

# Working Group 4: Quantifying and promoting diversity

- **Leader:** Marie-Catherine de Marneffe (Belgium)
- **Vice-leader:** Olha Kanishcheva (Germany/Ukraine)

## Workplan

This WG is transversal to WGs 1-3 and will focus on how the Action serves inter- and intra-linguistic diversity. Its activities will overlap with the 3 other WGs in:

1. **Networking** for diversity:
  - bringing together pre-existing groups dedicated to NLP-applicable universality,
  - integrating experts of (notably low-resourced) languages not yet covered by these groups,
  - integrating experts in linguistic typology;
2. **Quantifying** diversity:
  - designing **measures of inter- and intra-linguistic diversity** in language resources and tools,
  - using these measures to quantify diversity in UD and PARSEME corpora;
3. Promoting diversity:
  - procedures for **better use of the existing resources**, based on their estimated diversity,
  - **selecting new data** to be annotated, so as to favour intra-linguistic diversity,
  - designing **evaluation scenarios** which favour tools performing well on rare and diverse phenomena and low resourced languages,
  - integrating and **training** new experts dedicated to low-resourced and endangered languages,
  - validating the unified annotation **guidelines** (WG1) and **lexicon formats** (WG2) against newly included languages and defining new language-specific categories and extensions, if needed,
  - coordinating the creation and enhancement of annotated **corpora** and **lexica** for low-resourced languages,
  - discovering and analysing **rare** linguistic **phenomena**, and describing them in resources and tools,
  - coordination of the development of NLP **tools** (WG3) for low-resourced and endangered languages.

## Current Subtasks

- **Task 4.1:** Promoting low-resourced/endangered languages [co-leaders: Lucia Amoros, Abigail Walsh]
- **Task 4.2:** Survey of diversity measures [leader: Louis Estève]

## Documents

- WG4 Meeting 1 [slides and discussion summary](#) - **16-17 March 2023**, Paris-Saclay University, France (co-located with the [UniDive 1st general meeting](#))
- WG4 online meetings [summaries](#): September 4 2023 (full day), September 29 2023 (1h30), November 10 2023 (1h30)
- WG4 Meeting 2 [summary](#) - **9 February 2024**, University of Naples L'Orientale, Italy (co-located with the [UniDive 2nd general meeting](#))
- [Digital Language Survival Kits](#) by the [DLDP project](#)

From:

<https://unidive.lisn.upsaclay.fr/> - **Universality, diversity and idiosyncrasy  
in language technology  
CA21167 COST Action**

Permanent link:

<https://unidive.lisn.upsaclay.fr/doku.php?id=wg4:wg4&rev=1712257710>

Last update: **2024/04/04 21:08**

