The Story of the Translation Industry in 2022

A helicopter view of a five year innovation roadmap for the translation sector

@ Translating Europe Forum 2017
Today’s Headlines

Nearly Indistinguishable From Human Translation — Google Claims Breakthrough (September 26, 2016)

Systran Goes Neural, Lets You Try It (October 18, 2016)

Baidu Sells Human Translations at Rock-Bottom Rates to Gain Edge in AI (March 16, 2017)

Amazon Plans to Take On Google Translate (June 27, 2017)

A Convolutional Encoder Model for Neural Machine Translation - @ Facebook (July 25, 2017)

How SDL’s Secure Neural Machine Translation Helps Enterprises Reduce Risk (August 14, 2017)

Sogou Invests USD 6-7m in Shanghai-Based UTH in Scramble for Translation Data (August 22, 2017)

Linguee’s Founder Launches DeepL in Attempt to Challenge Google Translate (August 30, 2017)
**Getting the Pace Right**

**1954: Electric Brain**

“In five years we will translate all of the Soviet Union into the Queen’s English in just one week.”

**2017: Neural MT**

“The time has come. The future does not need translators.”
How do you know?

You can’t manage if you can’t measure.
Evolution of the Translation Technology Landscape

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TYPE OF CONTENT</td>
<td>Documents</td>
<td>Software</td>
<td>Simship</td>
<td>Integration in enterprise systems</td>
<td>Embedded in every app, on every screen</td>
<td>Ubiquitous</td>
</tr>
<tr>
<td>LANGUAGES</td>
<td>Paper</td>
<td>Digital</td>
<td>Static Web</td>
<td>Dynamic Web</td>
<td>Personalized</td>
<td>Unlimited</td>
</tr>
<tr>
<td>TECHNOLOGY</td>
<td>None</td>
<td>TM and terminology software</td>
<td>Workflow (GMS)</td>
<td>MT and Advanced Leveraging</td>
<td>Real-time customized MT</td>
<td>FAUT</td>
</tr>
<tr>
<td>DATA</td>
<td>Glossaries</td>
<td>Project TMs</td>
<td>Centralized TMs</td>
<td>Limited shared data</td>
<td>Web crawled</td>
<td>Plug and Play</td>
</tr>
</tbody>
</table>

- **Evolution of Translation Technology**

1980: Translation
- Focus: Documents
- Type of Content: Paper
- Languages: 1 > 10
- Technology: None
- Data: Glossaries

1990: Localization
- Focus: Software
- Type of Content: Digital
- Languages: 1 > 25
- Technology: TM and terminology software
- Data: Project TMs
- Communications: C2B

2000: Globalization
- Focus: Simship
- Type of Content: Static Web
- Languages: 1 > 40
- Technology: Workflow (GMS)
- Data: Centralized TMs
- Communications: C2C

2010: Integration
- Focus: Integration in enterprise systems
- Type of Content: Dynamic Web
- Languages: 6 <= 60
- Technology: MT and Advanced Leveraging
- Data: Limited shared data
- Communications: Social

2020: Convergence
- Focus: Embedded in every app, on every screen
- Type of Content: Personalized
- Languages: 60 <= 60
- Technology: Real-time customized MT
- Data: Web crawled
- Communications: Internet of Things

2030: Singularity
- Focus: Ubiquitous
- Type of Content: Unlimited
- Languages: 150 <= 150
- Technology: FAUT
- Data: Plug and Play
- Communications: M2M
## Globalization: Profile of 2000

### Connectivity rules!

#### Technology
- Client-server

#### Profiling
- Global span
- Full service

#### Value proposition
- Single source
- Simship

---

### 2000 Globalization

#### Industry Focus
- Simship

#### Type of Content
- Static Web

#### Languages
- 1 > 40

#### Technology
- Workflow (GMS)

#### Data
- Centralized TMs

#### Communications
- C2C

---

### HLT
- Server-based TM
- GMS & TMS
Globalization:
“42 steps to process a simple translation job”

The emergence of GMS and TMS: Idiom, GlobalSight, Uniscape
What Else Happened in 2000

MT is back!
Let a thousand MT systems bloom

TAUS Summit, Taos, New Mexico, March 15-17, 2007: A Historical Moment
Integration: Technology Profile of 2010

The age of web services

Technology
- APIs

LSP Profile
- Diversification of services
- Testing
- Digital marketing
- Consulting

HLT
- Web-based TM
- TM and MT as hybrid solution
- MT as an enabler for other business

Value proposition
- Technology integration
Integration: Technology Landscape of 2010

Move to the Cloud

More words translated by machines than by humans

Enterprise Language Strategy

“For the first time enterprises become strategic about translation.” Eight things to change!
Enterprise SWOT Analysis (2010 – 2011)

**S**
- High leverage from Translation Memories
- Well established process and management

**W**
- Quality inconsistent (local flavor missing)
- Lack of flexibility in landscape, reactive rather than creative
- Quality review is slow – bottleneck

**O**
- Opening new markets with MT
- Engaging with users & communities
- Convergence with video and speech
- Search engine optimization
- Translation of user generated content
- Use of mobile
- Content personalization

**T**
- Locked in to vendor base
- Not scalable to expand quickly
- Urgent requirement to support new markets
- Inability to ensure quality
- Opportunity loss due to lack of personalization
Static Cascaded Supply Chain

- Clients
  - Quality Assurance
  - Translation Memory
  - Project Management
  - Resources Management

- MLV’s
  - Quality Assurance
  - Account Management
  - Project Management
  - Resources Management

- In-country offices/partners
  - Quality Assurance
  - Translation Memory
  - Project Management

- Distributed translators/authors
  - Quality Assurance
  - Translation Memory

4 to 30 vendors
10 to 40 languages
100’ to 1000’s translators/authors

This slide may not be used or copied without permission from TAUS
Content Disruption

LOCALIZATION INDUSTRY

"BATTLE FOR WORDS"

WEB
OSD
CL PRODUCT CATALOGUE
USER MANUALS

SOCIAL MEDIA

CHAT/MOBILE
SUPPORT ARTICLES/KNOWLEDGE BASE
USER GENERATED CONTENT

NEW TECHNOLOGIES AND SOLUTIONS
Innovation Dilemma

S
• High leverage from Translation Memories
• Well established process and management

W
• Quality inconsistent (local flavor missing)
• Lack of flexibility in landscape, reactive rather than creative
• Quality review is slow – bottleneck
• Execution on innovation fails

O
• Opening new markets with MT
• Engaging with users & communities
• Convergence with video and speech
• Search engine optimization
• Translation of user generated content
• Use of mobile
• Content personalization

T
• Locked in to vendor base
• Not scalable to expand quickly
• Urgent requirement to support new markets
• Inability to ensure quality
• Lack of corporate awareness of new locales
• Opportunity loss due to lack of personalization

This slide may not be used or copied without permission from TAUS
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. Unidirectional

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

1. Quality differentiation
2. Long-tail of languages
3. Zettabytes of content—owned, shared, earned
4. Data is core
5. Continuous translation
6. Collaborative translation
7. User-driven
8. Multidirectional

Many big worlds in one small planet
Convergence: Technology Profile of 2020

The cloud is irresistible

**Technology**
- Mobile
- Free & open source

**LSP Profile**
- Start-ups competition
- Location not so relevant anymore
- Platforms instead of supply chains

**HLT**
- Real-time MT
- Speech-to-speech

**Value proposition**
- Subscriptions
- Innovative offerings: social
- Quality levels: transcreation
- Dashboards
Convergence of Markets: Translation Shifts Gears

From 10,000 customers who buy translation as a ‘luxury’ product to 6 billion users who consider translation ‘free’.

Publisher-driven translation industry
What is Happening Now

Convergence of:
• Technologies: MT & S2S
• Markets: Consumer & Business
• Business: Free and Premium

Drag and drop ...
• Real-time customization of MT

Plug and play ...
• Mobile apps for speech-to-speech

Translation as utility
• On every screen, in every app

Many new start-ups

Insider and invader innovators

MT as an API: pay as you go or free for data
The Story of the Translation Industry in ‘22 in ten mini-chapters

1. Algorithmic Management
2. Datafication of Translation
3. Quality Matters
4. Convergence
5. Speech Translation
6. Bite-Sized and Long-Read Translations
7. The Quantum Leap
8. The Long Tail
9. Pay-as-you-Go
10. Users First
#1 Algorithmic Management

- Robots enhance, expand and replace our work
- Counting words…. resource allocation
- Confidence scoring (based on 42 metrics)
- ROI by segment
- Self-driving translations
#2 Datafication of Translation

- Unreasonable effectiveness of data
- Very large quantities of data belong to the past
- Challenges for protectionists are opportunities for pirates
- CEF in Europe
- New ventures in China
- TAUS Data Market
#3 Quality Matters

- Don’t be blinded by technology
- ‘Dumb translations’ are not enough
- Stories need to be told and imagination needs to be triggered
- Data help to decide where to invest
#4 Convergence

- Confluence of technologies, business models and markets
- At these crossroads innovation happens
- Microsoft Hololens combining S2ST and video-conferencing to assist field engineers
- Alternatives to translations that nobody needs
#5  Speech Translation

- Skype Translators & Microsoft Live Translator integrated in PowerPoint
- Spoken translation in wearable technology
- Opportunities for new specialization: data, tools, workflows and apps
#6 Bite-Sized and Long-Read Translations

- Continuous delivery – smaller jobs
- Concept of project is blurring
- Translation is always on
- Disintermediation is a theme again
- Widening gap between automatic and creative translation
The Quantum Leap

- Five years of NMT = twenty years of SMT
- 5,000 lines of code – 200,000 lines of code
- Zero shot translation
- Thousand(s) decisions taken by machines
- Speech translation
- Everything is translated
Accelerated Improvement of Machine Translation

Statistical Machine Translation
- Translate, translate, translate...
- Wow, I can learn a lot of translation patterns!

Human-translated documents → New Source Text → Machine Translation Output

Quality


Rule-Based Statistical MT Neural MT

DeepL Translator
Try out the world’s best machine translation.
The Long Tail

- Globalization reaches smaller companies
- Pursuit of customers in new markets
- China One-Belt One-Road
- Zero shot translations adds MT for more languages
- Stimulates interest in new cultures and languages
#9 Pay-as-you-Go

- Ubiquitous, always-on translation triggers new business and pricing models
- Alternatives for fixed price per word: pay per view, differentiate by content profile
#10 Users First

- Usability and findability become crucial
- User-centric business models will prevail
- Users are the new talents
Conversations with Captains of the Translation Industry

Adolfo Hernandez, CEO SDL  
"From People-Driven to Data-Driven"

Aiman Copty, Vice President, International Solutions, Oracle  
"Making Translation Invisible"

Rory Cowan, Chairman Lionbridge  
"Getting the Pace Right"

Smith Yewell, CEO Welocalize  
"In the Business of Global Talent Search"
Key Take-Aways

1. Change – jobs, processes, technology
2. Control – measure, dashboard
3. Analyze – use your data
The Modern Translation Pipeline

- **Algorithmic management**
  - Content profiling
  - Matching resources
  - Confidence scores
  - UTS scoring

- **Quantum Leap**
  - MT
  - TM
  - CT

- **Datafication**

- **Dashboards**
  - DQF

**FAUT**: Fully Automatic Useful Translation

**Good Enough**: Review

**High Quality**: Transcreation

**MT Training**: Data loopback

**Machine Learning**: Data loopback
The *Modern* Translation Pipeline ...

1. made great progress in interoperability
2. is adopting DQF as a quality standard
... and where are the Jobs?!

A search for global multilingual talents to undertake a plethora of new value-added services:

- Brand Validation
- Multilingual Digital Marketing
- eDiscovery
- Usability Testing
- Campaign Management
- Content Curation
- Community Development
1. Jobs of post-editor/translator, quality evaluator, editor and reviewer are blurring

2. They feed the data that train the translation pipeline, to fork content in three levels:
   - FAUT
   - Good enough
   - High Quality

3. Their annotation and corrections train the machine translation engines.
Six Reasons for Reviewers to be Obsessed about Data

1. **Measurability:** You can’t manage what you don’t measure.
2. **Objectivity:** Data is unbiased.
3. **Efficiency:** Impossible to proofread everything.
4. **Benchmarking:** Comparing apples with apples.
5. **Quality levels:** Data bring structure and distinction.
6. **Predictions:** Be ready for machine learning.
How DQF and Quality Dashboard Work
DQF-MQM error typology
It’s time for a counterrevolution!

Locamaniacs Versus Globaloneyes

Localization industry has a mission:

- We must evangelize cultural diversity, increase understanding
- Drive globalization – grow our companies across borders
- Use technology to the maximum, work with the robots – treat them as ‘cobots’
- Measure, analyze, benchmark, control