

Helping communities share knowledge more effectively

across the language barrier

Goals

- Develop a user-friendly (“minimally intrusive”) strategy for pre-editing User-Generated Content for Statistical Machine Translation.
- Address the post-editing bottleneck by building volunteer communities of bilingual and monolingual subject-matter experts as post-editors.
- Improve learning and develop feedback loops to improve SMT results for community data.
- Improve reliability of forum content translation using Text Analytics.

WP2 Definition of Rules

- Improve source linguistic resources for SMT performance.
- Improve and develop new **pre-editing** and **post-editing** rules for forum text and NGO materials .

WP4 Improving SMT

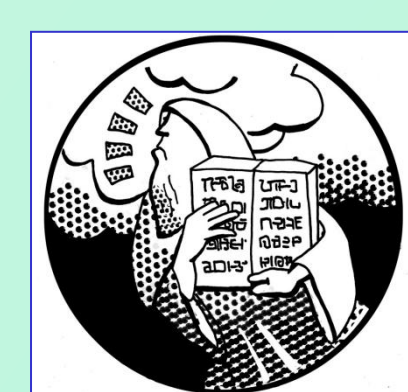
- Develop novel methods for SMT.
- Train baseline MOSES engines.
- Use in-domain & supplementary out-of-domain training data.
- Exploit user edits in later phases.

WP7 & WP8 Post-editing

- Establish how users perform when post-editing MT output.
- Conduct user studies with different set-ups.
- Compare monolingual to bilingual performance.

WP5 Portal Integration

- Develop software clients for monolingual and bilingual checking and editing.
- Develop API-based portal that facilitates source and target evaluation.

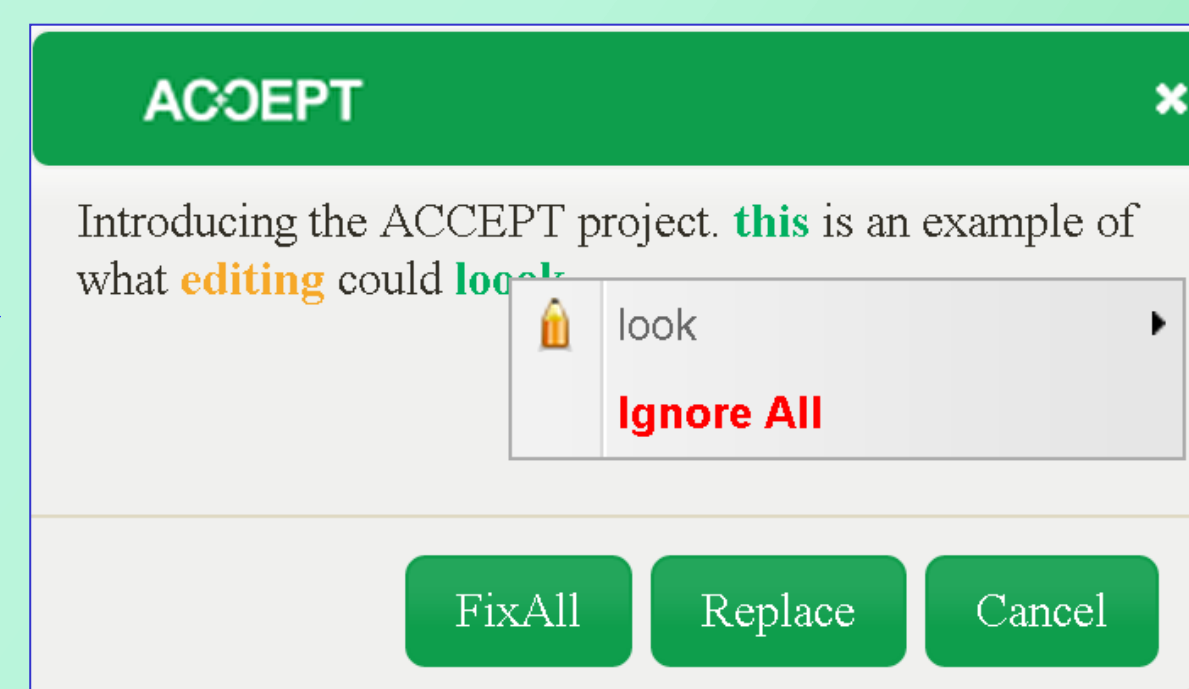
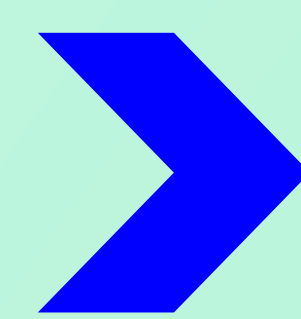


MOSES

+

```
avoid_word_with_slash
disambiguate_ing_words
disambiguate_that
```

rules



WP6 Community Development

- Build a community for post-editing.
- “Train” community members on the tool set.
- Collect data on membership, throughput, motivation and sentiment.

WP10 Dissemination

- Recruit organisations to a Special Interest Group (SIG)
- Ensure visibility of project results and discussions.
- Drive interest on social media.
- Develop audio visual demos for commercial and NGO use cases.

accept-project.eu

WP9 MT Evaluation

- Evaluate the impact of monolingual and bilingual pre- and post-editing rules, on output quality and ease of use.
 - Measure source and MT quality.
- BLEU
Meteor
TER
GTM

VS.



WP3 Text Analytics

- Develop taxonomy of forum content & NGO content.
- Develop rules for automatic **classification** and **sentiment analysis**.

a happy bit of feedback

01-15-2012 11:35 PM

my use of this forum interactive

thanks heaps

"Great forum"! Feedback :-)

12-15-2009 06:42 AM

Description

We hope to provide a mechanism where authors can check and edit their source content with a view to improving MT quality. Additionally, we seek to offer guidance to improve sources. Thus, post-editing can be performed by monolingual domain experts, rather than by the traditional route of pressing a translator to do post-editing in a domain where they are not power users.

The project materials are being drawn from Symantec’s security focussed Norton forum and from a range of NGOs, working in medical and nutrition fields, through Translators Without Borders. In this way, we hope to address a wide divergence of content types and communities. In the coming months we will engage with a SIG which will further inform the project with a variety of content and writing styles.