



# Tilde Language Technologies

**BIG solutions for small languages**

Indra Sāmīte



 **TAUS**  
[www.translationautomation.com](http://www.translationautomation.com)



# Tilde

- Largest language technology company in the Baltics
- Long-standing experience in language technologies and localization
- Language tool products for the local market + significant multi-national research and development activities.

# The Challenge

Languages of 10 mil speakers or less face the following hurdles in MT:

- Lack of parallel data
- Complex language structure
- Under developed terminology



# Example:

- 800 000 words in English = 20 000 000 words/word forms in LV & LT
- Need MORE parallel corpora, but there is less



# Proof of Concept

## **TAUS challenge/Adobe**

*Build a successful MT engine in 24 hours*

## **Language Shore/MS**

*The BEST LV>EN MT engine in the WORLD*



## Translator

[Home](#) | [Tools](#) | [Help](#)

Languages

Latvian



English

[Translate](#)

[Clear All](#)

*In partnership with Tilde*

### Enter text or webpage URL

Piektdien, 24. septembrī Eiropas Zinātnieku Nakts 2010 ietvaros Latvijas Nacionālā bibliotēka (LNB) rīko vairākus pasākumus tās centrālajā ēkā K. Barona ielā 14. Plkst. 16.00 paredzēta informatīva ekskursija vidusskolēniem un visiem, kuri padziļināti vēlas iepazīties ar prāta pētniecības vēsturi, tagadni un nākotnes tendencēm. Savukārt plkst. 17.30 Jaunākās periodikas lasītavā, 1. stāvā notiks diskusija par prāta pētniecību, galvenokārt akcentējot mūsdienu teorijas un eksperimentus.

Friday, 24. in September the (LNB) hosts a number of me informative tour of high school research history, present and room, 1. the discussion will t emphasizing modern theory a

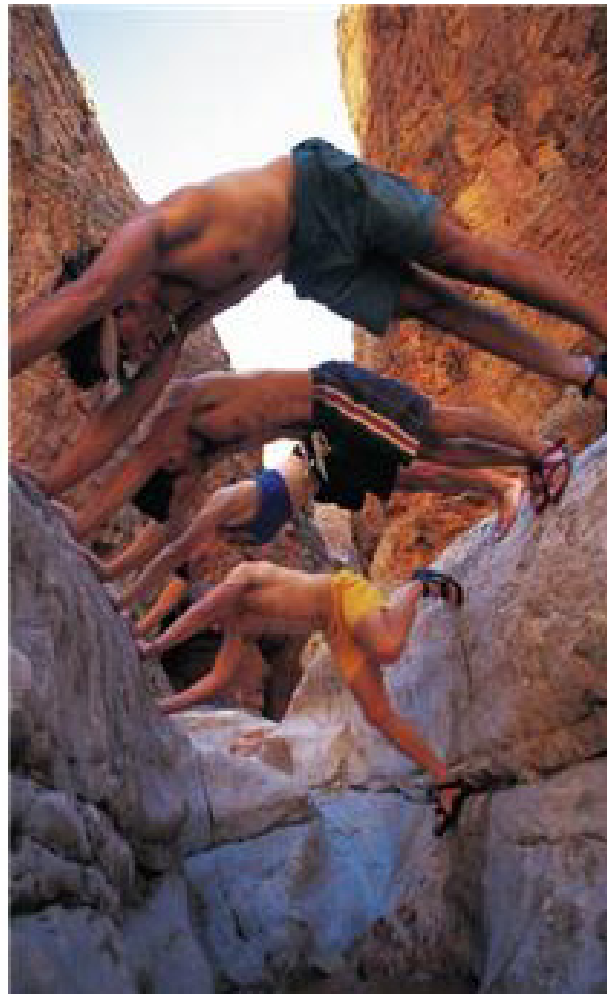
Rate this translation: ☆☆☆☆

[New: Translator V2 API & Widget announced at MIX 2010](#)

Powered by **Microsoft** Translator

# EU Agenda

Eliminating  
the digital  
divide to  
provide equal  
access to  
information



 **TAUS**  
[www.translationautomation.com](http://www.translationautomation.com)



# Multi disciplinary approach:

- Parallel corpora boosted with comparable corpora
- Rules and statistical hybrid
- Automated terminology mining





# Collaborate to Innovate



**Data**



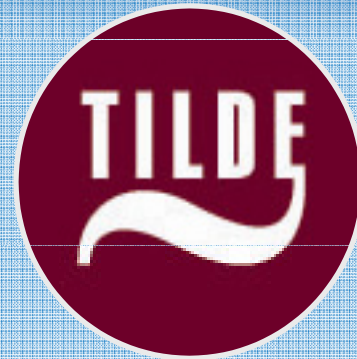
**Comparable  
Corpora**



**Terminology  
mining**



**Platform**



**Specialist  
knowledge**

**DIY MT in the cloud**





**TAUS**  
www.translationautomation.com



# Quo vadis?

- Expand language set
- Application scenarios
  - Integration into Localization process
  - CAT tools
- Collaboration
  - Resources + Research



