



LT-OBSERVATORY

## ASSESSING THE OPERATIONAL USABILITY OF LANGUAGE RESOURCES FOR THE PURPOSE OF MACHINE TRANSLATION

Venue:

**The International Auditorium**  
International Trade Union House  
Boulevard du Roi Albert II, No. 5 / 2  
B-1210 Brussels

**Date:** 4 December 2015  
**12.00-16.00 Hrs.**

### 1 BACKGROUND

The LT Observatory initiative ([www.lt-observatory.eu](http://www.lt-observatory.eu)) derives from the need expressed by enterprises and part of the research community to have better access to more **usable language resources** (LRs), to agree on a methodology to benchmark the operational usability of language resources and to provide for a mechanism to evaluate and rate them online.

Language resources are indispensable for the development of tools for machine translation (MT) or various kinds of computer-assisted translation (CAT). In particular language corpora, both parallel and monolingual are considered most important for MT/CAT. Numerous EU and national projects have created LR, which LR repositories such as META-SHARE or CLARIN-ERIC have attempted to pool. But information about the most useful/relevant LR remains scarce. In order to remedy this lack of overview for the professional user, it is important to **apply a user-driven approach towards the identification/mapping of best practices in terms of collecting relevant LR**. The LT Observatory has organised a first workshop involving LR stakeholders in June 2015 in Brussels and a first Workshop in July in Vienna.

Subsequently, the LT Observatory has started assembling information about LR deemed to be useful for MT in a dedicated LR Catalogue ([www.lt-innovate.org/lt-observe/resources-list](http://www.lt-innovate.org/lt-observe/resources-list)). In addition, a preliminary study has been carried out in order to determine which kind of LR are particularly useful for MT practitioners in an operational context and what the most relevant metadata and quality/usability criteria should be. The findings of the study are summarized in a Discussion Paper (<http://tiny.cc/6n6p5x>).

### 2 COMMITTEE OF EXPERTS

The LT Observatory has also convened a committee of experts who will participate in the initial assessment of the LR Catalogue. The committee will review and rate the first batch of LR compiled in the LR Catalogue during the second part of November 2015.



### 3 PURPOSE OF WORKSHOP

The purpose of this workshop is to convene the experts who have participated in the above mentioned initial LR “review & rating” process for a discussion in view of defining "best in class" LRs, objective usability criteria in an operational context and a benchmarking methodology against which the “operational usability” of LRs can be measured.

### 4 PROPOSED AGENDA

- 12.00 Welcome, summary of Workshop objectives and presentation of the LT Observatory’s LR Catalogue
- 12.30 Presentation by experts of individual feedback following the review & rating process of the first batch of LRs assembled in the LR Catalogue
- 13.15 Discussion:
- Assessment of the relevance of the LRs reviewed (in an operational MT context)
  - Relevance/completeness of the metadata collected in the LR Catalogue – improvements to be made to the Catalogue
  - Formulation of “operational usability” criteria
  - LR benchmarking methodology
  - Next steps for the improvement of the LR Catalogue
- 15.30 Conclusions
- 16.00 End of the Workshop

A light sandwich lunch will be served.

