

Moses from the point of view of an LSP: The Trusted Translations Experience

TAUS Open Source Machine Translation
Pre Conference Workshops
GALA MONACO 2012
Sunday 25 March 13:30 - 17:30

Gustavo Lucardi
COO Trusted Translations, Inc.
@glucardi



An On-going Story

- Not a Success Story (Yet)
- An On-going Story (Since October 2010)
- Moses from the Point of View of an LSP:
 - *Why Moses? Why Now?*
 - What exactly are we doing?
 - Our Expectations (What we are looking for)
 - Unexpected Experiences and Issues
 - Next Steps
 - Our Conclusions (Till Now)

Why Moses? Why Now?

- Because we have to
 - In Friedman's *The World is Flat*, the Refrigeration Industry was transformed with the advent of the electric refrigerator killing the Ice Block industry. Transform or die.
 - Every LSP has to start learn about MT or prepare to suffer the same fate as the Ice Block makers.
- We believe that the future of the professional translation industry will include MT as a part (central?) of the process.
- To expose all the members of our company to this new nascent model and its consequences in all the areas of the business cycle from sales to production, administration, in- house resources, freelancers, and of course, IT.

What exactly are we doing?

- We built two Moses engines and are running tests:
 - Legal English to Spanish (based on legal specialization of *many clients*)
 - Technical English to Spanish (based on technical experience from *one specific client*)
- We are also testing other MT solutions like:
 - Pangea MT (Pangeanic)
 - DoMY (Translation Precision Tools)
 - Language Studio (Asia Online)
 - Smart MATE (Applied Languages)
 - LetsMT
 - Specially Manuel Herranz from Pangeanic helped us understand MT

Our Expectations (What we are looking for)

- To be able to deliver higher volumes in less time while maintaining quality and consistency (Quality and Time, not Price)
- The price of doing MT+PE+E+P is similar to the price of T+P (i.e.: $PE+E+P \sim T+P = T*0.6+T*0,33+T*0.16 = T+T*0.16 = 1.09 T \sim 1.16 T$)
- Introduce MT culture and concepts within the entire company
- Be ready for the *New New* thing in the translation industry, which of course is MT (as CAT tools were the *New New* thing 20 years ago)
- Have working knowledge and hands-on experience to be able to choose the best MT solution for each particular project (Language Pair, Area of Specialization, Volume)

Unexpected Experiences and Issues

- The lack of information and documented procedures on how post-editing should be done (We only found one Paper of TAUS)
- The difficulty of measuring human post-editing quality and productivity
- The difficulty to measure the quality of an Engine utilizing BLEU (For example an Engine with a BLEU of 70 can have a worse output than Engines with a BLEU of 40)
- The difficulty of aligning existing Parallel DATA without human intervention
- The fear of translators to be displaced by MT

Next steps

- Scripts to Train (Automate through scripts the pre-processing of DATA to train Moses engines: Take out Placeables like Numbers, Acronyms, XML Tags, E-mail addresses and URLs / Remove Short Segments and Long Segments / Spell check on Source and Target / Remove segments where target and source have different structures)
- Scripts to Translate (Automate through scripts the pre and post-processing for translations with Moses engines: Go from TMX to TXT and from TXT to TMX)
- Plug-in for Moses interaction with CAT TOOLS (Create or find PLUG-INS to use Moses directly through CAT TOOLS, like the ones for Okapi or Globalsight. This could allow us to avoid to go to and from TMX and TXT. Plug-ins for Moses like the plug-ins for Trados to connect with FreeTranslation, Systran and others)
- Moses GUI (Create or find a user interface to allow the less tech savvy to interact directly with Moses)
- Keep testing more mature MT implementations

Our Conclusions (Till Now)

- Moses has great potential, but it is not yet optimized for LSPs
- Its interface is not user-friendly (in fact there is no GUI yet)
- There are no reliable sources of information for certain niches (a lot of work is still needed to obtain a reliable corpus in some niches with the tools that exist at the moment)
- We still have to wait for the development of the new features that LSPs require (glossaries, blacklist words, labels)
- It is necessary to optimize the internal processes of the LSPs to adapt to the needs of Moses (Management of Memories, product names and trademarks, languages)
- It's better to work with an engine for a specific client, than with one for a specific industry or area of specialization

Thank you!

Gustavo Lucardi

@glucardi

glucardi@trustedtranslations.com

<http://www.trustedtranslations.com>

<http://translation-blog.trustedtranslations.com>

