The 14th Conference of The Association for Machine Translation in the Americas

www.amtaweb.org

WORKSHOP PROCEEDING

Workshop on the Impact of Machine Translation

Organizers:
Sharon O'Brien (ADAPT, CTTS Dublin City University)
Michel Simard (National Research Council Canada)

©2020 The Authors. These articles are licensed under a Creative Commons 3.0 license, no derivative works, attribution, CC-BY-ND.
Contents

1 The Machine is Blind: Bottom-Up Feedback on the Impact of MT on Human Translation Performance
   Invited Speaker: Rhett Whitaker

18 Responsible ‘Gist’ MT Use in the Age of Neural MT
   Marianna J. Martindale

46 A Different, Ethical Machine Translation is Possible: English-Catalan Free/Open-Source Neural Machine Translation
   Vicent Briva-Iglesias

62 A Case Study of Natural Gender Phenomena in Translation: A Comparison of Google Translate, Bing Microsoft Translator and DeepL for English to Italian, French and Spanish
   Argentina Anna Rescigno, Johanna Monti, Andy Way and Eva Vanmassenhove

91 Empowering translators of marginalized languages through the use of language technology
   Alp Öktem, Manuel Locria, Eric Paquin and Grace Tang

123 Exploring Greater Impact: Business Translation Beyond Localization?
   Kirti Vashee

152 Predictive Translation Memory in the Wild - A Study of Interactive Machine Translation Use on Lilt
   Invited Speaker: Geza Kovacs
The Machine is Blind
Bottom-Up Feedback on the Impact of MT on Human Translation Performance

RHETT WHITAKER
OCTOBER 6, 2020
Blind spots and information flow
What do translators really think of MT?
The challenge facing translators
Impacts in the short and long term
Recommendations
Blind Spots and Information Flow
Know How Your Information Flows

Decision-makers
Executives and upper management

Facilitators
Middle management, service coordinators, project managers

Production
Translators, editors, and other linguists

Downstream
Upstream
What Do Translators Really Think of MT?
“MT frees me up for other, more valuable tasks.”
“This is more work than translating from scratch.”

“I don’t know why it’s making these errors.”
The Challenge Facing Translators
A Challenging Situation
TRANSLATORS BETWEEN A ROCK AND A HARD PLACE

LESS PAY
LSPs tend to prorate what they pay for MT post-editing services, sometimes to a significant degree.

MORE WORK
In particular, high-quality translators view poorly implemented MT as a hindrance to their work.

MORE TEDIOUS
Long intervals between retraining MT engines can lead to frustration on the translator’s part.

DEAF EARS
Bottom-up feedback that is ignored can act as a significant demotivating force.

“Is this kind of work still worth doing?”
Impacts in the Short and Long Term
Short term

Translators dealing with poorly implemented MT are often unfocused, unmotivated, and less effective. This could produce the following short term impacts:

- Translators increasingly reject MT post-editing jobs
- Translators raise rates to compensate for prorated pay
- Organizations see declining quality, similar overall costs, and diminished capacity
Long term

Sustained negative attitudes toward MT and frustration with the post-editing process can be a serious demotivating force for translators. This can produce the following long term impacts:

- Translators leave the talent pool permanently
- The availability of highly-skilled professionals drops below critical thresholds
- Organizations face serious challenges to profitability and ultimately an existential threat
Recommendations
Countermeasures

KEEPING YOUR EYES OPEN

- Be diligent in implementing MT
  Give yourself the best chance for successful MT integration and translator retention. Don’t rush implementation.

- Establish upstream feedback channels
  You can’t react to outcomes you don’t know about. Take measures to acquire reliable, actionable information.

- Don’t reduce rates prematurely
  Make sure your translators are fully on board with the concept of MT before you think about adjusting their pay.

- Act on upstream feedback
  Use feedback to improve your MT systems and processes. It can help keep translators at the top of their game and bolster profitability.
Thank You

We help customers manage their content and customer touchpoints to improve efficiency, increase revenue, reduce time to market and ensure quality and compliance.
Responsible Gist MT Use in the Age of Neural MT

Marianna J. Martindale, iSchool PhD Candidate
University of Maryland, College Park

Also: Computational Linguist, Center for Applied Machine Translation, USG

OBLIGATORY DISCLAIMER: Opinions in this talk are my own and not necessarily those of any part of the U.S. Government
What Makes Neural MT (NMT) Different?

• Scores well on automated metrics & human evaluations
• Improves many types of errors (especially fluency)
• More languages & platforms than ever

But...

Sometimes fails catastrophically
Humorous Catastrophic Failures

- Bark bark bark! Bark bark bark bark bark. BARK!
  - Good luck! God bless you. Good!
  - Automatically Translated

- bark bark bark. bark bark bark bark bark bark bark bark bark! bark...
  - Good luck. It's good for you! Good luck...
  - Automatically Translated
  - Just now  Like  Reply  More

- BARK bark bark bark bark barkbark bark bark!
  - Good morning! God bless you!
  - Automatically Translated

- BARK!! Barkbarkbark bark bark bark bark bark bark bark bark barkbark bark!
  - Blessed!! Bless you and bless you!
  - Automatically Translated

Facebook, 29 April 2020
(Semi-)Humorous Catastrophic Failures

INTERNET NEWS  JANUARY 18, 2020 / 12:27 PM / UPDATED 5 MONTHS AGO

Facebook says technical error caused vulgar translation of Chinese leader’s name

By Poppy McPherson  3 MIN READ

YANGON (Reuters) - Facebook Inc FB.O on Saturday blamed a technical error for Chinese leader Xi Jinping’s name appearing as “Mr Shithole” in posts on its platform when translated into English from Burmese, apologizing for any offense caused.
Dangerous Catastrophic Failures

https://www.haaretz.com/israel-news/palestinian-arrested-over-mistranslated-good-morning-facebook-post-1.5459427
Dangerous Catastrophic Failures

https://www.haaretz.com/israel-news/palestinian-arrested-over-mistranslated-good-morning-facebook-post-1.5459427
When are (N)MT Errors Dangerous?

• Output is believable (in context)
• Lack of means and/or motivation to verify
• Use case involves MT informing action
Believable Output

Believability = Fluency + Plausibility + Human Judgment

• Fluency: Does it “feel” like the target language?
  • Users more likely to trust fluent output (Martindale & Carpuat 2018)
  • NMT more likely to produce fluent but not adequate output (Martindale et al. 2019)

• Plausibility: Does it make sense?
  • MT output is more believable when it is plausible (Work in progress)

• Human: People use heuristics to judge credibility of information

---

When are NMT Errors Dangerous?

✓ Output is believable (in context)
- Lack of means and/or motivation to verify
- Use case involves MT informing action

Gist MT?
When are Gist MT Errors Dangerous?

Lack of means and/or motivation to verify?

Gist MT use characteristics

• High volume of foreign language text and/or tasks

• Impractical to translate everything or hire only bilinguals
  • Especially bilinguals with domain expertise

• Monolingual domain experts use MT to triage text or glean information

• Ideally: Bilinguals translate/evaluate documents/info monolinguals find
  • In practice: people may cut corners...
When are Gist MT Errors Dangerous?

Use case involves MT informing action?

Gist MT use examples

• Journalist looking for relevant, local Tweets after an event
• Business analyst monitoring press for info about foreign competitors
• Investigator checking social media as part of background check
“Information collected from social media, by itself, will not be a basis to deny refugee resettlement”

Official statement, September 2019
Example: USCIS Refugee Vetting

“Information collected from social media, by itself, will not be a basis to deny refugee resettlement” -- Official statement, September 2019

However...

• Incorrect MT could tip scales of suspicion (in either direction)
• Social media is out of domain from MT training
• Often low-resource languages
Is NMT for Gisting Worth the Risk?

• IMHO: Yes!

Good news:

• Truly misleading output is rare
• Faster to read, easier to understand
• Users like it

Just need to mitigate risk
How can we mitigate the dangers?

Dangers

• Output has errors
• Output is believable (in context)
• Lack of means and/or motivation to verify
• Use case involves MT informing action

Mitigation goals

• Error-free MT
• Encourage appropriate skepticism
• Make it easier to recognize potential errors
• Verify before acting
How can we mitigate the dangers?

**Dangers**

- Output has errors
- Output is believable (in context)
- Lack of means and/or motivation to verify
- Use case involves MT informing action

**Mitigation goals**

- Error-free MT
- Encourage appropriate skepticism
- Make it easier to recognize potential errors
- Verify before acting
Mitigation Strategies

Policy interventions
- Normative principles organizations with gist MT use cases should follow
- Changes to procedures and training

Technological interventions
- Changes to the technology environment or the technology itself
- Requires additional research and development
Policy Interventions

1. Independent, in-domain evaluation
2. Training for MT users
3. Workflows that require validation before action
P1: Independent, In-Domain Evaluation

Principle: An organization should not deploy or encourage the use of MT without independent evaluation in the domain(s) and language pair(s) it is intended to be used on.

- If the intended use shifts/expands, additional testing should be conducted

Why? MT quality varies by language/domain

Independent – Not conducted by the MT company

Domain – Style and/or topic

Evaluation – Formal or informal

- Evaluators should know source language
P2: Training for MT Users

Principle: Users should be trained to understand the technology well enough to expect variations in quality including dropped or hallucinated words and phrases.

Why? NMT is not intuitive! Hard to recognize what you don’t expect.

Example hands-on exercises:
- Change context window, capitalization, punctuation, etc and observe output changes
- Compare output from high- and low- resource languages
- Try to get the system to hallucinate (e.g. fake Hawaiian)
P3: Require Validation Before Action

Principle: Organizations with workflows that include critical decisions or actions informed by MT should require validation by someone who knows the source language before taking action.

Why? Establishing a consistent process deters corner-cutting.
  ◦ Even professional translation services rely on at least one level of quality control!

Considerations
  ◦ Level of validation proportionate to impact of action/decision
  ◦ E.g., Self-validation through other resources may be sufficient for minimal-impact actions/decision
Technological Interventions

1. Provide access to multiple MT outputs
2. Provide access to additional language resources
3. Build in “nudges” to help the user recognize quality issues
T1: Multiple MT Outputs

What: Display outputs from two or more MT systems/models

LOE: Moderate
- Obtain licenses and/or build models
- Modify/create interface to display

Why? Users can observe differences to flag possible errors

Anecdote: Users actually prefer this anyway!
T2: Additional Language Resources

What: Provide CAT-like tools to MT users

LOE: Low-Moderate

- Teach users features in existing services (e.g. Google Translate, Systran, Wiktionary, Linguee)
- Obtain access to resources (dictionaries/terminologies/TMs, etc)
- Integrate access alongside MT

Why?

- Individual word lookup can validate/clarify MT output
- Terminologies can resolve technical terms
- TM lookup can provide alternate contexts
T3: Nudges

What: Automatically flag questionable output
  ◦ Quality estimation
  ◦ Diff on multiple outputs

LOE: High
  ◦ QE is an open research area

Why? Draw user’s attention to problem areas
## Summary

<table>
<thead>
<tr>
<th>Dangers</th>
<th>Mitigation goals</th>
<th>Recommended Interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Output has errors</td>
<td>• Error-free MT</td>
<td>• (Continue improving)</td>
</tr>
<tr>
<td>• Output is believable (in context)</td>
<td>• Encourage appropriate skepticism</td>
<td>• P1 (Evaluation), P2 (Training), T3 (Nudges)</td>
</tr>
<tr>
<td>• Lack of means and/or motivation to verify</td>
<td>• Make it easier to recognize potential errors</td>
<td>• T1 (Multi-outputs), T2 (Lang resources), T3 (Nudges)</td>
</tr>
<tr>
<td>• Use case involves MT informing action</td>
<td>• Verify before acting</td>
<td>• P3 (Verify)</td>
</tr>
</tbody>
</table>
Conclusion

• There can be risks to gist MT use
• Steps can be taken to mitigate them
• These are just examples
• Stakeholders should be looking at these mitigations and others
  • Organizational leadership
  • MT integrators
  • MT researchers
• See also: AI Ethics
For further information or questions contact:
Marianna J. Martindale
mmartind@umd.edu
A Different, Ethical MT is Possible:

English-Catalan Free/Open-Source NMT

Vicent Briva-Iglesias

SFI Centre for Research Training in Digitally-Enhanced Reality (D-REAL), Dublin City University
Overview

01
Understanding the Problem
Current context of MT.

02
Variables
MT engines and methodology.

03
Results
Relative ranking, quality, and post-editing evaluation.

04
iMpacT
Effects and use-cases.
Understanding the Problem

01 Catalan context
Minoritized, stateless language. Low-resource.

02 NMT Requirements
Huge computational power (GPUs). Difficulty to find high-quality corpora for low-resource languages.

03 Literacy
You have the corpora. Now, how is an MT engine trained?

04 Data Privacy
Confidential information may be at stake.
What is the iMpacT of open-source MT for low-resource languages?

1. Which MT engine evaluated [Apertium, Softcatalà, Google] offers a higher translation quality?

2. Which MT engine evaluated offers a bigger productivity increase when introducing it into a translation workflow?

3. Can a free/open-source MT engine for a low-resource language beat the flagship MT engine for the English-Catalan language combination?
Variables – MT Engines

Apertium
- Free/Open-Source RBMT engine
- Originally developed for close languages (e.g. ES-CA)

Softcatalà Translator
- Free/Open-Source EN<>CA NMT engine (OpenNMT)
- Trained with TMs from the Softcatalà project («in-domain»)

Google Translate
- Flagship of commercial MT
- NMT from 2020
- Thousands of language combinations (including CA)
Variables – Text

HomeAssistant.io

• Open-source smart home software (GitHub)

• Preparation of the text with Okapi Framework

• Segments chosen randomly for the creation of the samples to be evaluated
Methodology – Human Evaluation 1

Relative Ranking
11 professional evaluators.
200 segments.

Rànquing de TA (Rank Comparison)

Source (English (United Kingdom))

Current: This entity does not have a unique ID, therefore its settings cannot be managed from the UI.

Next: The (platform) integration is not loaded.

Target (Catalan)

Aquesta entitat no té un ID únic, per tant la seva configuració no es pot gestionar des de la IU.

Aquesta entitat no té un ID únic, per tant, la seva configuració no es pot gestionar des de la interfície d’interès.

Aquesta entitat no té un únic ID, per tant no es poden abastar els seus paràmetres des del UI.

Comments

Characters left: 500
Methodology – Human Evaluation 2

Adequacy & Fluency

11 professional evaluators.
100 segments.

Precisió i fluïdesa S2, TA2

Source (English (United Kingdom))

Start

Current: This service is run by our partner, a company founded by the founders of Home Assistant and Hass.io.

Next: Go to the integrations page.

Target (Catalan)

Start

Current: Aquest servei el gestiona el nostre soci, una empresa fundada pels fundadors de Home Assistant i Hass.io.

Next: Vés a la pàgina d'integracions.

Fluency:

- Incomprehensible
- Difficult
- Good
- Flawless

(A More Info)

Adequacy:

- None
- Little
- Most
- Everything

(A More Info)
Methodology – Human Evaluation 3

Post-Editing Evaluation

6 evaluators (2 groups of study: professionals & volunteers).
2 texts of 100 segments.

Information

Required Level of Quality: Similar or equal to human translation
Content Type: User Interface Text
Filename: PE_Sample1_TANS_taus_xlsx_empty_prod-qual.xlsx
Segment: 1 of 100

Source: English (United Kingdom)

Start

Current: This entity does not have a unique ID, therefore its settings cannot be managed from the UI.

Next: The (platform) integration is not loaded.

Target: Catalan

Start

Current: Aquesta entitat no té un ID únic; per tant la seva configuració no es pot gestionar des de la UI.
## Post-Editing Evaluation (Explanation)

<table>
<thead>
<tr>
<th>Evaluator</th>
<th>Text 1, Engine 1</th>
<th>Text 1, Engine 2</th>
<th>Text 2, Engine 1</th>
<th>Text 2, Engine 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator 1</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Evaluator 2</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Evaluator 3</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Evaluator 4</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>
Results – MT Ranking

% of times an engine has received Ranking 1 evaluation

- Google Translate: 40.6%
- Apertium: 19.7%
- Softcatalà Translator: 39.7%

Ranking distribution per engine (in %)

- Softcatalà Translator: 62.27%, 30.36%, 7.36%
- Google Translate: 58%, 32.54%, 9.45%
- Apertium: 82.63%, 11.63%, 5.72%
Results – Fluency & Adequacy

Fluency (in %)

<table>
<thead>
<tr>
<th>Translator</th>
<th>0</th>
<th>0,5</th>
<th>5,5</th>
<th>25</th>
<th>45</th>
<th>24,5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Softcatalà Translator</td>
<td>65</td>
<td>31,5</td>
<td>3.5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google Translate</td>
<td>61</td>
<td>33</td>
<td>5,5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apertium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adequacy (in %)

<table>
<thead>
<tr>
<th>Translator</th>
<th>0</th>
<th>0</th>
<th>4,5</th>
<th>6,5</th>
<th>7,5</th>
<th>21</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Softcatalà Translator</td>
<td>60,5</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google Translate</td>
<td>48,5</td>
<td>45</td>
<td>0</td>
<td>6,5</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apertium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- Flawless (4)
- Good (3)
- Disfluent (2)
- Incomprehensible (1)
- Everything (4)
- Most (3)
- Little (2)
- None (1)
## Results – Post-Editing Productivity
(group of study 1: Softcatalà–Google)

<table>
<thead>
<tr>
<th></th>
<th>Softcatalà Translator</th>
<th>Google Translate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PE Time (s)</strong></td>
<td>Median</td>
<td>Median</td>
</tr>
<tr>
<td></td>
<td>3909.07</td>
<td>4131.64</td>
</tr>
<tr>
<td><em><em>Edit Distance</em> (segment)</em>*</td>
<td>9.79</td>
<td>10.35</td>
</tr>
</tbody>
</table>

222.563 seconds of difference; 5.69% productivity increase

<table>
<thead>
<tr>
<th></th>
<th>1-5 words</th>
<th>6-15 words</th>
<th>16 or &gt;16 words</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PE Time (s)</strong></td>
<td>Softcatalà</td>
<td>Median</td>
<td>8.15</td>
</tr>
<tr>
<td></td>
<td>Google</td>
<td>Median</td>
<td>9.41</td>
</tr>
<tr>
<td></td>
<td>Google</td>
<td>Median</td>
<td>18.44</td>
</tr>
<tr>
<td></td>
<td>Google</td>
<td>Median</td>
<td>20.08</td>
</tr>
<tr>
<td></td>
<td>Google</td>
<td>Median</td>
<td>34.08</td>
</tr>
<tr>
<td></td>
<td>Google</td>
<td>Median</td>
<td>33.67</td>
</tr>
</tbody>
</table>

|                     | Softcatalà | 5.34       | 11.53 | 9.79     |
|                     | Google     | 12.22      | 9.31  | 11.20    |

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 - 9, 2020, Workshop on the Impact of Machine Translation
## Results – Post-Editing Productivity (group of study 2: Softcatalà–Apertium)

<table>
<thead>
<tr>
<th></th>
<th>Softcatalà Translator</th>
<th>Apertium</th>
</tr>
</thead>
<tbody>
<tr>
<td><em><em>PE Time</em> (s)</em>*</td>
<td>Median</td>
<td>Median</td>
</tr>
<tr>
<td></td>
<td>1859.51</td>
<td>3743.41</td>
</tr>
<tr>
<td><em><em>Edit Distance</em> (segment)</em>*</td>
<td>6.81</td>
<td>24.85</td>
</tr>
</tbody>
</table>

1883.89 seconds of difference; 101.31 % productivity increase

<table>
<thead>
<tr>
<th>PE Time* (s)</th>
<th>Softcatalà</th>
<th>Apertium</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Median</td>
</tr>
<tr>
<td>1-5 words</td>
<td>5.95</td>
<td>14.18</td>
</tr>
<tr>
<td></td>
<td>14.11</td>
<td>28.64</td>
</tr>
<tr>
<td>6-15 words</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 or &gt;16 words</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Edit distance* (seg.)</th>
<th>Softcatalà</th>
<th>Apertium</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-15 words</td>
<td>6.21</td>
<td>10.73</td>
</tr>
<tr>
<td></td>
<td>40.65</td>
<td>37.76</td>
</tr>
<tr>
<td>16 or &gt;16 words</td>
<td>10.11</td>
<td>36.15</td>
</tr>
</tbody>
</table>
Normalisation
Low-resource languages gain presence on the Internet, society, etc.

Data Privacy
Confidential information is preserved.

Language Diversity
Avoid language shifts to predominant languages. And fosters language literacy.

Crisis Scenarios
Multilingual communication to reach everyone, e.g. COVID pandemic, natural disasters.

**Impact and Effects**
Thanks!

Do you have any questions?

Vicent Briva-Iglesias
D-REAL, Dublin City University
vicent.brivaiglesias2@mail.dcu.ie
@VicentBriva

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik and illustrations by Stories.
Gender bias in Neural Machine Translation

Argentina Anna Rescigno
Eva Vanmassenhove
Johanna Monti
Andy Way

6th October 2020

This work has been supported by the Dublin City University Faculty of Engineering & Computing and the University of Naples “L'Orientale” Department of Literary, Linguistic and Comparative Studies under the Erasmus+ Traineeship project number 2019-1-IT02-KA103-061753
Presentation Outline

- **Introduction**
  - A Note on Terminology
  - A Quick Problem Sketch

- **Experimental setup**
  - Compilation of Datasets
  - Description of the MT systems

- **Results & Analysis**

- **Three main points:**
  - Why does this kind of bias matter
  - What is its impact and on whom
  - Why we need to correct this bias

- **Conclusions and Future Work**
Introduction
Introduction: a note on terminology

Natural Gender

“Gender based on the sex or, for neuter, the lack of sex of the referent of a noun, as English girl (feminine) is referred to by the feminine pronoun she, boy (masculine) by the masculine pronoun he, and table (neuter) by the neuter pronoun it.”

http://www.collinsdictionary.com
## Introduction: a note on terminology

<table>
<thead>
<tr>
<th>Natural Gender</th>
<th>Grammatical Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Gender based on the <strong>sex</strong> or, for neuter, the lack of sex of the referent of a noun, as English girl (<strong>feminine</strong>) is referred to by the feminine pronoun she, boy (<strong>masculine</strong>) by the masculine pronoun he, and table (<strong>neuter</strong>) by the <strong>neuter</strong> pronoun it.”</td>
<td>“Gender based on arbitrary assignment, without regard to the referent of a noun, as in French ‘le livre’ (<strong>masculine</strong>), “the book,” and German ‘das Mädchen’ (<strong>neuter</strong>), “the girl.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Natural Gender</th>
<th>Grammatical Gender</th>
<th>Social Gender</th>
</tr>
</thead>
</table>
| “Gender based on the **sex** or, for neuter, the lack of sex of the referent of a noun, as English girl (**feminine**) is referred to by the feminine pronoun she, boy (**masculine**) by the masculine pronoun he, and table (**neuter**) by the **neuter** pronoun it.” | “Gender based on arbitrary assignment, without regard to the referent of a noun, as in French ‘le livre’ (**masculine**), “the book,” and German ‘das Mädchen’ (**neuter**), “the girl.” | - Embedded in the lexicon of many languages  
- Systematic structural bias.  
- Masculine forms the default for generic use. |
Introduction: a note on terminology

Romance Languages (e.g. ES, FR, IT)

- animate/persons/animals
  \[\text{grammatical gender} = \text{natural gender}\]

- inanimate objects
  \[\text{grammatical gender} = \text{arbitrary}\]
# Introduction: a note on terminology

<table>
<thead>
<tr>
<th>Romance Languages (e.g. ES, FR, IT)</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>● animate/persons/animals</td>
<td>● grammatical gender is not inflectional</td>
</tr>
<tr>
<td>↓ gramatical gender = natural gender</td>
<td>● <em>pronominal gender</em> → gender expressed through the pronouns = natural gender</td>
</tr>
<tr>
<td>● inanimate objects</td>
<td>● <em>gender-neutralization</em> of the language</td>
</tr>
<tr>
<td>↓ gramatical gender = arbitrary</td>
<td></td>
</tr>
</tbody>
</table>
Introduction: a quick problem sketch

A simple example:

I am happy!
lo sono contento!
lo sono contenta!

[Natural Gender]
[Grammatical Gender]

I am happy!
Je suis heureux!
Je suis heureuse!

[Natural Gender]
[Grammatical Gender]
# Introduction: a quick problem sketch

<table>
<thead>
<tr>
<th></th>
<th>Subject gender</th>
<th>Predicative nominative gender</th>
<th>Agreement?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English</strong></td>
<td>Mark is an efficient nurse.</td>
<td>M</td>
<td>covered</td>
</tr>
<tr>
<td><strong>Italian</strong></td>
<td>Mark è un’infermiera efficiente.</td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td><strong>French</strong></td>
<td>Mark est une infirmière efficace.</td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td><strong>Spanish</strong></td>
<td>Mark es una enfermera eficiente.</td>
<td>M</td>
<td>F</td>
</tr>
</tbody>
</table>

Nov 2019

➢ Lack of diversity → preference for masculine & gender-bias exemptions

➢ Agreement errors
Experimental Setup
Gender bias in MT

- personality adjectives
- profession nouns
- bigender nouns (in Italian)
  - minimal sentence “I am a(n)...
  - sentence with a referring adjective

<table>
<thead>
<tr>
<th></th>
<th>#</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjectives</td>
<td>136</td>
<td>(I, 2019a); (II, 2019a); (III, 2019)</td>
</tr>
<tr>
<td>Professions</td>
<td>107</td>
<td>(I, 2019b); (II, 2019b)</td>
</tr>
<tr>
<td>Bigender</td>
<td>30</td>
<td>(Cacciari et al., 1997); (Thornton and Anna, 2004)</td>
</tr>
</tbody>
</table>

Table 1: Overview of adjectives, profession and bigender nouns along with the sources from which they were retrieved.
## Compilation of Datasets

<table>
<thead>
<tr>
<th>Adjectives</th>
<th>#</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I, 2019a); (II, 2019a); (III, 2019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professions</td>
<td>107</td>
<td>(I, 2019b); (II, 2019b)</td>
</tr>
<tr>
<td>Bigender</td>
<td>30</td>
<td>(Cacciari et al., 1997); (Cacciari et al., 2011) (Thornton and Anna, 2004)</td>
</tr>
</tbody>
</table>

### Table 1: Overview of adjectives, profession and bigender nouns along with the sources from which they were retrieved

<table>
<thead>
<tr>
<th>English</th>
<th>Italian</th>
<th>French</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am an assistant.</td>
<td>Sono un assistente.</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>I am a beautiful assistant.</td>
<td>Sono una bellissima assistente.</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>I am an efficient assistant.</td>
<td>Sono un assistente efficiente.</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>I am a translator.</td>
<td>Sono un traduttore.</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>I am a beautiful translator.</td>
<td>Sono una bellissima traduttrice.</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>I am an efficient translator.</td>
<td>Sono un traduttore efficiente.</td>
<td>M</td>
<td>M</td>
</tr>
</tbody>
</table>
Description of MT systems

Google Translate

- 2003
- statistical MT system
- 2016 → neural MT system
- 2018 → double alternatives on word level
Description of MT systems

Google Translate

DeepL Translator

- 2017
- convolutional neural networks
- Linguee database (dictionary)
- nine languages supported
- provides not morphological alternatives
- serves also as glossary
Description of MT systems

- Google Translate
- DeepL Translator
- Bing Microsoft Translator

- originally a statistical MT system
- switched to a neural system
- does not provide alternatives but
- provides examples of usage
Results & Analysis
Results & Analysis

ADJECTIVES

<table>
<thead>
<tr>
<th>ADJ</th>
<th>GT</th>
<th>BMT</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>37.3</td>
<td>1.5</td>
<td>22.8</td>
</tr>
<tr>
<td>M</td>
<td>39.2</td>
<td>58.8</td>
<td>45.6</td>
</tr>
<tr>
<td>N</td>
<td>20.7</td>
<td>33.1</td>
<td>26.5</td>
</tr>
<tr>
<td>Other</td>
<td>2.8</td>
<td>6.5</td>
<td>5.1</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2: Results in % for male (M), female (F) and neutral (N) adjectives generated for EN → IT for GT, BMT and DL. The “Other” label includes all results obtained that do not correspond to the “adjective” category.
NOUNS

<table>
<thead>
<tr>
<th>NOUN</th>
<th>GT</th>
<th>BMT</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>35.8</td>
<td>0.9</td>
<td>7.5</td>
</tr>
<tr>
<td>M</td>
<td>46.1</td>
<td>60.4</td>
<td>60.4</td>
</tr>
<tr>
<td>N</td>
<td>17.6</td>
<td>28.3</td>
<td>28.3</td>
</tr>
<tr>
<td>Other</td>
<td>0.6</td>
<td>10.5</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Table 3: Results in % for male (M), female (F) and neutral (N) nouns generated for EN → IT for GT, BMT and DL. The “Other” label includes all results obtained that do not correspond to the “noun” category.
### Results & Analysis

<table>
<thead>
<tr>
<th></th>
<th>IT</th>
<th></th>
<th>FR</th>
<th></th>
<th>ES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>N</td>
<td>F</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>no adj.</td>
<td>10.0</td>
<td>86.7</td>
<td>Q*</td>
<td>10.0</td>
<td>63.3</td>
<td>26.7</td>
</tr>
<tr>
<td>beautiful</td>
<td><strong>63.3</strong></td>
<td>36.7</td>
<td>0.0</td>
<td>43.3</td>
<td><strong>56.7</strong></td>
<td>0.0</td>
</tr>
<tr>
<td>other adj.</td>
<td>13.3</td>
<td><strong>83.3</strong></td>
<td>Q*</td>
<td>3.3</td>
<td><strong>96.7</strong></td>
<td>0.0</td>
</tr>
<tr>
<td>DL</td>
<td>IT</td>
<td>FR</td>
<td>ES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no adj</td>
<td>30.0</td>
<td><strong>70.0</strong></td>
<td>0.0</td>
<td>20.0</td>
<td>63.3</td>
<td>16.7</td>
</tr>
<tr>
<td>beautiful</td>
<td><strong>83.3</strong></td>
<td>16.7</td>
<td>0.0</td>
<td><strong>73.3</strong></td>
<td>26.7</td>
<td>0.0</td>
</tr>
<tr>
<td>other adj</td>
<td>53.3</td>
<td>43.3</td>
<td>Q*</td>
<td>13.3</td>
<td><strong>83.3</strong></td>
<td>3.3</td>
</tr>
<tr>
<td>GT</td>
<td>IT</td>
<td>FR</td>
<td>ES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no adj</td>
<td>6.7</td>
<td><strong>93.3</strong></td>
<td>0.0</td>
<td>6.7</td>
<td><strong>90.0</strong></td>
<td>3.3</td>
</tr>
<tr>
<td>beautiful</td>
<td>43.3</td>
<td><strong>56.7</strong></td>
<td>0.0</td>
<td><strong>80.0</strong></td>
<td>20.0</td>
<td>0.0</td>
</tr>
<tr>
<td>other adj</td>
<td>3.3</td>
<td><strong>96.7</strong></td>
<td>0.0</td>
<td>3.3</td>
<td><strong>96.7</strong></td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 4: Results in % for male (M), female (F) and neutral (N) forms generated for EN → IT, FR and ES for BMT, DL and GT

- **beautiful**
- other adjectives:
  - **efficient**
  - **intelligent**
  - **sad**
  - **famous**
### Results & Analysis

<table>
<thead>
<tr>
<th>BMT</th>
<th>IT</th>
<th></th>
<th></th>
<th>FR</th>
<th></th>
<th></th>
<th>ES</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>N</td>
<td>F</td>
<td>M</td>
<td>N</td>
<td>F</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>no adj.</td>
<td>10.0</td>
<td>86.7</td>
<td>Q*</td>
<td>10.0</td>
<td>63.3</td>
<td>26.7</td>
<td>3.3</td>
<td>66.7</td>
<td>30.0</td>
</tr>
<tr>
<td>beautiful</td>
<td>63.3</td>
<td>36.7</td>
<td>0.0</td>
<td>43.3</td>
<td>56.7</td>
<td>0.0</td>
<td>66.7</td>
<td>33.3</td>
<td>0.0</td>
</tr>
<tr>
<td>other adj.</td>
<td>13.3</td>
<td>83.3</td>
<td>Q*</td>
<td>3.3</td>
<td>96.7</td>
<td>0.0</td>
<td>6.7</td>
<td>93.3</td>
<td>0.0</td>
</tr>
<tr>
<td>DL</td>
<td>IT</td>
<td></td>
<td></td>
<td>FR</td>
<td></td>
<td></td>
<td>ES</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>N</td>
<td>F</td>
<td>M</td>
<td>N</td>
<td>F</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>no adj.</td>
<td>30.0</td>
<td>70.0</td>
<td>0.0</td>
<td>20.0</td>
<td>63.3</td>
<td>16.7</td>
<td>3.3</td>
<td>76.6</td>
<td>20.0</td>
</tr>
<tr>
<td>beautiful</td>
<td>83.3</td>
<td>16.7</td>
<td>0.0</td>
<td>73.3</td>
<td>26.7</td>
<td>0.0</td>
<td>96.7</td>
<td>3.3</td>
<td>0.0</td>
</tr>
<tr>
<td>other adj.</td>
<td>53.3</td>
<td>43.3</td>
<td>Q*</td>
<td>13.3</td>
<td>83.3</td>
<td>3.3</td>
<td>6.7</td>
<td>93.3</td>
<td>0.0</td>
</tr>
<tr>
<td>GT</td>
<td>IT</td>
<td></td>
<td></td>
<td>FR</td>
<td></td>
<td></td>
<td>ES</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>N</td>
<td>F</td>
<td>M</td>
<td>N</td>
<td>F</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>no adj.</td>
<td>6.7</td>
<td>93.3</td>
<td>0.0</td>
<td>6.7</td>
<td>90.0</td>
<td>3.3</td>
<td>3.3</td>
<td>66.7</td>
<td>30.0</td>
</tr>
<tr>
<td>beautiful</td>
<td>43.3</td>
<td>56.7</td>
<td>0.0</td>
<td>80.</td>
<td>20.0</td>
<td>0.0</td>
<td>80.0</td>
<td>20.0</td>
<td>0.0</td>
</tr>
<tr>
<td>other adj.</td>
<td>3.3</td>
<td>96.7</td>
<td>0.0</td>
<td>3.3</td>
<td>96.7</td>
<td>0.0</td>
<td>3.3</td>
<td>96.7</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 4: Results in % for male (M), female (F) and neutral (N) forms generated for EN → IT, FR and ES for BMT, DL and GT

- beautiful
- other adjectives:
  - efficient
  - intelligent
  - sad
  - famous
- **From a linguistic point of view:**
  - Avoiding basic gender agreement mistakes

- **From a technological point of view:**
  - Solving these issues is not trivial (see attempts Google)
  - Black box of NLP (we have no/little control over the actual output that are being generated)

- **From a societal/ethical point of view:**
  - Identifying biases in current state-of-the-art systems is important so they don't end up getting mistaken for ‘objective’ translations
  - if an MT system is being used without human in the loop: real-world consequences
Break the cycle

→ more bias
  ○ social
  ○ gender
  ○ ethnic
  ○ ...
→ exclusion

which reflects human behaviour and prejudices

...acquire social prejudices
because they are built upon training data

from statistics

Algorithms...
Conclusion and Future Work
Conclusion and Future Work

Conclusion:
- Remove gender bias in training data
- Train algorithms to address the problem
- Stop using masculine “neutral” in machine learning texts
- Evaluation of gender phenomena is challenging

Future Work:
- Extend to other language pairs (different languages → different gender phenomena)
- Larger evaluation of more diverse set of words
- Create language specific challenge sets to evaluate how biased is an MT system
- Train our own MT system to verify whether machine bias influences the output of the translation
Thank you for your attention!
References


Contact info

Argentina A. Rescigno: argentina.res@gmail.com
Eva Vanmassenhove: vanmassenhove.eva@gmail.com
Johanna Monti: johmonti@gmail.com
Andy Way: andy.way@adaptcentre.ie
Empowering translators of marginalized languages through the use of language technology

Alp Öktem, Manuel Locria, Eric Paquin, Grace Tang
Language and COVID-19


- Where cases are rising fastest

% CHANGE (LAST 5 DAYS) OF COVID-19 CASES

- Active cases
- Per capita
- Language diversity

NUMBER OF LANGUAGES SPOKEN

1 - 540

47 AVG

TWB, OCHA, John Hopkins, TWB, OCHA, John Hopkins, TWB, OCHA, John Hopkins, © CARTO
Linguistic crisis response

STEP 1: Language Mapping
STEP 2: Recruitment
STEP 3: Training
STEP 4: Terminology
STEP 5: Translation
Linguistic crisis response
Hausa vs. French

# Volunteer translators

Word count/translator

Data from Kato: TWB’s translation platform during Covid-19 pandemic
How can language technology help to empower translators of marginalized languages?
Language data collection
parallel and audio data

MT model development
leveraging low-resource methodologies
Language data collection
parallel and audio data

MT model development
leveraging low-resource methodologies

Machine-assisted translation
tailored for non-professional translators
Language data collection
NMT for humanitarian impact

Language Data Disparity

Data has been consolidated from the OPUS collection of publicly available parallel corpora paired with English.

<table>
<thead>
<tr>
<th>Language</th>
<th>#Parallel sentences with English</th>
<th>Native speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>200.2m</td>
<td>76.8m</td>
</tr>
<tr>
<td>German</td>
<td>93.3m</td>
<td>90m</td>
</tr>
<tr>
<td>Dutch</td>
<td>75.1m</td>
<td>24m</td>
</tr>
<tr>
<td>Arabic (MSA)</td>
<td>69.2m</td>
<td>0</td>
</tr>
<tr>
<td>Turkish</td>
<td>52m</td>
<td>75.7m</td>
</tr>
<tr>
<td>Swahili</td>
<td>1.2m</td>
<td>150m</td>
</tr>
<tr>
<td>Swahili (Congo)</td>
<td>600k</td>
<td>22.26m</td>
</tr>
<tr>
<td>Hausa</td>
<td>400k</td>
<td>80m</td>
</tr>
<tr>
<td>Tigrinya</td>
<td>400k</td>
<td>9m</td>
</tr>
<tr>
<td>Kurmanji</td>
<td>300k</td>
<td>15m</td>
</tr>
<tr>
<td>Kanuri</td>
<td>300</td>
<td>8.6m</td>
</tr>
<tr>
<td>Arabic (Syria)</td>
<td>6</td>
<td>36.2m</td>
</tr>
<tr>
<td>Rohingya</td>
<td>0</td>
<td>1.8m</td>
</tr>
</tbody>
</table>
Gamayun kits

• Starting point for developing audio and text corpora for languages without pre-existing data resources.

• Four dataset versions:
  • Mini-kit - 5,000 sentences
  • Small-kit - 10,000 sentences
  • Medium-kit - 15,000 sentences
  • Large-kit - 30,000 sentences.

• Source sentences in English, Spanish, French

• Freely available from https://gamayun.translatorswb.org/
  • Currently mini-kits in Hausa, Kanuri, Rohingya, Swahili, Nande
MT model development
MT model development

- Languages: Levantine Arabic, Tigrinya, Congolese Swahili
- Main techniques employed:
  - Domain adaptation
  - Dialect adaptation
  - Cross-lingual transfer learning
  - Back-translation
Domain/dialect adaptation

- Levantine Arabic to English machine translation
- For social media content by Syrian refugees in Jordan
- Small in-domain data (5200 sentences)
- Modern Standard Arabic as base model
Domain/dialect adaptation

Manual evaluation of TWB’s Levantine Arabic MT for usability in social media monitoring
Domain/dialect adaptation

![Bar chart showing human preference]

- Human Baseline: 1.5
- TWB MT: 0.9
- Google MT: 0.6
Tigrinya NMT

• Semitic language with estimate # speakers of 7.9 million
• Refugee language in Europe and USA
• Hard-to-resource for translation
  • 3 active translators
  • %81 claimed in 2020
  • 72-day average delay
• Transfer learning from Amharic
Cross-lingual transfer learning and domain adaptation
Bidirectionality challenge:
- Tigrinya-to-English: 23.60 BLEU
- English-to-Tigrinya: 9.92 BLEU

More details on paper:

https://gamayun.translatorswb.org/
Machine-assisted translation
Interactive Machine Translation

• Proof-of-concept by Microsoft Research India
• Assisted translation through:
  • on-the-fly hints
  • suggestions
• Alternative to post-editing

Interactive Machine Translation

- Faster turnaround of document translations
  - compared to manual, and post-edited

<table>
<thead>
<tr>
<th>Word Coverage and Translation Gisting</th>
<th>Suggestions</th>
<th>Keystrokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similarly, knowledge for mental health is necessary.</td>
<td>Similarly,</td>
<td>Tab Tab Tab</td>
</tr>
<tr>
<td>In the same way, knowledge of knowledge is essential for mental health</td>
<td>In the knowledge</td>
<td>Tab Tab Tab Tab</td>
</tr>
<tr>
<td>In the same way, knowledge is essential for mental health</td>
<td>Thus, So the</td>
<td></td>
</tr>
<tr>
<td>In the same way, knowledge is essential for mental health</td>
<td>is essential</td>
<td></td>
</tr>
<tr>
<td>In the same way, knowledge is essential for mental health</td>
<td>is necessary</td>
<td>Enter ←</td>
</tr>
<tr>
<td>In the same way, knowledge is essential for mental health</td>
<td>is required to</td>
<td></td>
</tr>
</tbody>
</table>
Interactive Machine Translation

• Faster turnaround of document translations
  • compared to manual, and post-edited

• Human-machine collaboration to best leverage low-resource models

<table>
<thead>
<tr>
<th>Data Size</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>bn-en</td>
<td>1.1M</td>
<td>25.31</td>
<td>27.54</td>
<td>35.68</td>
</tr>
<tr>
<td>hi-en</td>
<td>1.5M</td>
<td>40.64</td>
<td>42.06</td>
<td>47.90</td>
</tr>
<tr>
<td>ml-en</td>
<td>897K</td>
<td>19.76</td>
<td>21.95</td>
<td>29.84</td>
</tr>
<tr>
<td>ta-en</td>
<td>428K</td>
<td>18.71</td>
<td>20.90</td>
<td>27.05</td>
</tr>
<tr>
<td>te-en</td>
<td>104K</td>
<td>11.92</td>
<td>14.57</td>
<td>21.17</td>
</tr>
</tbody>
</table>

Table 2: Multi-BLEU Score with x% of partial input

Interactive Machine Translation

• Faster turnaround of document translations
  • compared to manual, and post-edited
• Human-machine collaboration to best leverage low-resource models
• Boost for hard-to-source languages
  • for translation by non-experts
  • for crowdsourced data collection
Language data collection
parallel and audio data

MT model development
leveraging low-resource methodologies

Machine-assisted translation
tailored for non-professional translators
NMT for humanitarian impact

Diagram edited from Koehn and Knowles (2017)
Tigrinya NMT

SMT on 7 Ethiopian languages (Teferra Abate et al., 2018)

Parallel corpus of 300+ languages from jw.org (Agić and Vulić, 2019)

Available on OPUS repository (Tiedemann, 2012)

TWB’s translation memories

<table>
<thead>
<tr>
<th>Language</th>
<th>Ethiopian corpus</th>
<th>JW300</th>
<th>Bible-suedin</th>
<th>Global voices</th>
<th>GNOME</th>
<th>Tanzil</th>
<th>TWB</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amharic</td>
<td>66K</td>
<td>722K</td>
<td>61K</td>
<td>1.6K</td>
<td>57K</td>
<td>94K</td>
<td>-</td>
<td>1M</td>
</tr>
<tr>
<td>Ge’ez</td>
<td>11K</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11K</td>
</tr>
<tr>
<td>Tigrinya</td>
<td>36K</td>
<td>400K</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.5K</td>
<td>439K</td>
</tr>
</tbody>
</table>

Dataset sizes (#sentences) for Ge’ez scripted languages
## Gamayun kits

<table>
<thead>
<tr>
<th>Language</th>
<th>kit-5k</th>
<th>Audio</th>
<th>Language tech development goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hausa</td>
<td>✔</td>
<td></td>
<td>Machine-assisted data collection</td>
</tr>
<tr>
<td>Kanuri</td>
<td>✔</td>
<td></td>
<td>Machine-assisted data collection</td>
</tr>
<tr>
<td>Kurmanji Kurdish</td>
<td></td>
<td>✔</td>
<td>Machine-assisted survey transcription</td>
</tr>
<tr>
<td>Rohingya</td>
<td>✔</td>
<td>✔</td>
<td>Glossary with voice search</td>
</tr>
<tr>
<td>Coastal Swahili</td>
<td>✔</td>
<td>✔</td>
<td>MT and audio keyword detection</td>
</tr>
<tr>
<td>Congolese Swahili</td>
<td>✔</td>
<td></td>
<td>Interactive neural machine translation</td>
</tr>
<tr>
<td>Tigrinya</td>
<td></td>
<td></td>
<td>Interactive neural machine translation</td>
</tr>
</tbody>
</table>

**Proceedings of the 14th Conference of the Association for Machine Translation in the Americas**

October 6 - 9, 2020, Workshop on the Impact of Machine Translation
Interactive Machine Translation

How?

• Constrained decoding on top of OpenNMT models
• Latest development: BPE integration
• Work-in-progress: Evaluation with our volunteer translators

Demo

• https://microsoft.github.io/inmt/
BUSINESS TRANSLATION BEYOND LOCALIZATION

KIRTI VASHEE

AMTA 2020
THE GLOBAL VILLAGE IS A REALITY

We are connected as never before

Content increasingly defines the digital presence of the modern enterprise
CONTENT **REALLY** MATTERS IN THE DIGITAL MARKETPLACE

**DIGITAL TRANSFORMATION**
IS THE FUEL FOR ECONOMIC GROWTH

87% of companies believe digital transformation is a competitive opportunity

**GLOBALIZATION**
HAS GONE DIGITAL

50% of the world’s traded services are delivered digitally

**SECURITY**
REMAINS A TOP CONCERN

81% of companies expressed high levels of concern over data breaches
Since 2000, 52% of companies in the Fortune 500 have either gone bankrupt, been acquired, or ceased to exist as a result of digital disruption.

75% of today’s S&P 500 will be replaced by 2027

Innosight Research
Large volumes of multilingual data flows have created a huge and growing need for rapid translation
THE IMPACT OF DIGITAL TRANSFORMATION

Customers expect large volumes of relevant content available across all digital channels 24/7.

Content is the best salesperson for the active digitally savvy customer.

Rapid response with the right content is a requirement to be digitally relevant.

**Customer Journey**
- Awareness
- Consideration
- Decision
- Purchase
- Adoption
- Retention
- Expansion
- Advocacy

**Buyer Journey**
- Content drives CX
MT expands the reach of translation solutions into the heart of the enterprise

The potential to use unedited RAW MT continues to grow and increasingly enhances international business initiatives
MT makes all content instantly multilingual

Customers
- Listen
- Understand
- Communicate

Employees
- Collaborate
- Communicate
- Innovate

Partners
- Collaborate
- Leverage
- Co-create

MT works across ongoing data flows between stakeholders

Proceedings of the 14th Conference of the Association for Machine Translation in the Americas
October 6 - 9, 2020, Workshop on the Impact of Machine Translation
MT IN THE LOCALIZATION INDUSTRY

COST CONTAINMENT
PEMT EFFICIENCY
QUALITY MEASUREMENT

Ignores the transformational role of RAW MT when integrated with flowing enterprise content
Strategic MT use cases drive us to higher level discussions that are focused on mission-critical enterprise issues & C-Level concerns
Enterprise MT

Communication & Collaboration

Improved Global Agility & Responsiveness

Internal & External
Problem: Staff need to communicate and collaborate in real-time, globally, in their multiple languages, and listen and respond to global customers.

Where can translation be used in the Enterprise?

- Customer Support Content
- Product Design & Knowledge Sharing
- Customer Social Media Analysis
- Emails Chat Internal Reports
Content drives revenue and is critical to overall customer experience

Keep Customers
- Customer service
- Technical support
- Education + adoption
- Advice + best practices
- Personalized moments
- Personalized recommendations

Get Customers
- Thought leadership
- Brand awareness
- Buying research
- Sales Guidance
Enterprise MT

Global Customer Care & Support
Enhance the Global Customer Experience
Today, email and voice are top supported interactions; email and chat are to become top interactions within 12 months
(Any device, Any channel, Always on)

Contact Center 2.0 Research Report

This corresponds with the top challenges facing today’s contact centers, with companies ranking improving customer experiences and customer satisfaction in the top first and third spots, respectively.

“I love calling customer service!”
...said no customer ever.
QUALITY = DID IT SOLVE THE CUSTOMER PROBLEM

Easy
- 24/7
- Omni-channel access
- Multilingual

Fast
- Single interaction resolution
- Minimal Wait

Accurate
- Single source of truth
- Complete

Is support content available faster around the world?
Is it easily found?
Is it useful?
MT ENABLES BROAD GLOBAL REACH ACROSS ESCALATION TIERS

Self Service Knowledge Base
Interactive Chatbots
Multilingual Chat Enabled Live Agents

Translating millions of words in real-time without editing
Enterprise MT

eCommerce

Making Product Catalogues Global
eCommerce is one of the biggest transformations of commercial business practice in history.
Multilingual eCommerce

Online eCommerce Product Portfolios
• Allow rapid expansion of global buyers with multilingual Product Catalogues
• Rapidly expand global customer base

Expand into global markets in a cost effective way

Product Title
Product Description
Global User Reviews
Buyer <> Seller Communications
Transaction Related Pricing, Policies & Procedures
ECOMMERCE: THE FASTEST ACCESS TO THE GLOBAL MARKET

Top-Tier Markets
- United States
- United Kingdom
- China
- Japan
- South Korea
- Australia

Second Wave
- India
- Indonesia
- Mexico
- Brazil
- Saudi Arabia
- Sweden
- Switzerland

Wait and See
- Russia
- Argentina
- South Africa
- Nigeria

Source: Shopify
UNDERSTANDING MT QUALITY IN USE CONTEXT

Consumer Experience, Communication & Collaboration, eDiscovery

High translation volume: 10s of millions of words per day

Larger budgets > Accelerate global business agility & response

Limited post-editing possible

Linguistic steering and moderate customization produce positive outcomes

Localization

Low translation volume: 10s of thousands of words per day

Small budgets > Improve efficiency, reduce cost

Post-editing is critical

Requires deep, costly customization to enable positive PEMT outcomes
Linguistic Steering vs Post Editing

CX, Communication, Global Collaboration eCommerce eDiscovery use cases

Millions of words a day with little human touch: Real-time

Corpus and linguistic pattern level focus & linguistic feedback

Big Data Orientation

Localization Use Case

Thousands of words a day with multiple levels of human touch

Sentence level focus: Batch

PEMT focused culture

Published Content Orientation
<table>
<thead>
<tr>
<th><strong>CX, Communication, Collaboration eCommerce/eDiscovery use cases</strong></th>
<th><strong>Localization use case</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Millions of words a day</strong></td>
<td><strong>Thousands of words a day</strong></td>
</tr>
<tr>
<td>Massive volumes of unstructured content</td>
<td>Small volumes of structured and controlled content</td>
</tr>
<tr>
<td>Mission-critical data flow</td>
<td>Necessary for regulatory compliance-related data flows</td>
</tr>
<tr>
<td>Broad coverage encompassing all enterprise departments</td>
<td>Basic product documentation and high-level marketing and support content</td>
</tr>
</tbody>
</table>
The Translation Opportunity Beyond Localization

Develop large-scale translation ability

• Understand Linguistic Steering vs PEMT
• Understand how to solve dynamic, big-data translation challenges
• Understand corpus level linguistic profiling
• Identify internal and external high value content

Leverage multilingual content production
Looking at Opportunity Beyond Localization

Focus on the metrics that matter most

• Enhanced global communication and collaboration
• Expanded coverage & rapidity of response in global customer service/support scenarios
• Identify & Understand what customers care about across the globe
• Improved conversion rates in eCommerce

Improve the Customer Digital Experience
Thank You
Predictive Translation Memory in the Wild: A Study of Interactive Machine Translation Use on Lilt

Geza Kovacs
geza@lilt.com
Why Interactive MT?

• **Problem**: MT systems cannot guarantee correctness. Errors can affect business reputation

• A **human in the loop** is needed to ensure correctness

• **Interactive MT**: optimizing interactions between the translator and MT system
Post-editing: Translators edit MT output

An idea with a long history (Bisbey and Kay 1972)
Post-editing: Translators edit MT output

Source text

The physicist Arthur Eddington drew on Borel's image further in The Nature of the Physical World (1928), writing: If I let my fingers wander idly over the keys of a typewriter it might happen that my screed made an intelligible sentence.

MT suggestion

Le physicien Arthur Eddington a attiré sur l'image de Borel dans le caractère du monde physique (1928), écrit: Si je laisse mes doigts se promener les bras croisés sur les touches de la machine à écrire, il peut arriver que mon chape fait une phrase intelligible.

Image Source
Post-editing: Translators edit MT output

Pros

• Easy to implement (can use off-the-shelf MT system)
• Reduces translation time [1]

Cons

• Post-edited text is more similar to MT than unassisted translations [1]
• Translators can find post-editing frustrating [2]

Predictive Translation Memory

MT system suggests text predictions that complete the translation the user has already entered.

If the MT suggestion is correct, user can accept it; if it isn’t, user can type as normal.

MT suggestions update and improve as users type.
Transtype (Foster 2000)

The Canadian International Development Agency and the Canada Mortgage and Housing Corporation will be taking part in a conference which will deal with housing for the needy.

The conference will be held in the fall of 1987.

The Canada Mortgage and Housing Corporation is now looking into the possibility of financing further conferences and forums of this kind.

Opération africaine 2000 qui a été lancée par moi est une exemple de la détermination du Canada pour aider les gens des régions rurales d’Afrique à surmonter la famine et à briser le cycle de pauvreté.

L’agence canadienne de développement international

MT suggestion starting with “L’a”
Lilt’s Interactive MT

To enter text, user can:
1) Accept the MT-suggested word (Enter)
2) Accept the rest of the MT suggestion (Shift-Enter)
3) Just type normally

Source text

MT-suggested completion that continually updates so that it starts with the currently-entered translation
Interactive MT Implementation

Prefix-constrained MT model (based on Transformer architecture). Details at lilt.com/research
Interactive MT Implementation

Prefix-constrained MT model (based on Transformer architecture). Details at lilt.com/research
Interactive MT Implementation

Prefix-constrained MT model (based on Transformer architecture). Details at lilt.com/research
Interactive MT Implementation

Prefix-constrained MT model (based on Transformer architecture). Details at lilt.com/research

Source Text
“Fan Replacement Instructions”

Prefix
“Instruzioni per”

MT model

MT suggestion
“la sostituzione della ventola”
Interactive MT Implementation

Prefix-constrained MT model (based on Transformer architecture). Details at lilt.com/research
Interactive MT Implementation

Prefix-constrained MT model (based on Transformer architecture). Details at lilt.com/research
Interactive MT Implementation

Prefix-constrained MT model (based on Transformer architecture). Details at lilt.com/research
Interactive MT needs to be fast

New MT suggestion needs to be computed whenever the user’s entered text no longer matches the MT prediction.

90% of our MT requests are computed in less than 500ms
17 Click 1 or the topic for details:
Fare clic su 1 o sull'argomento per informazioni dettagliate:

18 Option 1: Receive and install a new fan.
Opzione 1: ottieni e installa una nuova ventola.

19 You can request a replacement fan kit and replace the fan yourself.
Puoi richiedere un kit di sostituzione della ventola e sostituirla.

20 Review the fan replacement instructions before deciding on this option.

21 Fan Replacement Instructions (PDF)

22 PDF icon
How helpful is Lilt’s Interactive MT?

• How often do translators use our MT suggestions?
• How often are our MT suggestions available and correct?
• How much do translators use our word-level suggestions, and how much do they post-edit?
• How do translators spend time on Lilt?
How often do translators use our MT suggestions?

- Check how much text is inserted via Enter and Shift-Enter
- Data is from August to September 2020
- We consider only newly-generated* segments

*newly-generated segments = no TM matches, no segments majority copy-pasted
Keys through which text is inserted

![Bar chart showing the percent of text inserted through different keys.]

- **Typed**: Approximately 60% of text inserted.
- **Enter/Tab**: Approximately 20% of text inserted.
- **Shift-Enter/Shift-Tab**: Approximately 10% of text inserted.
- **Ctrl-v** (Copy and Paste): A very small percentage of text inserted.
21% of entered text is MT suggestions accepted at the word-level

58% of entered text is manually typed

Keys through which text is inserted

<table>
<thead>
<tr>
<th>Key Type</th>
<th>Percent of Text Inserted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typed</td>
<td>58%</td>
</tr>
<tr>
<td>Enter/Tab</td>
<td>12%</td>
</tr>
<tr>
<td>Shift-Enter/Shift-Tab</td>
<td>12%</td>
</tr>
<tr>
<td>Ctrl-v</td>
<td>6%</td>
</tr>
</tbody>
</table>
Keys through which text is inserted

21% of entered text is MT suggestions accepted at the word-level

- Typed
- Enter/Tab
- Shift-Enter/Shift-Tab
- Ctrl-v
17% of entered text is MT suggestions accepted at the segment-level
Why aren’t translators using our MT suggestions more?

- Maybe translators aren’t aware they can press Enter?
- Maybe they aren’t editing at the end of the segment?
- Maybe the MT suggestion takes too long to show up?
- Maybe the MT suggestions don’t match what the translator wants to type?
How often are our MT suggestions available and correct?
How often are our MT suggestions available and correct?

MT suggestion matches what user enters for 46% of inserted text.
How often are our MT suggestions available and correct?

![Bar chart showing percent of text inserted by different methods](chart.png)

- **Typed**
- **Shift-Enter/Shift-Tab**
- **Enter/Tab**
- **Ctrl-v**

*Legend:*
- **Percent of text inserted (characters):**
  - 0
  - 10
  - 20
  - 30
  - 40
- **Categories:**
  - MT suggestion correct
  - MT suggestion up-to-date but incorrect
  - Not typing at end
  - Waiting for initial MT suggestion
  - Waiting for updated MT suggestion
How often are our MT suggestions available and correct?

When MT suggestion is correct, users accept word-level MT suggestions 45% of the time (21% of total text inserted).
How often are our MT suggestions available and correct?

When MT suggestion is correct, users accept segment-level MT suggestions 38% of the time (17% of total text inserted)
How often are our MT suggestions available and correct?

When MT suggestion is correct, users manually type it out 17% of the time (8% of total text inserted).

Percent of text inserted (characters)

- MT suggestion correct
- MT suggestion up-to-date but incorrect
- Not typing at end
- Waiting for initial MT suggestion
- Waiting for updated MT suggestion

Legend:
- Typed
- Shift-Enter/Shift-Tab
- Enter/Tab
- Ctrl-v
How often are our MT suggestions available and correct?

For 27% of inserted text, up-to-date MT suggestion was available, but didn’t match what user typed.
How often are our MT suggestions available and correct?

For 20% of inserted text, the user wasn’t inserting at the end, so no MT suggestions were available.
How often are our MT suggestions available and correct?

For 5% of inserted text, no MT suggestion was available yet.
How often are our MT suggestions available and correct?

For 3% of inserted text, the MT suggestion was out-of-date (still waiting for updated suggestion)
Are users using Lilt interactively, or as a post-editing system?

- We see a lot of users are using Shift-Enter (accept the entire remaining MT suggestion)

- We also see a lot of users making insertions outside the end of the segment

- Are more users using Lilt in an interactive, suffix-suggestion style, or post-editing?
Are users using Lilt interactively, or as a post-editing system?

Was Enter or Shift-Enter used when translating segment?
Are users using Lilt interactively, or as a post-editing system?

For 41% of segments, users don’t use our MT suggestions at all.
Are users using Lilt interactively, or as a post-editing system?

For 40% of segments, users accept word-level suggestions, but not segment-level.
Are users using Lilt interactively, or as a post-editing system?

For 10% of segments, users start by accepting the full MT suggestion, then make edits (post-editing).
Are users using Lilt interactively, or as a post-editing system?

For 7% of segments, users accept the full MT suggestion as-is (and don’t make any edits)
Histogram of users by percent of segments they post-edit
Histogram of users by percent of segments they post-edit

80% of users do not post-edit at all (0% of the segments they translated were entered starting with Shift-Enter followed by editing)
17% of users post-edit more than half of the segments they translate.
How do translators spend their time on Lilt?

Our efforts have focused on helping translators type translations faster via interactive MT. Is that actually most time-consuming part?

Data based on mouse and keyboard activity while using Lilt in translation mode, permitting up to 30 seconds of idle time between events.
How do translators spend their time on Lilt?
How do translators spend their time on Lilt?

Target Text Editor for Segment: 57% of time spent
1. Support for Tumblr Photo Sets

2. Support for Tumblr Photo Sets

3. Overview

4. Publishing Photo Sets to Tumblr is now supported.

5. Use this feature to publish a collection of photos to Tumblr in one post.

6. Note: /1 To learn more about this feature, contact your Success Manager.

Nota: Para obtener más información sobre esta función, póngase en contacto con su Administrador de Éxito.
How do translators spend their time on Lilt?

In browser, outside Lilt: 6% of time spent
How do translators spend their time on Lilt?

Sidebars (Lexicon, Find/Replace, Memory, etc): 5%
5. Use this feature to publish a collection of photos to Tumblr in one post.

6. **Note:** /1 To learn more about this feature, contact your Success Manager.

**Nota:** Para obtener más información sobre esta función, póngase en contacto con su Administrador de Éxito.

7. Publish a Photo Set to Tumblr
How do translators spend their time on Lilt?

Tag Editor for Segment: 3%
Use this feature to publish a collection of photos to Tumblr in one post.

Note: To learn more about this feature, contact your Success Manager.

Nota: Para obtener más información sobre esta función, póngase en contacto con su Administrador de Éxito.

Publish a Photo Set to Tumblr

1. Click on the icon in the top navigation bar and select Quick Publish

2. In the Select Accounts field, select the Tumblr account you would like to publish from

3. Click Photo to open the Add Photos window

4. In the Add Photos window, click on the photos you would like to include in your photo set. Selected images will have a check mark in the upper right corner.
Percent of time translators spend editing tags depends on what they’re translating - most segments have 0 tags.

So if 3% of all time is spent on tags, then among segments that have tags, editing tags take considerable effort.
Lilt does automatic tag placement when segments are confirmed, which users can correct.

See lilt.com/research for details (Zenkel 2020)
How do translators spend their time on Lilt?

Time spent on tasks that can benefit from machine assistance (in red): 65%
How do translators spend their time on Lilt?

Time spent on tasks that will not likely benefit from machine assistance (in black): 29%
Conclusion: A Study of Interactive Machine Translation Use on Lilt

• 57% of translator time is spent on actually writing the translation, which we can optimize with interactive MT.

• Our prefix-constrained interactive MT shows the correct suggestion to translators for 46% of the text they type. Of this, they use our autocompletion for 83% of the text.

• Main areas for improvement are MT quality and showing suggestions when user isn’t typing at the end. Latency is very good (< 500ms).

• While Lilt is used in an interactive style 4x more than post-editing, 17% of our users primarily use it for post-editing.
Backup slides
Keys through which text is inserted, broken down by MT state
Keys through which text is inserted, broken down by MT state

32% of manually typed text isn’t inserted at the end, so we don’t show MT suggestions (18% of total)
Keys through which text is inserted, broken down by MT state

13% of manually typed text matched MT suggestions (8% of total)
Keys through which text is inserted, broken down by MT state

8% of manual typing occurs when initial MT suggestion is not available yet (4% of total)
Keys through which text is inserted, broken down by MT state

- Waiting for updated MT suggestion
- Waiting for initial MT suggestion
- Not typing at end
- MT suggestion up-to-date but incorrect
- MT suggestion correct

5% of manual typing occurs when MT suggestion is outdated and waiting for updated suggestion (3% of total)
Keys through which text is inserted, broken down by MT state

43% of text is typed when up-to-date MT is available, but MT didn’t match what user typed (25% of total)