

UniDive core group meeting 10

17 October 2023

See also the [CG master document](#) and the [UniDive master document](#)

Attendants	1
Agenda/minutes	1
TODO	3
INTEGRATING EXPERTS OF LOW-RESOURCED LANGUAGES	4

Attendants

Atul, ~~Beata~~, Bruno, Dan, Alina, Joakim, ~~Marie~~, ~~Nurit~~, Olesea, Verginica, Agata

Agenda/minutes

- [AGATA] LREC/COLING [paper](#) - harmonization is ongoing
 - Are the figures final?
 - Overlapping language names in Fig. 1
- [AGATA] Goals of the Work and Budget Plan for year 2
 - Year 2 draft budget - see [here](#)
 - Those defined by WG leaders
 - Increasing the convergence of PARSEME and UD in treebanking
 - The notion of “syntactic word” across languages: taking stock and developing recommendations for future annotation projects
 - Designing a lexicon-corpus interface for interlinking MWE lexicon entries with their occurrences in corpora
 - Setting up a framework and workflow for documenting multilingual language technology tools within the scope of UniDive
 - Devising a diversity measure promoting universality and diversity in language technology
 - Launching at least one multilingual evaluation campaign promoting universality and diversity in language technology
 - Enhancing the visibility of the Action on social media
 - Reinforcing the strategy for the development of language resources and tools for under-resourced languages
 - Link with the request from Beata and Martin - how can experts of low-resourced languages actively participate in UniDive – see [here](#).
 - Strong links with AUTOGRAMM - TODO
 - Part of the training school?
- Organization of the 2nd General Meeting in Naples:
 - the WG1 Meeting on **Wednesday 7 Feb**

- the 2nd General Meeting on **Thursday-Friday 8-9 Feb** - Posters? Invited talk? WG meetings in parallel?
 - JN: Liked the format in Paris-Saclay, posters are a good occasion to share work and discuss
 - BG: More submissions than expected, people like the idea: posters increase participation vs. passive attendance
 - OC: The reviews were not always understandable; can we
 - JN: Review complete work, work in progress and planned work in different pools - 3-4 categories (to check at the submissions stage)
 - OC: Template for posters prepared by the communication team
- PC
 - Nominating PC chairs
 - WG chairs submit names of active people from WGs who could play this role - Thursday 20/10
 - Extended Core Group - OK
 - Call for PC members - trust people to be competent
 - Submission platform - Open Review
- **Timeline:**
 - 25 Oct: Call
 - 3rd week of November: Submission deadline
 - See who is concerned with visa issues, deal with invitations
 - 10 December: Reviews
 - mid-December: Notification
 - List of presenters
 - Early January: Official invitation
- Visa issues:
 - In the submission form ask if the presenter (if any) will need a visa
 - Same in the EoI form
- Invited talk
 - Sylvain Kahane (AUTOGRAMM)
 - Martin & Beata - African languages
 - Other ideas?
- Plenary session about communication (Olesea)
- Invitation letter sent to Anna Danielyan (Armenia) to let her get her visa on time
 - CG - OK to invite her anyway, ask Johanna for an invitation
- Organization of the 1st Training School in Jerusalem
 - Dates: July 8-12 2024 OR July 15-19 2024
 - Do we want to maintain the venue, given the current situation in Israel and Gaza?
 - 2 options:
 - Maintain Jerusalem + find a backup venue (Malta, Bucharest - no lab room?), ideally in an ITC
 - Change to another venue, keep Jerusalem and Budapest for other events
 - Plan

- Wait to hear from Chaya and Giuseppe
 - If Chaya is OK, go on with Jerusalem
 - Look for a backup venue
 - Start talking to potential trainers (for the transfer course)
 - [AGATA] Ideas of topics and trainers
 - Annotating a UD treebank for a new language (Marie? Joakim?) - an occasion to enhance the annotation guidelines for newcomers
 - Annotating a PARSEME corpus for a new language
 - Git for newcomers + UD/PARSEME software infrastructure (Dan? Agata?)
 - Transfer learning for low-resourced languages (Ivan? Lonneke? Tanja? Myriam?)
 - Grew-match and Arborator-grew (Bruno et al.?)
 - Other ideas?
 - Labs or personal computers?
 - Working with personal laptops would be most useful
 - Offer a few PCs for those who can't work on their laptops
- [OLESEA] Communication - current issues → delay to the next meeting
 - Invite members of WG1, WG3 and WG4 to join Telegram
- [5 min.] Forthcoming deliverable(s) - content, WGs involvement, deadlines → delay to the next meeting
- Next meeting: **Nov 21, 2023 11:00-12:30.**

TODO

- [AGATA] Publish the minutes from this meeting
- LREC/COLING paper
 - [AGATA: 18/10] Finalize the draft
 - [ALL: 19/10] Send feedback
- [AGATA] WBP
 - Finish editing
 - Wait for Giuseppe and Chaya to confirm Jerusalem
- [AGATA] Newsletter:
 - Need for reviewers from the LRE journal
 - Naples
 - Workshop proposal submitted
- Naples:
 - [AGATA] [Google doc](#) for the event
 - [WG LEADERS: Thursday 19 Oct] Ideas of PC chairs
 - [AGATA] Propose PC chairs, contact them, help with the CfP. Anna's abstract pre-selected
 - [AGATA] Create the event's page, create an account for Johanna
 - [AGATA] Talk to Johanna
 - Event's page
 - Sending out invitations (but only an e-cost invitation confirms the reimbursement), including Anna Danielyan

- [ALL] Ideas of invited speakers
- [JOHANNA] Prepare to send out invitations
- [OLESEA: mid-December] Prepare a poster template with the communication team
- UD-MWE workshop
 - [AGATA] Check with the proposers: the COST conditions for co-organising the event
- Next CG meeting
 - [AGATA] Agenda with the 2 last items + integrating experts of low-resourced languages + Jerusalem backup venue + courses/trainers
 - [ALL] Think about invited speakers for Naples
 - [ALL] Think about a backup venue for the training school (preferably in an ICT)

INTEGRATING EXPERTS OF LOW-RESOURCED LANGUAGES

Beata and Martin are experts of African languages. They would like to enhance the NLP coverage of these languages in UniDive but they do not know where to start.

Could it be a new task (maybe in WG4?) to integrate, inform and train such experts?

Here are first ideas:

1. Existing resources

A) Look at the [UD](#) and [AUTOGRAMM](#) pages and check if there already exist treebanks of the languages that you know.

B) Check if these languages have repositories of senses like Wordnet.

C) Check if these languages have MWE lexicons.

2. Ongoing UniDive tasks

- If the answer to 1A is yes, you could participate in the 2.1 task on [what is a word](#).

- If the answer to 1B is yes, you could participate in the 2.2 task on a [sense-annotated corpus](#)

- If the answer to 1C is yes, you could participate in the 2.3 task on [MWE lexicons](#)

- There are other ongoing tasks, in WG1, WG2 and WG4, but these are more "universal", in the sense of looking at many languages at the same time and defining unified guidelines and measures. There, you can of course also participate but the direct impact on African languages will be less directly visible.

In any case, possible contributions would be to:

3. Create new UD treebanks (annotated for morpho-syntax) or PARSEME corpora (annotated for idiomatic expressions), even small, for the languages you know, maybe with help from your students/colleagues

The June 2023 UniDive webinar provided tutorials for newcomers to these two frameworks.

The talks are recorded [here](#).

The training school in Jerusalem will likely also touch upon these issues.

4. Create data for future WG3 evaluation campaigns. We are currently discussing what this data should look like but your participation would certainly be VERY precious , since strong impact should be put on heavily under-resourced languages.