

UniDive MC meeting 4

24 April 2024, 10:00-12:00 CEST, online

Agenda	1
Minutes	2
TODO	10
TOP-UP BUDGET SUMMARY	11

Agenda

1. Welcome to participants, verification of the quorum and adoption of agenda
2. Information to the MC
 - a. Recap of the minutes of the previous MC meeting
 - b. Action Membership
 - c. Grant Awarding by the Action
 - d. Budget status: summary from the Grant Holder
3. Follow up and discussion
 - a. Events in Grant Period 2
 - b. Implementation of the COST [Excellence and Inclusiveness Policy](#)
 - c. Progress of each working group
 - d. Progress on [MoU](#) Objectives, deliverables, and goals for the current GP
4. Planned events in Grant Period 3
5. AOB
6. Next MC meeting

Minutes

Attendance:

Absent: Beata Wójtowicz

Minuting: Agata Savary, Alina Wróblewska

7. [AGATA SAVARY, ALINA WRÓBLEWSKA] Welcome to participants, verification of the quorum and adoption of agenda
 - a. Fill in the [attendance list](#) (put 1 next to your name)
 - b. The quorum is 2/3 of COST Full and Cooperating Member countries (one representative per country is enough)
 - c. Confirming the quorum: 33/37 countries represented (89.2%)
 - d. MC vote rule: if two representatives of the same country cast different votes (approved vs. objected), the vote for this country is considered void
8. [AGATA SAVARY] Recap of the [minutes](#) of the MC meeting on 2 November 2023, e-votes and matters arising since the last meeting
 - a. 1 vote: including Beata Wójtowicz, the Grant Holder Manager, in the Core Group
 - b. We had 16 (Extended) Core Group meetings so far. All past and upcoming [CG meetings](#), with discussions on the arising issues, are documented at the Action's website.
 - c. The MC delegated the following powers to the CG:
 - i. Approving/refusing WG applications - see below
 - ii. Selecting event participants who will be reimbursed - see below
 - iii. Adjusting the location and dates of activities (in close consultation with the Local Organizer) - see below
9. [DANIEL ZEMAN] Action Membership:
 - a. New COST Members represented in the MC
 - i. 36 countries were represented in the MC in March 2023 (out of all 43 COST, Cooperating and Partner Members - see the [list](#))
 - ii. 23 of those (64%) are Inclusiveness Target Countries
 - iii. Now 37 countries (Denmark has joined); 23 countries (62%) are ITC; + special status of Israel ("cooperating member"). Not in the action: only 2 ITC (Cyprus, Montenegro); 3 non-ITC (Austria, Iceland, Luxembourg); 1 "partner member" (South Africa).
 - b. Action Participation: WG membership and applications
 - i. 330 individual participants in working groups (as of 2024-04-11), many of them are members of more than one WG
 - ii. 58% female, 42% young, 49% based in ITC
 - iii. WG1: 200 members
 - iv. WG2: 158 members
 - v. WG3: 206 members
 - vi. WG4: 128 members
 - vii. As reported during the previous MC meeting, in the first year there were some applicants whose background was so unrelated that they were put

on hold (and did not get back to us) or even rejected. No new such cases since then; all new applicants were accepted. We still have several new applications each month.

- c. No questions or comments from the MC
10. [NURIT MELNIK] Grant Awarding by the Action
 - a. Year 2, First round
 - i. Total granted: 14,500 EURO
 - ii. Accept 9 (14 applications)
 - iii. Total grants: 15,700 EUR (with adjustments to requests)
 - iv. Gender: 4 female (of 8 applications), 5 males (of 6 applications)
 - v. Seniority: 2 students, 2 recent PHDs, 5 senior
 - vi. ITC countries: 3 (of 6 applications)
 - b. Year 2, Second round
 - i. Remaining budget: 14,265 EURO
 - ii. Call publication: 17th March 2024
 - iii. Application deadline: **7th May, 2024**
 - iv. Notification: 14th May, 2024
 - c. Hosting proposal
 - i. **United Kingdom** at the University of Sheffield
 - ii. **Italy** at the CIRCSE Research Centre Università Cattolica del Sacro Cuore
 - iii. **Czechia** at the ÚFAL (Institute of Formal and Applied Linguistics)
 - iv. **Bulgaria** at the Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences
 - d. ITC conference grants, Year 2: deadline: **15 June 2024**
 - e. No questions or comments from the MC
11. [ALINA WRÓBLEWSKA on behalf of BEATA WÓJTOWICZ] Budget status: summary from the Grant Holder
 - a. 2nd General Meeting in Naples
 - i. budget: 78800
 - ii. real expenses: 69644,53
 - iii. savings: **9155,47 EUR** => we propose to assign it to the **MWE-UD workshop**
 - b. STSMs (see also above):
 - i. budget: 28765
 - ii. real expenses so far: 2750
 - iii. grants allocated: 11750
 - iv. budget still to allocate: 14265
 - c. Top-up possibility from COST for year 2: **25,000 EUR** (Giuseppe Lugano's email on 8/04/2024)
 - i. Proposed allocation:
 - MWE-UD workshop in Turin (Italy) - 6,250 EUR (10 participants initially, +12 with the savings from Naples, +8 with the top-up budget, 30 in total)

- 1st UniDive Training School in Chişinău (Moldova) - 12,000 EUR (10 extra trainees, 55 in total)
 - ITC or Dissemination Conference Grants - 3,000 EUR (2-3 more grants)
 - See justification [here](#)
 - ii. Separate MC vote on the top-up budget online in the 2 upcoming weeks
 - d. No questions or comments from the MC
 - e. **Vote:** assigning the 9155 EUR savings from Naples to the MWE-UD workshop
 - i. 0 against, 0 abstaining, all in favor
12. Grant Period 2 (Nov 2023-Oct 2024) events:
- a. [AGATA SAVARY] General Meeting in **Naples 7-9 February 2024**
 - i. Day 1: a WG1 meeting with plenary presentations and parallel hands-on sessions of 4 ongoing tasks.
 - ii. Day 2: the 2nd UniDive Workshop with 2 invited talks and 40 posters, an Extended Core group meeting and social events
 - iii. Day 3: parallel WG sessions and plenary reporting
 - iv. All material from the event is available [online](#).
 - v. Number of reimbursed participants: 69 based on leadership roles, poster submissions, balance criteria (geography, age, gender)
 - vi. Lessons learned: leave more room for self-funded participants
 - vii. Comments and questions from the MC
 - b. [VOULA GIOULI] [MWE-UD workshop](#), 25 May 2024
 - i. Co-organized by UniDive: the Organizing/Program Committee chairs belong UniDive
 - ii. Co-funded by UniDive (funding for an invited speaker, if needed, and for about 10+12+8 participants)
 - iii. Most e-cost invitations sent out last Monday
 - iv. 52 submissions (44 archival + 8 non-archival), 3 reviews per paper, 26 selected archival papers, acceptance rate 59%, 2 invited talks, program
 - v. 29 participants will receive funding
 - vi. Modalities: hybrid event
 - vii. About eligibility: It can be a voluntary affiliation, e.g. registered as a volunteer with a charity - see annotated rules
 - viii. No comments or questions from the MC
 - c. [VICTORIA BOBICEV, AGATA SAVARY] Training School:
 - i. History:
 - Call for Local Organizers on 9 June 2023
 - Proposals received
 - a. 3 for the 2nd general meeting in February 2024: Jerusalem, Naples, Malta
 - b. 3 for the training school in summer 2024: Budapest, Jerusalem, Tbilisi
 - Initial selection (by the Core Group): general meeting in Jerusalem, training school in Budapest

- [Measures from COST](#) on 4 July 2023 about suspending cooperation with the Hungarian legal entities implied by [the Council Decision 2022/2056](#)
 - Advice from Giuseppe Lugano, our COST Officer, to consider a different host for the 2024 training school and keep Budapest as an option for the following grant periods
 - Change of plans by the Core Group: 2nd general meeting in [Naples](#), training school in [Jerusalem](#)
 - Tragical events in Israel and Gaza since 7 October 2023
 - The Local Organiser (Chaya Liebeskind) and the other Representative of Israel in UniDive (Nurit Melnik) propose to relocate the event to another venue.
 - Proposal from Chişinău, Moldova, Victoria Bobicev as the Local Organizer
 - MC vote validating the new venue and daily allowance on 27/02/2024
- ii. Current developments
- Courses:
 - a. Dependency syntax, Surface-Syntactic UD and UD - by Sylvain Kahane and Francis Tyers (to confirm)
 - b. Annotation of multiword expressions for newcomers - by Verginica Mititelu and Voula Giouli
 - c. Corpus annotation infrastructure- by Bruno Guillaume, Daniel Zeman and Agata Savary
 - d. Brainstorming hackaton based on open questions submitted by the trainees
 - e. Posters for selected trainee projects
 - Potential trainees should submit a project on topics related to the TS courses, deadline: **29 April**
 - Selection based on the projects, balance criteria, resourcedness of the language concerned
 - Invitations to be sent out in May
 - Call for reviewers of the student projects: 32 volunteers so far, more are needed.
 - 2nd calls for projects and reviewers should be sent out.
- iii. Trainee projects: 4 submissions so far
- iv. Comments and questions from the MC
- [VICTORIA BOBICEV] An Ukrainian student cannot come and asks for online participation
 - [ANNA DANIELYAN] Is it possible to attend the training school online without submitting a proposal?
 - a. Trainers find it hard to have a fully hybrid format but the courses will be recorded

- [VICTORIA BOBICEV] Student for Philippines, upcoming affiliation in Europe. Can he participate?
 - a. The affiliation, not the nationality, determines eligibility for funding. With an affiliation in Europe, anyone is eligible.
- [VICTORIA BOBICEV] Discussion about safety in Moldova: internal situation: political parties are very active, inhabitants feel safe, even more now after the vote in the US Senat
- Should we issue a statement about safety in Moldova?
 - a. Pros: Reassuring people
 - b. Cons: A statement itself will probably make people question the safety (Victoria, Dan)
 - c. Ideas:
 - i. [DAVE SAYERS]: link to government recommendations like [in UK](#)
 - ii. [MIRYAM DE LHONEUX] Have this information somewhere, maybe not at the front page, in case people ask

13. [AGATA SAVARY: 10 min.] Grant Period 3 events

- a. 3rd General Meeting + WG2 meeting:
 - i. Proposals: Budapest, Malta, Yerevan
 - ii. Core Group pre-selection: Budapest
 - iii. Proposed dates: **28-30 January 2025** (Tuesday-Thursday)
 - iv. Daily allowance
 - [LOCAL ORGANIZER] *If we consider that most people will choose the Benczúr Hotel, which will also be the main venue for the conference, the cost there is around 56-62 EUR per room including breakfast. Dining options are also affordable, as is transportation.*
 - [AGATA SAVARY] We should also consider that some participants might fail to book in Benczur.
 - Proposed rate: **140 EUR/day**
 - v. No comments or questions from the MC
 - vi. **MC vote:** 0 against, 0 abstaining, all in favor
- b. Other events:
 - i. Culminating event of the WG3 shared task on morpho-syntactic parsing (see WG3 below) - possible venues are being looked at

14. Follow up and discussion

- a. Implementation of the [COST Excellence and Inclusiveness Policy](#)
 - i. [AGATA SAVARY: 3 min.] Policy for selecting the participants of the general meeting in Naples - see above
 - ii. [NURIT MELNIK/STELLA MARKANTONATOU: 5 min.] Statistics from STSMs - among the accepted STSM applications
 - Women
 - a. Year 1: 12 of 17

- b. Year 2: 4 of 9
 - Inclusiveness target countries
 - a. Year 1: 5 of 17
 - b. Year 2: 3 of 9
 - Early-stage Researchers and Investigators
 - a. Year 1: 11 of 17
 - b. Year 2: 4 of 9
 - iii. Questions or comments from the MC
 - Call to the MC: Look for potential candidates in your networks, forward the calls, approach particular candidates
 - Eligible participants: those affiliated in COST Countries and Near-Neighbor Countries (see the [map](#))
- b. Progress of each working group
 - i. [BRUNO GUILLAUME/KAJA DOBROVOLJC: 5 min.] [WG1](#)
Corpus annotation (Next meeting: 11th June)
 - **Task 1.1: Linguistic typology and multilingual corpus annotation**
 - **Task 1.2: MWE annotation guidelines and UD-PARSEME unification**
 - **Task 1.3: Extensions and updates to morphosyntactic annotation guidelines**
 - **Task 1.4: Sharing tools, formats, and infrastructure**
 - **Task 1.5 (starting): Annotation of Spoken data**
 - ii. [VERGINICA MITITELU/VOULA GIOULI: 5 min.] WG2:
Lexicon-corpus interface (next monthly meeting: 26th April)
 Ongoing tasks:
 - **T2.1.1. Harmonizing the definition of a “syntactic word” across languages** - a second survey (created in a LimeSurvey instance) is being prepared, designed according to the insights gained in the first one; expected to be sent out in May
 - focused on clarifications of key concepts,
 - answers collected in a structured way (tables with example, transliterations, glosses, and translations)
 - Interest extend also to languages not yet present in UD
 - **T2.1.2. Harmonizing lemmatization rules (for words and MWEs) and lexical features across languages** - An initial list of lemmatization challenges has been compiled and presented in the general meeting in Naples. It will be enriched further by investigating UD treebanks and PARSEME corpora.
 - **T2.2. Design of a lexicon-corpus interface** - regular monthly meetings (next meeting: 9th May); focus on extending the ELEXIS corpus
 - New annotation levels: UD parsing, MWEs annotation, NEs annotation

- 11 new languages declared interest - the active ones are in the checking the translation phase;
- A demo with the INCEpTION tool was made in Naples - for correcting the automatic parsing
- **T2.3.1. Survey on MWEs lexicons** - regular bimonthly meetings (next meeting: 15th May); progressing with structuring the information about the MWEs lexicons
- **T2.3.2. Survey on MWE lexica-aware NLP tasks/applications** - regular bimonthly meetings (next meeting: 26th April):
 - Literature survey
 - Survey created to be sent out

To start soon:

- **T2.1.3. Standardizing lists of lexemes for auxiliaries, pronouns and determiners**
- **T2.3.3. Standardizing lexicons of MWEs** - already a presentation in Naples, envisaging the work

Publications:

- The volume “Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives” (eds. V. Giouli, V. Barbu Mititelu), Language Science Press (forthcoming)
- a paper (co-authored by the tasks co-leaders and WG co-leaders) accepted to be presented at the MWE-UD WS.
-

No comments or questions from the MC

iii. [JOAKIM NIVRE/GULSEN ERYIGIT: 5 min.] WG3

Multilingual language technology. Ongoing tasks:

- **T3.1** Documentation of multilingual tools and resources
- **T3.2** Multilingual evaluation campaign: Morphosyntactic parsing
 - a. Possible venues for a culminating event: IWPT workshop , Syntax Fest
- **T3.3** Conceptions of multilinguality

Special request to all MC members: Find a representative of each language spoken in the country to participate in subtask

T3.1.1 Assessing the “discoverability” of NLP tools.

- Task 3.1.1 is described in this shared slide deck (slides 1-8).
https://docs.google.com/presentation/d/1B5H3IQ1CF8Ka_WwIVP_Wcug-aOzLV7w0iCnuMWtug0l/edit#slide=id.p
- If volunteers are interested in participating, the spreadsheet template that they should use is here:

https://docs.google.com/spreadsheets/d/17_5jhUWeYy7WD6OY79Kdyn3ngoB-3Y82YF0yaRpQSv8/edit#gid=0

- The ReadMe tab explains the task clearly.
- The volunteer should enter their details on this first tab.
- The volunteer should append their name to the filename before uploading to the shared drive:
<https://drive.google.com/drive/u/1/folders/1u6hnAAPchSddPIMav1U0jHzzF2o4Ha6A>
- Questions can be sent to:
Teresa Lynn (teresa.m.lynn@gmail.com),
Maria Giagkou (Mgiagkou@athenarc.gr) and
A. Seza Dođruöz (as.dogruoz@ugent.be).

Discussion:

- [ALINE VILLAVICENCIO] Promote these initiatives in local networks

iv. [MARIE-CATHERINE DE MARNEFFE/OLHA KANICHSHEVA: 5 min.]
[WG4](#) Quantifying and promoting diversity

Two current subtasks:

Task 4.1. Promoting diversity, especially low-resource languages

Leaders: Abigail Walsh (abigail.walsh@adaptcentre.ie) and Lucia Amorós (lamoros@um.es)

- Bridge with other WGs
- Report on SOTA on low-resource languages, in parallel with the Digital Language Equality metric
- **Call for a "project manager"**, skilled at organizing the individual subtasks in the subtask. Any takers? If yes, contact Abigail and Lucia.
- Contact Abigail and Lucia to **share any of your endeavors that promote** low-resource languages!

Task 4.2. Quantifying diversity

Leader: Louis Estève

Survey that focuses on diversity measures proposed in ecology (that use the concept of variety (how many types), balance (how are types distributed), and disparity (how different types are)) and see which ones are applicable to linguistics. Right now the focus is on textual data, and intra-linguistic diversity.

Upcoming task: **Task 4.3. Measuring NLP shared-task diversity**

Looking for a leader! Contact Marie

(marie-catherine.demarneffe@uclouvain.be) and Olha

(kanichshevaolga@gmail.com) if interested

v. Comments or questions from the MC

- [ALINE VILLAVICENCIO] What is the readiness of a possible candidate to join task 4.1?
- [MARIE-CATHERINE] Any help in subtask management is welcome.

15. [ALINA WRÓBLEWSKA: 5 min.] Progress on MoU Objectives, deliverables, and goals for the current GP.

- a. See the [deliverables diagram](#)
- b. Comments and questions from the MC
 - i. Reporting achievements promoting UniDive on <https://unidive.lisn.upsaclay.fr/doku.php?id=outcomes:dissemination>

16. AOB

- a. [Vladimir BENKO] How can an invitee from a non-COST and non-NNC participate in UniDive activities? Example of an Uzbek colleague, willing to construct a UD treebank:
 - i. If the person has an affiliation in a COST or NNC, she/he is eligible for all activities
 - ii. Otherwise she/he can
 - Be member of any WG
 - Participate in any WG meetings online
 - Be funded for one onsite meeting participation, once within the lifetime of the action
 - iii. She/he cannot:
 - Be funded for the training school
 - Apply for STSMs or conference grants

17. Next MC meeting

a. Thursday 24 October 10:00-12:00 CET

18. Summary of MC decisions

- a. Assigning the 9155 EUR savings from Naples to the MWE-UD workshop
- b. Organizing the 3rd general meeting jointly with WG2 meeting

TODO

- [MC MEMBERS] Help us with the ongoing tasks
 - Advertise the training school and STSMs in your national networks, approach particular candidates
 - Contribute yourself and/or find as many contributors as possible to the T3.1.1 about Assessing the “discoverability” of NLP tools (see above), for languages spoken in your countries or those you work on
 - Submit proposals for a venue for the culminating event of the WG3 shared task
 - Consider being part of Task 4.1 and contribute with management skills, or approach candidates available to do so

- Consider venues to disseminate outcomes of the action towards a larger public (see examples [here](#))
- [TRAINING SCHOOL ORGANIZERS] Send of 2nd call for (i) trainee projects, (ii) reviewers

TOP-UP BUDGET SUMMARY

Type of Activity benefiting from the top-up (e.g. WG Meeting, Workshop/Conference, Grants, Dissemination)	MWE-UD workshop in Turin (Italy)	1st UniDive Training School in Chişinău (Moldova)	ITC or Dissemination Conference Grants
Specify if activity is already planned (included in the work and budget plan) or New (not included in the plan)	Planned (for 25 May 2024)	Planned (for 8-12 July 2024)	Planned
Added value of the activity towards Action objective(s) and/or deliverable(s)	MWE-UD is a union of two established workshops, central to the topics of UniDive. Additional budget would directly contribute to 2 Grant Period Goals: (i) Increasing the convergence of PARSEME (whose main venue is the MWE workshop) and UD (whose main venue is the UD workshop) in treebanking, (ii) Reinforcing the strategy for the development of language resources and tools for under-resourced languages (presenters of work on under-resourced languages are a priority for UniDive grants).	The training school serves the aim of reinforcing the strategy for the development of language resources and tools for under-resourced languages. With more trainees invited, more under-resourced languages will potentially be covered.	The initial budget for these grants is rather low (1500 EUR). Additional budget will reinforce dissemination to new communities.
Estimated budget (FSAC excl.) to allocate to the proposed activity	6,250 EUR (8 more participants)	12,000 EUR (10 extra trainees)	3,000 EUR (2-3 more grants)

Total for the scientific activity: 21250

FSAC: 3750 (15%)

TOTAL: 25,000 EUR