

# PEDAL: Post-Editing with Dynamic Active Learning

Science Foundation Ireland  
Technology Development Innovation Award (TIDA) Feasibility Study  
Grant: 12/TIDA/I2438

## List of partners



Dublin City University, Ireland



CNGL Centre for Global Intelligent Content, Ireland

**Project duration: July 2013 — June 2014**

## Summary

Machine translation, in particular statistical machine translation (SMT), is making big inroads into the localisation and translation industry. In typical workflows (S)MT output is checked and (where required) manually post-edited by human translators. Recently, a significant amount of research has concentrated on capturing human post-editing outputs as early as possible to incrementally update/modify SMT models to avoid repeat mistakes. Typically in these approaches, MT and post-edits happen sequentially and chronologically, following the way unseen data (the translation job) is presented. In this project, we add to the existing literature addressing the question whether, and if so, to what extent, this process can be improved upon by Active Learning, where input is not presented chronologically but dynamically selected according to criteria that maximise performance with respect to (whatever is) the remaining data. The criteria we use are novel and allow the **MT system to improve its performance earlier**. Because these criteria are computationally cheap and language independent, our technology, together with incremental retraining, can be **easily integrated into the industry workflows**.