SMT HAS REACHED INDUSTRIAL MATURITY – DO’S AND DON’T’S FOR MANAGERS
AGENDA

» IT-Managers ...
» Key factors
» Do’s and Don’ts
» Future developments

'IF I AM SELLING TO YOU, THEN I SPEAK YOUR LANGUAGE, ...
GEEK & POKE'S WEEKLY TIP FOR PROJECT MANAGERS

THERE IS NOOOOO WAY THAT WE'LL SHIP ON MARCH 1 AND IT'S WAAAAAY MORE THAN $100,000
SMT HAS REACHED INDUSTRIAL MATURITY – DO’S AND DON’T’S FOR MANAGERS
IT MANAGERS BRIDGE GAPS

TIP 2: ONLY MEMORIZE THE IMPORTANT FACTS

- MARCH 1
- $100,000

Source: http://geek-and-poke.com
SMT HAS REACHED INDUSTRIAL MATURITY

Key factors

Do's and Don'ts

Future developments

... ABER WENN DU MIR ETWAS VERKAUFS, ...
**KEY FACTORS**

- Value adding technology and features
- Dedicated inhouse teams and great partners
- Language Concept, OLA & budget (constraints) in place
- Organisational readiness and continuous improvement
- Focused business case
Our translation assets are precious. We are an SME with 400 employees and limited resources.

Every year, Finnova translates 6,000 customer support messages within software release letters from German to English.

The aim of the project was to develop a translation system tailored to translate those texts automatically.

ROI within two years and cost savings of more than 30%.
KEY FACTORS

» Prof. Dr. Martin Volk
» Dr. Rico Sennrich
» Dr. Mark Fishel
» Katrin Affolter
» Mathias Müller

» Andreas Schmidt
» Dedicated translation team

» Sandra Roth
» Kate O’Dwyer
» Michaela Schnetzer

Dedicated inhouse teams and great partners
KEY FACTORS

Value adding technology and features

» Private in-house tool hosting on SUSE Linux
» Customisation of an open source SMT system
» Simple integration into a CAT tool workflow
» Implementation of an algorithm that gives more weight to in-domain data
» Clever customised preprocessing and post-processing heuristics
» Flexible output formats (TXT, HTML, XLIFF or TMX)
KEY FACTORS

Value adding technology and features

<table>
<thead>
<tr>
<th>Corpora</th>
<th>Segments</th>
<th>Tokens EN</th>
<th>Tokens DE</th>
</tr>
</thead>
<tbody>
<tr>
<td>out-domain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europarl v7</td>
<td>1.9 m</td>
<td>52 m</td>
<td>50 m</td>
</tr>
<tr>
<td>JRC-Acquis version 3.0</td>
<td>1.2 m</td>
<td>29 m</td>
<td>27 m</td>
</tr>
<tr>
<td>OpenSubtitles v2013</td>
<td>10.9 m</td>
<td>90 m</td>
<td>83 m</td>
</tr>
<tr>
<td></td>
<td>14.0 m</td>
<td>171 m</td>
<td>160 m</td>
</tr>
<tr>
<td>in-domain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help desk tickets</td>
<td>172,000</td>
<td>2.3 m</td>
<td>2.1 m</td>
</tr>
<tr>
<td>ECL</td>
<td>119,000</td>
<td>0.8 m</td>
<td>0.7 m</td>
</tr>
<tr>
<td>BHB</td>
<td>56,000</td>
<td>0.6 m</td>
<td>0.5 m</td>
</tr>
<tr>
<td></td>
<td>347,000</td>
<td>3.7 m</td>
<td>3.3 m</td>
</tr>
</tbody>
</table>
KEY FACTORS

Language Concept, OLA & budget (constraints) in place

Organisational readiness and continuous improvement

Pretranslation

Post-editing by human translator

Delivery/Result

source text

TMX

source/target segments

TMX_in

source/target segments

TMX_out

target text

SMT HAS REACHED INDUSTRIAL MATURITY – DO’S AND DON’T’S FOR MANAGERS
Instead of translating the target segments directly, we chose TMX as the exchange format between the SMT and CAT workflow. This allows for maximum flexibility and easy integration with CAT tools.

Language Concept, OLA & budget (constraints) in place.

Organisational readiness and continuous improvement.

Target segments are generated with the SMT TMX.

Quality assurance: Target segments are post-edited with TMX made by human translators.
## KEY FACTORS – USER FEEDBACKS

<table>
<thead>
<tr>
<th></th>
<th>Trans1</th>
<th>Trans8</th>
<th>Trans11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td>Better than the previous one.</td>
<td>Still very Dinglish, quite literal and lots of untranslated segments or things that appear to have been assembled from fragments.</td>
<td>Could be better.</td>
</tr>
<tr>
<td><strong>Retranslate</strong></td>
<td>This time, mostly segments longer than 3-4 lines had to be retranslated.</td>
<td>Large proportion</td>
<td>Lots of retranslation needed.</td>
</tr>
<tr>
<td><strong>Word order</strong></td>
<td>Improved. I have noticed an improvement compared to previous projects.</td>
<td>Still very German, or rearranged in some totally inexplicable way</td>
<td>Very frequently incorrect, and lots of missing words.</td>
</tr>
<tr>
<td><strong>not EN words</strong></td>
<td>a few French words, but they were part of the project</td>
<td>Still quite a few German words</td>
<td>As usual- mostly in words with hyphen.</td>
</tr>
<tr>
<td><strong>Terminology</strong></td>
<td>OK, as usual</td>
<td>Not always consistent with termbase</td>
<td>Sometimes inconsistent, different translation used for the same term.</td>
</tr>
<tr>
<td><strong>Speed</strong></td>
<td>Due to the improvement regarding the placement of the verbs in the machine-translated text, I perceived a 10-15% improvement in my translation speed.</td>
<td>Same as usual, does not get faster</td>
<td>Slower than last time.</td>
</tr>
<tr>
<td><strong>Short sentences</strong></td>
<td>In most of the cases, they could be confirmed as they were.</td>
<td>Occasionally OK</td>
<td>Several of them had to be retranslated, too.</td>
</tr>
<tr>
<td><strong>Overall compared to previous finnova MT</strong></td>
<td>Due to the improvement regarding the placement of the verbs in the machine-translated text, better than the previous Finnova MT</td>
<td>I fail to see much, if any, improvement.</td>
<td>Much worse than in July. Missing words, incorrect word order, lots of editing needed.</td>
</tr>
</tbody>
</table>
### KEY FACTORS – USER FEEDBACKS

<table>
<thead>
<tr>
<th></th>
<th>TransA</th>
<th>TransB</th>
<th>TransC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>Better than previous ones.</td>
<td>Still very Dinglish, quite literal and lots of untranslated segments</td>
<td>Could be better.</td>
</tr>
<tr>
<td>Retranslate</td>
<td>Large proportion needed to be retranslated.</td>
<td>Very frequently incorrect, and lots of missing words</td>
<td></td>
</tr>
<tr>
<td>Word order</td>
<td>Improved. I have noticed an improvement compared to previous projects</td>
<td>Very frequently incorrect, and lots of missing words</td>
<td></td>
</tr>
<tr>
<td>not EN words</td>
<td>a few French words, but they were part of the project</td>
<td>As usual- mostly in words with hyphen.</td>
<td></td>
</tr>
<tr>
<td>Terminology</td>
<td>OK, as usual</td>
<td>Sometimes inconsistent, different translation used for the same term.</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>Due to the improvement regarding the placement of the verbs in the</td>
<td>Slower than last time.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>machine-translated text, I perceived a 10-15% improvement in my</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short sentences</td>
<td>In most of the cases, they could be confirmed as they were.</td>
<td>Several of them had to be retranslated, too.</td>
<td></td>
</tr>
<tr>
<td>Overall compared to previous finnova MT</td>
<td>Due to the improvement regarding the placement of the verbs in the machine-translated text, better than the previous Finnova MT</td>
<td>I fail to see much, if any, improvement.</td>
<td>Much worse than in July. Missing words, incorrect word order, lots of editing needed.</td>
</tr>
</tbody>
</table>
KEY FACTORS

Expenditures before and after the SMT system integration

SMT integration
AGENDA

IT-Managers …

Key factors

Do’s and Don’ts

Future developments
# DO’S AND DON’TS

<table>
<thead>
<tr>
<th>Questions that were helpful (do’s)</th>
<th>Questions that were not helpful (don’ts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you already have a language concept, OLA’s and a budget for translations?</td>
<td>Do we need to cut down our translation volume or our costs?</td>
</tr>
<tr>
<td>Did you already optimise along the translation value chain?</td>
<td>We could go SMT all-in, couldn’t we?</td>
</tr>
<tr>
<td>Do you have an application owner that can take care of the project and understand the needs of all stakeholders?</td>
<td>So, a Linux server, the Moses software and a little bit of magic is necessary, isn’t it?</td>
</tr>
<tr>
<td>Can I afford a linguistic preconditioning of my translators?</td>
<td>What kind of texts are especially fit for SMT?</td>
</tr>
<tr>
<td>What is an acceptable sentence? Did you ask the translator?</td>
<td>Did the BLEU score as only benchmark improve now?</td>
</tr>
<tr>
<td>How much quality assurance/correctness can we afford?</td>
<td>The system will produce the same average quality that the training set had – why should we post-edit?</td>
</tr>
<tr>
<td>Is post-editing the same as proof-reading/reviewing? Do you need the same skill set?</td>
<td>Could you please proof-read the text as usual after the SMT translation?</td>
</tr>
</tbody>
</table>
AGENDA

IT-Managers …

Key factors

Do's and Don’ts

Future developments

... Dann musst du Deutsch sprechen’.
Willy Brandt
Managers bridge gaps

A cross-functional end-to-end view on the whole language value chain helps

SMT goes industrial because one basic solution
- can be used for many real-life language challenges
  - for any language pair
  - for any text type
  - for any domain
- can be integrated into many applications like a
  - self-service in-house hosted translation desk
  - plug-in for automatic pretranslation in CAT tools
  - translation portal on demand

Given that the organisation has already optimised its processes and workflows, i.e. a language concept, OLA and budgets are in place, terminology DB, glossaries, CAT tools, etc. are available, then 20 to 30% of cost savings are a realistic scenario.
The Institute of Computational Linguistics of the University of Zurich conducts research and provides teaching in the domains of (theory-oriented) computational linguistics and (applications-oriented) language technology. It offers undergraduate and postgraduate degrees in both domains.

Our research focuses on the development of intelligent methods to access text (answer extraction, relation extraction, question answering), statistics-based machine translation, and parsing (automatic syntax analysis).
Syntax Übersetzungen AG is the most prestigious owner-run company in Switzerland for integrated foreign language management services. Years of experience working closely with customers from all industries and disciplines have enabled us to develop the philosophy of the “made-to-measure language solution” to create the Syntax method, incorporating modular services that are fine-tuned to the customer’s own special wishes and requirements. We’d be happy to arrange a “made-to-measure language solution” for you, too.
Finnova is a leading provider of banking software in the Swiss financial centre. We help banks and outsourcing providers to realise growth in the banking sector, even in challenging times, thanks to efficient and innovative IT solutions compliant with regulatory requirements. “Smarter banking” with Finnova - that is what we stand for. And that is why over 100 banks have already put their trust in us.

Founded in 1974, Finnova AG employs around 400 people at its headquarters in Lenzburg and at other branch offices in Chur, Seewen and Nyon. Its standard solution, Finnova Banking Software, is currently used by around 80 universal banks and 20 private banks. They benefit from the software’s extensive functional breadth and depth, combined with the lowest total cost of ownership (TCO) in the industry. Finnova’s outstanding scalability and parameterisation capability enable banks to implement their strategic requirements efficiently, within budget and on schedule.
THANK YOU VERY MUCH!