Working Group 3: Multilingual and cross-lingual language technology

- **Leader**: Joakim Nivre (Sweden)
- **Vice-leader**: Gülşen Cebiroğlu Eryiğit (Turkey)

Workplan

Unified modelling helps solve NLP tasks with higher accuracy and better awareness of diversity. Therefore, this WG will be dedicated to NLP coordinating the development of tools leveraging universality and promoting diversity:

1. Multilingual and cross-lingual **syntactic parsers** which:
     - pay attention to hard and underrepresented phenomena (unbounded dependencies, MWEs,…),
     - leverage transfer of annotations or models in order to cope with data scarceness;
2. Prototypes of multilingual and cross-lingual **semantic parsers** which:
     - derive bi-lexical semantic dependencies from syntactic trees,
     - resolve idiosyncrasies in the syntax-semantics interface;
3. Multilingual **MWE discovery** tools which:
     - exploit large non-annotated data to compensate the sparseness of MWEs in annotated corpora,
     - are coupled both with lexicons and MWE identifiers;
4. Multilingual **MWE identifiers** which:
     - are coupled with MWE discovery and lexica to better handle unseen data,
     - pay attention to underrepresented phenomena, e.g., discontinuity/variability of MWEs;
5. Prototypes of tools for automatic identification of idiosyncratic **constructions**.

The tools themselves will be funded at the national level. WG3 will bring the federating effect to these activities, notably by organizing multilingual **evaluation campaigns** on parsing and MWE identification. Diversity-based evaluation measures from WG4 will be promoted. The outcomes should validate the computational tractability of the terminologies unified in WG1.

Documents

wg3:test