

## Work And Budget Plan

CA21167 Grant Agreement Period 2

01/11/2023 to 31/10/2024

### Action Profile

#### Action General Information

Action Code	CA21167	MC Chair	Prof Agata Savary
Action Title	CA21167 - Universality, diversity and idiosyncrasy in language technology		
MOU	081/22	Draft MOU	OC-2021-1-25588
CSO Approval Date	2022-05-27		
Action Start Date	23/09/2022	Action End Date	22/09/2026
Science Officer	Dr Giuseppe Lugano	Administrative Officer	Ms Andrea Tortajada

#### Participation in the Action:

Number of	COST Full or Cooperating Members	COST Partner Members	Specific Organisations	Near Neighbour Countries	Third States	Total
<b>Countries</b>						
COST Members / Specific Organisations represented in the MC	36	0	0	n.a.	n.a.	36
	ITC 64%					
	Non-ITC 36%					
Countries represented in the Working Groups	39	0	0	2	5	46
	ITC 62%					
	Non-ITC 38%					
<b>Individuals</b>						
Nominated MC Members / MC Observers	67	0	0	n.a.	n.a.	67
	ITC 63%					
	Non-ITC 37%					
Approved Working Group members	279	0	0	2	9	290
	ITC 58%					
	Non-ITC 42%					

Submitted : 2023-11-07 3:03 PM

Generated : 2023-11-09 12:20 PM

## Working Groups

	WG Title	WG Leader	Number of WG members
WG1	WG1: CORPUS ANNOTATION	Mr Bruno Guillaume	177
WG2	WG2: LEXICON-CORPUS INTERFACE	Dr Verginica MITITELU	141
WG3	WG3: MULTILINGUAL AND CROSS-LINGUAL LANGUAGE TECHNOLOGY	Prof Joakim Nivre	185
WG4	WG4: QUANTIFYING AND PROMOTING DIVERSITY	Dr Marie-Catherine de Marneffe	114

## MoU objectives, Action deliverables and Grant Agreement Period Goals

### Action Objectives from MoU

Aim/primary Objective
reconcile language diversity with rapid progress in language technology
Secondary objectives
<ol style="list-style-type: none"> <li>1. To develop methods for quantifying inter- and intra-linguistic diversity.</li> <li>2. To develop a common understanding of language universals across 70 languages represented in the Action.</li> <li>3. To coordinate the diversity-driven creation, merging and enhancement of language resources unified across over 100 languages from the UD and PARSEME collections.</li> <li>4. To coordinate efforts towards a better coverage of inter-/intra-linguistic diversity in NLP tools.</li> <li>5. To raise awareness of the international community about the importance of diversity preservation in language technology.</li> <li>6. To disseminate the Action outcomes to stakeholders.</li> <li>7. To create a network of experts in a large number of languages working on modelling and processing of morphological, syntactic and semantic phenomena within a common framework.</li> <li>8. To foster the capacities of Young Researchers and Innovators (YRIs), with special focus on COST ITC participants.</li> <li>9. To coordinate and boosting universality-driven initiatives worldwide.</li> <li>10. To set up a long-term roadmap for the joint efforts of the universality-driven NLP community.</li> </ol>

## Action Deliverables

Deliverable	Month
1. Website: describing the Action's organisation, and gathering the links to its events and outcomes.	3
2. A scientific publication describing measures of inter- and intra-linguistic diversity in language resources and tools.	12
3. Centralized documentation of unified file formats and conventions for corpora and lexica.	18
4. Unified, enhanced and enlarged versions of existing annotation guidelines for lexical features, morphology, syntax and MWEs, together with the criteria for applying these unified guidelines to specific languages.	24
5. Centralized documentation of the nationally funded software infrastructures coordinated in WG1 and WG4, to support universality and diversity in language resources.	24
6. Centralized documentation of (new or enhanced) annotated corpora for at least 100 languages.	36
7. Documentation of the prototypes of NLP-applicable lexica of MWEs and idiosyncratic constructions.	42
8. Centralized documentation of multilingual and cross-lingual NLP tools coordinated in WP3: syntactic and semantic parsers, MWE discovery tools, MWE identifiers, prototypes of identifiers of idiosyncratic constructions.	45
9. Diversity benchmarks for NLP: diversity-driven evaluation scenarios for NLP resources and tools; infrastructure for evaluation campaigns of NLP tools; evaluation results of at least 2 evaluation campaigns and focused on inter/intra-linguistic diversity in 100 languages.	45
10. Other dissemination material: reports from STSMs; material from training schools; proceedings of workshops; joint papers in Open Access journals, conferences and books; dissemination material dedicated to a large audience (e.g., demonstrations of tools and Wikipedia entries about diversity in NLP, MWEs and idiosyncratic constructions, and interesting syntactic phenomena).	48

## Grant Agreement Period

Grant Agreement Period Start Date	01/11/2023	Grant Agreement Period End Date	31/10/2024
-----------------------------------	------------	---------------------------------	------------

## Grant Agreement Period Goals

Number	Grant Agreement Period Goal	MoU Objective(s) it relates to
GAPG 1	Increasing the convergence of PARSEME and UD in treebanking	<ul style="list-style-type: none"> <li>• Secondary objective 2</li> <li>• Secondary objective 3</li> <li>• Secondary objective 9</li> </ul>
GAPG 2	The notion of “syntactic word” across languages: taking stock and developing recommendations for future annotation projects	<ul style="list-style-type: none"> <li>• Secondary objective 2</li> <li>• Secondary objective 3</li> <li>• Secondary objective 9</li> <li>• Secondary objective 10</li> </ul>
GAPG 3	Designing a lexicon-corpus interface for interlinking MWE lexicon entries with their occurrences in corpora	<ul style="list-style-type: none"> <li>• Secondary objective 3</li> <li>• Secondary objective 7</li> <li>• Secondary objective 9</li> </ul>
GAPG 4	Setting up a framework and workflow for documenting multilingual language technology tools within the scope of UniDive	<ul style="list-style-type: none"> <li>• Secondary objective 1</li> <li>• Secondary objective 4</li> <li>• Secondary objective 5</li> <li>• Secondary objective 6</li> </ul>
GAPG 5	Devising a diversity measure promoting universality and diversity in language technology	<ul style="list-style-type: none"> <li>• Challenge</li> <li>• Secondary objective 1</li> <li>• Secondary objective 3</li> <li>• Secondary objective 5</li> <li>• Secondary objective 6</li> <li>• Secondary objective 10</li> </ul>
GAPG 6	Launching at least one multilingual evaluation campaign promoting universality and diversity in language technology	<ul style="list-style-type: none"> <li>• Challenge</li> <li>• Secondary objective 1</li> <li>• Secondary objective 3</li> <li>• Secondary objective 4</li> <li>• Secondary objective 5</li> <li>• Secondary objective 7</li> <li>• Secondary objective 8</li> <li>• Secondary objective 10</li> </ul>
GAPG 7	Enhancing the visibility of the Action on social media	<ul style="list-style-type: none"> <li>• Secondary objective 3</li> <li>• Secondary objective 4</li> <li>• Secondary objective 5</li> <li>• Secondary objective 6</li> <li>• Secondary objective 7</li> <li>• Secondary objective 8</li> </ul>
GAPG 8	Reinforcing the strategy for the development of language resources and tools for under-resourced languages	<ul style="list-style-type: none"> <li>• Challenge</li> <li>• Secondary objective 2</li> <li>• Secondary objective 3</li> <li>• Secondary objective 4</li> <li>• Secondary objective 7</li> <li>• Secondary objective 8</li> </ul>

## Work and Budget Plan for the Grant Agreement Period

### Work and Budget Plan Summary

<b>A. COST Networking Tools</b>	EUR
(1) Meetings	86,000.00
(2) Training Schools	72,800.00
(3) Mobility of Researchers and Innovators	28,765.00
(4) Presentation at Conferences organised by Third Parties	1,500.00
(5) Dissemination and Communication Products	0.00
(6) Other Expenses Related to Scientific Activities (OERSA)	500.00
<b>B. Total Science Expenditure (sum of (1) to (6))</b>	189,565.00
<b>C. Financial and Scientific Administration and Coordination (FSAC) (max. of 15% of B)</b>	28,434.75
<b>Total Grant (B+C)</b>	217,999.75

## Meetings

### Overview

Meeting Title	Meeting Type	Dates	Location	ITC	Total Cost (EUR)
UniDive 2nd General Meeting	Core Group Meeting, Working Group Meeting, Workshops/Conferences	07/02/2024 - 09/02/2024	Naples (Italy)	No	78,800.00
Management Committee meeting 4	Management Committee Meeting	24/04/2024 - 24/04/2024	N/A	N/A	0.00
MWE-UD workshop	Workshops/Conferences	20/05/2024 - 20/05/2024	Torino (Italy)	No	7,200.00
Management Committee Meeting 5	Management Committee Meeting	24/10/2024 - 24/10/2024	N/A	N/A	0.00
				<b>Total</b>	<b>86,000.00</b>

### Details

Title of the Meeting	UniDive 2nd General Meeting		
Meeting Type(s)	Core Group Meeting, Working Group Meeting, Workshops/Conferences		
Grant Period Goal(s) it will address	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, Reinforcing the strategy for the development of language resources and tools for under-resourced languages		
Description	Day 1: Dedicated exclusively to a WG1 meeting, given that WG1 deliverables are the earliest in the Action's Gantt diagram in the MoU. Program: progress on the ongoing WG1 tasks. Day 2: General workshop, with an invited talk, plenary poster boosters, and classical poster sessions. Day 3: Parallel WG sessions. An Extended Core Group meeting. Note that the Core Group, the WGs and the WG subtasks also regularly meet online. - Core Group meetings take place on each 3rd Tuesday in a month from 11:00 to 12:30 CEST/CET. - WGs typically meet once a month to structure the WG into subtasks and report on the progress withing the subtasks (but these meetings might become less frequent as the WG subtasks are launched and the need for WG coordination becomes lower) - WG subtasks groups meet regularly online, depending on the agenda at hand. We do not encode all these CG, WG and subtask meetings in the WBP but they are documented on our website ( <a href="https://unidive.lisn.upsaclay.fr/">https://unidive.lisn.upsaclay.fr/</a> ).		
Output(s)	Minutes and technical documentation from the WG meetings; posters and slides from scientific presentations; minutes from the ECG meeting.		
Location	Naples (Italy)	ITC	No
Start Date	2024-02-07 09:00:00	End Date	2024-02-09 18:00:00
Duration	3 days	Attendance Type	Hybrid
Total number of expected participants	70	Number of participants expected to be reimbursed from COST funds	70
Daily allowance (EUR)	160.00	Average number of daily allowances per participant	4

Average Long-Distance Costs (EUR)	400.00		
Average reimbursement (per participant) (EUR)	1,040.00		
Total Travel, Accommodation and Subsistence Costs (EUR)	72,800.00		
Total unique participants to be accounted for LOS grant	70	Local Organiser Support (EUR)	6,000.00
Total cost of the meeting (EUR)	78,800.00		

Title of the Meeting	Management Committee meeting 4		
Meeting Type(s)	Management Committee Meeting		
Grant Period Goal(s) it will address	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		
Description	Management Committee meeting, including reporting from the Core Group to the MC about the decisions related to the delegation of powers and about the progress in MoU objectives.		
Output(s)	Minutes from the meeting Agenda		
Start Date	2024-04-24 10:00:00	End Date	2024-04-24 12:00:00
Duration	1 day	Attendance Type	Virtual
Total number of expected participants	60	Number of participants expected to be reimbursed from COST funds	0
Daily allowance (EUR)	0.00	Average number of daily allowances per participant	2
Average Long-Distance Costs (EUR)	0.00		
Average reimbursement (per participant) (EUR)	0.00		
Total Travel, Accommodation and Subsistence Costs (EUR)	0.00		
Total unique participants to be accounted for LOS grant	0	Local Organiser Support (EUR)	0.00
Total cost of the meeting (EUR)	0.00		

Title of the Meeting	MWE-UD workshop		
Meeting Type(s)	Workshops/Conferences		



Grant Period Goal(s) it will address	Increasing the convergence of PARSEME and UD in treebanking, Reinforcing the strategy for the development of language resources and tools for under-resourced languages		
Description	<p>A workshop organised jointly with the Universal Dependencies community and the MWE section of SIGLEX (ACL Special Interest Group on the Lexicon), to be co-located with the LREC-COLING 2024 conference. These both communities and their topics largely overlap with UniDive. The MWE workshop and the UD workshop are largely established events with world-wide participation and affiliation with the major association (ACL) in the scientific domain of UniDive. Thus, UniDive will benefit from this co-organized event by spreading its objectives and outcomes to a broader international community. The workshop co-chairs are all members of UniDive. Most if not all publications are expected to contribute to the objectives of UniDive. Good part of the participants will likely also be from UniDive. At least one of the sessions might therefore be explicitly labelled as a UniDive session. It was already agreed on that COST shall be visible as the organizer on the event's website, COST Action website and publication and dissemination materials. The event's website is not yet available, since the workshop is at the stage of a submitted proposal, being evaluated by the chairs of the main conference. The notification should be received on 17 November 2023. In case of rejection, the proposal will likely be resubmitted to another venue or the event will take place independently of any conference.</p>		
Output(s)	Proceedings from the conference published under an open license.		
Location	Torino (Italy)	ITC	No
Start Date	2024-05-20 09:00:00	End Date	2024-05-20 18:00:00
Duration	1 day	Attendance Type	Face to face
Total number of expected participants	40	Number of participants expected to be reimbursed from COST funds	10
Daily allowance (EUR)	160.00	Average number of daily allowances per participant	2
Average Long-Distance Costs (EUR)	400.00		
Average reimbursement (per participant) (EUR)	720.00		
Total Travel, Accommodation and Subsistence Costs (EUR)	7,200.00		
Total unique participants to be accounted for LOS grant	0	Local Organiser Support (EUR)	0.00
Total cost of the meeting (EUR)	7,200.00		

Title of the Meeting	Management Committee Meeting 5
Meeting Type(s)	Management Committee Meeting
Grant Period Goal(s) it will address	<p>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</p>

Description	Management Committee meeting, including reporting from the Core Group to the MC about the decisions related to the delegation of powers and about the progress in MoU objectives		
Output(s)	Minutes from the meeting		
Start Date	2024-10-24 10:00:00	End Date	2024-10-24 12:00:00
Duration	1 day	Attendance Type	Virtual
Total number of expected participants	60	Number of participants expected to be reimbursed from COST funds	0
Daily allowance (EUR)	0.00	Average number of daily allowances per participant	2
Average Long-Distance Costs (EUR)	0.00		
Average reimbursement (per participant) (EUR)	0.00		
Total Travel, Accommodation and Subsistence Costs (EUR)	0.00		
Total unique participants to be accounted for LOS grant	0	Local Organiser Support (EUR)	0.00
Total cost of the meeting (EUR)	0.00		

## Training Schools

### Overview

Title of the Training School	Dates	Location	ITC	Total Cost (EUR)
UniDive 1st Training School	08/07/2024 - 12/07/2024	Jerusalem (Israel)	No	72,800.00
			Total	72,800.00

### Details

Title of the Training School	UniDive 1st Training School		
Grant Period Goal(s) it will address	Reinforcing the strategy for the development of language resources and tools for under-resourced languages		
Description	Training School dedicated mainly you YRIs, especially those working on low-resourced languages. Planned courses: - Annotating a UD treebank and a PARSEME corpus for a new language - UD and PARSEME infrastructure (Grew-match, Arborator-grew, Git, ...) tools for newcomers - Transfer learning for low-resourced languages The courses should be in the form of lectures and practical hands-on sessions. The Jerusalem venue has been intensely discussed due to the current situation in the Middle East. Due to the reasons presented in the MC3 meeting on 2 November 2023 (see the minutes on the Action's page), we propose to go on with these plans, while also looking for a back-up plan.		
Output(s)	Training material (slides, labs, guidelines, etc.)		
Location	Jerusalem (Israel)	ITC	No
Start Date	2024-07-08 09:00:00	End Date	2024-07-12 18:00:00
Duration	5 days	Attendance Type	Face to face
Participant information	Trainers	Trainees	
• Daily allowance (EUR)	170.00	170.00	
• Average number of daily allowances per participant	6	6	
• Average Long-Distance Costs (EUR)	650.00		
• Average reimbursement (EUR)	1,670.00	1,670.00	
• Total number of expected trainers/trainees	5	40	
• Trainers/trainees expected to be reimbursed from COST funds	5	35	
Total Travel, Accommodation and Subsistence Costs (EUR)	66,800.00		
Total unique participants to be accounted for LOS grant	45		
Local Organiser Support (EUR)	6,000.00		
Total cost of the Training School (EUR)	72,800.00		

## Mobility of Researchers and Innovators

Grant Period Goal(s) it will address:	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, Reinforcing the strategy for the development of language resources and tools for under-resourced languages	
Description:	STSMs dedicated mainly (but not only) to YRIs for the implementation of the action's aims.	
Budget (EUR)	Short Term Scientific Missions (STSM) grants	28,765.00
	This budget would allow for approx. 14 STSM grants (based on the average costs per STSM grant spent by all Actions).	
	Virtual Mobility grants	0.00
	This budget would allow for approx. 0 Virtual Mobility grants (based on the average costs per VM grant spent by all Actions).	
	Total	28,765.00

**Presentations at Conferences organised by Third Parties**

Description:	2 ITC Conference grants (especially for YRIs) and 1 Dissemination Conference grant	
Budget (EUR)	ITC Conference grants	1,000.00
	This budget would allow for approx. 1 ITC Conference grants (based on the average costs per ITC Conference grant spent by all Actions).	
	Dissemination Conference grants	500.00
	This budget would allow for approx. 0 Dissemination Conference grants (based on the average costs per DC grant spent by all Actions).	
	Total	1,500.00

## Other Expenses Related to Scientific Activities (OERSA)

### Overview

Item	Total Cost (EUR)
Bank charges	500.00
Total	500.00

### Details

Type:	Bank charges
Total (EUR): (without VAT)	500.00