



# SMT HAS REACHED INDUSTRIAL MATURITY – DO'S AND DON'T'S FOR MANAGERS

ROBERTO NESPECA – HEAD DOCUMENT MANAGEMENT

# AGENDA

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

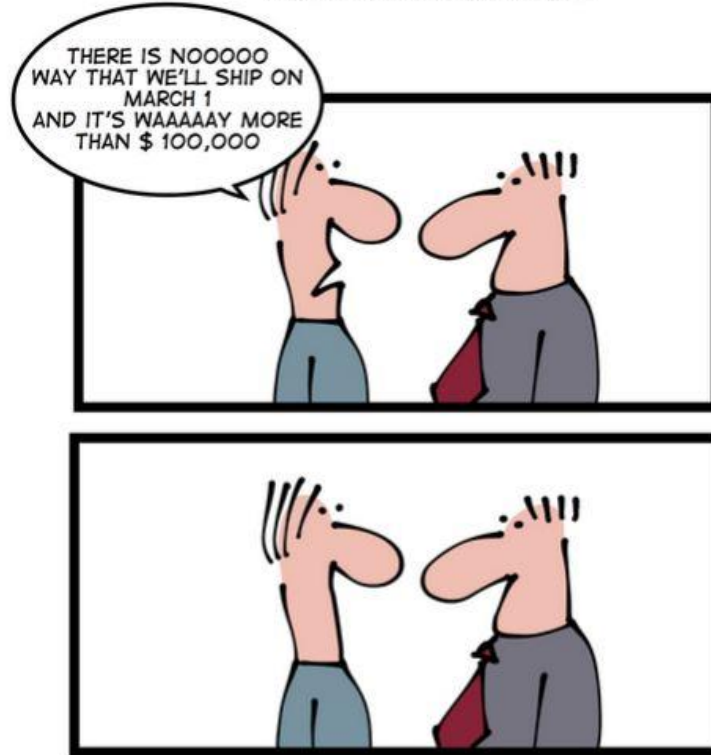


**‘IF I AM SELLING TO YOU  
THEN I SPEAK YOUR  
LANGUAGE, ...**

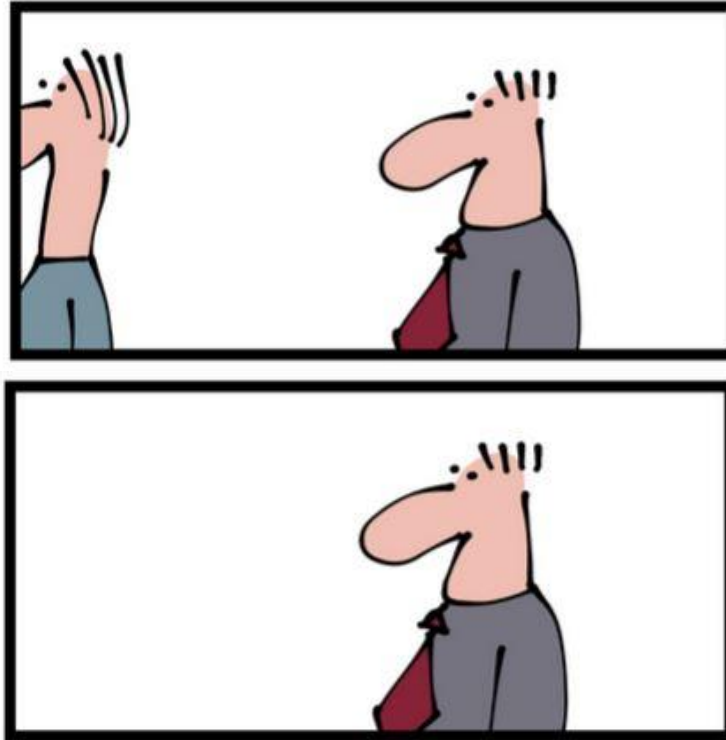


# IT MANAGERS...

## GEEK & POKE'S WEEKLY TIP FOR PROJECT MANAGERS



# IT MANAGERS...



# IT MANAGERS...



**TIP 2: ONLY MEMORIZE THE IMPORTANT FACTS**

Source: <http://geek-and-poke.com>

# IT MANAGERS BRIDGE GAPS



**TIP 2: ONLY MEMORIZE THE IMPORTANT FACTS**

Source: <http://geek-and-poke.com>

# AGENDA

IT-Managers ...

## ► Key factors

Do's and Don'ts

Future developments



... ABER WENN  
DU MIR ETWAS  
VERKAUFST, ...

# KEY FACTORS





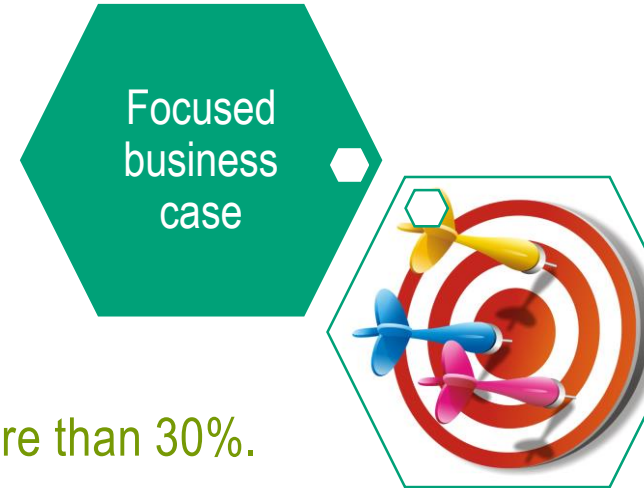
# KEY FACTORS

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



Dedicated  
inhouse  
teams and  
great partners



**University of  
Zurich**<sup>UZH</sup>

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

**s y n t a x**

- » Andreas Schmidt
- » Dedicated translation team



- » Sandra Roth
- » Kate O'Dwyer
- » Michaela Schnetzer

# KEY FACTORS



**MOSES**  
statistical  
machine translation  
system

- » 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

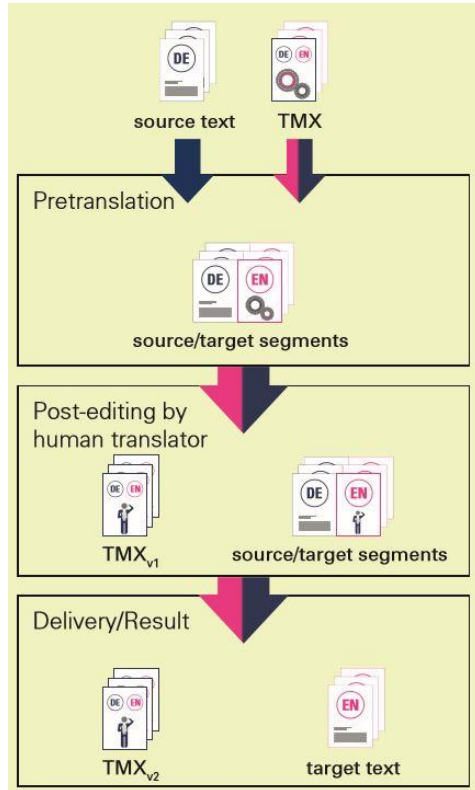


**MOSES**  
statistical  
machine translation  
system

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



# KEY FACTORS



Language  
Concept, OLA  
& budget  
(constraints)  
in place

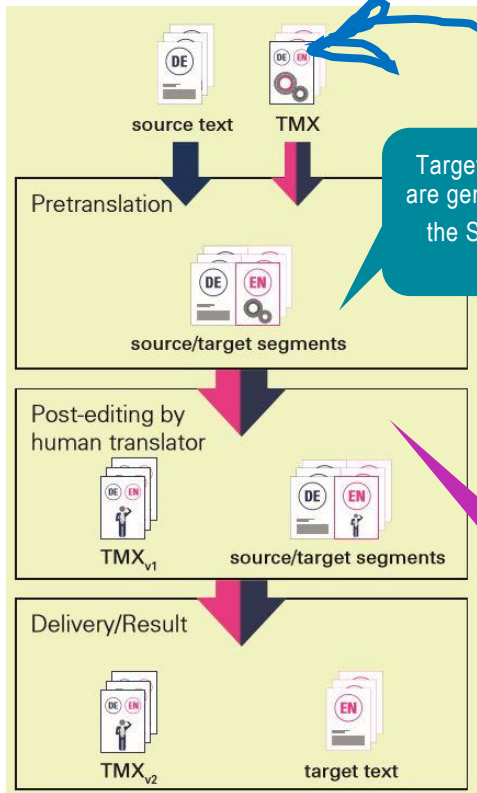
Organisational  
readiness  
and continous  
improvement



**MOSES**  
statistical  
machine translation  
system

# KEY FACTORS

*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.*



Target segments are generated with the SMT TMX.

Language Concept, OLA & budget (constraints) in place

Organisational readiness and continuous improvement



Quality assurance: Target segments are post-edited with TMX made by human translators.



**MOSES**  
statistical machine translation system

# KEY FACTORS – USER FEEDBACKS

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

# KEY FACTORS – USER FEEDBACKS

	TransA	TransB	TransC
<b>Overall</b>	Better than the previous Finnova MT	Still very Denglish, quite literal and lots of untranslated segments	Could be better.
<b>Retranslation</b>	Had to be retranslated	Large proportion of text needed to be retranslated	Several sentences needed.
<b>Word order</b>	Improved compared to previous Finnova MT. Verbs should be placed	Still very German, or rearranged in some totally inexplicable way	Very frequently incorrect, and lots of missing words.
<b>not EN words</b>	a few words that they were part of the project	Some words	As usual- mostly in words with hyphen.
<b>Terminology</b>	Overall good	Some use	Sometimes inconsistent, different translation used for the same term.
<b>Speed</b>	Due to the improvement regarding the placement of the verbs in the machine-translated text, I perceived a 10-15% improvement in translation speed.		Slower than last time.
<b>Short sentences</b>	In most of the cases, they could be confirmed as they were.		Several of them had to be retranslated, too.
<b>Overall compared to previous finnova MT</b>	Due to the improvement regarding the placement of the verbs in the machine-translated text, better than the previous Finnova MT	I fail to see much, if any, improvement.	Much worse than in July. Missing words, incorrect word order, lots of editing needed.

Due to the improvement regarding the placement of the verbs in the machine-translated text, better than the previous Finnova MT

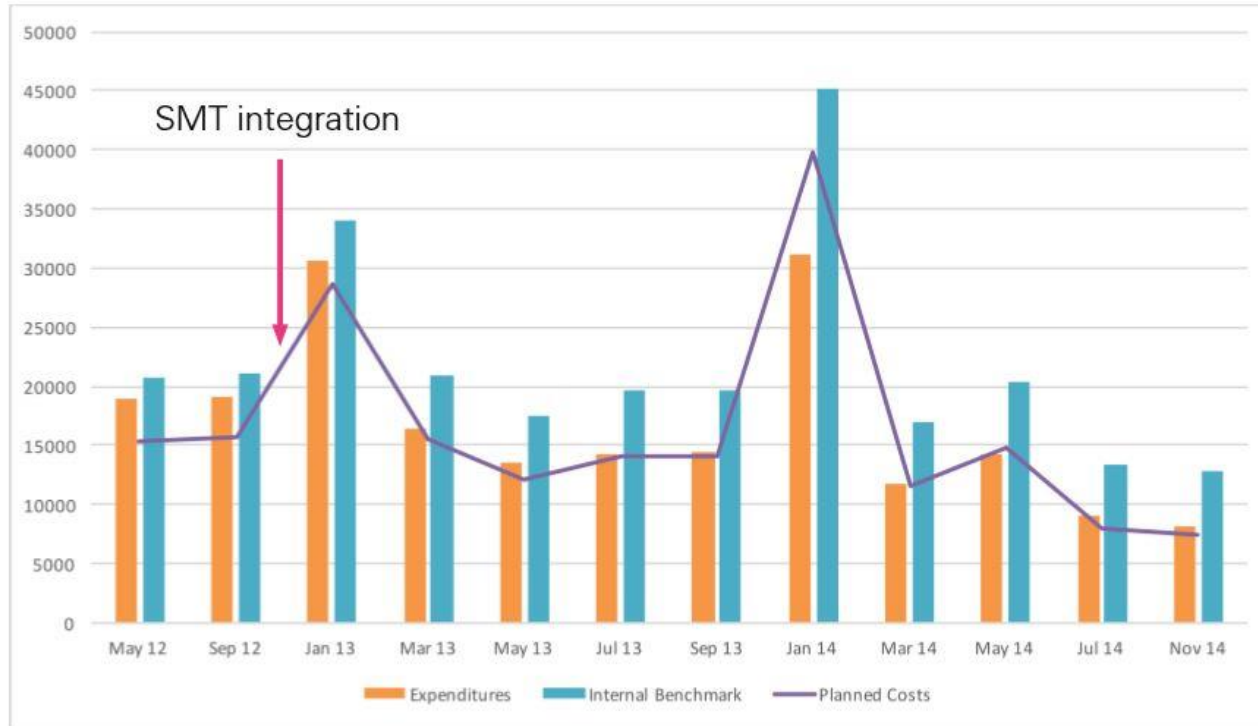
I fail to see much, if any, improvement.

Much worse than in July. Missing words, incorrect word order, lots of editing needed.



# KEY FACTORS

## Expenditures before and after the SMT system integration



# AGENDA

IT-Managers ...

Key factors

► **Do's and Don'ts**

Future developments



# DO'S AND DON'TS

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

# AGENDA

IT-Managers ...

Key factors

Do's and Don'ts

➤ **Future developments**



**... DANN MUSST  
DU DEUTSCH  
SPRECHEN'.**

*WILLY BRANDT*



# TAKE-AWAYS

- » 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.

# PARTNERS



**University of  
Zurich** <sup>UZH</sup>

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).

# PARTNERS

## s y n t a x

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.

# PARTNERS



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!