

PARSEME 2.0 Multilingual Shared Task on Identification and Paraphrasing of Multiword Expressions

- **Event title:** culminating workshop of the PARSEME 2.0 shared task
- **Proposal:** submitted to [SemEval 2026](#)
- **Location:** TBA
- **Dates:** TBA
- **Data:** to be provided by the ongoing [PARSEME/UniDive annotation campaign on multiword expressions](#)
- **Shared task organizers:**
 - Manon Scholivet, Université Paris-Saclay, France
 - Takuya Nakamura, Université Paris-Saclay, France
 - [Agata Savary](#), Université Paris-Saclay, France
 - Eric Bilinski, Université Paris-Saclay, France
 - [Carlos Ramisch](#), Aix-Marseille Université, France



Subtask 1: MWE identification

This subtask is an extension of [PARSEME shared tasks](#) on automatic identification of verbal MWEs.

- Task: Given a raw text, automatically underline MWEs in it
- Data: PARSEME 2.0 annotated corpora (not necessarily all the texts from release 1.3)
- Language teams willing to participate with PARSEME data
 - Albanian
 - Egyptian (ca. 2700-2000 BC): MWEs from the UD-EUJA treebank.
 - Georgian
 - Greek (Modern)
 - Greek (Ancient)
 - Hebrew
 - Japanese
 - Lithuanian
 - Persian (Farsi)
 - Polish
 - Romanian
 - Serbian
 - Slovene
 - Swedish
 - Ukrainian
- Minimum annotation effort: 2000 annotated MWEs

Subtask 2: MWE paraphrasing

- Task: Given a sentence with a MWE, rephrase a sentence so that there is no MWEs but the meaning is the same
- Examples:
 - *She made up her mind to...* $\Rightarrow$ *She finally decided to...*
 - *He kicked the bucket* $\Rightarrow$ *He died* (But not *He passed away*)
- Data:
 - Selected sentences from PARSEME annotated corpora
 - The same sentences manually paraphrased
 - One to several hundred examples per language
- Language teams willing to participate
 - Albanian
 - French
 - Greek (Modern)
 - Japanese
 - Hebrew
 - Lithuanian
 - Persian (Farsi)
 - Polish
 - Brazilian Portuguese
 - Romanian
 - Serbian
 - Slovene
 - Swedish
 - Ukrainian

Timeline

- ~~31 March 2025 – SemEval proposal submission~~
- 19 May 2025 - SemEval notification
- 15 July 2025: Sample data ready
- 1 September 2025: Training data ready
- 1 December 2025: Evaluation data ready (internal deadline; not for public release)
- 10 January 2026: Evaluation start
- 31 January 2026: Evaluation end (latest date; task organizers may choose an earlier date)
- February 2026: Paper submission
- March 2026: Notification to authors
- April 2026: Camera ready
- Summer 2026: SemEval workshop (co-located with a major NLP conference)

From:

<https://unidive.lisn.upsaclay.fr/> - **Universality, diversity and idiosyncrasy
in language technology
CA21167 COST Action**

Permanent link:

<https://unidive.lisn.upsaclay.fr/doku.php?id=other-events:parseme-st>

Last update: **2025/04/15 15:17**

