WG3 2nd Meeting Minutes – 2023-09-08

The second WG meeting was mainly devoted to the two tasks/deliverables outlined in the MoU, documentation of multilingual language tools and multilingual evaluation campaigns. In addition, we discussed ways of working in the WG.

Documentation of Multilingual Language Technology Tools

The morning session on documentation of multilingual tools was led by the task leaders A. Seza Doğruöz, Kristina Kocijan and Teresa Lynn and featured the following presentations:

- Introduction to European Language Grid (ELG) (Teresa) (slides)
- Survey on preferences concerning documentation platforms (Kristina) (slides)
- Overview of ELG (invited presentation by Maria Giagkou) (slides)
- Exploration of data in ELG (Kristina) (slides)
- Other platforms and future directions (Seza) (slides)

There was a lively discussion covering many topics but focusing on the choice of documentation platform and the scope of the documentation task. Here is a summary of the main points and some of the issues raised in relation to them:

- It makes sense to use on an existing platform because (a) we don't have the resources to develop our own platform (or maintain it after the end of the action) and (b) we don't need to start from scratch because many tools will already be documented.
- Of the available platforms, ELG seems to be the best fit for what we want to do, although all platforms have pros and cons. For example, while ELG harvests information other platforms and sources, it does not provide direct access to tools and resources like CLARIN.
- We need to define (and narrow) the scope of our documentation task along several dimensions, including the follows:
  - What is meant by multilinguality? Only strictly multilingual tools, or monolingual tools for many languages?
  - What languages do we include? ELG focuses on languages in Europe, but could easily be extended. In practice, we may be limited by the expertise available within the action. How do we deal with dialects and language variants?
  - How do we take users into account?
  - Do we restrict ourselves to tools or do we also include (other) resources?
  - Should we try to integrate typology resources? How?