

PARSEME 2.0 shared task

November 27, 2024

Have you ever heard about... PARSEME ?

What was the PARSEME shared task ?

- Tokens identification that belong to MultiWord Expressions (MWEs)
- ... And what is a MWE ?
 - Paris is a rainy city, it *rains cats and dogs* all the time!
 - The child is *crying wolf*.

Have you ever heard about... PARSEME ?

- 3 editions (2017, 2018 and 2020)
- Multilingual (26 languages)
- UD compatibility
- 9.3M tokens, 456K sentences, more than 127K annotated verbal MWEs

Goals of PARSEME 2.0

- Evaluate the systems' ability to **predict** MWEs
- Evaluate the **diversity** of the MWEs predicted
- Assess the ability of Large Language Models (LLMs) to **understand and/or predict** MWEs

Ongoing work (WG1 task 1.2)

- Annotation **guidelines** of all MWEs
 - (soon) **Extension** of current PARSEME corpora
- ~1-10K MWEs expected for each language

Diversity evaluation

A more diverse system is a system that predicts more *Categories / Species* (increasing the **Variety**)



A more diverse system is a system that predicts more *Categories* / *Species* (increasing the **Variety**)

Can predict the MultiWord Expressions :

- *take place*
- *take fire*

Can predict **more** MultiWord Expressions :

- *take place*
- *take fire*
- *cry wolf*

- Two other aspects of diversity: the **balance** and the **disparity**
- To compute the diversity, we need:
 - Categories (species)
 - Elements (individuals)

Diversity evaluation : example on MWEs

- She *cries wolf*
- He often *cries wolf*
- The child has been *crying wolf*
- You should never *cry wolf*

Here : 1 *Category*, 4 *Elements*

Diversity ... But why?

- Continuity of **unseen MWEs** evaluation (which prove critically hard for the task)
- Training on more diverse data covers more diverse phenomena
- To promote **rare** phenomena

Idea 1 : Idiomatic language **generation** by LLMs

Given a one paragraph context on the task, what will the LLM predict ?

- PROMPT: “It’s *raining cats* and ...”
- PROMPT: “The child has been *crying* ...”

Idea 2 : Idiomatic language **comprehension** by LLMs

Given a one paragraph context on the task, can the LLM predict if the tokens ***pulling the strings*** are literal or idiomatic ?

- PROMPT: “[...] who is going to be ***pulling the strings*** in the Europe of tomorrow”
- PROMPT: “[...] the artist was ***pulling the strings*** to move the marionette”

And can the LLM explain the meaning of an MWE?

- CodaBench
- SemEval 2026
- Creation of a diversity benchmark

- CodaBench
- SemEval 2026
- Creation of a diversity benchmark
- (Carbon Impact Evaluation too ?)

Tentative timeline

- January - August 2025 : PARSEME corpus annotation for MWEs of all syntactic types
- Spring 2025 : SemEval 2026 shared task proposal
- Autumn 2025 : Training data publication
- Winter 2026 : Systems evaluation
- Summer 2026 : SemEval Workshop

Thank you for your attention!