Translation in the 21st Century
January 1954

ELECTRONIC
"BRAIN"

Translates
RUSSIAN to ENGLISH

PARAMOUNT NEWS
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
TM 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 Volunteer
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

Closed (Competitive)

Open (Collaborative)

Human & Machine

?
## SWOT for Enterprise Language Service

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

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

Localization industry

“Battle for words”

Social media

New technologies and solutions

User generated content

Support

Knowledge Base

Web

UI

Sales

Manuals

Content Disruption

Localization industry

“Battle for words”

Social media

New technologies and solutions

User generated content

Support

Knowledge Base

Web

UI

Sales

Manuals
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

<table>
<thead>
<tr>
<th>Old Model</th>
<th>New Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. One translation fits all</td>
<td>1. Quality differentiation</td>
</tr>
<tr>
<td>2. Selecting locales</td>
<td>2. Long tail of languages</td>
</tr>
<tr>
<td>3. Counting words</td>
<td>3. Unlimited content</td>
</tr>
<tr>
<td>4. TM is core</td>
<td>4. Data is core</td>
</tr>
<tr>
<td>5. Project based</td>
<td>5. Continuous translation</td>
</tr>
<tr>
<td>7. Publisher driven</td>
<td>7. User driven</td>
</tr>
<tr>
<td>8. One directional</td>
<td>8. Multi-directional</td>
</tr>
</tbody>
</table>
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)

- Quality differentiation
- Selecting locales
- One translation quality fits all
- 2011

Closed

- One-directional
- Project based
- TM is core
- Counting words

Open (Collaborative)

- Continuous translation
- Multi-directional
- Collaborative translation
- Unlimited content
- Long-tail of languages
- One translation quality fits all
- 2013

Human & Machine

- Cascaded supply chain
- Project based
- TM is core
- Counting words

2012
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 
\textit{delays}, and \textit{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

Research
Consulting
Events - workshops

www.translationautomation.com