Translation in the 21st Century
Tower of Babel

Divide & Unite

Language as Instinct

Translation is very difficult …
October 2004

Reality Check

… TAUS founded.
Let a Thousand MT Systems Bloom!

Effectiveness of Data
Profit from sharing
(“Pirate’s Dilemma”)
MT is here to stay

Translation as a Utility

... as a Human Right

Five years ahead ...
Technology in 5 Years

Hybrid Systems

Targeted Correction
(communities, games)

Real-time Training
TMIs in 5 Years

Cleaning
Semantic Clustering
Corpus Linguistics
Preserve Endangered Languages

Exciting new perspectives …
Profession in 5 Years

Choices

End to Repetitive Tasks

Productivity 5 to 10 Times Higher

Supplemented by Non-Professional Volunteers
Applications in 5 Years

Translation out of the Wall
Virtuous Circle
Spoken Translation
Enterprises in 5 Years

Need a Language Strategy *not just reducing word rates*

“Dual Linguaspheres”
Industry in 5 Years
Thinking about drivers/trends

From TAUS Copenhagen Forum (May 2010)

**Certain**
- Explosion in new content
- Shift from text to text and multi-media (word counts go down)
- Mobile user, hand held devices
- Real time/Just in time demand
- Cross-lingual translation challenges
- Balance of cost, timeliness and quality

**Uncertain**
- Open (collaborative) vs Closed (competitive)?
- Fee vs free?
- Human vs Machine? (incremental step or technology breakthrough)
Industry in 5 Years

SWOT
Content disruption
Innovation dilemma
Embedding technology

Machines
Open (Collaborative)
Closed (Competitive)
Human & Machine
## SWOT for Enterprise Language Service

<table>
<thead>
<tr>
<th><strong>S</strong></th>
<th></th>
<th><strong>W</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• High leverage from TM</td>
<td></td>
<td>• Quality inconsistent (local flavor missing)</td>
</tr>
<tr>
<td>• Well established process and management</td>
<td></td>
<td>• Lack of flexibility, reactive rather than creative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>O</strong></th>
<th></th>
<th><strong>T</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Opening new markets with MT</td>
<td></td>
<td>• Rigid landscape (vendor lock-in)</td>
</tr>
<tr>
<td>• Engaging with users &amp; communities</td>
<td></td>
<td>• Not scalable to expand quickly</td>
</tr>
<tr>
<td>• Convergence with video and speech</td>
<td></td>
<td>• Inability to ensure quality in new markets</td>
</tr>
<tr>
<td>• Search engine optimization</td>
<td></td>
<td>• Lack of corporate awareness of new locales</td>
</tr>
<tr>
<td>• Translation of user generated content</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Content Disruption

Localization industry

"Battle for words"

Social media

New technologies and solutions

User generated content

Support

Knowledge Base

Web

UI

Sales
Innovation Dilemma

High leverage from TM
Well established process and management
Quality inconsistent (local flavor missing)
Lack of flexibility (reactive, rather than creative)
Opening new markets with MT
Community/user feedback
Convergence with video and speech
Search engine optimization
Translation of user generated content
Rigid landscape (vendor lock-in)
Not scalable to quickly support new markets
Inability to ensure quality in new markets
Lack of corporate awareness of new locales
Innovation Dilemma

**S**
- High leverage from TM
- Well established process and management

**W**
- Quality inconsistent (local flavor missing)
- Lack of flexibility (reactive, rather than creative)

**O**
- Opening new markets with MT
- Community/user feedback
- Convergence with video and speech
- Search engine optimization
- Translation of user generated content

**T**
- Rigid landscape (vendor lock-in)
- Not scalable to quickly support new markets
- Inability to ensure quality in new markets
- Lack of corporate awareness of new locales
## Business Model Attributes

### Old Model

1. One translation fits all
2. Selecting locales
3. Counting words
4. TM is core
5. Project based
6. Cascaded supply chain
7. Publisher driven
8. One directional

### New Model

1. Quality differentiation
2. Long tail of languages
3. Unlimited content
4. Data is core
5. Continuous translation
6. Collaborative translation
7. User driven
8. Multi-directional
Vision Statement

- We envision translation as a standard feature, a ubiquitous service. Like the internet, electricity, and water, translation is one of the basic needs of human civilization.
20th Century Translation

Top-down globalization
Export mentality – pushing out

1. One translation quality fits all
2. Selecting locales – limited languages
3. Counting words – owned content
4. TM is core
5. Project-based translation
6. Cascaded supply chain
7. Publisher-driven
8. One directional

One big world
21st Century Translation

Bottom-up and top-down globalization
Information is omnipresent – people are connecting

1. Quality differentiation
2. Long-tail – unlimited languages
3. Unlimited content – owned, shared, earned
4. Data is core
5. Continuous translation
6. Collaborative translation
7. User-driven
8. Multi directional

Many big worlds in one small planet
Strategic Direction

Machines

Translation as utility

Closed (Competitive)

- 7. Cascaded supply chain
- 6. One-directional
- 5. Project based
- 4. TM is core
- 3. Counting words
- 2. Selecting locales
- 1. One translation quality fits all

Open (Collaborative)

- 7. Collaborative translation
- 6. Multi-directional
- 5. Continuous translation
- 4. Data is core
- 3. Unlimited content
- 2. Long-tail of languages
- 1. Quality differentiation.

2011 2012 2013
Imagine we have 100 Billion Translated Words at our Disposal
Imagine we have 100 Billion Translated Words at our Disposal

Terminology mining and dictionary building

Now
Selection and validation of term candidates is a highly manual task, causing delays, and inconsistencies.

New
Automatic extraction using advanced statistical and alignment tools.

Opportunities
Faster and more efficient process
Capture the most common used term candidates
Imagine we have 100 Billion Translated Words at our Disposal

Customize automated translation

Now
MT is not always good due to lack of domain ‘knowledge’.

New
Rapid MT training by selecting the right and good data from the cloud using semantic clustering and cleaning tools.

Opportunities
Much better MT
More efficient
Imagine we have 100 Billion Translated Words at our Disposal

Global market and customer analysis

Now
Translation is an isolated function and cost center.

New
Applying text analytics to our vast corpus of multilingual data allows us to provide strategic and valuable services.

Opportunities
Provide top line oriented services such as SEO, market intelligence.
Imagine we have 100 Billion Translated Words at our Disposal

Quality management

Now
Quality assurance is a highly manual function, subjective, slow and costly.

New
Apply automatic statistical tools and linguistic intelligence to clean source texts and translations.

Opportunities
Faster, more efficient
Measurable and objective
Industry Data Resources

Billions of Words
Translation Matching
TM Cleaning
Matching Scores

www.tausdata.org
Imagine we have 100 Billion Translated Words at our Disposal

**Stakes are high!**  **Risks are low.**

**Only fear can stop us**
Discussion Topics

Machine translation technology is good for the translation profession. Yes or no?

Sharing of translation memories in a super cloud leads to greater translation efficiency and quality. Yes or no?

Terminology 'lives' in the actual translation corpora (i.e. translation memories). If we have good terminology mining and extraction tools, we do not need separate terminology management. True or false?
TAUS in 5 Years

Industry Think Tank & Innovation Partner

Consulting

Events - workshops

www.translationautomation.com