



# 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



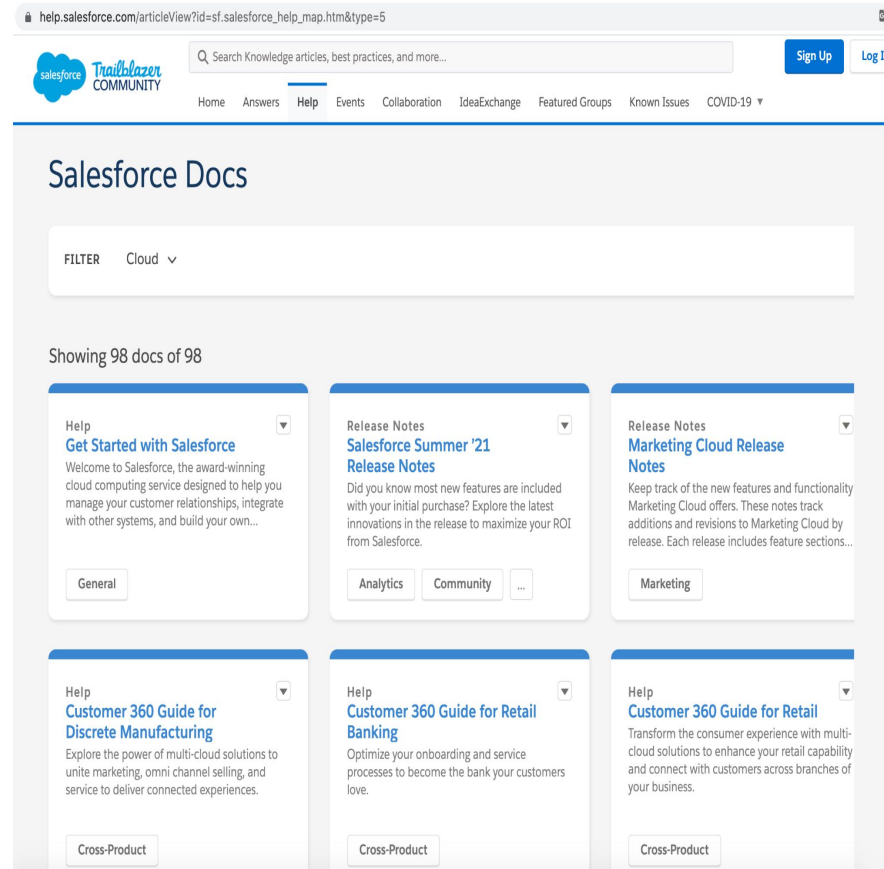
**US Patent No. 10,963,652 for  
"Structured Text Translation"**

# 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)

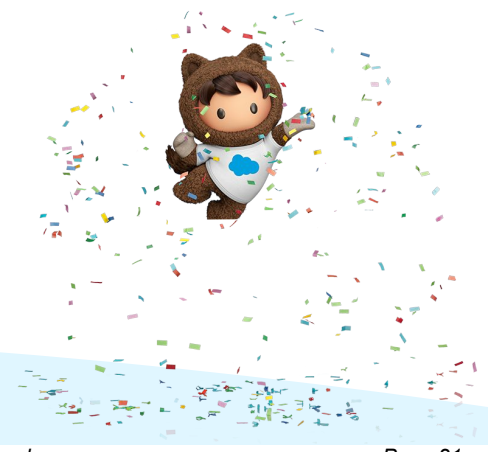


# What's Done

## 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

## Manual

### 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.

# MT Quality - Part 2

Automated



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



Language	R1	R2	R3	R4	R5
ja-JP	64.22%	74.30%	81.60%	81.94%	81.35
da-DK		86.12%	89.87%	92.77%	91.67
de-DE		72.55%	82.76%	82.25%	82.71
es-MX		89.08%	95.74%	89.18%	89.55
fr-FR		81.84%	86.58%	86.12%	86.58
nb-NO		81.13%	84.39%	87.73%	88.51
pt-BR		86.67%	93.73%	94.15%	93.97
sv-SE		84.57%	90.18%	93.36%	94.61
ko-KR				81.84%	88.09
fi-FI				87.90%	77.71
it-IT					85.99
nl-NL					81.07
ru-RU					86.18
zh-CN					85.39
zh-TW					81.93

Average  
all languages:  
**86.35%**



# MT Quality: Unedited Segments

Average for 5 releases

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



Average  
all languages:  
**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?





# Thank You