

CROSS-LINGUAL LINK DISCOVERY FOR UNDER-RESOURCED LANGUAGES

LREC 2022, MARSEILLES



**TASK 1.3 CROSS LINGUAL
INTERLINKING**

M. Rosner, S. Ahmadi, E. Apostol, J. Bosque-Gil,
C. Chiarcos, M. Dojchinovski, K. Gkirtzou,
J. Gracia, D. Gromann, C. Liebeskind, G. Valunaite
Oleškevičienė, G. Sérasset, C. Octavian Truica



ABSTRACT:

In this paper, we provide an overview of current technologies for cross-lingual link discovery, and discuss challenges, experiences and prospects of their application to under-resourced languages. We first introduce the goals of cross-lingual linking and associated technologies, and in particular, the role that the Linked Data paradigm applied to language data can play in this context. We define under-resourced languages with a specific focus on languages actively used on the internet, i.e., languages with a digitally versatile speaker community, but limited support in terms of language technology. We argue that languages for which considerable amounts of textual data and (at least) a bilingual word list are available, techniques for cross-lingual linking can be readily applied, and that these enable the implementation of downstream applications

This work is supported by COST Action NexusLinguarum (CA18209), supported by COST (European Cooperation in Science and Technology) (www.cost.eu) and partially supported by the I+D+i project PID2020-113903RB-I00, funded by MCIN/AEI/10.13039/501100011033, by DGA/FEDER, and by the Agencia Estatal de Investigación of the Spanish Ministry of Economy and Competitiveness and the European Social Fund through the Ramón y Cajal programme (RYC2019-028112-I).