

WG1 - Task 1.2 - subtask meetings: MWE annotation guidelines (former 1.2) + UD-PARSEME unification (former 1.3)

Task co-leaders:

- Voula Giouli
- Stella Markantonatou
- Agata Savary
- Carlos Ramisch
- Sara Stymne

20 March 2025	3
19 March 2025	3
February 2025	3
Questions from LLs:	3
TIMELINE (all dates in 2025)	3
Call for expressions of interest in PARSEME/UniDive annotation campaign on multiword expressions	4
Budapest task 1.2 session - 30 January 2025	5
Co-leaders meeting 22 January 2025	5
Co-leaders meeting 15 January 2025	6
Co-leaders meeting 9 December 2024	7
Co-leaders meeting 18 November 2024	7
Co-leaders meeting 24 October 2024	8
Co-leaders meeting 17 October 2024	9
Co-leaders meeting 11 October 2024	9
Co-leaders meeting 2 October 2024	9
Co-leaders meeting 13 March 2024	10
Task sessions in person, Naples, 7 Feb	11
Data for WG1 leaders about the task sessions in Naples	11
Subtask meeting in person, Athens	11
Task meeting online, 18 December 2023, 10:00-11:00	12
Subtask meeting online, 23 November 2023, 11:00-12:00	15
Zoom:	15
Minutes:	15
Subtask meeting, 20 October 2023, 10:00-11:00	16
Subtask co-leaders' meeting, 22 and 26 September 2023	18
Useful links	19
Minutes	19

Useful links:

- PARSEME 2.0 annotation campaign
 - [Master](#) file
 - [Technical infrastructure of the PARSEME guidelines](#) - Agata-Carlos meeting 30.01.2024
 - [Implementing PARSEME guidelines 2.0](#) - TODO list (Agata, Takuya)
- [EoI](#) from Language Leaders
- Files from Budapest
 - Examples for [FuncMWEs](#)
 - Examples for [AMWEs](#)
 - [Slides](#) about task 1.2
- [SemEval 2026](#)
- Notes from Athens and Naples about **nominal** and **modifier** MWE guidelines
 - [dedicated document](#) from Athens
 - [draft guidelines](#)
 - [tests with examples](#)
 - [issues](#) from pilot annotation, issues in Gitlab
 - [template](#) for pilot annotation without FLAT
- **Functional** MWEs
 - Draft [guidelines](#)
 - Meeting [minutes](#)
 - [Examples of functional MWEs](#) in EN, FR, RO, EL, SL, PL
- Guidelines [issues](#) on Gitlab
- [White paper](#) on UD-PARSEME unification
- PARSEME VMWE annotation [guidelines](#) (version 1.3)
- Multilingual [examples](#) of MWEs from the subtask 1.2 participants
- UniDive [webinar](#) on UD, PARSEME and Grew-match
- Permanent Zoom link for this subtask:

- <https://cnrs.zoom.us/j/92794488497?pwd=Smtmdm4rTCs1S3hFdjZsUk1rZlU1dz09>
 - Meeting ID: 927 9448 8497
 - Passcode: kwFM5k
 - Existing annotation guidelines for non-verbal MWEs
 - [French](#) guidelines (Candito et al.)
 - Irish guidelines (Walsh et al.)
-

20 March 2025

Kick-off with the LLs - see [agenda/minutes](#)

19 March 2025

Voula, Sara, Takuya, Agata

Kick-off preparation

- Moving NMWE [DEVERB] and [PronID] just before [CRAN] ⇒ agreed
- ~~Tagset size too big ⇒ eliminate subtypes for NV and AV; no, keep them but place at the bottom of the list~~
- VPC ⇒ IVPC : changed in the guidelines but not in the corpora (still to do)

TODO:

- ~~Guidelines:~~
 - ~~Move NMWE [DEVERB] and [PronID] just before [CRAN]~~
- Check where the number of the already annotated MWEs is available
- FLAT:
 - ~~Keep only 3 active configurations:~~
 - ~~The 2 top ones + ALL MWE TYPES 2025~~
 - Make sure which files are your most recent versions
 - Ideally, .cupt files before the split
 - If .folia.xml files, then a change of an XML attribute must be done manually - the updated [FLAT user's guide](#) explains the process
 - Make sure that annotators (who are not LLs) cannot upload their files themselves

- LLS should clean up the annotators' files (to avoid confusion with the previous campaigns)
 - Create 2.0 folders in all annotations/user/ folders
 - Which FLAT accounts are active?
-

February 2025

Questions from LLS:

- What are the desired genres?
 - Lithuania - is a corpus of propaganda texts appropriate?
-

TIMELINE (all dates in 2025)

- [task leaders: ~~mid-February~~~~late January~~] call for expression of interest
 - [task leaders: **mid-February**] finalizing the online guidelines (see [here](#))
 - [task leaders: **mid-February**] updating the PARSEME wiki
 - [language leaders: **end-February**] expression of interest
 - [] updating FLAT configuration, updating FLAT user's guide
 - [task leaders: **late-February**] creating FLAT accounts, adding people to the team lists and to the parseme-corpora list
 - [ALL: early March] Annotation campaign **kick-off meeting**
 - [language leaders: **mid-March** ~~late-February~~] reading guidelines, reading the LL guide, filling in MWE examples, recruiting annotators, selecting corpora
 - [all: **March**] Pilot annotation
 - [WG3 shared task leaders: **31 March**] [SEMEVAL](#) shared task proposal
 - [language teams: **April-May**] Annotation (including a double-annotated sample)
 - [SEMEVAL: **19 May**] Notification about the selected shared task
 - [LLs: **June**] Consistency checks and inter-annotator agreement estimation
 - [WG3 shared task leaders: 15 July] sample data
 - [task leaders: **September**] consolidating and splitting the corpora
 - [WG3 shared task leaders: **1 September**] Training data for SEMEVAL (protecting the test)
 - [2026-] Synergies with Universal Dependencies (following the [white paper](#))
-

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). We call for expression of interest from current or future Language Leaders, who wish to propose a language team. If you are interested, please, fill in the [EoI form](#), best before **27 February 2025**.

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.

The current annotation campaign will cover MWEs of **all syntactic types**. It follows the spirit of **universality**. Namely, the annotation guidelines are unified across all participating languages, whenever possible, still leaving room for truly language-specific phenomena. This approach is expected to promote meaningful cross-language comparisons. The resulting corpus will be used in a **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 documentation and tools (including the online FLAT annotation platform) will be available to participating language teams.

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
- [all: **March**] Pilot annotation
- [shared task leaders: **31 March**] SEMEVAL shared task proposal
- [language teams: **April-May**] Annotation (including a double-annotated sample for inter-annotator agreement estimation)
- [SEMEVAL: **19 May**] Notification about the selected shared task
- [language leaders: **June**] Consistency checks and inter-annotator agreement estimation
- [shared task leaders: **15 July**] Sample data ready
- [task 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](#).

Feel free to contact us for any questions you might have.

UniDive task 1.2 co-leaders

Voula Giouli, Stella Markantonatou, Carlos Ramisch, Agata Savary, Sara Stymne

Budapest task 1.2 session - 30 January 2025

Agenda:

- Outlining the plans for the UniDive MWE annotation campaign scheduled for spring 2025, including its goals, timeline, and recommendations for getting started. [link TBA]
- Presenting the current state of the new MWE guidelines, which now encompass verbal, nominal, modifier, and functional MWEs. [link TBA]
- Demonstrating the use of [GitLab issues](#) to report challenges or propose amendments encountered during the annotation process.
- Presenting the tests for modifier and functional MWEs and enriching them with multilingual examples through onsite brainstorming with participants. [link TBA]

Co-leaders meeting 22 January 2025

Stella, Nathan, Voula, Verginica, Agata

Agenda:

- Tests with examples: [FuncMWEs](#), [ModifMWEs](#)
- Presenting the new guidelines
- Completing the organisation of the task 1.2 session in Budapest
- Should we have a pilot?

Co-leaders meeting 15 January 2025

Verginica, Nathan, Kaja, Voula, Agata

Agenda:

- Organisation of the task 1.2 session in Budapest

- Description of the session (5-10 lines with links to specific document if needed) for Bruno - before 17 Jan

Minutes:

- On day 3 - how to occupy the 45 min. session
 - Explaining gitlab issues
 - Add examples to the tests in google docs for functional MWEs and then modifier MWEs (together or in groups)
 - Use Kaja's examples for 5 languages
 - Can't be in the database because it should be open only for the LLs
- TODO:
 - [VERGINICA, VOULA: Monday 20th Jan] Make copy of the modifier examples files. Create a similar document for functional MWE examples. Make sure to have at least one English example per test.
 - [NATHAN: 27th Jan] Nathan's students: next week. Assignment 21th Jan, results 27th. Review the English examples.
 - Instructions for : look into the tests, provide positive and negative examples
 - [KAJA: 17 Jan] Outline for WG1 - description of task 1.2 subsession
 - Presenting the state of the guidelines
 - Showing how to use gitlab issues
 - Timeline of the annotation campaign
 -
 - Collecting EoI from language teams to annotate their corpora
 - Collecting examples
 - [] Timeline (EoI)
 - [JANUARY 15th] Call for EoI, with the timeline
 - [JANUARY] Expression of interest
 - [FEBRUARY-MAY] Annotation (including a double-annotated sample)
 - [JUNE] Consistency checks and IAA estimation
 - [SEPTEMBER] training data for SEMEVAL; protecting the test
- On day 2 there will be interest in the new guidelines, expect questions.
- TODO next week: Preparing and sending out the call for EoI.

Co-leaders meeting 9 December 2024

Sara, Agata

Agenda:

- Finish “closing” the issues
 - 115 - solved in #119
 - 116 - we proposed a “drastic” solution (eliminating the PronID category); we also made the [DIST] test more precise
 - 117 - solved
 - Issue 118/127 is unsolved and hard
 - All other unsolved issues are about verbal MWEs (to solve later)
- Discuss the timeline for the shared task
- Discuss what we would like to do in Budapest
 - Agata: have a pilot annotation of functional MWEs
 - Stella: a list of fixed/functional MWEs per language is needed
 - Kaja: prepare the guidelines for all MWE types; have a pilot with these guidelines in Budapest; explain the Gitlab issues to all
 - Sara: likes Kaja’s proposal
 - Agata: this was submitted to WG1
- TODO:
 - [All] Look into issues 118 and 127 (DIST test - how to define it for modifiers)
 - [All] Apply the tests to [functional MWEs](#) in other languages
 - [Agata] Find the next slot

Co-leaders meeting 18 November 2024

Voula, Verginica, Kaja, Agata, Sara

- We defined the status of issues 109-111 to 113-114 as almost closed. We should continue until 119
- Kaja looked into examples of Slovene [functional MWEs](#) from UD, to check if the current guidelines are enough - see [these issues](#). It seems like the current tests cover most functional MWEs

- If possible apply the tests to [functional MWEs](#) in other languages.
 - Next meeting **9 December 3 p.m. CEST**
-

Co-leaders meeting 24 October 2024

Agata, Nathan, Kaja, Stella

- Issue #116 - ask Abigail if DIST solves the issue
- New issue [#127](#) - ask Voula and Verginica if they agree to delete the hypoCat
- Long discussion on issue #118. Distribution does not seem easy to define in linguistics. Maybe we can check if this poses a problem in annotation?
- Functional MWEs - we are not sure the current tests are enough, especially for those that contain only functional components
- One issue still to discuss #119 (issues #120-126 do not concern the new guidelines)
- Kaja constructed a resource to test [functional MWEs](#) in several languages. Selection criterion: DEPREL=case|cc|mark|det|discourse and there is a non-trivial tree below (not a single leaf) -

TODO

- ~~[AGATA] Issues~~
 - ~~Contacted authors of issues #109 to #112.~~
 - ~~Issue #113 is solved, it has to be applied to the guidelines.~~
 - ~~Issue #115 seems in good progress. We do not know who the original author is.~~
 - ~~Issues #116, #117, #119 - Abigail, Carlos, Manon and Louis contacted~~
 - [ALL] Use Kaja's examples above to test the functional MWE guidelines
 - [Verginica, Voula and Sara] Check issues #116, #117, #118, #127
 - ~~[AGATA] Doodle for the next meeting~~
-

Co-leaders meeting 17 October 2024

Agata, Verginica, Stella, Voula, Nathan

- Issue [#114](#) - we need Verginica's answer
- We commented on issue [#115](#)
- Next meeting 24 Oct 17:30 CEST

Co-leaders meeting 11 October 2024

Voula, Sara, Agata

- Starting to edit the new guidelines online
 - Try the same website for all cats
 - Categorise - with subsection per POS
 - Cross-lingual tests - separate section for each category
 - Help
- TODO:
 - Find a clear yet precise definition of a semantic head
 - [VOULA] Write to the people
 - Ask Jaka if he is fine with the answer to issue [#110](#).
 - Ask Abigail to validate the proposal in issue [#111](#) . If validated, apply the modifications to the VMWE guidelines on meaning-preserving variants.
 - Ask Kadri to comment on issue [#112](#)
 - Ask Verginica to answer issue [#113](#)

Co-leaders meeting 2 October 2024

Stella, Kaja, Verginica, Sara, Agata

Agenda:

- Issues submitted during the summer
- Toward integration of the guidelines and their publication

Minutes:

- Solving issues
 - 109-111 - Verginica; 109: The semantic head is the component that determines the selectional restrictions of the constructions.
 - 112-114 - Kaja
 - 115-116 - Stella
 - 118-121 - Sara
 - 117, 122 (help needed), 124 (easy to solve), 125 (help needed by an English LL), 126 (more complex to implement, concerns VMWEs, lower priority?) - Agata

- **TODO:**
 - [9 OCTOBER] We all comment on issues 109-126
 - [11 OCTOBER] The person “in charge” of an issue tries to consolidate the discussion and propose a solution.
 - Next meeting **Friday 11/10 1h30 p.m. CEST**
-

Co-leaders meeting 13 March 2024

Agenda - data about the task for WG1 leaders:

- Objectives
 - Model and annotate multiword expressions in a way which is unified across many languages. Make UD and PARSEME initiatives converge in this respect.
- Current activities
 - Extending the PARSEME annotation guidelines to non-verbal MWEs
- Milestones
 - [summer 2024]: Consolidated version of the nominals&modifier MWE guidelines
 - [mid-May] Transforming feedback from Naples into Gitlab issues, adding solution proposals, inviting Language Leaders to react
 - [mid-June] Discussion ⇒ [mid-July]
 - [mid-July] Consolidated version of the nominals&modifier guidelines ⇒ [mid-September]
 - [end-July]: Draft of the guidelines for functional MWEs
 - [end October]: Pilot annotations and feedback from annotators
 - [December 2024]: Consolidated version of the all-MWE guidelines (in a web interface?)
 - [January-March 2025]: 3rd UniDive General Meeting; annotation campaign (with an extension to end April)
 - [2025 or 2026]: shared task of automatic identification of MWEs of all syntactic types (SemEval, MWE workshop - maybe with RANLP in September, UD workshop)
- Next meetings - TBA
- Contacts: Agata, Stella, Voula, Verginica, Kaja

- **TODO:**
 - ~~[KAJA] Share the data with Bruno~~
 - ~~[mid-April] Transform [issues](#) on nominals into [Gitlab issues](#), **propose our own solutions**~~
 - ~~Agata & Stella: issues 1-5~~
 - ~~Issue 1 - processed on 22/03~~
 - ~~Agata - issues 2 & 3~~
 - ~~Stella - issues 4 & 5~~
 - ~~Voula & Verginica: issues 6 to 10~~
 - ~~[AGATA] Understand Gitlab notifications~~
 - ~~Email to the parseme corpora about the new open issue~~
 - ~~Wait for reactions (votes or other suggestions)~~

Task sessions in person, Naples, 7 Feb

Agenda:

- 2 parallel sessions for pilot annotation + reporting in the plenary (to confirm)
- Task leader (to replace Agata)

Data for WG1 leaders about the task sessions in Naples

1. **Pilot PARSEME annotation** on non-verbal MWEs - see [program](#)
2. Presentation of the **Parseme pipeline** to **Language Leaders** - see the [program](#)

Subtask meeting in person, Athens

See a [dedicated document](#), and [draft guidelines](#), [tests with examples](#), [issues](#) from pilot annotation, [template](#) for pilot annotation without FLAT

Agenda:

- Drafting guidelines for non-verbal MWEs
- Call for participants of the pilot annotation in Naples; they need to prepare their corpora

Task meeting online, 18 December 2023, 10:00-11:00

Zoom:

<https://cnrs.zoom.us/j/96624166890?pwd=N0wrS3RhLzVWMC9zUmJSNEgwYWVVFUT09>

Meeting ID: 966 2416 6890

Passcode: Km7qxJ

The meeting was be **recorded** If you oppose, please, put a note below next to your name.

Here is the [recording](#) of the meeting.

Participants (please, highlight or add your name) :

- Voula Giouli
- Neslihan Ondeer-Ozdemir
- Stella Markantonatou
- Sara Stymne
- Carlos Ramisch
- Agata Savary
- Verginica Mititelu
- Veronika Lipp
- László Simon
- Lirim Shala
- Vladimir Cvetkoski
- Manjola Z. Lumani
- Besim Kabashi
- Hiwa Asadpour
- Purificação Silvano
- Irina Lobzhanidze
- Ivana Brac
- Yuval Pinter
- Cvetana Krstev
- Carole Tiberius
- Victoria Bobicev
- Giedre Valunaite Oleskeviciene
- Omer Goldman
- Bruno Guillaume
- Enriketa Sogutlu
- Gosse Bouma
- Polona Gantar

- Alexandra Vella
- Noémi Vadász
- Vahide Tajalli
- Cvetana Krstev
- Aleksandra Markovic
- Kristina Kocijan
- Roberto A. Díaz Hernández
- Amalia Todirascu
- Jolanta Kovalevskaite
- Joakim Nivre
- Mariia Shvedova
- Chaya Liebeskind
- Kadri Muischnek
- Frantisek Forgac

Agenda:

- Existing efforts towards annotation guidelines for non-verbal MWEs by:
 - Carlos Ramisch, Agata Savary - PARSEME-FR guidelines for annotating MWEs of all syntactic types - [slides](#), [paper](#), [online guidelines](#)
- Discussion
 - Neslihan: Necessity to translate the PARSEME-FR guidelines into English
 - Gosse - other criteria
 - Archaic inflection *zum* (a special case of cranberry words?)
 - MWE from other languages: *per se*
 - Stella:
 - Expressions coming from the diachrony of the language (treated foreign words)
 - ID and PRED - often it is not necessary to apply them both
 - Carlos&Agata: The tests are sufficient, so they are not ordered LEX criterium
 - Agata: what in UD says that there is irregular syntax:
 - Gosse: *fixed* relation (cannot be modified, split etc.)

- Stella: Agreement violation
 - Carlos: Maybe we need to add more fine-grained tests that implement this “irregular syntax” test
- Next steps towards annotating non-verbal MWEs
- Answers about Naples, needed by the WG1 leaders - see [here](#)
 -
- Feedback:
 - Stella: pilot annotation might show how the guidelines apply
 - Voula: Do we want a continuation of what we have in PARSEME? What would be useful for other WGs?
 - Carole: The guidelines look applicable to WG2 but it’s too early to say yes.
 - In Dutch there are many compounds - special tests?
 - NEs also reliable
 - Simon: In 2.1 task, parallel corpus of over 2000 sentences, NE identified (in English), which would transfer to other languages, and conceptual parts of NEs (e.g. “assembly” in “general” assembly).
 - Agata: Applicability of syntactic irregularity test.
 - Neslihan: Wants to contribute to corpus creation (selecting the material).
 - Jolanta: Make the Naples session hybrid
 - Jolanta: Do we want to also annotate NEs?
 - Verginica: Create a group of people to work on functional MWEs (to start with conjunctions)
 - Bruno (post-meeting): Do we want to favor annotating upon already UD annotated data? It may be useful to have both annotation layers on the same data.
- People interested in functional MWEs
 - Verginica
 - Carlos
 - Stella
 - Chaya
 - Voula
 - Amalia

- Who would like to annotate NEs next to non-verbal MWEs and if so - also include one-word NEs?
 - Bruno: As far as I know, for PARSEME-FR, NEs were not planned but the interaction MWE/NE makes it difficult to avoid them.

TODO:

- [ALL] Express your interest in participating in the subtask on functional MWE (add your name in the list above)
- [ALL] Express your will to annotate NEs next to non-verbal MWEs (add your name in the list above)
- [CO-LEADERS] Describe the requirements for the Naples session.
- [CO-LEADERS] Make the session in Naples hybrid

Subtask meeting online, 23 November 2023, 11:00-12:00

Zoom:

<https://cnrs.zoom.us/j/96624166890?pwd=N0wrS3RhLzVWMC9zUmJSNEgwYWVFUT09>

Meeting ID: 966 2416 6890

Passcode: Km7qxJ

Minutes:

- Part of the meeting was [recorded](#).
- Existing efforts towards annotation guidelines for non-verbal MWEs by:
 - Abigail Walsh
 - Compositionality annotation in nominal compound of type noun+modifier in running text - [slides](#)
 - Aline Villavicencio - [slides](#), Google doc with first steps towards PARSEME [guidelines for nominal MWEs](#)
 - Two tasks
 - Extending PARSEME guidelines to nominals (started a Google doc)
 - Compositionality annotation in 3 languages in nominal compounds
 - Carlos Ramisch and Agata Savary - [slides](#) - delayed to the next sessions

- Discussion
 - Could we have a new parallel task in WG1 about compositionality score annotation (like Aline and Abigail do it)?
 - [TOM] How to annotate nominal MWEs embedded in VMWEs, e.g. [*open a [can of worms]*]
- Next meeting: **Monday 18 December 10:00-11:00 CET**
- Co-leaders discussion (Voula, Agata):
 - Answers to Bruno
 - How about people wanting to annotate verbal MWEs (Carole-Gosse-NL; Victoria Fendel - ancient Greek; Voula will help)?
 - Thessaloniki could host a UniDive event

TODO:

- ~~[AGATA] Trigger the addition of Dutch and ancient Greek to FLAT; configuration (?)~~
- ~~[AGATA, CARLOS] Same for the VMWE guidelines (examples, accounts)~~

Subtask meeting, 20 October 2023, 10:00-11:00

Participants (please add your name):

- Voula Giouli
- Stella Markantonatou
- Agata Savary
- Verginica Mititelu
- Veronika Lipp
- Lirim Shala
- Vladimir Cvetkoski
- Manjola Z. Lumani
- Besim Kabashi
- Hiwa Asadpour
- Purificação Silvano
- Sara Stymne
- Irina Lobzhanidze
- Ivana Brac

- Yuval Pinter
- Cvetana Krstev
- Carole Tiberius
- Victoria Bobicev
- Giedre Valunaite Oleskeviciene
- Omer Goldman
- Bruno Guillaume
- Enriketa Sogutlu
- Gosse Bouma
- Polona Gantar
- Alexandra Vella
- Noémi Vadász
- Vahide Tajalli
- Cvetana Krstev
- Aleksandra Markovic
- Kristina Kocijan

Agenda:

- Personal introductions
- Merging sub-tasks 1.2 and 1.3 in WG1, at least at the beginning (see the [white paper](#))
- Planned timeline for this subtask (see below)
- Revisiting PARSEME VMWE annotation [guidelines](#) (version 1.3) with multilingual [examples](#)
 - Several participants wish to add : Dutch (Carole, Gosse), Georgian (Irina?), Albanian (Lirim), Macedonian (Vladimir Cvetkoski), Armenian (Anna Danielyan?)
 - Cvetana:
 - LVC.full and LVC.cause are hard to distinguish (enhance the examples in the guidelines)

- Verbal nouns (nominalisations) - nominals
- Task co-leaders
 - Who will report to the WG1 on 25 Oct? Voula
 - What activities in Naples on 7/02 from our subtask? Pilot annotation for the new guidelines

TODO:

- [NEWCOMERS] Watch the UniDive [webinar](#) on PARSEME
- [ALL] Prepare examples to discuss in the [spreadsheet](#) (nominals or verbals)
 - [CVETANA] Problems with nominals in Serbian
 - [~~VAIHIDE~~] Problems with Farsi LVCs
- ~~● On Nov 23rd, bring people who worked on guidelines for nominals, for Thursday 23 November 11:00-12:00 - 10 min. each~~
 - ~~○ [AGATA] Contact Abigail Walsh - agreed~~
 - ~~○ [AGATA] Contact Marie Candito - Marie is NOT available, Carlos and Agata will explain these guidelines~~
 - ~~○ [STELLA] Aline Villavicencio~~
- ~~● [VOULA] Announce the dates to the WG1 list~~
- ~~● [VOULA] Answer Bruno about the 25th October~~

Subtask co-leaders' meeting, 22 and 26 September 2023

Subtask leaders: Agata, Stella, Voula, Carlos, Sara

Useful links

- [WG1 coordination](#) document
- [EOIs](#) in WG1 tasks, [original survey](#)
- [Zoom link](#)
- [Spreadsheet](#) with multilingual VMWEs examples

Minutes

- People involved in guidelines writing
- Tentative schedule (1 hour session, once or twice a month)
 - Preparatory sessions (state of the art)

- Session 1 (October): PARSEME guidelines - Q&A and hands-on
 - (pre-requisite) Watch PARSEME webinar
 - Bring examples in your languages and illustrate how you go through the decision diagram with them
 - LVC.full vs. LVC.cause
 - Session 2-3 (November-December): Existing guidelines and other formal descriptions of **non-verbal content MWEs** (nominal & pronouns, adjectival, adverbial, function words)
 - Abigail, Aline, Marie?, Carlos, Verginica
 - Working sessions (incl. pilot)
 - Session 4 (January): Draft guidelines
 - Week **22-26 January 2024** in Athens
 - Maybe a talk at the same occasion at the NLP seminar
 - Session 5 (7th of February in Naples): Pilot annotation with many languages. Feedback on the guidelines.
- Fundamental question
 - Syntactic categories (e.g. verbal vs. nominal) as a structuring notion in extended MWE guidelines
- Misc:
 - [STELLA] A survey about what people mean by nominal?

TODO:

- ~~— [AGATA] Contact Sara for co-leadership~~
- ~~[TODAY: VOULA] Email to WG1 about Task 1.2 "MWE guidelines extension", deadline **2 Oct 4 Oct**, forward to Carole, Vahide, Sara, Johanna~~
 - Plan for online sessions until January
 - ~~— Doodles (fill in at least the 1st)~~
 - ~~— 16-20 Oct~~
 - ~~— 20-24 Nov~~
 - ~~— 18-22 Dec~~
 - ~~— 22-26 Jan~~
- ~~— [TASK LEADERS: CARLOS 3rd Oct] Prepare a spreadsheet 1-2 examples (with glosses), send an email:~~
 - ~~— 1 tab per language, sentences with EN glosses aligned below~~

- ~~— MWEs or interesting parts are highlighted in bold~~
- ~~— we add examples from EL, PL, PT, FR, EN~~
- [13 OCT] People add examples for their languages
- ~~— During the 1st session:~~
 - ~~— participants explain how the guidelines apply to examples in spreadsheet~~
- ~~— Short name of the task 1.2: *MWE guidelines*~~
- ~~— [NEXT MEETING] Find slots for sessions 1 and 2-3~~
- ~~— Find a 9-days slot in January for Athens (hard work on the draft guidelines + 2 nice weekends?) - **22-26 Jan**~~

WG1 - subtask: MWE annotation guidelines

12 September 2023

Subtask leaders: Stella, Voula, Carlos

Paris-Saclay meeting:

- New categories to address
 - **Nominal** expressions
 - **Inherently adpositional verbs** (enhancement)
 - Others (adverbials, functional...)

Things to have in mind and discuss with UD-MWE subtask and morphosyntactic annotation subtask:

- Lemmatisation
- Tokenisation
- Compounding
- Rich morphology

How

- Shared online **documents**
- **Webinars** and invited talks

Items to be presented at WG1 meeting on Sep 13 2023

- Present PARSEME verbal MWE guidelines so that everyone is familiar with them
 - The guiding principles: tests, decision trees...
 - Scope (of verbal MWEs)
 - Subcategories (VID vs. LVC, etc...)

- Methodology to develop and enhance the current verbal MWE guidelines
 - Pilot + writing + pilot + writing...
- THIS IS SOMETHING WE PROPOSE KEEPING
- Define theoretical foundations of the new guidelines (interaction with WG2 subtask 1)
 - Definitions of words, nominals (vs. other categories like adjectives, etc.), syntactic background, etc...
- Take inspiration from people who already described **non-verbal** MWE annotation guidelines - maybe focus on nominal (and maybe adjectival)
 - On invitation from relevant initiatives we're aware of
 - Call for people interested in sharing
 - Monthly meetings with presentations from authors
- Group of people interested in making concrete contributions - from WG1 form
 - Make it into a survey? Now or later?
 - What could people contribute? How much energy?
- How to organise ourselves?
 - Once a month: presentations from people who have non-verbal MWE annotation experiences to share
 - Subtask leaders meet every 2 weeks to organise the work
 - Later: maybe meet more often to make concrete progress on the annotation guidelines and pilots
 - Sync with other subtasks, other WGs