Salesforce NMT System: A Year Later

Raffaella Buschiazzo
Director, Localization @ Salesforce

Virtual MT Summit 2021: Building MT Capacity and Competence in-house
Agenda

Salesforce MT Overview
Primary Use Case
What's Done
MT Quality
2021 Roadmap
Future Applications
Salesforce MT Overview

A 4-year collaboration between R&D Localization and Salesforce Research teams

NMT system:

- Built on Salesforce domain
- Language-agnostic architecture with models per language
- Leveraging Salesforce data and publicly available pretrained models (mBART, XLM-R, etc.)

Goals:

- Reduce translation time by enhancing translators’ productivity
- Increase content accuracy/freshness by publishing updates on-time/more frequently > Increase case deflection
- For selected markets, eliminate translation cost by publishing raw MT or reduce cost through light PE
- Reinvesting savings into high-value content/products

Primary Use Case: Salesforce Online Help

Languages

English
Français
Deutsch
Italiano
日本語
Español (México)
Español
中文 (简体)
中文 (繁體)
한국어
Русский
Português (Brasil)
Suomi
Dansk
Svenska
Nederlands
ภาษาไทย
Norsk

3 major releases
Translated 3 x year
New feature/product terminology per major release
Authored in DITA XML (200+ tags)
What’s Done

Achievements in the last year

Implemented SF MT as a standard localization process for Core Help:

- 100% of Salesforce Help is MTed and PEd for all 16 languages.
- Developed plugin to track MT quality systematically.
- Trained our translators on MTPE best practices.
- Reduced training time for the MT models from 1 day to 2/3 hours per language.
MT Quality - Part 1

Initially

- BLEU score
- Conduct human evaluations at each MT system iteration. Translators evaluated 500 MTed strings using 1/2/3 categorization:
  - 1 - Translation is ready for publication
  - 2 - Translation is useful but needs human post-editing
  - 3 - Translation is useless
- Plus overall feedback provided by our translators after post-editing 100K new words + 300K fuzzies per major release.
In 2020-21

- We started using **PEMp (Post-Editing Modification %)** on every PEd segment.
- Calculated using an algorithm respecting the 'Damerau–Levenshtein' edit distance
- Counts the minimum number of operations needed to transform one string into the other where an operation is defined as an insertion, deletion, or substitution of a single character, or a transposition of two adjacent characters.
MT Quality: PEMp Scores/5 Releases

<table>
<thead>
<tr>
<th>Language</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ja-JP</td>
<td>64.22%</td>
<td>74.30%</td>
<td>81.60%</td>
<td>81.94%</td>
<td>81.35</td>
</tr>
<tr>
<td>da-DK</td>
<td>86.12%</td>
<td>89.87%</td>
<td>92.77%</td>
<td>91.67</td>
<td></td>
</tr>
<tr>
<td>de-DE</td>
<td>72.55%</td>
<td>82.76%</td>
<td>82.25%</td>
<td>82.71</td>
<td></td>
</tr>
<tr>
<td>es-MX</td>
<td>89.08%</td>
<td>95.74%</td>
<td>89.18%</td>
<td>89.55</td>
<td></td>
</tr>
<tr>
<td>fr-FR</td>
<td>81.84%</td>
<td>86.58%</td>
<td>86.12%</td>
<td>86.58</td>
<td></td>
</tr>
<tr>
<td>nb-NO</td>
<td>81.13%</td>
<td>84.39%</td>
<td>87.73%</td>
<td>88.51</td>
<td></td>
</tr>
<tr>
<td>pt-BR</td>
<td>86.67%</td>
<td>93.73%</td>
<td>94.15%</td>
<td>93.97</td>
<td></td>
</tr>
<tr>
<td>sv-SE</td>
<td>84.57%</td>
<td>90.18%</td>
<td>93.36%</td>
<td>94.61</td>
<td></td>
</tr>
<tr>
<td>ko-KR</td>
<td>81.84%</td>
<td>88.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fi-FI</td>
<td>87.90%</td>
<td>77.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>it-IT</td>
<td></td>
<td>85.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nl-NL</td>
<td></td>
<td>81.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ru-RU</td>
<td></td>
<td>86.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zh-CN</td>
<td></td>
<td>85.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zh-TW</td>
<td></td>
<td>81.93</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average all languages: 86.35%
## MT Quality: Unedited Segments

<table>
<thead>
<tr>
<th>Language</th>
<th>Unedited segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ja-JP</td>
<td>28.67%</td>
</tr>
<tr>
<td>da-DK</td>
<td>46.17%</td>
</tr>
<tr>
<td>de-DE</td>
<td>29.04%</td>
</tr>
<tr>
<td>es-MX</td>
<td>51.78%</td>
</tr>
<tr>
<td>fr-FR</td>
<td>29.28%</td>
</tr>
<tr>
<td>nb-NO</td>
<td>32.36%</td>
</tr>
<tr>
<td>pt-BR</td>
<td>47.60%</td>
</tr>
<tr>
<td>sv-SE</td>
<td>53.17%</td>
</tr>
<tr>
<td>ko-KR</td>
<td>41.32%</td>
</tr>
<tr>
<td>fi-FI</td>
<td>39.34%</td>
</tr>
<tr>
<td>it-IT</td>
<td>31.86%</td>
</tr>
<tr>
<td>nl-NL</td>
<td>28.39%</td>
</tr>
<tr>
<td>ru-RU</td>
<td>30.96%</td>
</tr>
<tr>
<td>zh-CN</td>
<td>26.87%</td>
</tr>
<tr>
<td>zh-TW</td>
<td>23.71%</td>
</tr>
</tbody>
</table>

Average for 5 releases: 36.03%
2021 Roadmap

MT API for Continuous Localization

Test current model to MT Help from acquisitions

Publish raw MT for Help in four Nordic languages.
  - Track page view #, MT disclaimer in H&T, thumbs up/down report, PE most viewed pages.

Knowledge Articles:
Increase number of translated KAs
Reduce cost

Video subtitles
Future applications

Internal to SFDC

- Extend MT to 34 languages
- MT for UI label testing (like pseudo loc)
- MTPE on software localization
- Other content (ex: developer’s guides)

Customer-Facing/Product

- Case feed
- Experience Cloud
- Slacks apps

Make Salesforce MT API available for customers

- OOTB
- Trainable?