## Action Profile

### Action General Information

| | | | |
|---|---|---|---|
| Action Code | CA21167 | Action 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 Nathalie Warenghien |

### Participation in the Action:

| Number of | COST Full or Cooperating Members | | COST Partner Members | Specific Organisations | Near Neighbour Countries | Third States | Total |
|---|---|---|---|---|---|---|---|
| **Countries** | | | | | | | |
| COST Members / Specific Organisations represented in the MC | 38 | | 0 | 0 | n.a. | n.a. | 38 |
| | ITC | 61% | | | | | |
| | Non-ITC | 39% | | | | | |
| Countries represented in the Working Groups | 40 | | 0 | 0 | 2 | 7 | 49 |
| | ITC | 60% | | | | | |
| | Non-ITC | 40% | | | | | |
| **Individuals** | | | | | | | |
| Nominated MC Members / MC Observers | 71 | | 0 | 0 | n.a. | n.a. | 71 |
| | ITC | 61% | | | | | |
| | Non-ITC | 39% | | | | | |
| Approved Working Group members | 357 | | 0 | 0 | 3 | 13 | 373 |
| | ITC | 53% | | | | | |
| | Non-ITC | 47% | | | | | |

Submitted : DRAFT NOT YET SUBMITTED
Generated : 2024-11-13  12:43 PM

**COST Association AISBL**
Avenue du Boulevard – Bolwerklaan 21 | 1210 Brussels, Belgium
T +32 (0)2 533 3800 | office@cost.eu | www.cost.eu

**Funded by the European Union**

## Working Groups

| | WG Title | WG Leader | Number of WG members |
|---|---|---|---|
| WG1 | WG1: CORPUS ANNOTATION | Mr Bruno Guillaume | 238 |
| WG2 | WG2: LEXICON-CORPUS INTERFACE | Dr Verginica MITITELU | 184 |
| WG3 | WG3: MULTILINGUAL AND CROSS-LINGUAL LANGUAGE TECHNOLOGY | Prof Joakim Nivre | 244 |
| WG4 | WG4: QUANTIFYING AND PROMOTING DIVERSITY | Dr Marie-Catherine de Marneffe | 150 |

# 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** |
| 1. To develop methods for quantifying inter- and intra-linguistic diversity.<br>2. To develop a common understanding of language universals across 70 languages represented in the Action.<br>3. To coordinate the diversity-driven creation, merging and enhancement of language resources unified across over 100 languages from the UD and PARSEME collections.<br>4. To coordinate efforts towards a better coverage of inter-/intra-linguistic diversity in NLP tools.<br>5. To raise awareness of the international community about the importance of diversity preservation in language technology.<br>6. To disseminate the Action outcomes to stakeholders.<br>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.<br>8. To foster the capacities of Young Researchers and Innovators (YRIs), with special focus on COST ITC participants.<br>9. To coordinate and boosting universality-driven initiatives worldwide.<br>10. To set up a long-term roadmap for the joint efforts of the universality-driven NLP community. |

## Action Deliverables

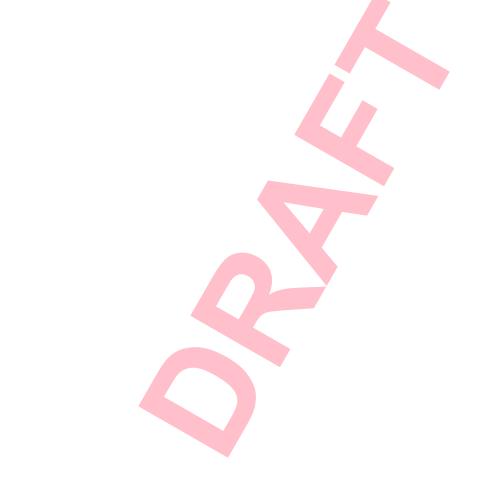| Deliverable | Month |
|---|---|
| 1. Website: describing the Action's organisation, and gathering the links to its events and outcomes. | 3 |
| 2. A scientific publication describing measures of inter- and intra-linguistic diversity in language resources and tools. | 12 |
| 3. Centralized documentation 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 | 01/11/2024 | Grant Agreement Period End Date | 31/10/2025 |

## Grant Agreement Period Goals

| Number | Grant Agreement Period Goal | MoU Objective(s) it relates to |
|---|---|---|
| GAPG 1 | Coordinating efforts towards better morphosyntactic corpus annotation, accounting for new languages, typology, constructions and spoken language (tasks 1.1, 1.3 and 1.5) | <ul><li>Secondary objective 2</li><li>Secondary objective 3</li><li>Secondary objective 4</li><li>Secondary objective 7</li><li>Secondary objective 10</li></ul> |
| GAPG 2 | Including new languages and new categories in multiword expression annotation (task 1.2) | <ul><li>Secondary objective 2</li><li>Secondary objective 3</li><li>Secondary objective 7</li><li>Secondary objective 10</li></ul> |
| GAPG 3 | Sharing tools, formats and infrastructures for the annotation of morphosyntax and multiword expressions (task 1.4) | <ul><li>Secondary objective 3</li><li>Secondary objective 7</li><li>Secondary objective 10</li></ul> |
| GAPG 4 | Cross-language unification of the definition of a syntactic word, a lemma and lemmatisation rules (tasks 2.1 and 2.4) | <ul><li>Secondary objective 2</li><li>Secondary objective 7</li><li>Secondary objective 8</li><li>Secondary objective 10</li></ul> |
| GAPG 5 | Coordinating the construction of a parallel word-sense annotated corpus and lexicon-corpus interface (task 2.2) | <ul><li>Secondary objective 2</li><li>Secondary objective 4</li><li>Secondary objective 7</li><li>Secondary objective 8</li><li>Secondary objective 10</li></ul> |
| GAPG 6 | Developing a proof-of-concept lexicon encoding for multiword expressions (task 2.3) | <ul><li>Secondary objective 2</li><li>Secondary objective 3</li><li>Secondary objective 7</li><li>Secondary objective 10</li></ul> |
| GAPG 7 | Organising a multilingual shared task on morphosyntactic analyso-parsing (task 3.2) | <ul><li>Secondary objective 1</li><li>Secondary objective 2</li><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><li>Secondary objective 9</li></ul> |
| GAPG 8 | Defining the concepts of multilingual, cross-lingual, and translingual in the context of language technology (task 3.3) | <ul><li>Secondary objective 4</li><li>Secondary objective 7</li></ul> |
| GAPG 9 | Structuring MWE resources regarding language resourcedness (task 4.1) | <ul><li>Secondary objective 1</li><li>Secondary objective 3</li><li>Secondary objective 4</li><li>Secondary objective 7</li><li>Secondary objective 8</li></ul> |
| GAPG 10 | Survey of diversity measures in NLP (task 4.2) | <ul><li>Secondary objective 1</li><li>Secondary objective 4</li><li>Secondary objective 5</li><li>Secondary objective 7</li><li>Secondary objective 8</li><li>Secondary objective 9</li><li>Secondary objective 10</li></ul> |

| GAPG 11 | Developing methods for measuring diversity of shared-tasks (task 4.3) | • Secondary objective 1<br>• Secondary objective 4<br>• Secondary objective 7<br>• Secondary objective 8<br>• Secondary objective 9<br>• Secondary objective 10 |
|---|---|---|
| GAPG 12 | Disseminating outcomes of the Action to stakeholders (task 4.1) | • Secondary objective 6 |

# Work and Budget Plan for the Grant Agreement Period

## Work and Budget Plan Summary

| A. COST Networking Tools | EUR |
|---|---|
| (1) Meetings | 154,080.00 |
| (2) Training Schools | 0.00 |
| (3) Mobility of Researchers and Innovators | 40,000.00 |
| (4) Presentation at Conferences organised by Third Parties | 2,820.00 |
| (5) Dissemination and Communication Products | 0.00 |
| (6) Other Expenses Related to Scientific Activities (OERSA) | 491.00 |
| **B. Total Science Expenditure (sum of (1) to (6))** | 197,391.00 |
| **C. Financial and Scientific Administration and Coordination (FSAC) (max. of 15% of B)** | 29,608.65 |
| **Total Grant (B+C)** | 226,999.65 |

**Meetings**

Overview

| Meeting Title | Meeting Type | Dates | Location | ITC | Total Cost (EUR) |
|---|---|---|---|---|---|
| 3rd General Meeting | Core Group Meeting, Working Group Meeting, Workshops/Conferences | 28/01/2025 - 30/01/2025 | Budapest (Hungary) | Yes | 123,280.00 |
| Management Committee meeting 6 | Management Committee Meeting | 02/04/2025 - 02/04/2025 | N/A | N/A | 0.00 |
| UniDive shared task on Morphosyntactic analyso-parsing | Workshops/Conferences | 26/08/2025 - 26/08/2025 | Ljubljana (Slovenia) | Yes | 30,800.00 |
| Management Committee Meeting 7 | Management Committee Meeting | 23/10/2025 - 23/10/2025 | N/A | N/A | 0.00 |
| | | | | Total | 154,080.00 |

Details

| Title of the Meeting | 3rd General Meeting | | |
|---|---|---|---|
| Meeting Type(s) | Core Group Meeting,Working Group Meeting,Workshops/Conferences | | |
| Grant Period Goal(s) it will address | Coordinating efforts towards better morphosyntactic corpus annotation, accounting for new languages, typology, constructions and spoken language (tasks 1.1, 1.3 and 1.5),Including new languages and new categories in multiword expression annotation (task 1.2),Sharing tools, formats and infrastructures for the annotation of morphosyntax and multiword expressions (task 1.4),Cross-language unification of the definition of a syntactic word, a lemma and lemmatisation rules (tasks 2.1 and 2.4),Coordinating the construction of a parallel word-sense annotated corpus and lexicon-corpus interface (task 2.2),Developing a proof-of-concept lexicon encoding for multiword expressions (task 2.3),Organising a multilingual shared task on morphosyntactic analyso-parsing (task 3.2),Defining the concepts of multilingual, cross-lingual, and translingual in the context of language technology (task 3.3),Structuring MWE resources regarding language resourcedness (task 4.1),Survey of diversity measures in NLP (task 4.2),Developing methods for measuring diversity of shared-tasks (task 4.3) | | |
| Description | 3rd General Meeting of the Action featuring invited talks, a workshop with posters, an parallel WG meetings with dicussions on the ongoing tasks 29-30 Jan 2025, Budapest, Hungary | | |
| Output(s) | Abstracts and posters from the workshop, slides from the invited talks, minutes from WG meetings and the Core Group meeting | | |
| Location | Budapest (Hungary) | ITC | Yes |
| Start Date | 2025-01-28 09:00:00 | End Date | 2025-01-30 17:30:00 |
| Duration | 3 days | Event Attendance Type | Face to face |
| Total number of expected participants | 130 | Number of participants expected to be reimbursed from COST funds | 118 |
| Daily allowance (EUR) | 140.00 | Average number of daily allowances per participant | 4 |
| Average Long-Distance Costs (EUR) | 400.00 | | |
| Average reimbursement (per participant) (EUR) | 960.00 | | |

| Total Travel, Accommodation and Subsistence Costs (EUR) | 113,280.00 | | |
|---|---|---|---|
| Total unique participants to be accounted for LOS grant | 130 | Local Organiser Support (EUR) | 10,000.00 |
| Total cost of the meeting (EUR) | 123,280.00 | | |

| Title of the Meeting | Management Committee meeting 6 | | |
|---|---|---|---|
| Meeting Type(s) | Management Committee Meeting | | |
| Grant Period Goal(s) it will address | Coordinating efforts towards better morphosyntactic corpus annotation, accounting for new languages, typology, constructions and spoken language (tasks 1.1, 1.3 and 1.5),Including new languages and new categories in multiword expression annotation (task 1.2),Sharing tools, formats and infrastructures for the annotation of morphosyntax and multiword expressions (task 1.4),Cross-language unification of the definition of a syntactic word, a lemma and lemmatisation rules (tasks 2.1 and 2.4),Coordinating the construction of a parallel word-sense annotated corpus and lexicon-corpus interface (task 2.2),Developing a proof-of-concept lexicon encoding for multiword expressions (task 2.3),Organising a multilingual shared task on morphosyntactic analyso-parsing (task 3.2),Defining the concepts of multilingual, cross-lingual, and translingual in the context of language technology (task 3.3),Structuring MWE resources regarding language resourcedness (task 4.1),Survey of diversity measures in NLP (task 4.2),Developing methods for measuring diversity of shared-tasks (task 4.3),Disseminating outcomes of the Action to stakeholders (task 4.1) | | |
| 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 | 2025-04-02 10:30:00 | End Date | 2025-04-02 12:30:00 |
| Duration | 1 day | Event Attendance Type | Online |
| Total number of expected participants | 70 | 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 | UniDive shared task on Morphosyntactic analyso-parsing |
|---|---|
| Meeting Type(s) | Workshops/Conferences |

| | |
|---|---|
| Grant Period Goal(s) it will address | Coordinating efforts towards better morphosyntactic corpus annotation, accounting for new languages, typology, constructions and spoken language (tasks 1.1, 1.3 and 1.5),Sharing tools, formats and infrastructures for the annotation of morphosyntax and multiword expressions (task 1.4),Organising a multilingual shared task on morphosyntactic analyso-parsing (task 3.2),Developing methods for measuring diversity of shared-tasks (task 4.3) |
| Description | Culminating event of the UniDive shared task on Morphosyntactic analyso-parsing, co-located with the SyntaxFest conference. The event will feature analysis of the shared task results, and presentations of the participating systems. |
| Output(s) | Proceedings from the workshop |

| | | | |
|---|---|---|---|
| Location | Ljubljana (Slovenia) | ITC | Yes |
| Start Date | 2025-08-26 09:00:00 | End Date | 2025-08-26 18:00:00 |
| Duration | 1 day | Event Attendance Type | Face to face |
| Total number of expected participants | 60 | Number of participants expected to be reimbursed from COST funds | 40 |
| 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) | 28,800.00 | | |
| Total unique participants to be accounted for LOS grant | 60 | Local Organiser Support (EUR) | 2,000.00 |
| Total cost of the meeting (EUR) | 30,800.00 | | |

| | |
|---|---|
| Title of the Meeting | Management Committee Meeting 7 |
| Meeting Type(s) | Management Committee Meeting |
| Grant Period Goal(s) it will address | Coordinating efforts towards better morphosyntactic corpus annotation, accounting for new languages, typology, constructions and spoken language (tasks 1.1, 1.3 and 1.5),Including new languages and new categories in multiword expression annotation (task 1.2),Sharing tools, formats and infrastructures for the annotation of morphosyntax and multiword expressions (task 1.4),Cross-language unification of the definition of a syntactic word, a lemma and lemmatisation rules (tasks 2.1 and 2.4),Coordinating the construction of a parallel word-sense annotated corpus and lexicon-corpus interface (task 2.2),Developing a proof-of-concept lexicon encoding for multiword expressions (task 2.3),Organising a multilingual shared task on morphosyntactic analyso-parsing (task 3.2),Defining the concepts of multilingual, cross-lingual, and translingual in the context of language technology (task 3.3),Structuring MWE resources regarding language resourcedness (task 4.1),Survey of diversity measures in NLP (task 4.2),Developing methods for measuring diversity of shared-tasks (task 4.3),Disseminating outcomes of the Action to stakeholders (task 4.1) |
| 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 | 2025-10-23 10:00:00 | End Date | 2025-10-23 12:00:00 |
| Duration | 1 day | Event Attendance Type | Online |

| | | | |
|---|---|---|---|
| Total number of expected participants | 70 | 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 | | |

**Mobility of Researchers and Innovators**

| | |
|---|---|
| Grant Period Goal(s) it will address: | Coordinating efforts towards better morphosyntactic corpus annotation, accounting for new languages, typology, constructions and spoken language (tasks 1.1, 1.3 and 1.5),Including new languages and new categories in multiword expression annotation (task 1.2),Sharing tools, formats and infrastructures for the annotation of morphosyntax and multiword expressions (task 1.4),Cross-language unification of the definition of a syntactic word, a lemma and lemmatisation rules (tasks 2.1 and 2.4),Coordinating the construction of a parallel word-sense annotated corpus and lexicon-corpus interface (task 2.2),Developing a proof-of-concept lexicon encoding for multiword expressions (task 2.3),Organising a multilingual shared task on morphosyntactic analyso-parsing (task 3.2),Defining the concepts of multilingual, cross-lingual, and translingual in the context of language technology (task 3.3),Structuring MWE resources regarding language resourcedness (task 4.1),Survey of diversity measures in NLP (task 4.2),Developing methods for measuring diversity of shared-tasks (task 4.3),Disseminating outcomes of the Action to stakeholders (task 4.1) |
| 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 | 40,000.00 |
|---|---|---|
| | This budget would allow for approx.19 STSM grants (based on the average costs of EUR 2100 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 of EUR 1300 per VM grant spent by all Actions). | |
| | Total | 40,000.00 |

**Presentations at Conferences organised by Third Parties**

| Description: | Major conferences in NLP and linguistics will be targeted. Only 2 such grants are foreseen, due to a low popularity of these grants in the past years. | |
|---|---|---|
| Budget (EUR) | Inclusiveness Target Country Conference grants | 2,820.00 |
| | This budget would allow for approx.2 Inclusiveness Target Country Conference grants (based on the average costs of EUR 1200 per Inclusiveness Target Country Conference grant spent by all Actions). | |
| | Dissemination Conference grants | 0.00 |
| | This budget would allow for approx.0 Dissemination Conference grants (based on the average costs of EUR 1250 per Dissemination Conference grant spent by all Actions). | |
| | Total | 2,820.00 |

**Other Expenses Related to Scientific Activities (OERSA)**

Overview

| Item | Total Cost (EUR) |
|------|------------------|
| Bank charges | 491.00 |
| Total | 491.00 |

Details

| Type: | Bank charges |
|-------|--------------|
| Total (EUR): (without VAT) | 491.00 |

**Other Expenses Related to Scientific Activities (OERSA)**