Neural Machine Translation for Increased Human Translation Efficiency at Migros Bank

SwissText 2019
Winterthur, 18 June 2019

Stephan Wick
COO, Migros Bank AG

Samuel Läubli
Partner, CTO, TextShuttle AG
Business Case
<table>
<thead>
<tr>
<th></th>
<th>Balance Sheet Total</th>
<th>Annual Profit</th>
<th>Customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.7 Billion CHF</td>
<td>204 Million CHF</td>
<td>801’500</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Employees (FTE)</th>
<th>Branches</th>
<th>Standard &amp; Poor’s Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1’344</td>
<td>67</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>
Need for Translation

- Translation is vital (and in some areas required by law) in multilingual markets.
- In Switzerland, a translated page of text costs between CHF 50 and 115 when purchased from a language services provider (LSP).
Translation at Migros Bank – 2016

- Translation of ~6,000 standard pages per year
- German to French and Italian (CH variants)
- Completely outsourced to a language services provider (LSP)
- Decision to build up an in-house translation team (insourcing)
Reasons for Insourcing

- Cost saving through internal translation
- Improved quality
- Accelerated translation
Translation at Migros Bank – 2017

- In-house translation team with 2.8 FTEs
- Introduction of “CAT tool” (translation memory, termbases)
- 40% of translation volume handled by in-house team
- Goal: Increase volume handled by in-house team
- Decision: Use machine translation to increase translator productivity
Problem: Data Privacy

« [...] you give Google a perpetual, irrevocable, worldwide, sublicensable, royalty-free, and non-exclusive license to Use content submitted, posted, or displayed to or from the APIs through your API Client. "Use" means use, host, store, modify, communicate, and publish. Before you submit content to our APIs through your API Client, you will ensure that you have the necessary rights (including the necessary rights from your end users) to grant us the license. »

Source: Google APIs Terms of Service
https://developers.google.com/terms
Requirements

● Data privacy
  ● No text to leave company network
  ● On-premise training and deployment

● Quality
  ● Comparable to market leaders
  ● Swiss language variants
  ● In-house terminology

● Productivity
  ● Significant time savings for in-house translation team
Technology
Neural Machine Translation

1 training data / 2 hyperparameters / 3 decoding
Neural Machine Translation at Migros Bank

Model architecture:

- WMT17-style bi-RNN models (Sennrich et al., 2017)

Training data:

- In-domain translations (FR: 385k, IT: 187k)
- Out-of-domain translations (~6M sentences per language pair)
- Heuristic filtering (Zwahlen et al., 2016)
- Oversampling of in-domain translations
On-premise Deployment and Scalability

Diagram showing the flow of data processing with multiple DE-FR and DE-IT decoders connected to a Load Balancer and API.
Quality
Evaluation of Quality

- Blind comparison with DeepL
- Initial version of DE–FR and DE–IT systems without retraining
- 400 sentences
- 4 translators per target language
- Relative ranking: A is better / B is better / equal
As you can see we are standing here in front of the Shopping centre Kaufhaus Tyrol.

Wählen Sie die bessere Übersetzung oder gleichwertig, wenn beide Übersetzungen ungefähr gleich gut sind:

A

Wie Sie sehen, stehen wir hier vor dem Kaufhaus Tirol.

X

gleichwertig

B

Wie Sie sehen können, stehen wir hier vor dem Einkaufszentrum Kaufhaus Tyrol.
Results

System is comparable in quality to market leaders.

- 59.67% same or better than DeepL (DE–FR)
- 60.17% same or better than DeepL (DE–IT)
Productivity
How to Save Time with Machine Translation?

Ihre Abos sind so gut, dass ich sie auf jeden Fall weiterempfehlen werde.

Machine Translation

Their subscriptions are so good that I will totally recommend them.
How to Save Time with Machine Translation?

Ihre Abos sind so gut, dass ich sie auf jeden Fall weiterempfehlen werde.

Machine Translation

Their subscriptions are so good that I will totally recommend them.

Post-editing

Your subscriptions are so good that I will totally recommend them.
Evaluation of Translator Productivity

- Primary measure of interest
- MT Quality ≠ translator productivity
- Evaluation in two conditions:
  - TM-Only: Regular software and translation aids
  - Post-Edit: Regular software and translation aids + machine translation
- 2 Migros Bank translators per target language
- Translators
  - had been working with the machine translation system for 3 months, and
  - had received 4 hours of post-editing training
Procedure

<table>
<thead>
<tr>
<th></th>
<th>FR-1</th>
<th>FR-2</th>
<th>IT-1</th>
<th>IT-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 Minuten</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>60 Minuten</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>60 Minuten</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>60 Minuten</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
</tbody>
</table>

- **Without MT (TM-Only)**
- **With MT (Post-Edit)**

- 60 Minuten
Results: Speed

German to French

German to Italian

Note: Different X axis
## Results: Quality

<table>
<thead>
<tr>
<th>Criterion</th>
<th>French TM-Only</th>
<th>French Post-Edit</th>
<th>Italian TM-Only</th>
<th>Italian Post-Edit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Impression</td>
<td>4.50</td>
<td>4.50</td>
<td>4.75</td>
<td>4.88</td>
</tr>
<tr>
<td>Coherence</td>
<td>4.75</td>
<td>5.25</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Cohesion</td>
<td>4.75</td>
<td>4.50</td>
<td>5.25</td>
<td>5.00</td>
</tr>
<tr>
<td>Grammar</td>
<td>4.75</td>
<td>4.75</td>
<td>4.75</td>
<td>4.88</td>
</tr>
<tr>
<td>Style</td>
<td>4.50</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Cultural Adequacy</td>
<td>4.50</td>
<td>4.75</td>
<td>4.50</td>
<td>4.75</td>
</tr>
</tbody>
</table>

Each text scored blindly by two translation experts (ZHAW). 1 = worst, 6 = best.
## Summary of Results

<table>
<thead>
<tr>
<th>Measure</th>
<th>TM-Only</th>
<th>Post-Edit</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words / h</td>
<td>584.81</td>
<td>934.14</td>
<td>59.74 %</td>
</tr>
<tr>
<td>Quality</td>
<td>4.50</td>
<td>4.50</td>
<td>0.00</td>
</tr>
</tbody>
</table>

German to French

<table>
<thead>
<tr>
<th>Measure</th>
<th>TM-Only</th>
<th>Post-Edit</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words / h</td>
<td>452.67</td>
<td>494.57</td>
<td>9.26 %</td>
</tr>
<tr>
<td>Quality</td>
<td>4.75</td>
<td>4.88</td>
<td>0.13</td>
</tr>
</tbody>
</table>

German to Italian
Results: Discussion

- Human translation is faster with machine translation (post-editing)

- Productivity varies between translators:
  - Some use speed-up to increase quality
  - Others use speed-up to finish jobs faster
  - Translator training in post-editing is vital

- Productivity varies between target languages:
  - DE–FR better than DE–IT system
  - Similar results when compared to DeepL
  - Less training material available (both in- and out-of-domain)
Conclusions
Conclusions

Custom machine translation systems

- are equal or better than DeepL for 60% of translated sentences;
- make human translation 35% faster.

Use at Migros Bank:

- In-house translation team now handles 60% of translation volume (vs. 40% without machine translation).
- Discounts negotiated with external service providers for remaining volume (only post-editing required).
- Quality improvements since initial evaluation through re-training.
Thanks for your attention.
References
