LT07 – Machine Translation Cannot Live on its Own
Evolution of Machine Translation

- **1970**: Rule-Based
- **2002**: Statistical MT
- **2016**: Neural MT

Quality


**Companies**
- SYSTRAN
- Google
- SDL
- Bing
- Facebook
- KantanMT.com
Rule-Based MT

- Analysis: SL text → SL repr.
- Transfer: SL repr. → SL→TL lexical and structural rules
- Synthesis: TL repr. → TL text

SL lexicon and grammars

Phrase-Based Statistical MT

He goes to the curry restaurant
彼は行くにカレー屋

Neural MT – The emergence of AI

1970

Over
the
line
!
<eos>

Çizgiyi
geçtin
!
<eos>
Over
the
line
!
MT Usage Cases

- Legal
- Medical
- eCommerce
- eMail
- Social Media
- Chat
- Travel
- Help
- Tech Pubs
- Knowledge Base
- Web Content
- UI
- Blogs
- Support
- eMail

Quality

Volume

Post-edit (PEMT)

FAUT (Fully Automated Usable Translation)
Factoid #1

- 500 Jobs on the KantanMT platform
- Breakdown of TM versus MT translations
KantanAnalytics – Tracks MT & TM Usage

Summary of MT & TM Usage:

<table>
<thead>
<tr>
<th>Segment</th>
<th>Segments</th>
<th>Words</th>
<th>Matched (%)</th>
<th>Characters</th>
<th>Tags</th>
<th>Placeholders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1,998</td>
<td>17,478</td>
<td>66%</td>
<td>108,237</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Matched</td>
<td>1,994</td>
<td>17,442</td>
<td>34%</td>
<td>108,024</td>
<td>0</td>
<td>38</td>
</tr>
</tbody>
</table>

KantanAnalytics Matched Words:

<table>
<thead>
<tr>
<th>Quality Estimation Score</th>
<th>Total</th>
<th>100%</th>
<th>90-99%</th>
<th>85-89%</th>
<th>70-84%</th>
<th>55-69%</th>
<th>40-54%</th>
<th>&lt;40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>1,120</td>
<td>9,568</td>
<td>54.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90-99%</td>
<td>89</td>
<td>857</td>
<td>4.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85-89%</td>
<td>128</td>
<td>1,272</td>
<td>3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70-84%</td>
<td>639</td>
<td>5,592</td>
<td>32%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-69%</td>
<td>12</td>
<td>84</td>
<td>0.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-54%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>4</td>
<td>21</td>
<td>0.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>1,998</td>
<td>17,478</td>
<td></td>
<td>108,237</td>
<td>0</td>
<td>38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Engine Scores:

- Word Count: 319,624
- Unique WC: 22,813
- Mono WC: N/A
- F-Measure: 48.5%
- BLEU: 47.8%
- TER: 38.5%
MT Technology Evolution

Technology S curve

Performance

Tech Improvement

New innovation

Time

Mature Technology

KantanMT.com
Technology Innovations

- **Rules -> Statistical -> Neural**
  - Complexity -> Simplicity
  - Data driven approaches (Factoid #2 - Neural MT uses 60% less data!)

- **Tags/Placeholder supported**
  - 92% of KantanMT content contains tags/placeholders

- **Terminology (Level 1) supported**
  - Similar approach to CAT Tools

- **Adaptive Models**
  - Continuous learning models

- **Sophisticated Connectors**
  - To CAT/TMS systems
MT Cannot Live on its Own

$2^n$

Power of Two
Optimal Deployment

- Translation Memory
  - Precision
  - Consistency
- Machine Translation
  - Flexibility
  - Coverage
- Impact
  - Faster Translation Cycles
  - Reduced Project Costs

Power of Two
Thank you...
Your opinion is important to us! Please tell us what you thought of the lecture. We look forward to your feedback via smartphone or tablet under http://lt07.honestly.de or scan the QR code.
Your opinion is important to us! Please tell us what you thought of the lecture. We look forward to your feedback via smartphone or tablet under http://lt07.honestly.de

or scan the QR code