

TransCenter: Web-Based Translation Research Suite

Michael Denkowski and Alon Lavie

www.cs.cmu.edu/~mdenkows/transcenter

Overview

TransCenter is a software suite for translation research. Deploying a TransCenter server allows translators to log into a web-based translation environment to complete translation, post-editing, monolingual editing, and evaluation tasks from any computer with an Internet connection. Extensive user tracking facilitates detailed analysis of translator activity, including precise measurement of the amount of time and effort required to translate or edit each sentence. TransCenter is free, open source, and platform independent software.

Motivating Applications

In addition to using human translations as training data for machine translation systems, the practice of using MT to assist human translators has recently gained attention in both academia and industry. Rather than translating from scratch, translators correct initial MT output through the process of post-editing. While shared evaluations have examined the amount of work required to post-edit MT output (Olive et al., 2011; Callison-Burch et al., 2012), the type of data collected is somewhat limited. Typically, only final translations, either directly from source sentences or post-edited from MT output, are collected. In some cases, additional translators provide assessments of how much work appears to have been done to edit each sentence. Although the goal of incorporating MT into translator workflows centers on reducing translation time and effort as perceived by the actual translator, this information is entirely absent when only final translations are collected. The need for more detailed information has led some research groups to design their own in-house post-editing tools (Aziz et al., 2012).

TransCenter Software

TransCenter provides an easy-to-deploy, easy-to-use, translation analysis suite. The browser-based editor interface displays source sentences and translations in a two-column format that emphasizes clarity and simplicity. Fast keyboard navigation and automatic highlighting provide a smooth user experience. During editing, all user actions (key presses and mouse clicks) are logged so that the full translation or post-editing process can be replayed and analyzed. For post-editing tasks, translators are asked to rate the amount of work required to post-edit each sentence immediately after completing it, yielding maximally accurate feedback. TransCenter also records the number of seconds each sentence is focused, allowing for exact timing measurements. A pause button is available if translators need to take breaks. TransCenter can generate reports of translation and post-editing effort as measured by (1) keystroke, (2) exact timing, (3) actual translator post-assessment. Final translations are also available for calculating edit distance. TransCenter server is written in Python and can be run on Linux, Microsoft Windows, and Mac OS. The web-based interface supports major browsers including Google Chrome, Mozilla Firefox, and Microsoft Internet Explorer. All data is stored in open formats for maximum interoperability.

References

- Wilker Aziz, Sheila Castilho Monteiro de Sousa, and Lucia Specia. 2012. PET: a Tool for Post-editing and Assessing Machine Translation. In *Proc. of LREC 2012*.
- Chris Callison-Burch, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia. 2012. Findings of the 2012 Workshop on Statistical Machine Translation. In *Proc. of NAACL WMT 2012*.
- Joseph Olive, Caitlin Christianson, and John McCary. 2011. *Handbook of Natural Language Processing and Machine Translation: Darpa Global Autonomous Language Exploitation*.