

Joint MWE-UD 2024

Hybrid Workshop at LREC-COLING 2024 (Turin)

SIGLEX - MWE section and UD

May 28, 2024



Best Paper Award

Under the **ACL Awards Policy**

The **Best Paper Award** of the Joint MWE-UD 2024 Workshop goes to...

Best Paper Award

Under the **ACL Awards Policy**

The **Best Paper Award** of the Joint MWE-UD 2024 Workshop goes to...

*Sign of the Times: Evaluating the use of Large Language Models for
Idiomatcity Detection*

Dylan Phelps, Thomas M. R. Pickard, Maggie Mi, Edward Gow-Smith, Aline
Villavicencio

ACL SIGLEX announcements

- Call for nominations for ACL SIGLEX board members
 - Vice President-Elect, Secretary-Elect, Representatives of the sections (MWE and SEMEVAL)
- Call for nominations for [ACL SIGLEX-MWE section](#) SC members (forthcoming)

Stay tuned! [ACL SIGLEX](#) news!

Phraseology and multiword expressions

Book series

- Language Science Press
 - Open access, collaborative proofreading
- 5 published volumes, 1 in press, 2 in preparation
- Open call for book proposals
 - Edited volumes or monographs
 - Details: <https://langsci-press.org/catalog/series/pmwe>



UniDive COST Action

- European research network (2022-2026)
 - **Universality**, diversity and **idiosyncrasy** in language technology
 - Reconcile language diversity with rapid progress in language technology
- 4 working groups - including MWE annotation, processing, etc.
- New members welcome
- <https://unidive.lisn.upsaclay.fr/> for webinars, news, calls, etc.
- Golden sponsor of the Joint MWE-UD 2024 Workshop (28 participants funded): THANK YOU!



Funded by
the European Union

UniDive COST Action 1st Training School

Chişinău, Moldova (relocated from Jerusalem, Israel)

8-12 July 2024

- Location: Technical University of Moldova
- Local Organizer: Victoria Bobicev
- Programme Committee:
 - Victoria Bobicev
 - Olesea Caftanatov
 - Atul Kr. Ojha
 - Agata Savary
- Programme already available



Funded by
the European Union



MWE-related announcements

- PARSEME VMWE corpus v1.3 annotated
- Extension of MWE annotation guidelines to other categories beyond verbal MWEs (within UniDive WG1)
- issues from Naples UniDive meeting will be available on GitLab
- Continuous corpus development infrastructure (work in progress)
 - UD updates → PARSEME corpora
 - Pilot annotation
- Open call for language leaders









Funded by
the European Union

UD-related developments

- May 15, UD release 2.14, 161 languages, 283 treebanks, 39 treebanks have grown, now 1.9 M sentences, 32 M syntactic words
- UD at LREC-Coling main program: at least 27 papers
- UD as a foundation for or addition to other annotation layers: Universal Anaphora, Parseme (MWE), Universal NER, DISRPT (Discourse relations), Parlamint (rich metadata for parlement proceedings), etc.
- Next (8th) UD workshop: TBD

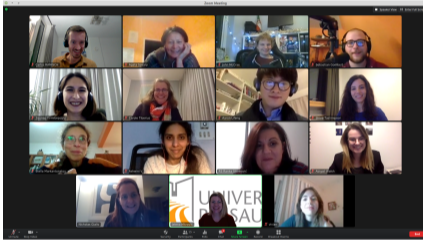
2 decades of MWE workshops

*

-  promoting research on MWEs in CL
-  collocated with major events
-  synergies with other communities
-  Shared Tasks, datasets, projects (PARSEME COST Action)
-  for the full history of previous events see the website
- 



2 decades of MWE workshops



Discussion: outlook

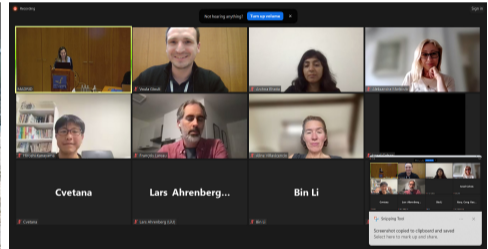
- Future directions of the MWE-related research and workshop
 - Extension of the MWE-UD joint aims and tasks
 - Joint synergies with other NLP communities
- LLMs and MWE research
 - What types of resources are needed?
 - What are the challenges?
- Shared Tasks and datasets
 - Shared Task on the identification of MWEs
 - New Shared Task?

Discussion: outlook

- Future releases of the UD corpora
 - Improved metadata for parallel corpora
 - Improved, standardized and finegrained, genre metadata
 - COstruction-level annotations
 - Morpheme segmentation and annotation
 - ...

*

Group photo



Thanks!

...see you next year!