RBMT - SMT - NMT
SUMMARY

✓ Intro

✓ MT: how is it applied?
  ✓ Process
  ✓ Customers

✓ RBMT – SMT – NMT
  ✓ Samples
  ✓ Demo

✓ Summary
# FACTS IN BRIEF

<table>
<thead>
<tr>
<th>COMPANY NAME</th>
<th>Yamagata Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESTABLISHED</td>
<td>1998 in Gent Belgium</td>
</tr>
</tbody>
</table>
| FOCUSED SERVICES | Human & Machine Translation (since 2009)  
                     Technical Writing  
                     Audio & Multimedia Localization  
                     Translation Check Tool: QA Distiller  |
| EMPLOYEES        | 40 in Gent, 10 in Bratislava          |
MT is used @ Yamagata Europe since 2009

Multiple technologies: RBMT (Systran), SMT (Moses, Systran), NMT (Systran)
MACHINE TRANSLATION | WHY ?

- Integrate it in your existing translation process
- No-other-choice-but-MT-scenario’s
MT | IN THE TRANSLATION PROCESS

Without MT:
MT | IN THE TRANSLATION PROCESS

Without MT:

With MT:
MT QUALITY | MAKE IT MEASURABLE

✓ Quality measurement:

✓ FTR (First Time Right Segment):
  ✓ No corrections needed, MT output = human translation

✓ AR (Almost Right Segment):
  ✓ 1-2 words need correction, MT output is almost correct

✓ HE (High Effort Segment):
  ✓ Retranslate is necessary, MT output is of bad quality
MT I PRESENT STATISTICS

SEGMENTS CHANGED: 508/846 (60.05%)
WORDS DELETED (RED): 1528/6790 (22.5%)
WORDS ADDED (GREEN): 1795/6790 (26.44%)
WORDS DELETED PE ONLY (RED): 1528/5494 (27.81%)
WORDS ADDED PE ONLY (GREEN): 1795/5494 (32.67%)
FTR: 338/846 (39.95%)
AR: 457/846 (54.02%)
HE: 51/846 (6.03%)

3) ID: # 2679856
File 00193950_ger.tmx
Source (% 0): Error message / screen display
MT TARGET (0/2): Fehlermeldung / Bildschirmzeige
PE TARGET (0/2): Fehlermeldung / Bildschirmzeige

4) ID: # 2679858
File 00193950_ger.tmx
Source (% 0): Remove the SD card and insert it again.
MT TARGET (1/8): Die SD-Karte entfernen und sie erneut einführen.
PE TARGET (0/7)(AR): Die SD-Karte entfernen und erneut einführen.

13) ID: # 2679867
File 00193950_ger.tmx
Source (% 0): You attempted to update a module (application) in use.
MT | TRANSLATOR ACCELERATOR

- TM + MT as pretranslation
- Increase speed
- Reduce translation costs
- Applicable to: technical documents & manuals
MT | NO-OTHER-CHOICE

✓ Raw MT, no human translator involved
✓ Quick access to information
✓ Access to information that was previously unavailable
✓ Access to huge amount of information
✓ Applicable to: user generated content, database content, company internal information, warranty data, user feedback, tweet, service documentation …
MT | TECHNOLOGIES

- Rule based
- Statistical
- Neural MT
MT | RULE BASED MT

- First development in 50’s, 6 rules from Russian into English – IBM+UNI Georgetown, Cold war

- First commercial systems on the market 60’s
RBMT technology generates translations on the basis of morphological, syntactic and semantic analysis of source and target language.
Translate English words into French

- a = indef.article -> un/une
- girl = noun -> fille
- eats = verb -> manger (infinitive)
- a = indef.article -> un/une
- green = adjective -> vert (male form)
- apple = noun -> pomme

- Mapping dictionary entries into appropriate inflected forms & reordering.

- Final result: *Une fille mange une pomme verte.*
**RBMT PRO & CON**

**Pros:**
- Output is consistent – Terminology
- Stable quality
- Engine improvement via glossary updates
- No need for huge bilingual corpus

**Cons/Requirements:**
- Limited number of supported languages
- Output is more mechanical
MT | STATISTICAL MT

- In the 80’s the foundations are build for commercial SMT
  - Trados founded in 1984

- In the 90’s two important factors:
  - Increased of capacity of computational power + storage
  - large volumes of bilingual data.

- 2000-2010:
  - Commercialization of MT
  - 2007: MOSES is launched
SMT technology generates translations on the basis of statistical models derived from bilingual text corpora.

= splitting every sentence into N-grams; where N is a number predefined by the user.

This is a very simple example sentence (split into 2-grams)

<s> This
This is
is a
a very
very simple
simple example
example sentence
sentence </s>
Translating word groups instead of single words

“finding the most probable translation”

For example:

<table>
<thead>
<tr>
<th>Original</th>
<th>Translation</th>
<th>Possibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is</td>
<td>das ist</td>
<td>0.8</td>
</tr>
</tbody>
</table>

There’s an 80% chance that “das ist” will be translated as “this is”

In other words:

“this is” is translated as “das ist” in 80% of the occurrences in your training data
MT | STATISTICAL MT

✓ Pros:
  ✓ Any language can be handled
  ✓ Output is often more fluent (no literal translations)
  ✓ Fast engine building

✓ Cons/Requirements:
  ✓ You need a bilingual corpus (200,000 segments)
  ✓ In-domain data
  ✓ Corpus should be of high quality (golden data)
  ✓ Quality deterioration possible when training engine with low quality data
NMT | NEURAL MT - WHY

✓ Framework is available:
  ✓ RBMT, SMT, Hybrid MT is widely accepted and used
  ✓ Deep learning algorithms are available
  ✓ Data to learn is available
  ✓ Computational power available (GPU’s vs. CPU)

✓ OpenNMT, Nematus, GoogleNMT, SystranNMT

✓ Uses artificial intelligence to provide translations
ONLINE SELF-SERVICE PORTAL | YMTS

Machine Translate Files

- **Engine**: General_EN_FR
- **File Upload**
  - Attach File 1
  - Attach File 2
  - Attach File 3
- **Awaiting Translation**
  - Date: 28/01/2016
    - File: Sample1.docx
  - Date: 28/01/2016
    - File: test.docx
- **Translated Files**
  - Date: 28/01/2016
    - File: test.docx
    - Engine: General_EN_FR

www.yamagata-europe.com
Listening to music on a portable audio device, etc., by connecting with an audio cable (AUDIO IN).

When the indicator flashes slowly in orange, the battery is running low and needs to be charged.

“Noise suppression” and “Echo cancellation” enable phone calls with high-quality sound.

Musik auf einem tragbaren Audiogerät, einem etc., durch das Anschließen an ein Audiokabel hören (AUDIO HEREIN).

Wenn der Indikator langsam in der Orange blitzt, läuft die Batterie niedrig und muss aufgeladen werden.

„Rauschunterdrückungs-“ und „Echoannullierung“ ermöglichen Telefonanrufen mit hochwertigem Ton.
This sample reveals itself as MT right away. Grammar not correct, not fluent, mechanical.

The second sentence is funny: “Wenn der Indikator langsam in der Orange blitzt” (When the indicator slowly flashes in the orange (the FRUIT, not the color 😊)).

But: For people with an understanding on the subject, this might be of any help when no other translation is available.
Listening to music on a portable audio device, etc., by connecting with an audio cable (AUDIO IN)
When the indicator flashes slowly in orange, the battery is running low and needs to be charged.
“Noise suppression” and “Echo cancellation” enable phone calls with high-quality sound.

Wiedergeben von Musik von einem tragbaren Audiogerät o. Ä. Über eine Verbindung mit einem Audiokabel (AUDIO IN)
Wenn die Anzeige langsam orange blinkt, wird der Akku schwächer und muss aufgeladen werden.
Dank der Rausch- und „Echounterdrückung“ die Telefonieren mit hoher Qualität.
SMT is much better and only shows minor mistakes, except for the last sentence, which is clearly MT.

But: understandable

With some extra training, suitable for PE.
Listening to music on a portable audio device, etc., by connecting with an audio cable (AUDIO IN)
When the indicator flashes slowly in orange, the battery is running low and needs to be charged.
“Noise suppression” and “Echo cancellation” enable phone calls with high-quality sound.

Wiedergabe von Musik auf einem tragbaren Audiogerät usw. durch Anschließen eines Audiokabels (AUDIO IN)
Wenn die Anzeige langsam orange blinkt, ist die Batterie fast leer und muss geladen werden.
"Rauschunterdrückung" und "Echounterdrückung" ermöglichen Telefonate mit hochwertigem Ton.
NMT ist by far the best to read and really fluently translated.

The engine just did not realize that “battery” should be translated as “Akku” (as SMT did), as a “Batterie” in German is something that cannot be recharged. But as the English doesn’t use “rechargeable battery” either, this might be forgiven.
## Comparision | RBMT – SMT - NMT

<table>
<thead>
<tr>
<th></th>
<th>RBMT</th>
<th>SMT</th>
<th>NMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluency</td>
<td>*</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Contextual Accuracy</td>
<td>*</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Consistency</td>
<td>***</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Languages</td>
<td>*</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Terminology</td>
<td>***</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Amount of data needed for training</td>
<td>***</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>Retraining</td>
<td>***</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Short sentences</td>
<td>***</td>
<td>***</td>
<td>**</td>
</tr>
</tbody>
</table>
MT | WHICH TECHNOLOGY TO SELECT?

✓ The technology that fits your purpose, taking into account:

✓ The volume of data that you have available for training the software

✓ The volume of data you wish to translate using MT

✓ The content you wish you translate using MT

✓ We are convinced that with these 3 technologies, a lot is possible
Q&A
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://lt12.honestly.de

or scan the QR code
CONTACT

Tiene Vertriest
Tiene.vertriest@yamagata-europe.com

Or visit our website! www.yamagata-europe.com
Or visit our booth at the Asian Pavilion in the entrance hall!
WE MAKE EVERYBODY UNDERSTAND YOUR MESSAGE

www.yamagata-europe.com