,

with equal weighting across languages. Three participating teams submitted system descriptions and achieved scores significantly above baselines but below their reported validation metrics. The different approaches are described in the shared task's summary paper, and the outcomes not only demonstrate the effectiveness of current approaches but also underscore the need for further research into large language models, ensemble techniques, and task-related strategies. Future studies should also explore the broader impact of PETs on model behavior and the potential connections to other linguistic tasks.
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.

Finally, we acknowledge NSF for their generous grant (grant #2226006) with which we are able to support registrations as well as travel and accommodation of a few individual.

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, Smaranda Muresan, Anna Feldman, Tuhin Chakrabarty, Emmy Liu, Workshop Co-Chairs

Organizing Committee

Workshop Organizers

Debanjan Ghosh, Educational Testing Service, USA
Smaranda Muresan, Columbia University, USA
Anna Feldman, Montclair State University, USA
Tuhin Chakrabarty, Columbia University, USA
Emmy Liu, Carnegie Mellon University, USA

Euphemisms Shared Task Organizers

Patrick Lee, Amazon
Anna Feldman, Montclair State University, USA

Multimodal Shared Task Organizers

Arkadiy Saakyan, Columbia University, USA
Shreyas Kulkarni, Columbia University, USA
Tuhin Chakrabarty, Columbia University, USA
Smaranda Muresan, Columbia University, USA

Program Committee

Program Chairs

Tuhin Chakrabarty, SalesForce Research
Anna Feldman, Montclair State University and Montclair State University
Debanjan Ghosh, Educational Testing Service and Massachusetts Institute of Technology
Emmy Liu, School of Computer Science, Carnegie Mellon University
Smaranda Muresan, Columbia University

Area Chairs

Tuhin Chakrabarty, SalesForce Research
Anna Feldman, Montclair State University and Montclair State University
Debanjan Ghosh, Educational Testing Service and Massachusetts Institute of Technology
Emmy Liu, School of Computer Science, Carnegie Mellon University
Smaranda Muresan, Amazon and Columbia University

Reviewers

Arav Kapish Agarwal, Khalid Alnajjar

Yulia Badryzlova, Yuri Bizzoni

Tuhin Chakrabarty, Jiale Chen

Verna Dankers, Rahul Divekar, Jonathan Dunn

Anna Feldman, Michael Flor

Lingyu Gao, Debanjan Ghosh

Nicholas Hankins, Mika Hämäläinen

Loukas Ilias

Hyeju Jang

Raghav Kapoor, Sedrick Keh, Ilker Kesen, Guneet Singh Kohli, Valia Kordoni, Shreyas Kulkarni

Yash Kumar Lal, Mark G. Lee, Patrick Lee, Els Lefever, Emmy Liu, Ziqian Luo

Smaranda Muresan, Elena Musi

Geetanjali Rakshit, Ellen Riloff, Anthony Rios, Andrés Torres Rivera

Arkadiy Saakyan, Farig Sadeque, Eyal Sagi, Sabine Schulte Im Walde, Meghdut Sengupta, Egon Stemle, Kevin Stowe, Carlo Strapparava, Tomek Strzalkowski, Stan Szpakowicz

Yufei Tian, Xiaoyu Tong

Tony Veale, Fedor Vitiugin

Peratham Wiriyathammabhum

Qihao Yang

Keynote Talk

Did language models "solve" figurative language?

Vered Shwartz
University of British Columbia
2024-06-21 –

Abstract: Figurative expressions, such as idioms, similes, and metaphors, are ubiquitous in English. For many years, they have been considered a pain in the neckfor NLP applications, due to their non-compositional nature. With LLMs excelling at understanding and generating English texts, it's time to ask: did LLMs solvefigurative language? Is it possible that the sheer amount of exposure to figurative language in their training data equipped them with the ability to understand and use figurative language? I will discuss the state of LLMs in recognizing figurative usage, interpreting figurative expressions in context, and usage of figurative language in generated text.

Bio: Vered Shwartz is an Assistant Professor of Computer Science at the University of British Columbia. Her research interests include commonsense reasoning, computational semantics and pragmatics, and multiword expressions. Previously, Vered was a postdoctoral researcher at the Allen Institute for AI (AI2) and the University of Washington, and received her PhD in Computer Science from Bar-Ilan University. Vered's work has been recognized with several awards, including The Eric and Wendy Schmidt Postdoctoral Award for Women in Mathematical and Computing Sciences, the Clore Foundation Scholarship, and an ACL 2016 outstanding paper award.

Table of Contents

Program

Friday, June 21, 2024

08:50 - 09:00 *Opening Remarks*

09:00 - 10:30 *Research Track*

Context vs. Human Disagreement in Sarcasm Detection
Hyewon Jang, Moritz Jakob and Diego Frassinelli

The Register-specific Distribution of Personification in Hungarian: A Corpus-driven Analysis
Gabor Simon

Comparison of Image Generation Models for Abstract and Concrete Event Descriptions
Mohammed Abdul Khaliq, Diego Frassinelli and Sabine Schulte Im Walde

Cross-Lingual Metaphor Detection for Low-Resource Languages
Anna Hülsing and Sabine Schulte Im Walde

A Hard Nut to Crack: Idiom Detection with Conversational Large Language Models
Francesca De Luca Fornaciari, Begoña Altuna, Itziar Gonzalez-Dios and Maite Melero

The Elephant in the Room: Ten Challenges of Computational Detection of Rhetorical Figures
Ramona Kühn and Jelena Mitrović

10:30 - 11:00 *Coffee Break*

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

Report on the Multilingual Euphemism Detection Task
Patrick Lee and Anna Feldman

A Report on the FigLang 2024 Shared Task on Multimodal Figurative Language
Shreyas Kulkarni, Arkadiy Saakyan, Tuhin Chakrabarty and Smaranda Muresan

Friday, June 21, 2024 (continued)

Optimizing Multilingual Euphemism Detection using Low-Rank Adaption Within and Across Languages
Nicholas Hankins

Guidelines for the Annotation of Intentional Linguistic Metaphor
Stefanie Dipper, Adam Roussel, Alexandra Wiemann, Won Kim and Tra-my Nguyen

Evaluating the Development of Linguistic Metaphor Annotation in Mexican Spanish Popular Science Tweets
Alec Misael Sánchez Montero, Gemma Bel-Enguix, Sergio-Luis Ojeda-Trueba and Marisela Colín Rodea

Can GPT4 Detect Euphemisms across Multiple Languages?
Todd E Firsich and Anthony Rios

12:30 - 14:00 *Lunch Break*

14:00 - 15:00 *Keynote Talk 1: Vered Shwartz: Did language models "solve" figurative language?*

15:00 - 15:30 *Research Track + Shared Tasks*

A Textual Modal Supplement Framework for Understanding Multi-Modal Figurative Language
Jiale Chen, Qihao Yang, Xuelian Dong, Xiaoling Mao and Tianyong Hao

An Expectation-Realization Model for Metaphor Detection: Within Distribution, Out of Distribution, and Out of Pretraining
Oseremen Oscar Uduehi and Razvan Bunescu

15:30 - 16:00 *Coffee Break*

16:00 - 16:30 *Shared Tasks*

Ensemble-based Multilingual Euphemism Detection: a Behavior-Guided Approach
Fedor Vitiugin and Henna Paakki

FigCLIP: A Generative Multimodal Model with Bidirectional Cross-attention for Understanding Figurative Language via Visual Entailment
Qihao Yang and Xuelin Wang