

We propose the following timeline: \* [language leaders: 27 February] Expression of interest from Language Leaders [task leaders: late February] Creating FLAT accounts [language leaders: mid-March] Reading guidelines, reading the Language Leader guide, filling in MWE examples, recruiting annotators, selecting corpora [shared task leaders: 31 March] Semeval shared task proposal [language teams: April-May] Annotation (including a double-annotated sample of inter-annotator agreement estimation) [Semeval: 1-2 May] Identification and selection of shared task [language leaders: June] Consistency checks and inter-annotator agreement estimation [shared task leaders: 30 July] Applying the shared task [language leaders: July-August] Consolidating and splitting the corpora [WG3 shared task leaders: 1 September] Training data for Semeval More details about the role of the Language Leader can be found in the PARSEME Language Leader guide.

**Call for expressions of interest in PARSEME/UniDive annotation campaign on multiword expressions**

The UniDive COST action (task 1.2) and the PARSEME community are happy to announce the upcoming **multilingual corpus annotation campaign** dedicated to multiword expressions (MWEs). Feel free to contact us for any questions you might have.

We call for expression of interest from current or future Language Leaders, who wish to propose a UniDive task 1.2 co-leaders Voula Giouli, Stella Markantonatou, Carlos Ramisch, Agata Savary, Sara Stymne

Three past PARSEME annotation campaigns were dedicated exclusively to verbal MWEs (VMWEs) and resulted in 4 editions of the **PARSEME corpus**, which jointly covers **26 languages**. Three **PARSEME shared tasks** on automatic identification of VMWEs have been organized on the basis of this corpus and set the state of the art in the task.

<https://unidive.lisn.upsaclay.fr/> - **Universality, diversity and idiosyncrasy**

The **UniDive annotation campaign** will cover MWEs of **all syntactic types**. It follows the **shared task** **universality**. Namely, the **annotation guidelines** are unified across all participating languages.

Permanent link: <https://unidive.lisn.upsaclay.fr/docs/propid-wg3-1-task1.2-call-for-language-leaders-rev-1-30954108>

When possible, still leaving room for truly language-specific phenomena. This approach is expected to promote meaningful cross-language comparisons. The resulting corpus will be

**PARSEME/UniDive shared task** on identifying and understanding MWEs, to be proposed for

**SemEval 2026**.



For the languages already present in the PARSEME corpus, the agenda is to:

- Re-annotate the existing corpus with MWEs other than verbal. Annotating only part of the existing corpus is an option. In this case we recommend a minimum of 3500 annotated MWEs (so that each selected text is exhaustively annotated for all syntactic types of MWEs). A lower number of annotations can do but the system results are expected not to be representative.
- Add some new texts annotated from scratch (to counterbalance language model contamination from previously published data)

For new languages, corpora will be annotated for all syntactic types at once.

A language team should consist of **at least 2 annotators** (including the Language Leader), for the sake of inter-annotator agreement estimation. It is possible to start annotating alone and recruit more annotators at a later stage (May at latest).

Centralized

<https://gitlab.com/parseme/corpora/-/wikis/PARSEME-Language-Leader-Guidedocumentation> and tools (including the online FLAT annotation platform) are available.