Deploying novel MT technology to raise the bar for quality:
Key advantages and challenges

Johann Roturier
August 2009
Overview

2. SPE: A Case Study in Deploying MT technology
3. Challenges and Next Steps
Chapter 1

Localization Requirements
Localization Requirements

Translation of ‘user interface’ into Japanese

ユーザインターフェイス
ユーザインターフェース
ユーザインターフェース
ユーザーユーザーインターフェース
ユーザーユーザーインターフェース
ユーザーアユーザーアユーザーアユーザーアユーザーア
Why use MT?

• Increase product terminology consistency
  – Focus on Features and Product Names (which can be enforced through terminology Preparation and machine-translation)
  – Correct Software References (which can be enforced by using a specific MT user dictionary)

• Reduce time to market (TTM) through increased productivity
  – Pre-Translate all content through TM and MT

• Lower localization cost
Production Requirements

- Must be used in conjunction with TM system
- Must support tagged input
- Must be easy to deploy globally (7 languages)
- Input should be controlled (acrocheck project score)
Defining a Production MT Workflow

- SW Strings
- Documentation
- Localisation with MT

TERMINOLOGY
Production MT Workflow

- SW Writing
- Controlled Documentation Authoring
- Terminology Harvesting and Definition
- MT Terminology Encoding (including SW strings)
- MT Tuning using Automated Post-Editing
- Machine-Translate Segments using custom UD and STS
- Export Unknown Segments (with Context)
- Source Files Analysis (against Translation Memory)
- Post-Editing using TM and MT
- Evaluation
- Update TM
Evaluation

- Identify suitable MT system
- Monitor performance of MT system
- Monitor PE work
- Determine PE adjustments
- Evaluation Methodology
**Timeline**

<table>
<thead>
<tr>
<th>YEAR</th>
<th>STEP</th>
<th>ACHIEVEMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>Research investigations (Desktop products)</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>Initial MT usage for Technical Support translations</td>
<td>Translation Productivity Proven (in-house)</td>
</tr>
<tr>
<td>2006</td>
<td>Enterprise system (SYSTRAN v5) and Controlled Authoring Environment (acrocheck 3) deployed</td>
<td>Global access by linguists and writers</td>
</tr>
<tr>
<td>2006</td>
<td>User documentation in production (EMEA) with reduced TTM and costs</td>
<td>First large enterprise product (500K words) we localised using MT in 2007 shipped in 7 days (15 days with previous version). Fewer bug fixes to implement in the Help system of this product.</td>
</tr>
<tr>
<td>2008</td>
<td>User documentation in production (APJ) with reduced TTM and costs</td>
<td>Reduced costs because PE task is faster than translation task.</td>
</tr>
<tr>
<td>2009</td>
<td>Controlled Authoring Environment (acrolinx IQ) deployed</td>
<td>Opening of ad-hoc translation opportunities</td>
</tr>
<tr>
<td></td>
<td>Enterprise MT system (SYSTRAN v6) deployed</td>
<td>Savings have outperformed the investment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality of output has reached a high level for some languages</td>
</tr>
</tbody>
</table>

SPE: A case study in deploying MT technology
Background

• Current MT output
  – Requires repetitive post-editing
  – Lacks fluency
• Need to show continuous improvement to post-editors
  – Current customization strategies:
    – User Dictionaries
      • Difficult and time-consuming to add exceptions (e.g. context clue)
    – Automated Post-Editing
      • Hand-coded (prone to errors)
      • High precision if well-crafted but low coverage
• Idea behind Statistical Post-Editing
  – Learn from Post-Editing activity to build statistical models
New Architecture

Source: Based on Laporte, P. LocWorld (2009)
Case Study

• Collaboration Project with Systran
  – PE Objective
    • Improve overall PE experience in all languages
    • Allow use of SPE for production project
  – Technical Objective
    • Fit into existing workflow
    • Do not deteriorate performance too much
  – Linguistic Objectives
    • Re-case output
    • Preserve key terminology
    • Support for tagged input
    • Show positive Improvement/Degradation
Case Study

Steps

– Cleaned TMs
  • Removed Segments with comments
  • Checked Key Terminology

– Sent TMs and resources to SYSTRAN
  • TMs: 40 K translation units
  • User Dictionaries (around 2 x 10K per language pair)

– Received SPE models
  • Deployed on staging server
  • Used Rules + SPE as Hybrid System

– Validated with:
  • Automatic Scores on test set (5K translation units)
  • Human Evaluation on subset of test set (100 segments)
  • PE pilot project
## Validation Results (Automatic)

<table>
<thead>
<tr>
<th>Language/Score</th>
<th>Baseline</th>
<th>Systran 6.06 (with default Symantec UD)</th>
<th>Systran Hybrid (with default Symantec UD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Systran 6.06</td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>33.84 (BLEU 0.1553)</td>
<td>46.42 (BLEU 0.3014)</td>
<td>56.24 (BLEU 0.4442)</td>
</tr>
<tr>
<td>French</td>
<td>40.92 (BLEU 0.2399)</td>
<td>51.91 (BLEU 0.3890)</td>
<td>56.72 (BLEU 0.4606)</td>
</tr>
<tr>
<td>Japanese</td>
<td>39.46 (BLEU 0.1703)</td>
<td>45.94 (BLEU 0.2336)</td>
<td>58.88 (BLEU 0.4110)</td>
</tr>
<tr>
<td>Simp. Chinese</td>
<td>44.40 (BLEU 0.2344)</td>
<td>52.67 (BLEU 0.3201)</td>
<td>58.14 (BLEU 0.3721)</td>
</tr>
<tr>
<td>German</td>
<td>29.26 (BLEU 0.1338)</td>
<td>37.67 (BLEU 0.2224)</td>
<td>45.94 (BLEU 0.3223)</td>
</tr>
</tbody>
</table>
## Validation Results (Human)

<table>
<thead>
<tr>
<th>Category/Lang.</th>
<th>JA</th>
<th>CS</th>
<th>FR</th>
<th>IT</th>
<th>DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Found Words</td>
<td>0.7</td>
<td>2.8</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Simple Terms</td>
<td>5.3</td>
<td>3.9</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Phrases</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Meaning</td>
<td>0.54</td>
<td>1.23</td>
<td>1</td>
<td>0</td>
<td>0.74</td>
</tr>
<tr>
<td>Determiners</td>
<td>-</td>
<td>5</td>
<td>7</td>
<td>0.5</td>
<td>1.12</td>
</tr>
<tr>
<td>Prepositions</td>
<td>1.34</td>
<td>1.85</td>
<td>5.75</td>
<td>6</td>
<td>1.19</td>
</tr>
<tr>
<td>Pronouns</td>
<td>-</td>
<td>0.5</td>
<td>1</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Tense</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Number</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>Gender</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>0.5</td>
</tr>
<tr>
<td>Other Grammar</td>
<td>2.75</td>
<td>-</td>
<td>3</td>
<td>26</td>
<td>0.25</td>
</tr>
<tr>
<td>Punctuation/Case</td>
<td>0.2</td>
<td>-</td>
<td>0</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>Word Order(Short)</td>
<td>0.17</td>
<td>-</td>
<td>1</td>
<td>0.7</td>
<td>9.5</td>
</tr>
<tr>
<td>Word Order (Long)</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>0.86</td>
</tr>
<tr>
<td>Tags</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>1.45</td>
<td>2.2</td>
<td>3</td>
<td>3</td>
<td>2.24</td>
</tr>
</tbody>
</table>
Validation Results (PE)

• PE Results
  – 5 K words

• Simplified Chinese
  – Progress in fluency and meaning

• French and Italian
  – Throughput was improved slightly
  – Overall experience was a little better
  – Use SPE models for large production project (200K+)

SPE: A case study in deploying novel MT technology

Challenges and Next Steps
Challenges

• Japanese PE feedback
  – Quality was impacted in places, yet high number of unchanged segments
  – Missing words in output sentence (for example, important negative words such as “not” disappeared) and words that were not in original sentence were used in output

• German PE feedback
  – Errors seem less predictable so less possible to think in terms of making global corrections
  – Still 35% faster than translating from scratch

• Explanations
  – Use of “free” translations in training data
  – Not enough model filtering
Challenges

• Synch MT resources
  – TM
  – SPE Models
  – MT UDS
  – Automated PE module

• Analyze changes brought by SPE

• Maintain performance
Challenge: MT Workflow Update

- SW Writing
- Controlled Documentation Authoring
- Terminology Harvesting and Definition
- MT Terminology Encoding (including SW strings)
- Export Unknown Segments (with Context)
- Source Files Analysis (against Translation Memory)
- SPE Training
- TM Cleaning
- Machine-Translate Segments using custom UD, STS and SPE
- MT Tuning using Automated Post-Editing
- Post-Editing using TM and MT
- Evaluation
- Update TM
SPE Developments

• Reduce degradations to a minimum
  – Investigate tuning based on human judgments AND automated scores
  – If not possible, relying on a pure rules-based output may be preferable for certain types of sentences
    • This decision should be made by the MT engine based on its confidence of the MT outputs it can produce.

• Investigate further TM cleanup and management
  – To isolate segments that are not worth re-using at a sub-segment level.
  – Selecting quality data rather than large data sets seems preferable when the objective is to improve PE experience.
Summary

• Good return on investment from MT and SPE
  – Good output quality for most languages
  – Statistical element has already helped raise the bar for quality for some languages (in large production projects)

• Additional opportunities for ad-hoc content localization using MT
  – Technical Support Documentation (including UGC)
  – Training materials

• Close collaboration with post-editors is key
  – Understand better PE task
  – Use PE activity and MT error analysis to optimize MT systems
Thank You!

Johann Roturier

johann_roturier@symantec.com