- [Home](#)
- [About TAUS Data](#)
- [Join TAUS Data](#)
- [Contact TAUS Data](#)

[RSS Feed](#)

TAUS Data Association Blog

Helping the world communicate better

You are reading

- October 7, 2010
- by Andrew Joscelyne
- In [Our Mission](#)
- One comment

TDA Members doing business with Moses

The translation industry is steadily appropriating the [Moses translation engine](#), an open source system available as a kit on the web. At the TAUS User Conference 2010 in Portland (Oregon) TDA members from major corporations and service vendors gave first-time accounts to the community of their experience with this MT engine. Here are the highlights.

Jeff Rueppel from **Adobe** got started with Moses using one engineer and off the shelf computing resources. He needed just over a week to move from zero to a good understanding of how the system works. He then moved through three phases to rollout.

Firstly, he had to "tame" his corpus. He had a corpus of 24 million words in 36 languages in large files that pushed the limits of desktop machines. These had to be converted into the Moses format using a special tool, and then tokenized and cleaned up in various ways. This involved serious but not difficult work.

Second came the training phase, which anyone with a basic grasp of Unix can do by following the information available inside the Moses kit. It took him 2 hours training and 12 hours of tuning. Lastly came the real deployment, which involved decisions about how you organize your servers, and security and redundancy processes. And lastly you have to evaluate the output, for which Moses provides no special software.

The bottom line is that users can build a Moses engine that works, but the system does not come as a complete package: You will especially need to develop or borrow tools to clean and prepare your corpus, evaluate your results, and maintain your system.

Elia Yuste from the LSP **Pangeanic** and **Salomé Lopez-Lavado** from **Sony** reported on their successful deployment of a Moses solution integrated into a TM and postediting workflow accessible online. They use TDA data to help train the engines. They use TMX and XLIFF standards throughout. Their main challenge was to see how the system handled the heavily formatted and marked-up content used in Sony documentation. This required the development of new tag-handling modules by Pangeanic. End users at Sony can now upload a file for translation, select a domain and a special tag processing option, and receive the translated file back within minutes.

Serge Gladkoff from **Logrus** walked the conference through his Moses rollout. Starting from scratch with no experience of the system, it took him 360 person hours to achieve full implementation, using a good programmer and a "skeptical" project manager. They use TDA data to help train the engines. He found that the advantages of Moses are that it is free and delivers pretty good quality right from scratch. The disadvantage is the "last mile" – users have to put quite a bit of effort into making it truly productive.
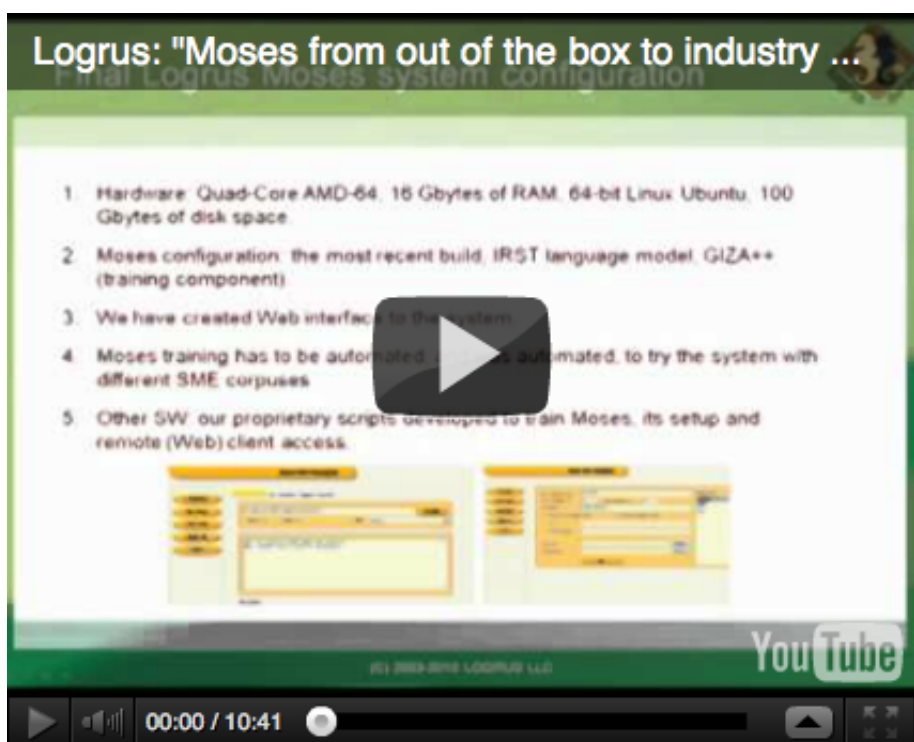
Anna Simpkins from **Applied Language Solutions** reported on a number of case studies in which Moses was used with clients. In one pilot, the sample used was too small to get good results. In another, they needed a lot of file engineering due to the image-heavy content. Overall, LSPs using Moses need to educate their clients and closely manage end-user expectations about MT output. And more guidelines

will be needed by the Moses user community to ensure a true quality guarantee.

Mirko Plitt from **Autodesk** reckons that there is a community opportunity to measure post editor productivity in statistical machine translation output. He has found that post editors typically overestimate how long they take to repair translations. And when the output is particularly poor, there is not much productivity loss. So he would like to see more shared insight into exactly what happens in the post-editing process, so that the effort devoted to measuring increases in productivity and the usability of MT output can be clarified and simplified.

In the general discussion, most Moses users said they tend to test their raw output against Google or Bing Translate output, but the results naturally varied with the type of content. Nearly all are now using Moses in production, and in many cases they use TDA data to help train the engines. Moses makes most sense when used for very high volumes of throughput. ROI (the investment is mostly labor) can be obtained in about two weeks.

Comments for this entry

Machine Translation using Moses | Pineapple Donut

[...] have recently posted a list of the highlights from a recent conference discussion about businesses that use it. As noted in the first account (from Adobe, no less), Moses doesn't take that many resources [...]

Categories

- Our Mission
- Your Technology
- Challenges
- User Experiences

Share

in f ⚏ ✉ ⭐ 🖨 | ➕ More

TAUS Search

TAUS Search enables translators, developers, support professionals, and in fact everyone to find and compare sector specific terminology by looking into the live translations for an industry. Go to Search.

Connect, Communicate, and Collaborate

in   Join the TAUS group on LinkedIn.

📶   Subscribe to TAUS Articles in a reader.

📶   Subscribe to TAUS Reports in a reader.

Recently on Twitter

-

- # Member Blogs

    - Applied Language Solutions Translation Blog
    - Carmazzi Global Solutions Blog
    - Cisco – The Platform
    - Dell Community
    - eBay Connect
    - Jensen Localization Blog
    - L10NBRIDGE Translation Blog
    - Microsoft Translator (and Bing Translator) Official Team Blog
    - Oracle Blog Center
    - Pangeanic Translation Technologies & News
    - SDL Blog
    - Smith's Blog
    - Sybase Blog Center
    - Tilde Machine Translation blog
    - Uppsala University
    - Xerox Blogs

- # Useful Links

    - TAUS
    - TAUS Data Association

- # Tags

animation APi armenian automated translation Challenges communication contributors desktop calendar dutch french greek issues italian language business innovation leveraging machine translation membership moses open source open translation platforms portland portuguese russian services sharing spanish TAUS taus data TAUS Search tda technology TM sharing TMX transcribe translate translation translation automation translation industry translation matching translation memories translation memory translators user conference 2010 wallpaper XLIFF

RSS Feed. This blog is proudly powered by Wordpress and uses Modern Clix, a theme by Rodrigo Galindez.