

**COST Action CA21167: Universality, diversity and idiosyncrasy in language technology (UniDive)  
Permanent Call for STSM Hosting Institutions**

Name and address of the hosting institution

CIRCSE Research Centre

Università Cattolica del Sacro Cuore

Largo Gemelli, 1

20123, Milan, Italy

<https://centridiricerca.unicatt.it/circse-home?rdeLocaleAttr=en>

Contact information of the host

Prof. Marco Carlo Passarotti

[marco.passarotti@unicatt.it](mailto:marco.passarotti@unicatt.it)

+39-02-72342380

Approximate duration and dates of the STSM:

- From 5 calendar days to 3 months
- All the yearlong, except on August and December. Lesson semesters go from early October to early December, and from late February to late May

Research topic, including explanation of its importance and the potential contribution of the STSM to the action's objectives

The research work pursued at the CIRCSE Research Centre accounts for and fosters language diversity both on the theoretical side (by discussing annotation issues in the light of assumptions of theoretical linguistics), and on the practical side (by building Universal Dependencies treebanks and by making resources interoperable using language-independent vocabularies and data categories).

In particular:

- *Language Resources*: the CIRCSE has expertise in building and distributing both textual and lexical resources. As for the former, the research topics of the Centre focus on dependency treebanks for the Latin language (*Index Thomisticus* Treebank, UDante, LLCT), all available in the Universal Dependencies collection. As for the latter, the CIRCSE developed two valency lexica (IT-VaLex and Latin Vallex) and a derivational lexicon for Latin (Word Formation Latin). Furthermore, the CIRCSE hosts the LiLa: Linking Latin ERC Consolidator Grant project (<http://lila-erc.eu/>), which built a Linked Open Data knowledge base of interoperable resources for Latin
- *Natural Language Processing*: the CIRCSE is a world leader in Natural Language Processing for Latin. Together with Rachele Sprugnoli (University of Parma, Italy) organizes the biannual evaluation campaign of NLP tools for Latin EvaLatin (<https://circse.github.io/LT4HALA/2022/EvaLatin.html>)
- *Language Universals and Idiosyncrasy*: the members of the CIRCSE team (in particular, Marco C. Passarotti and Flavio M. Cecchini) are active in the Universal Dependencies community, particularly focusing on issues related to the annotation guidelines of the treebanks

We would be particularly happy to host both scholars with a background in theoretical linguistics applied to linguistic resources and scholars interested in multi-lingual Natural Language Processing. However, scholars with other backgrounds and interests are welcome, as well.

The CIRCSE team includes more than 10 members and covers a wide range of expertise, including Computational Linguistics, Latin Linguistics, Natural Language Processing, Classical Philology and Humanities Computing.