MWE 2026

The 22nd Workshop on Multiword Expressions (MWE 2026)

Proceedings of the Workshop

March 28, 2026

The MWE organizers gratefully acknowledge the support from the following organizations.

Gold

©2026 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 979-8-89176-363-0

Introduction

The 22nd Workshop on Multiword Expressions (MWE 2026) took place on 28th March 2026, in Rabat, Marocco, collocated with the 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026).
MWE 2026 is supported by the Special Interest Group on the Lexicon (SIGLEX) of the Association for Computational Linguistics (ACL), SIGLEX's Multiword Expressions Section (SIGLEX-MWE), and UniDive COST Action CA21167.
The notion of multiword expressions (MWEs) encompasses a range of closely related phenomena: idioms, compounds, light-verb constructions, phrasal verbs, rhetorical figures, collocations, institutionalized phrases, etc. They exhibit lexical, syntactic, semantic, pragmatic, and/or statistical idiosyncrasies. Given their irregular nature, MWEs often pose complex problems in various natural language processing (NLP) tasks such as language understanding and machine translation, hence still represent an open issue for computational linguistics.
For the 22nd edition of the workshop, our call for papers focused particularly on the following topics:

- Annotation (expert, crowdsourcing, automatic) and representation in resources such as corpora, treebanks, e-lexicons, WordNets, constructions (also for low-resource languages);
- Processing in syntactic and semantic frameworks (e.g. CCG, CxG, HPSG, LFG, TAG, UD, etc.);
- Evaluation of annotation and processing techniques;
- Discovery and identification methods, including for specialized languages and domains such as clinical or biomedical NLP;
- Interpretation of MWEs and understanding of text containing them;
- Language acquisition, language learning, and non-standard language (e.g. tweets, speech);
- Computationally-applicable theoretical work in psycholinguistics and corpus linguistics;
- Processing for end-user applications (e.g. MT, NLU, summarisation, language learning, etc.);
- Implicit and explicit representation in pre-trained language models and end-user applications;
- Evaluation and probing of pre-trained language models;
- Resources and tools (e.g. lexicons, identifiers) and their integration into end-user applications;
- Adaptation and transfer of annotations and related resources to new languages and domains including low-resource ones.

For this edition, 75% of the submitted papers were accepted.
The dominating trend in MWE research presented at the workshop focuses on large language models, particularly examining their capabilities and limitations across diverse tasks. Multiple studies compare performance of task-specific fine-tuned models to general-purpose generative LLMs on MWE identification tasks. Domain-specific applications (in the technical domain, bioinformatics, etc.) reveal how MWE extraction methods can be tailored to specialised contexts. Several papers identify systematic features that current LLMs fail to capture with respect to MWEs: cognitive studies reveal distinct processing signatures for different MWE types that computational models miss, while semantic interpretation tasks expose fundamental limitations in achieving fine-grained understanding of compound meaning and implicit relations between constituents.
The workshop demonstrates substantial multilingual coverage, addressing languages generally underrepresented in MWE research and exposing cross-linguistic variation in MWE behavior. Papers present

research on Swedish, Ukrainian, Romanian, Galician, Marathi, Turkish, Korean, Sinhala, and Chinese. Several studies explicitly employ the PARSEME annotation framework to enable cross-linguistic comparison, discussing the methodological challenges in ensuring annotation consistency and inter-annotator agreement across languages.
Research focuses on language-specific phenomena that challenge existing annotation schemes and identification methodologies, particularly in morphologically complex and typologically distinct languages. Studies on agglutinative and postpositional languages identify constructions that are difficult to categorise within existing frameworks. Low-resource language studies discuss limitations in handling culturally-specific phenomena, rare expressions and MWEs absent in training corpora, a challenge that persists despite advances in LLMs.
The MWE 2026 Workshop hosted two shared tasks: PARSEME 2.0, whose objective is to identify and paraphrase MWEs in written text, and AdMIRe 2 (Advancing Multimodal Idiomaticity Representation), which explores the comprehension ability of multimodal models for MWEs in a variety of languages.

Verginica Barbu Mititelu, Mathieu Constant, A. Seza Doğruöz, Atul Kr. Ojha, Alexandre Rademaker, Ivelina Stoyanova (MWE-2026 Organizers and Co-Chairs)

Organizing Committee

Workshop Chairs

Mathieu Constant, Université de Lorraine, CNRS, ATILF
A. Seza Doğruöz, Ghent University
Ivelina Stoyanova, Institute for Bulgarian Language, Bulgarian Academy of Sciences
Verginica Barbu Mititelu, Romanian Academy Research Institute for Artificial Intelligence
Atul Kr. Ojha, Insight Research Ireland Centre for Data Analytics, DSI, University of Galway, Ireland and Panlingua Languague Processing LLP, India
Alexandre Rademaker, School of Applied Mathematics of Getulio Vargas Foundation

Program Committee

Abigail Walsh, Dublin City University
Agata Savary, Université Paris-Saclay
Ahmet Erdem, Istanbul Technical University
Alberto Barrón-Cedeño, Università di Bologna
Ali Azmoudeh, Istanbul Technical University
Andrea Horbach, Leibniz Institute for Science and Mathematics Education
Andrei Tiberiu Carp, Tomorrow University of Applied Sciences
Anna Hülsing, Christian-Albrechts-Universität Kiel
Atakan Site, Istanbul Technical University
Barış Bilen, Istanbul Technical University
Beata Trawinski, Leibniz Institute for the German Language
Bora Şenceylan, Istanbul Technical University
Carlos Ramisch, LIS - Laboratoire d'Informatique et Systèmes
Chikara Hashimoto, Rakuten Institute of Technology
Cristea Alexandru-Marian, University of Bucharest
Cvetana Krstev, University of Belgrade, Faculty of Philology
David Cotigă, University of Bucharest
Debora Ciminari, University of Bologna
Doğukan Arslan, Istanbul Technical University
Elif Bayraktar, Istanbul Technical University
Emmanuele Chersoni, The Hong Kong Polytechnic University
Eric G C Laporte, Université Gustave Eiffel
Gaël Dias, University of Caen Normandy
Gražina Korvel, Vilnius University
Gülşen Eryiğit, Istanbul Technical University
Irina Lobzhanidze, Ilia Chavchavadze State University
Irina Moise, University of Bucharest
Ismail El Maarouf, Imprevicible
Ivelina Stoyanova, Deaf Studies Institute
Jan Odijk, Utrecht University
John Philip McCrae, University of Galway
Kenneth Church, Northeastern University
Kubilay Kağan Kömürcü, Istanbul Technical University
Laura A. Michaelis, University of Colorado at Boulder
Le Qiu, The Hong Kong Polytechnic University
Manfred Sailer, Johann Wolfgang Goethe Universität Frankfurt am Main

Manon Scholivet, Université Paris-Saclay
Maria Mitrofan, Research Institute for Artificial Intelligence
Mathieu Constant, Université de Lorraine, CNRS, ATILF
Matthew Shardlow, The Manchester Metropolitan University
Meghdad Farahmand, University of Genoa
Mehmet Utku Colak, Istanbul Technical University
Miriam Butt, Universität Konstanz
Monika Czerepowicka, University of Wamia and Masuria
Muhammed Abdullah Gümüş, International Technological University
Nina Hosseini-Kivanani, RTL
Oğuz Ali Arslan, Istanbul Technical University
Özge Umut, Istanbul Technical University
Oguzhan Karaarslan, Istanbul Technical University
Paul Cook, University of New Brunswick
Petya Osenova, Sofia University St. Kliment Ohridski
Ranka Stanković Stanković, University of Belgrade
Rares-Alexandru Roscan, University of Bucharest
Sabine Schulte im Walde, University of Stuttgart
Sergiu Nisioi, University of Bucharest
Shiva Taslimipoor, University of Cambridge
Stan Szpakowicz, University of Ottawa
Stella Markantonatou, ATHENA RIC
Tugce Temel, Istanbul Technical University
Tiberiu Boros, Adobe Systems
Tunga Gungor, Bogazici University
Veronika Vincze, University of Szeged
Yu-Yin Hsu, The Hong Kong Polytechnic University
Yunus Karatepe, International Technological University

Table of Contents

Large Language Models Put to the Test on Chinese Noun Compounds: Experiments on Natural Language Inference and Compound Semantics
Le Qiu, Emmanuele Chersoni, He Zhou and Yu-Yin Hsu 1

SinFoS: A Parallel Dataset for Translating Sinhala Figures of Speech
Johan Nevin Sofalas, Dilushri Pavithra, Nevidu Jayatilleke and Ruvan Weerasinghe 8

Swedish Multiword Expression Corpora in PARSEME
Sara Stymne, Astrid Berntsson Ingelstam and Eva Pettersson 27

Ukrainian Multiword Expressions Corpus: Creation, Annotation, and Linguistic Analysis
Hanna Sytar, Maria Shvedova and Olha Kanishcheva 38

Cognitive Signatures of Multi-Word Expressions: Reading-Time and Surprisal
Diego Alves, Sergei Bagdasarov and Elke Teich 48

Cheese it up: CamemBERT Outperforms Large Language Models for Identification of French Multiword Expressions
Sergei Bagdasarov, Diego Alves and Elke Teich 54

Extracting Multi-Word Expressions Representing Technical Terms and Proper Nouns in Log Messages
Kilian Dangendorf, Sven-Ove Hänsel, Jannik Rosendahl, Felix Heine, Carsten Kleiner and Christian Wartena 61

Two Birds with One Stone: Annotating Romanian Multiword Expressions with an Eye to the PARSEME 2.0 Guidelines Applicability
Verginica Mititelu, Mihaela Cristescu, Elena Irimia and Carmen Mîrzea Vasile 66

Incorporating Multiword Expressions in Galician Neural Machine Translation: Compositionality, Efficiency, and Performance
Daniel Solla, Paula Pinto-Ferro, Laura Castro, Pablo Gamallo and Marcos Garcia 75

Beyond Single Words: MWE Identification in Bioinformatics Research Articles and Dispersion Profiling Across IMRaD
Jurgi Giraud and Andrew Gargett 86

The Lock, Stock, and Barrel of Marathi Multiwords
Aakanksha Padhye and Ashwini Vaidya 96

An Idiom Benchmark for Turkish
Ebru Çavuşoğlu and Cagri Coltekin 103

Diversity patterns run deep: Impact of diversity intake on multiword expression identification
Mathilde Deletombe, Manon Scholivet, Louis Estève, Thomas Lavergne and Agata Savary . . 110

A Curious Class of Adpositional Multiword Expressions in Korean
Junghyun Min, Na-Rae Han, Jena D. Hwang and Nathan Schneider 117

PolyFrame at MWE-2026 AdMIRe 2: When Words Are Not Enough: Multimodal Idiom Disambiguation
Nina Hosseini-Kivanani 127

IdiomRanker-X at MWE-2026 AdMIRe 2: Multilingual Idiom-Image Alignment via Low-Rank Adaptation of Cross-Encoders
Mehmet Utku Colak 134

alexandru412 at MWE-2026 AdMIRe 2.0: Advancing Multimodal Idiomaticity Representation
Cristea Alexandru-Marian 139

BeeParser at MWE-2026 PARSEME 2.0 Subtask 1: Can Cross-Lingual Interactions Improve MWE Identification?
Ahmet Erdem and Oguzhan Karaarslan 144

VisAffect at MWE-2026 AdMIRe 2: IMMCAN Idiom Multimodal Cross-Attention Network
Barış Bilen, Ali Azmoudeh, Hazım Kemal Ekenel and Hatice Kose 149

Sahara Tokenizers at PARSEME 2.0 Subtask 1: Combining Contextual Embeddings with Structural Decoding for Multi-Word Expression Detection
Yunus Karatepe, Mert Sülük, Zeynep Tuğçe Kırımlı and Begüm Özbay 154

3K2T at MWE-2026 AdMIRe 2: CARIM– Category-Aware Reasoning for Idiomatic Multimodality
Kubilay Kağan Kömürcü and Tugce Temel 160

PMI MWE Scorer at PARSEME 2.0 Subtask 1: identifying multi-word expressions using pointwise mutual information and universal dependencies
Anna Bogdanova and Ileana Bucur 165

tiberiucarp at MWE-2026 AdMIRe 2: GLIMMER-Gloss-based Image Multiword Meaning Expression Ranker
Andrei Tiberiu Carp 170

IPN at MWE-2026 PARSEME 2.0 Subtask 1: MWE Identification via Related Languages and Harnessing Thinking Mode
Anna Hülsing, Noah-Manuel Michael, Daniel Mora Melanchthon and Andrea Horbach 177

Semantic Stars at MWE-2026 PARSEME 2.0 Subtask 2: Alternative Approaches for MWE Paraphrasing
Elif Bayraktar, Vedat Doğancan, Muhammed Abdullah Gümüş and Nusret Ali Kızılaslan . . . 187

MorphoFiltered-Gemini at MWE-2026 PARSEME 2.0 Subtask 1: Tackling LLM Overgeneration via Universal POS-based Constraints
Irina Moise and Sergiu Nisioi 196

LST at MWE-2026 AdMIRe 2: Advancing Multimodal Idiomaticity Representation
Le Qiu, Yu-Yin Hsu and Emmanuele Chersoni 203

UniBO at MWE-2026 PARSEME 2.0 Subtask 2: A Cross-lingual Approach to Multiword Expression Paraphrasing
Debora Ciminari and Alberto Barrón-Cedeño 208

DCSN-NLP at MWE-2026 AdMIRe 2: Bridging Literal and Figurative Meaning Through Hierarchical Multimodal Reasoning
David Cotigă and Sergiu Nisioi 217

ITUNLP at MWE-2026 AdMIRe 2: A Zero-Shot LLM Pipeline for Multimodal Idiom Understanding and Ranking
Atakan Site, Oğuz Ali Arslan and Gülşen Eryiğit 226

Archaeology at WE-2026 PARSEME 2.0 Subtask 1 and 2: Parsing is for Encoders, Paraphrasing is for LLMs
Rares-Alexandru Roscan and Sergiu Nisioi 237

ITUNLP2 at MWE-2026 AdMIRe 2: Modular Zero-Shot Pipelines for Multimodal Idiom Grounding and Ranking
Özge Umut and Bora Şenceylan 248

Edition 2.0 of the PARSEME shared task on multilingual identification and paraphrasing of multiword expressions

Manon Scholivet, Agata Savary, Carlos Ramisch, Eric Bilinski, Takuya Nakamura, Maria Mitrofan and Vasile Pais 254

MWE-2026 Shared Task: AdMIRe 2 Advancing Multimodal Idiomaticity Representation

Doğukan Arslan, Rodrigo Wilkens, Wei He, Dilara Torunoglu Selamet, Thomas Pickard, Aline Villavicencio, Adriana Silvina Pagano and Gülşen Eryiğit 276

Program

Saturday, March 28, 2026

09:00 - 09:15 *Welcome and Introduction to 22nd MWE Workshop*

09:15 - 09:45 *Findings of the MWE 2026 Shared Tasks*

Edition 2.0 of the PARSEME shared task on multilingual identification and paraphrasing of multiword expressions
Manon Scholivet, Agata Savary, Carlos Ramisch, Eric Bilinski, Takuya Nakamura, Maria Mitrofan and Vasile Pais

MWE-2026 Shared Task: AdMIRe 2 Advancing Multimodal Idiomaticity Representation
Doğukan Arslan, Rodrigo Wilkens, Wei He, Dilara Torunoglu Selamet, Thomas Pickard, Aline Villavicencio, Adriana Silvina Pagano and Gülşen Eryiğit

09:45 - 10:30 *Oral Session*

Swedish Multiword Expression Corpora in PARSEME
Sara Stymne, Astrid Berntsson Ingelstam and Eva Pettersson

Cognitive Signatures of Multi-Word Expressions: Reading-Time and Surprisal
Diego Alves, Sergei Bagdasarov and Elke Teich

Diversity patterns run deep: Impact of diversity intake on multiword expression identification
Mathilde Deletombe, Manon Scholivet, Louis Estève, Thomas Lavergne and Agata Savary

10:30 - 11:00 *Coffee Break*

11:00 - 12:00 *Poster Session*

Large Language Models Put to the Test on Chinese Noun Compounds: Experiments on Natural Language Inference and Compound Semantics
Le Qiu, Emmanuele Chersoni, He Zhou and Yu-Yin Hsu

SinFoS: A Parallel Dataset for Translating Sinhala Figures of Speech
Johan Nevin Sofalas, Dilushri Pavithra, Nevidu Jayatilleke and Ruvan Weerasinghe

Ukrainian Multiword Expressions Corpus: Creation, Annotation, and Linguistic Analysis
Hanna Sytar, Maria Shvedova and Olha Kanishcheva

Cheese it up: CamemBERT Outperforms Large Language Models for Identification of French Multi-word Expressions
Sergei Bagdasarov, Diego Alves and Elke Teich

Extracting Multi-Word Expressions Representing Technical Terms and Proper Nouns in Log Messages
Kilian Dangendorf, Sven-Ove Hänsel, Jannik Rosendahl, Felix Heine, Carsten Kleiner and Christian Wartena

Two Birds with One Stone: Annotating Romanian Multiword Expressions with an Eye to the PARSEME 2.0 Guidelines Applicability
Verginica Mititelu, Mihaela Cristescu, Elena Irimia and Carmen Mîrzea Vasile

Incorporating Multiword Expressions in Galician Neural Machine Translation: Compositionality, Efficiency, and Performance
Daniel Solla, Paula Pinto-Ferro, Laura Castro, Pablo Gamallo and Marcos Garcia

Beyond Single Words: MWE Identification in Bioinformatics Research Articles and Dispersion Profiling Across IMRaD
Jurgi Giraud and Andrew Gargett

The Lock, Stock, and Barrel of Marathi Multiwords
Aakanksha Padhye and Ashwini Vaidya

An Idiom Benchmark for Turkish
Ebru Çavuşoğlu and Cagri Coltekin

A Curious Class of Adpositional Multiword Expressions in Korean
Junghyun Min, Na-Rae Han, Jena D. Hwang and Nathan Schneider

PolyFrame at MWE-2026 AdMIRe 2: When Words Are Not Enough: Multimodal Idiom Disambiguation
Nina Hosseini-Kivanani

IdiomRanker-X at MWE-2026 AdMIRe 2: Multilingual Idiom-Image Alignment via Low-Rank Adaptation of Cross-Encoders
Mehmet Utku Colak

alexandru412 at MWE-2026 AdMIRe 2.0: Advancing Multimodal Idiomaticity Representation
Cristea Alexandru-Marian

BeeParser at MWE-2026 PARSEME 2.0 Subtask 1: Can Cross-Lingual Interactions Improve MWE Identification?
Ahmet Erdem and Oguzhan Karaarslan

VisAffect at MWE-2026 AdMIRe 2: IMMCAN Idiom Multimodal Cross-Attention Network
Barış Bilen, Ali Azmoudeh, Hazım Kemal Ekenel and Hatice Kose

Sahara Tokenizers at PARSEME 2.0 Subtask 1: Combining Contextual Embeddings with Structural Decoding for Multi-Word Expression Detection
Yunus Karatepe, Mert Sülük, Zeynep Tuğçe Kırımlı and Begüm Özbay

3K2T at MWE-2026 AdMIRe 2: CARIM– Category-Aware Reasoning for Idiomatic Multimodality
Kubilay Kağan Kömürcü and Tugce Temel

PMI MWE Scorer at PARSEME 2.0 Subtask 1: identifying multi-word expressions using pointwise mutual information and universal dependencies
Anna Bogdanova and Ileana Bucur

tiberiucarp at MWE-2026 AdMIRe 2: GLIMMER-Gloss-based Image Multiword Meaning Expression Ranker
Andrei Tiberiu Carp

IPN at MWE-2026 PARSEME 2.0 Subtask 1: MWE Identification via Related Languages and Harnessing Thinking Mode
Anna Hülsing, Noah-Manuel Michael, Daniel Mora Melanchthon and Andrea Horbach

Semantic Stars at MWE-2026 PARSEME 2.0 Subtask 2: Alternative Approaches for MWE Paraphrasing
Elif Bayraktar, Vedat Doğancan, Muhammed Abdullah Gümüş and Nusret Ali Kızılaslan

MorphoFiltered-Gemini at MWE-2026 PARSEME 2.0 Subtask 1: Tackling LLM Overgeneration via Universal POS-based Constraints
Irina Moise and Sergiu Nisioi

LST at MWE-2026 AdMIRe 2: Advancing Multimodal Idiomaticity Representation
Le Qiu, Yu-Yin Hsu and Emmanuele Chersoni

UniBO at MWE-2026 PARSEME 2.0 Subtask 2: A Cross-lingual Approach to Multiword Expression Paraphrasing
Debora Ciminari and Alberto Barrón-Cedeño

Saturday, March 28, 2026 (continued)

DCSN-NLP at MWE-2026 AdMIRe 2: Bridging Literal and Figurative Meaning Through Hierarchical Multimodal Reasoning
David Cotigă and Sergiu Nisioi

ITUNLP at MWE-2026 AdMIRe 2: A Zero-Shot LLM Pipeline for Multimodal Idiom Understanding and Ranking
Atakan Site, Oğuz Ali Arslan and Gülşen Eryiğit

Archaeology at WE-2026 PARSEME 2.0 Subtask 1 and 2: Parsing is for Encoders, Paraphrasing is for LLMs
Rares-Alexandru Roscan and Sergiu Nisioi

ITUNLP2 at MWE-2026 AdMIRe 2: Modular Zero-Shot Pipelines for Multimodal Idiom Grounding and Ranking
Özge Umut and Bora Şenceylan

12:00 - 12:20 *Community discussion*

12:20 - 12:30 *Concluding Remarks*