EMNLP 2021 | Sixth Conference on Machine Translation (WMT21)

SHARED TASK
European Low-Resource Multilingual Translation

MULTILINGUAL SETTING

★★★
HIGH-QUALITY TEST SETS

⇄
INTRA-FAMILY LANGUAGE COMBINATIONS

Evaluation Period: 29 June to 6 July 2021

Dear all,

we are happy to invite you to participate in our **WMT21 shared task** on **European low-resource multilingual translation**, which will focus on **multilinguality** in the **cultural heritage** domain for two Indo-European language families, i.e. **North-Germanic** and **Romance**.

Massively multilingual machine translation has shown impressive capabilities, including zero and few-shot translation of low-resource languages. However, these models are often evaluated (and trained) from or into English, where the most data is available, and assuming that models generalise to other pairs and low-resource languages.

With this shared task, we want to explore how information in one language can be transferred to other related languages by evaluating translation quality in low-resourced language pairs, but explicitly encourage the use of data of the high-resourced language pairs in the same family. In doing so, we want to shed some light on the question if English and/or Spanish are required for high-quality translation of their related languages. And if this holds true, we want to identify the best ways of combining the data. The shared task will be divided into two subtasks: **Europeana thesis abstracts** translation (**North-Germanic languages from/to Icelandic, Norwegian Bokmål** and **Swedish**) and **Wikipedia cultural heritage articles** translation (**Romance languages from Catalan to Occitan, Romanian** and **Italian**). It will be organised by DFKI in cooperation with ELRC (SMART 2019/1083) and LT-Bridge (H2020, 952194) and has been supported by the Directorate-General for Language Policy, Ministry of Culture, Government of Catalonia. Further information is also provided here.

Are you up to this challenge?

Join the WMT google group and/or send an email to cristinae@dfki.de. Please feel free to share this information with anyone who might be interested in participating!

All the best,

The project teams of ELRC & LT-Bridge