BIG solutions for small languages
Indra Sāmīte
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.
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
Friday, 24. in September the (LNB) hosts a number of mea informative tour of high scho research history, present and room, 1. the discussion will t: emphasizing modern theory a
EU Agenda

Eliminating the digital divide to provide equal access to information
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
Quo vadis?

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