Evaluation of Document-level Machine Translation

Marzena Karpinska

Large language models effectively leverage document-level context for literary translation, but critical errors persist

Marzena Karpinska   **Mohit Iyyer**

Manning College of Information and Computer Sciences

University of Massachusetts Amherst

`{mkarpinska, miyyer}@cs.umass.edu`

Flaubert, 1856 (French)
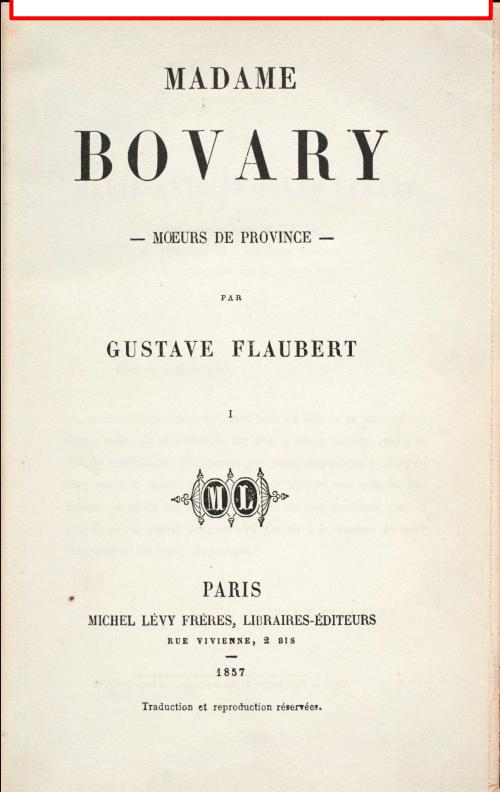
20 different English translations, 1886-2011
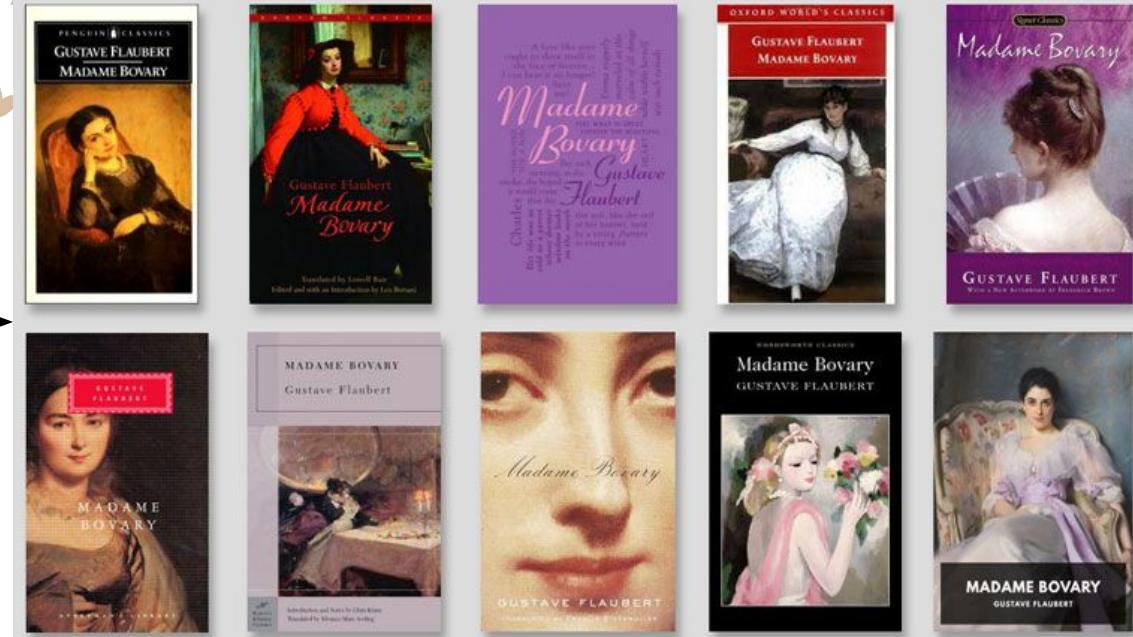
The Art of Literary Translation

Literary translator has to:

- Respect author's individual voice;

- Interpret the source with no objective reality to measure against;

- Make the translation enjoyable to the readers in the target language.

The Art of Literary Translation

→ **A translator applies different techniques:**

- ○ additions & omission;

- ○ cultural adaptation;

- ○ sentence splitting, merging, and content rearrangement.

Seiichi Shiojiri, 1947

Gael looked more self-complacent than usual, and with a bright smile all over his face talked about the Quorax Cabinet then in power.

'Quorax' is only a meaningless interjection like 'Oh.'

But it was the name of a political party whose primary concern was supposed to be the promotion of the welfare of the kappa race.

"The Quorax Party is under the control of Loppé," said Gael.

"As you know, he is a famous kappa statesman.

Bismarck said that honesty is the best diplomacy, but Loppé is honest not only in his diplomacy but also in his management of home affairs ….."

"But that speech of Loppe's—"

Geoffrey Bownas, 1971

Gael seemed even more elated and self-satisfied than usual, a broad smile on his face all the time. Among other things, he was telling me about the Quorax Party which had just come to power.

This word 'Quorax' is yet one more of the interjections so common as proper names in Kappaland: it has no meaning, and can only be translated by some such phrase as 'Good Heavens' or 'Bless Me'.

Be that as it may, the Quorax Party stands, above all, for 'the welfare of the whole 'Kappa State'.

'The man at the head of the Quorax Party is a distinguished statesman by the name of Loppe.

You will remember, no doubt, Bismarck's dictum—"Honesty is the best foreign policy." Well, this Loppe extends honesty even to home policy.'

'But, in Loppe's public utterances….'

Kappa (1927)
by Ryūnosuke Akutagawa

Gael looked more self-complacent than usual, and with a bright smile all over his face talked about the Quorax Cabinet then in power.

'Quorax' is only a meaningless interjection like 'Oh.'

But it was the name of a political party whose primary concern was supposed to be the promotion of the welfare of the kappa race.

"The Quorax Party is under the control of Loppé," said Gael.

"As you know, he is a famous kappa statesman.

Bismarck said that honesty is the best diplomacy, but Loppé is honest not only in his diplomacy but also in his management of home affairs ….."

"But that speech of Loppe's—"

Gael seemed even more elated and self-satisfied than usual, a broad smile on his face all the time. Among other things, he was telling me about the Quorax Party which had just come to power.

This word 'Quorax' is yet one more of the interjections so common as proper names in **Kappaland**: it has no meaning, and can only be translated by some such phrase as 'Good Heavens' or 'Bless Me'.

Be that as it may, the Quorax Party stands, above all, for 'the welfare of the whole 'Kappa State'.

'The man at the head of the Quorax Party is a distinguished statesman by the name of Loppe.

You will remember, no doubt, Bismarck's dictum—"Honesty is the best foreign policy."

Well, this Loppe extends honesty even to home policy.'

'But, in Loppe's public utterances….'

Seiichi Shiojiri, 1947

Gael looked more self-complacent than usual, and with a bright smile all over his face talked about the Quorax Cabinet then in power.
'Quorax' is only a meaningless interjection like 'Oh.'
But it was the name of a political party whose primary concern was supposed to be the promotion of the welfare of the kappa race.
"The Quorax Party is under the control of Loppé," said Gael.
"As you know, he is a famous kappa statesman.
Bismarck said that honesty is the best diplomacy, but Loppé is honest not only in his diplomacy but also in his management of home affairs ....."
"But that speech of Loppe's—"

kappa language

kappanese

Geoffrey Bownas, 1971

Gael seemed even more elated and self-satisfied than usual, a broad smile on his face all the time. Among other things, he was telling me about the Quorax Party which had just come to power.
This word 'Quorax' is yet one more of the interjections so common as proper names in **Kappaland**: it has no meaning, and can only be translated by some such phrase as 'Good Heavens' or 'Bless Me'.
Be that as it may, the Quorax Party stands, above all, for 'the welfare of the whole 'Kappa State'.
'The man at the head of the Quorax Party is a distinguished statesman by the name of Loppe.
You will remember, no doubt, Bismarck's dictum—"Honesty is the best foreign policy."
Well, this Loppe extends honesty even to home policy.'
'But, in Loppe's public utterances....'

Literary translation is *hard*

I have spent **the last three years** translating Flaubert's Madame Bovary into English... There were times when I tumbled into the crevice between the two languages, **lost all sight of a natural English sentence,** felt myself turning into the constituent molecules of a linguistic object – a pattern of auxiliaries, participles, pronouns.

Adam Thorpe, 2011
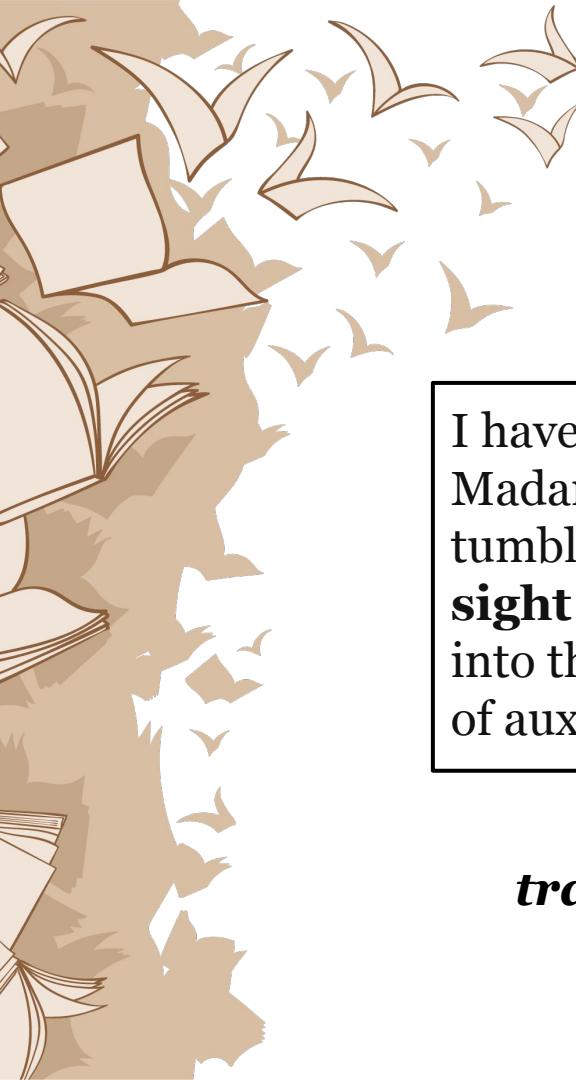
Literary translation is *hard*

I have spent **the last three years** translating Flaubert's Madame Bovary into English... There were times when I tumbled into the crevice between the two languages, **lost all sight of a natural English sentence,** felt myself turning into the constituent molecules of a linguistic object – a pattern of auxiliaries, participles, pronouns.

Adam Thorpe, 2011

translationese
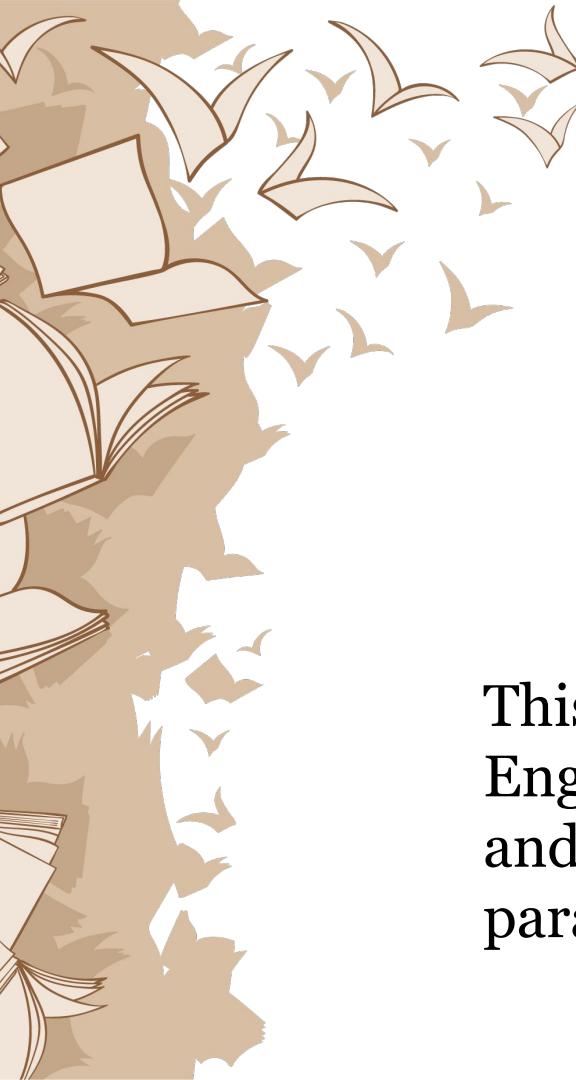
Literary MT

Extracting Paraphrases from a Parallel Corpus

Regina Barzilay and **Kathleen R. McKeown**
Computer Science Department
Columbia University
10027, New York, NY, USA
{regina,kathy}@cs.columbia.edu

This ACL 2001 paper proposes to collect multiple English translations of literary works, align them, and then extract word, phrase, and sentence-level paraphrases.

Literary MT

Extracting Paraphrases from a Parallel Corpus

Regina Barzilay and **Kathleen R. McKeown**
Computer Science Department
Columbia University
10027, New York, NY, USA
{regina,kathy}@cs.columbia.edu

**Exploring Document-Level Literary Machine Translation
with Parallel Paragraphs from World Literature**

Katherine Thai[★◇]   Marzena Karpinska[★◇]   Kalpesh Krishna[◇]   William Ray[◇]
Moira Inghilleri[♠]   John Wieting[♣]   Mohit Iyyer[◇]

[◇]Manning College of Information and Computer Sciences, UMass Amherst
[♠]Department of Languages, Literatures, and Cultures; UMass Amherst
[♣]Google Research
{kbthai,mkarpinska,kalpesh,miyyer}@cs.umass.edu
minghilleri@complit.umass.edu, jwieting@google.com
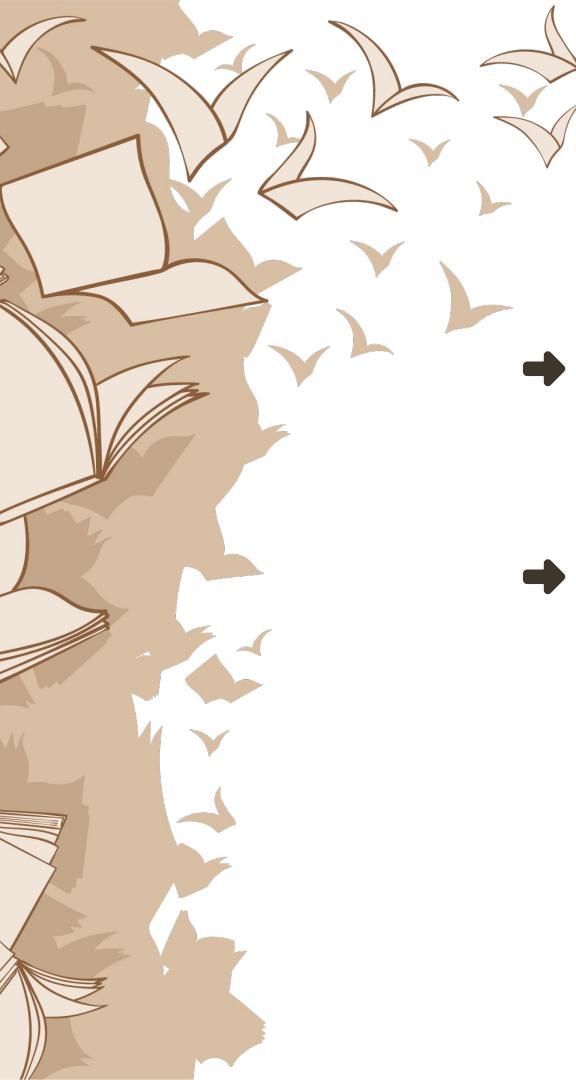
Research Question

➡ How good are LLMs at translating *literary* works (e.g., novels, short stories)?

➡ Can LLMs **leverage paragraph-level context** to improve translation quality compared to the standard sentence-by-sentence translation?

Document-level MT

→**Statistical MT:** vocab consistency (Carpuat, 2009); cross-sentence features (Ture et al., 2012), cross-sentence decoding (Hardmeier et al., 2012)

→**Neural MT:** concatenation-based (Jean et al., 2017, Tiedemann & Scherrer, 2017), hierarchical (Miculicich et al., 2018), Doc2Doc (Sun et al., 2020), transfer learning (Zhang et al., 2022)

Escaping the sentence-level paradigm in machine translation

Matt Post and **Marcin Junczys-Dowmunt**
Microsoft
Redmond, Washington
{mattpost,marcinjd}@microsoft.com

Document-level MT

➔ **Statistical MT:** vocab consistency (Carpuat, 2009);
cross-sentence features (Ture et al., 2012), cross-sentence
decoding (Hardmeier et al., 2012)

➔ **Neural MT:** concatenation-based (Jean et al., 2017,
Tiedemann & Scherrer, 2017), hierarchical (Miculicich et al.,
2018), Doc2Doc (Sun et al., 2020), transfer learning (Zhang
et al., 2022)

➔ **Evaluation: ???**
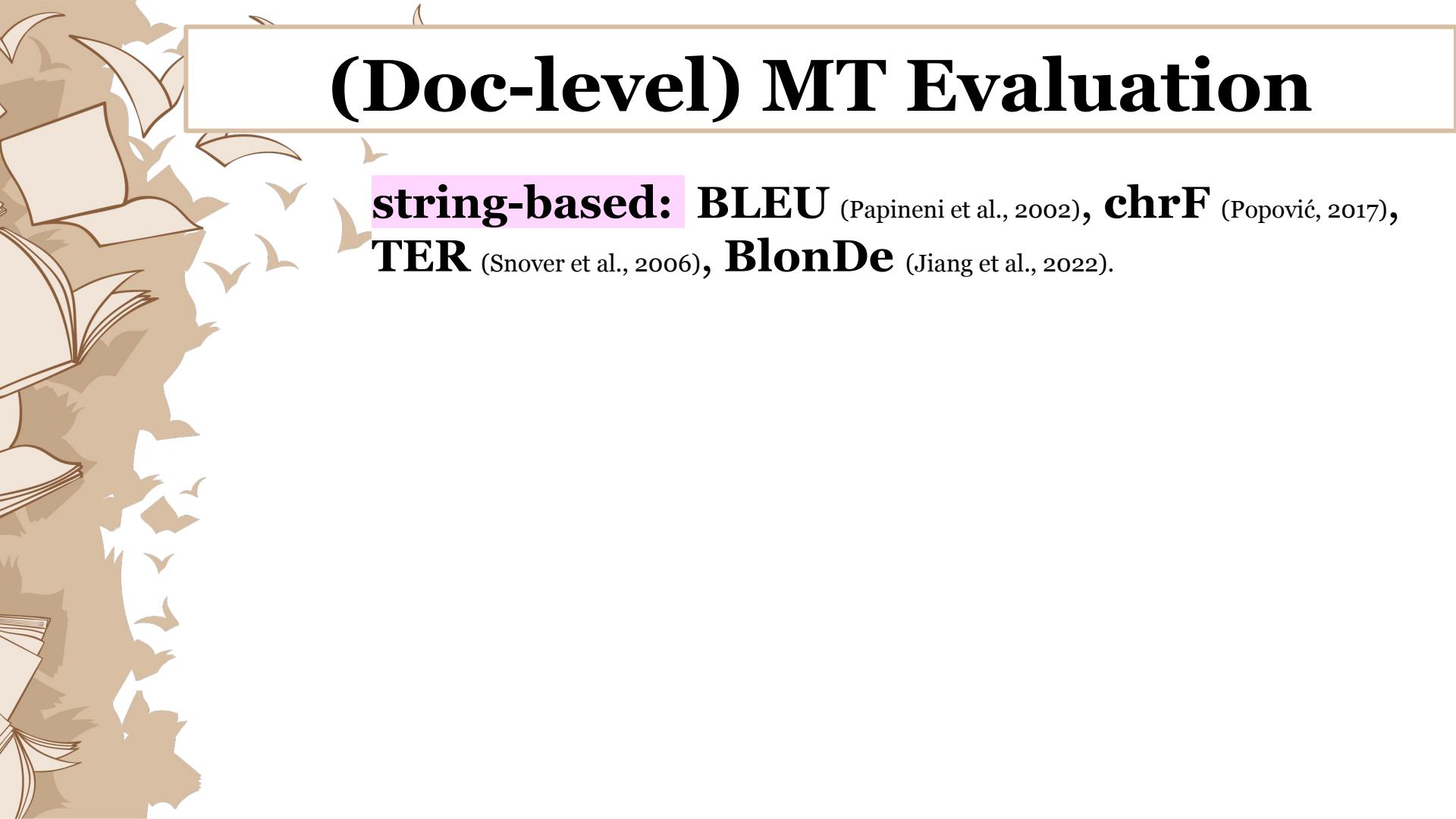
(Doc-level) MT Evaluation

string-based: BLEU (Papineni et al., 2002), chrF (Popović, 2017), TER (Snover et al., 2006), BlonDe (Jiang et al., 2022).

(Doc-level) MT Evaluation

string-based: **BLEU** (Papineni et al., 2002), **chrF** (Popović, 2017), **TER** (Snover et al., 2006), **BlonDe** (Jiang et al., 2022).

learned: **BLEURT** (Sellam et al., 2020), **COMET** (Rei et al., 2022), **PRISM** (Thompson and Post, 2020), **COMET-QE** (Rei et al., 2021) , **PRISM-QE** (Thompson and Post, 2020).

Embarrassingly Easy Document-Level MT Metrics:
How to Convert Any Pretrained Metric Into a Document-Level Metric

Giorgos Vernikos[*]   **Brian Thompson**   **Prashant Mathur**   **Marcello Federico**
EPFL + HEIG-VD   AWS AI Labs   AWS AI Labs   AWS AI Labs
georgios.vernikos@epfl.ch brianjt@amazon.com   pramathu@amazon.com marcfede@amazon.com

(Doc-level) MT Evaluation

string-based: BLEU (Papineni et al., 2002), **chrF** (Popović, 2017), TER (Snover et al., 2006), **BlonDe** (Jiang et al., 2022).

learned: BLEURT (Sellam et al., 2020), COMET (Rei et al., 2022), PRISM (Thompson and Post, 2020), COMET-QE (Rei et al., 2021), PRISM-QE (Thompson and Post, 2020).

human evaluation:
- **direct assessment** (DA),
- **preference choice** (ranking),
- **MQM error annotation.**

Document-level MT Evaluation
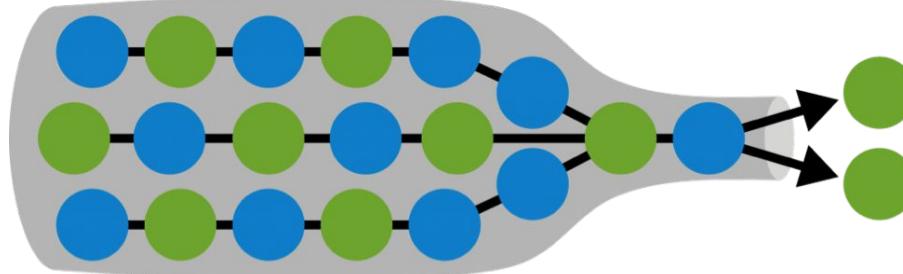
Escaping the sentence-level paradigm in machine translation

Matt Post and Marcin Junczys-Dowmunt
Microsoft
Redmond, Washington
{mattpost,marcinjd}@microsoft.com

You can't improve what you can't measure.
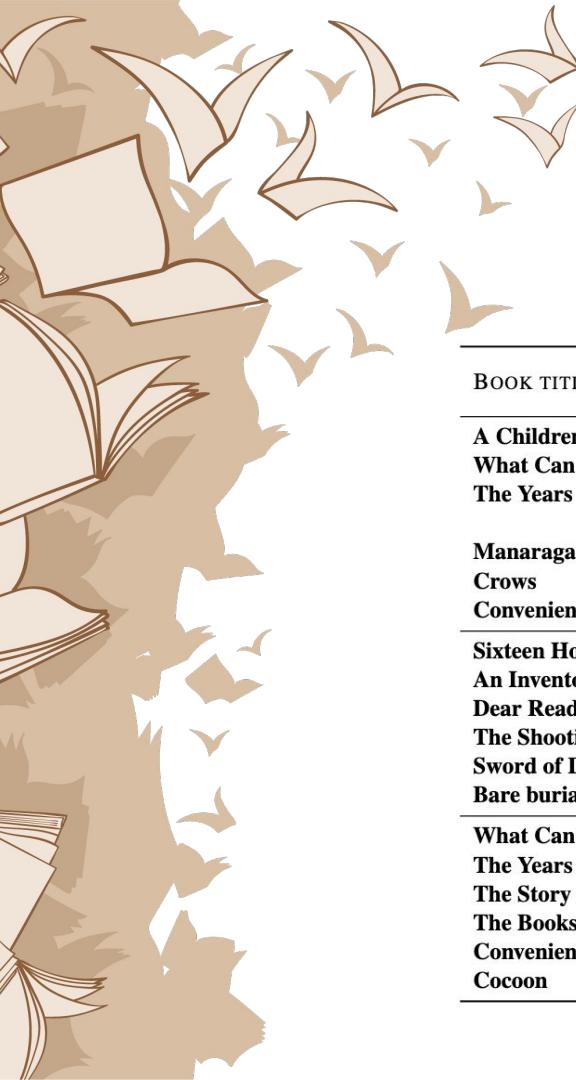
Data collection

→ identified **18** *recently* published translations

Book title	Author	Translator(s)	Language Source	Language Target	Year Published Translation	Year Published Original
A Children's Bible	Lydia Millet	Aga Zano	*en*	*pl*	2022	2020
What Can You See From Here	Mariana Leky	Agnieszka Walczy	*de*	*pl*	2021	2017
The Years	Annie Ernaux	Krzysztof Jarosz & Magdalena Budzińska	*fr*	*pl*	2022	2008
Manaraga	Wladimir Sorokin	Agnieszka Lubomira Piotrowska	*ru*	*pl*	2018	2017
Crows	Petra Dvorakova	Mirosław Śmigielski	*cs*	*pl*	2020	2020
Convenience Store Woman	Sayaka Murata	Dariusz Latoś	*ja*	*pl*	2019	2016
Sixteen Horses	Greg Buchanan	Fuji Yoshiko	*en*	*ja*	2022	2021
An Inventory of Losses	Judith Schalansky	Naoko Hosoi	*de*	*ja*	2022	2018
Dear Reader	Paul Fournel	Kei Takahashi	*fr*	*ja*	2022	2011
The Shooting Party	Anton Chekhov	Takuya Hara	*ru*	*ja*	2022	1884
Sword of Destiny	Andrzej Sapkowski	Yasuko Kawano	*pl*	*ja*	2022	1992
Bare burial	Fang Fang	Shin'ichi Watanabe	*zh*	*ja*	2022	2016
What Can You See From Here	Mariana Leky	Tess Lewis	*de*	*en*	2021	2017
The Years	Annie Ernaux	Alison L. Strayer	*fr*	*en*	2017	2008
The Story of a Life	Konstantin Paustovsky	Douglas Smith	*ru*	*en*	2022	1956
The Books of Jacob	Olga Yokarczuk	Jennifer Croft	*pl*	*en*	2022	2014
Convenience Store Woman	Sayaka Murata	Ginny Tapley Takemori	*ja*	*en*	2018	2016
Cocoon	Zhang Yueran	Jeremy Tiang	*zh*	*en*	2022	2018

Data collection

➡ manually extracted **20** aligned paragraphs

Data collection

→ manually extracted **20** aligned paragraphs
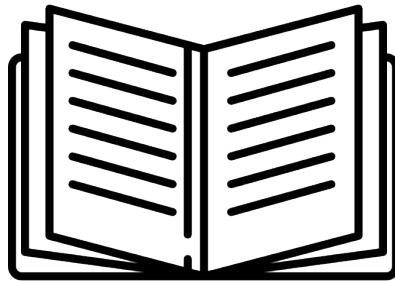
Source Text

Data collection
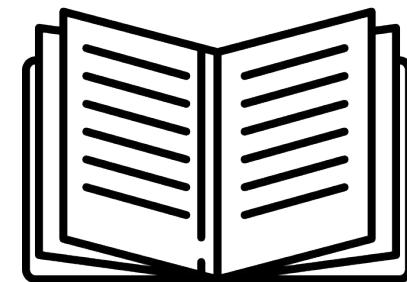
➔ manually extracted **20** aligned paragraphs

Source Text

Target Text

Data collection

➡ manually extracted **20** aligned paragraphs

Source Text   **Target Text**

Data collection

➡ manually extracted **20** aligned paragraphs

Source Text
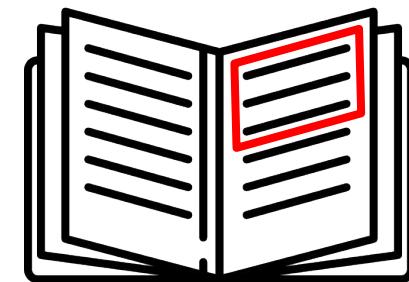
Target Text

Data collection

→ manually extracted **20** aligned paragraphs

Source Text

Target Text
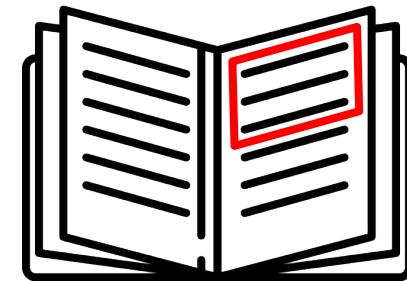
Po śmierci Jakuba nastał dla mnie czas wielkiego spokoju. Głównym moim zajęciem stało się studiowanie merkawy, do tego doszły jeszcze nasze rozmowy z Jeruchimem Dembowskim, jako że zamieszkaliśmy w jednym pokoju, przez co zbliżyliśmy się do siebie. Jemu, Jeruchimowi, (...)

–Polish Source (from *The Books of Jacob* by Olga Tokarczuk, 2014)

After Jacob's death, I entered into a period of peace. My main occupation became studying Merkavah, which my conversations with Yeruhim Dembowski also managed to cover, as we lived in the one room, which meant we got closer to each other. To him, to Yeruhim, (...)

–English Target (translated by Jennifer Croft, 2022)

Data collection

Source Text
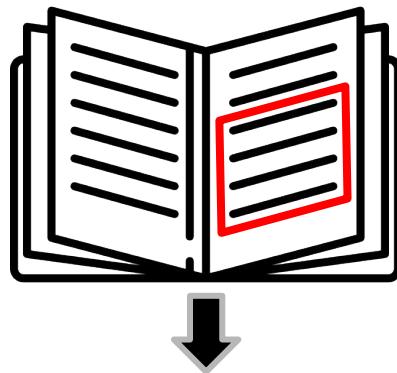
Target Text

Po śmierci Jakuba nastał dla mnie czas wielkiego spokoju. Głównym moim zajęciem stało się studiowanie merkawy, (...)
–Polish Source (from *The Books of Jacob* by Olga Tokarczuk, 2014)
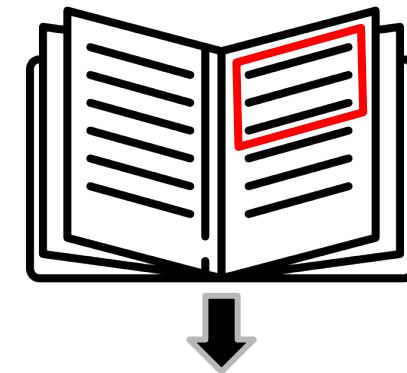
After Jacob's death, I entered into a period of peace. My main occupation became studying Merkavah, (...)
–English Target (translated by Jennifer Croft, 2022)

360 aligned paragraphs

Source and Target Languages

🌐 **_8_ Source Languages:**

- **Japanese**
- **Russian**
- **Polish**
- **Czech**
- **French**
- **German**
- **Chinese**
- **English**

Source and Target Languages

8 Source Languages:

- **Japanese**
- **Russian**
- **Polish**
- **Czech**
- **French**
- **German**
- **Chinese**
- **English**

3 Target Languages:

- **Japanese**
- **Polish**
- **English**

Translating with Large Language Models

- **5-shot translation** with 5 demonstrations from novels *different* than the translated novel

- translation with **GPT-3.5** (`text-davinci-003`) in three setups:

  - **SENT** - sentence-level translation *without* context

  - **PARA_SENT** - sentence-level translation *with* context

  - **PARA** - paragraph-level translation

- GoogleTranslate (**GTr**) as a commercial baseline

Sentence-level without Context (SENT)

Source Language

Sentence-level without Context (SENT)

Source Language
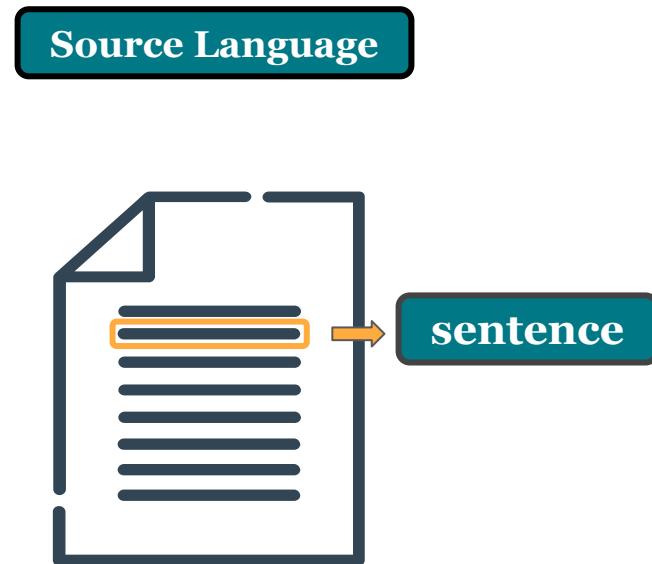
sentence

Sentence-level without Context (SENT)
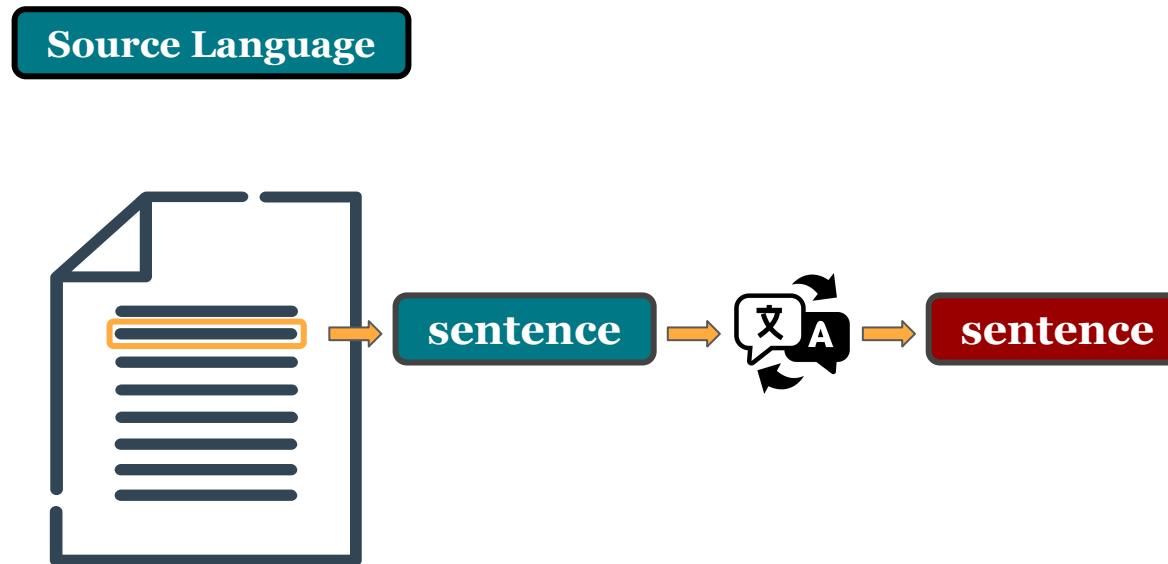
Sentence-level without Context (SENT)
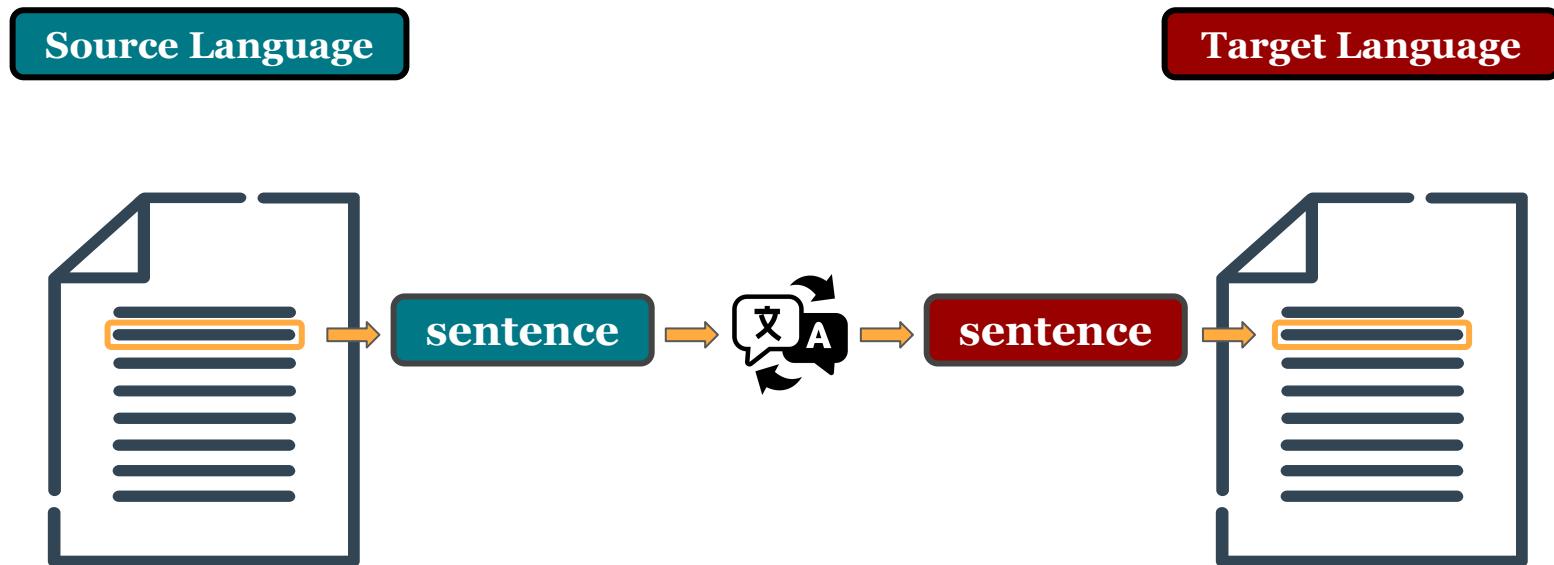
Sentence-level without Context (SENT)

demonstrations →

> **Original text in Japanese:**
> 「そういうのは実際には起こらないの？」
>
> **Translation into Polish:**
> – To się w rzeczywistości nie zdarza?
>
> **(...)**

translation →

Original text in Japanese:
「いらっしゃいませ！」

**Translation into Polish:**

Sentence-level with Context (PARA_SENT)

Source Language

Sentence-level with Context (PARA_SENT)

Source Language

Sentence-level with Context (PARA_SENT)

Sentence-level with Context (PARA_SENT)

Source Language

sentence in paragraph

Sentence-level with Context (PARA_SENT)

Source Language


sentence in paragraph

translation so far

Sentence-level with Context (PARA_SENT)

Sentence-level with Context (PARA_SENT)

Sentence-level with Context (PARA_SENT)

demonstrations →
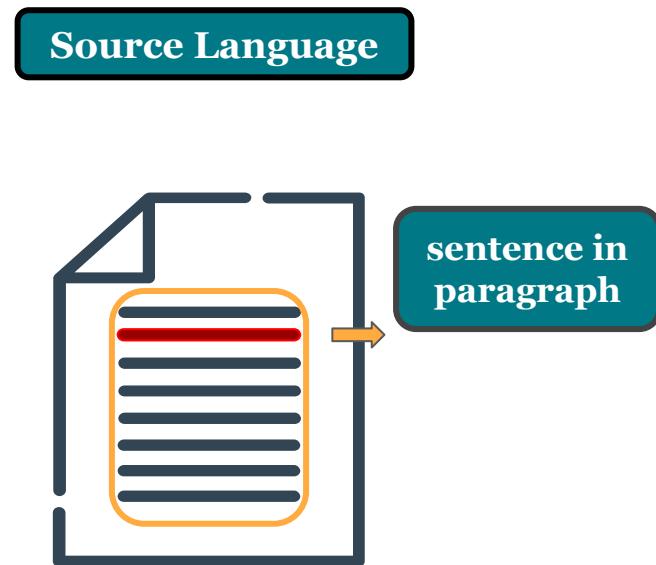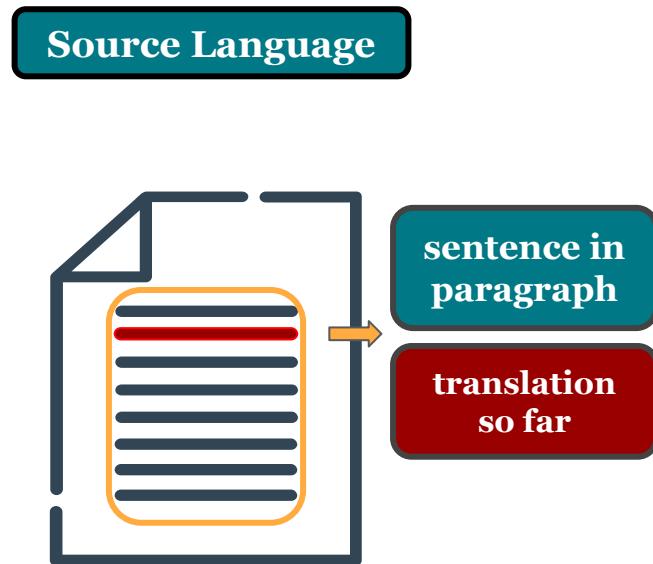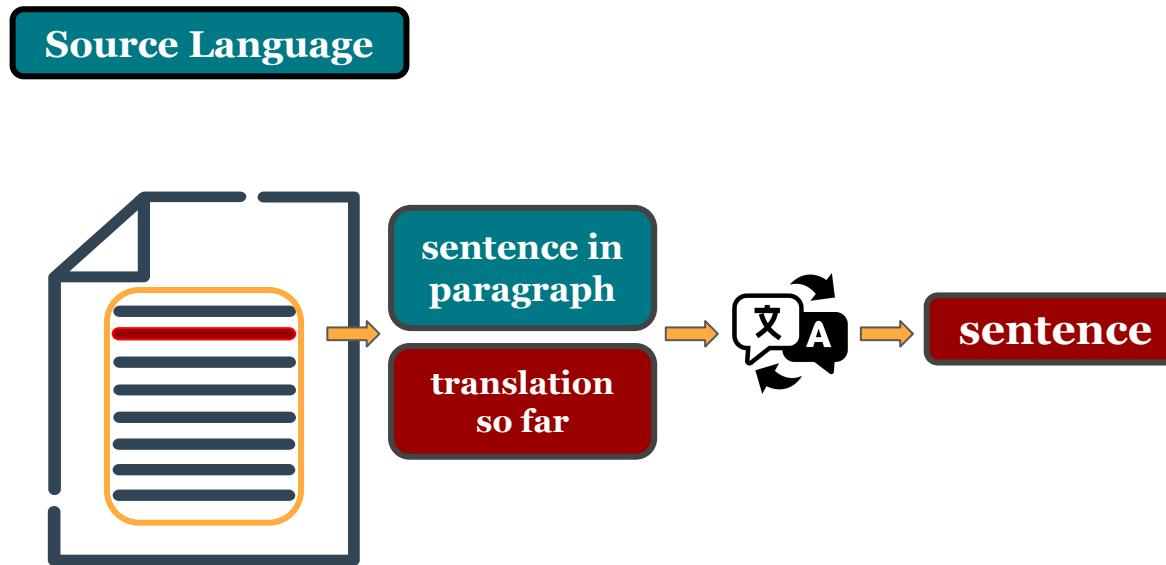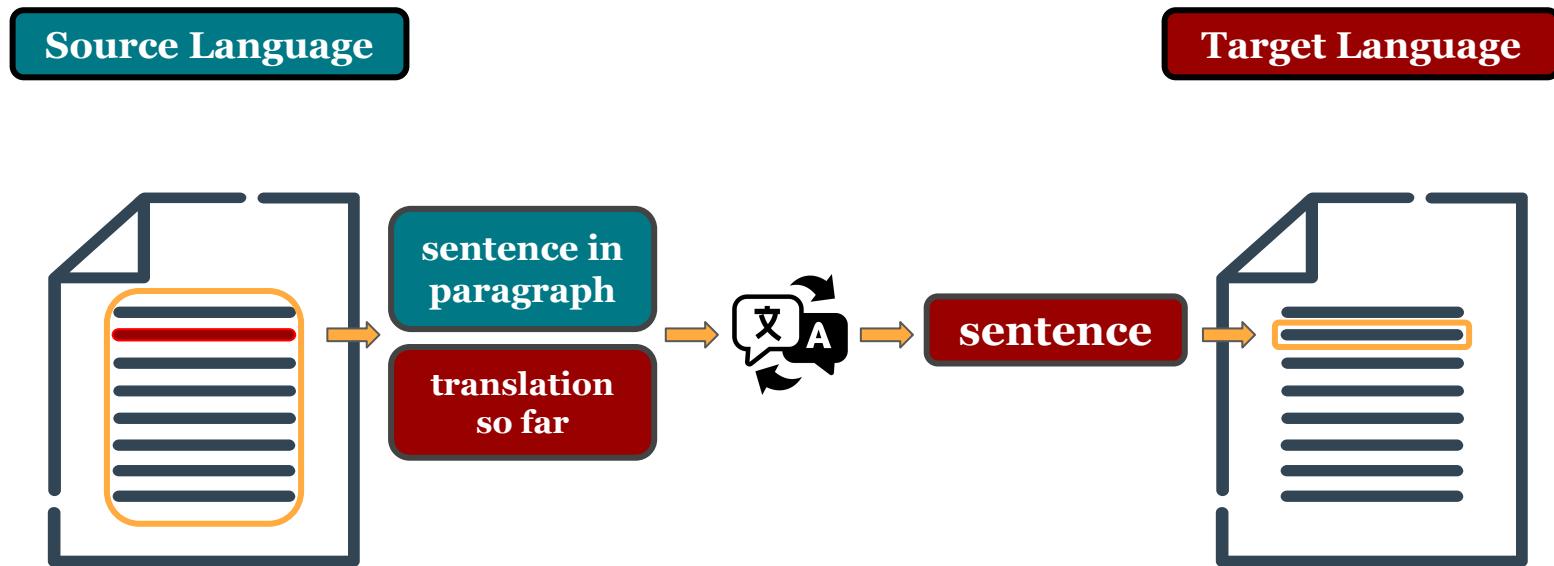
Original text in Czech:

V hospodě U kalicha seděl jen jeden host. Byl to civilní strážník Bretschneider, stojící ve službách státní policie. **<translate>**Hostinský Palivec myl tácky a Bretschneider se marně snažil navázat s ním vážný rozhovor.**</translate>**

Translation into Polish:

W gospodzie „Pod Kielichem" siedział tylko jeden gość. Był to wywiadowca Bretschneider, będący na służbie policji państwowej. **<translated>**Gospodarz Palivec zmywał podstawki, a Bretschneider daremnie usiłował wyciągnąć go na poważną rozmowę.**</translated>**

(...)

translation →

Original text in Czech:

„Je to fakt dobrý," řekl mi Frodo, když se na můj výkres koukal. **<translate>**A skoro to vypadalo, že je z toho obrázku překvapenej.**</translate>** Pak se ptal, jestli maluju i doma, tak jsem odpověděla, že jo. Nejradši bych mu řekla i to, že chci být malířkou, ale mamka pořád opakuje, že je to blbost, že člověk musí dělat něco pořádnýho, jako třeba ona dělá knihovnici v knihovně, tak jsem raději mlčela, aby si nemyslel, že mám blbý nápady. Chvíli na výkres ještě koukal a otáčel ho ze všech stran a pak šel dál a koukal na obrázek Lindy, která nakreslila takový docela pěkný jabko.

Translation into Polish:

– To jest naprawdę dobre – powiedział Frodo, patrząc na moją pracę. **<translated>**

Paragraph-level Translation (PARA)

Paragraph-level Translation (PARA)

Paragraph-level Translation (PARA)

Source Language

paragraph
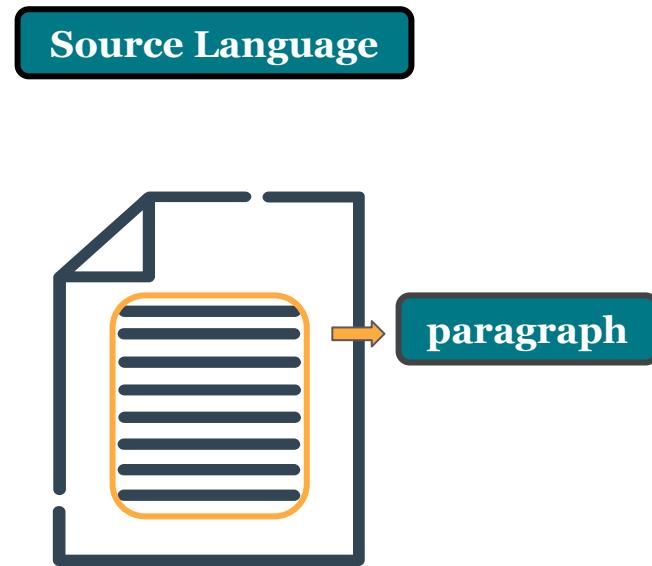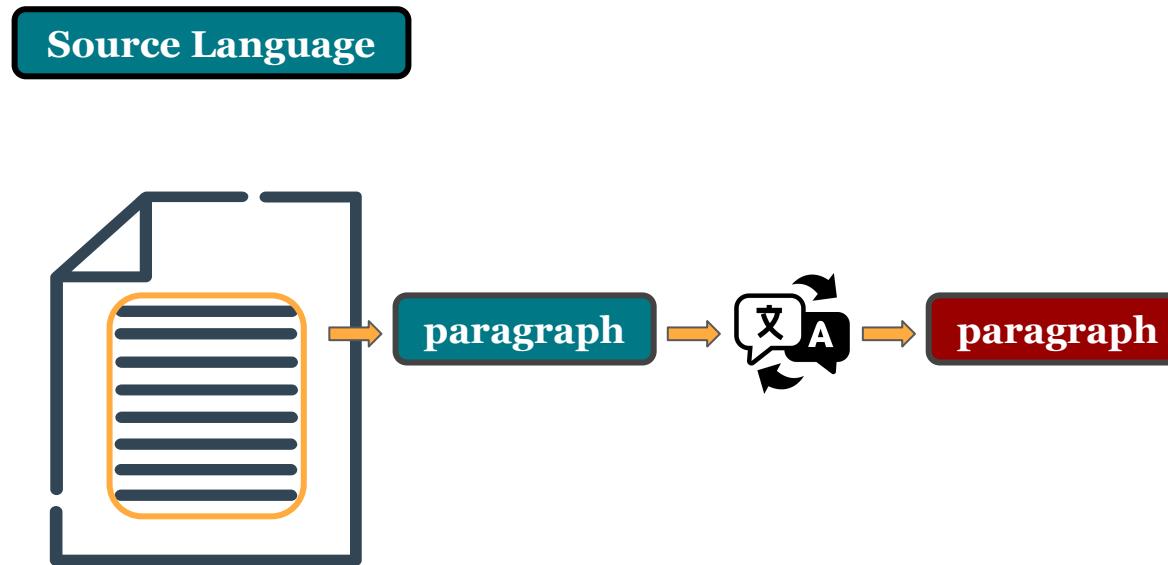
Paragraph-level Translation (PARA)

Paragraph-level Translation (PARA)

Paragraph-level Translation (PARA)

demonstrations →
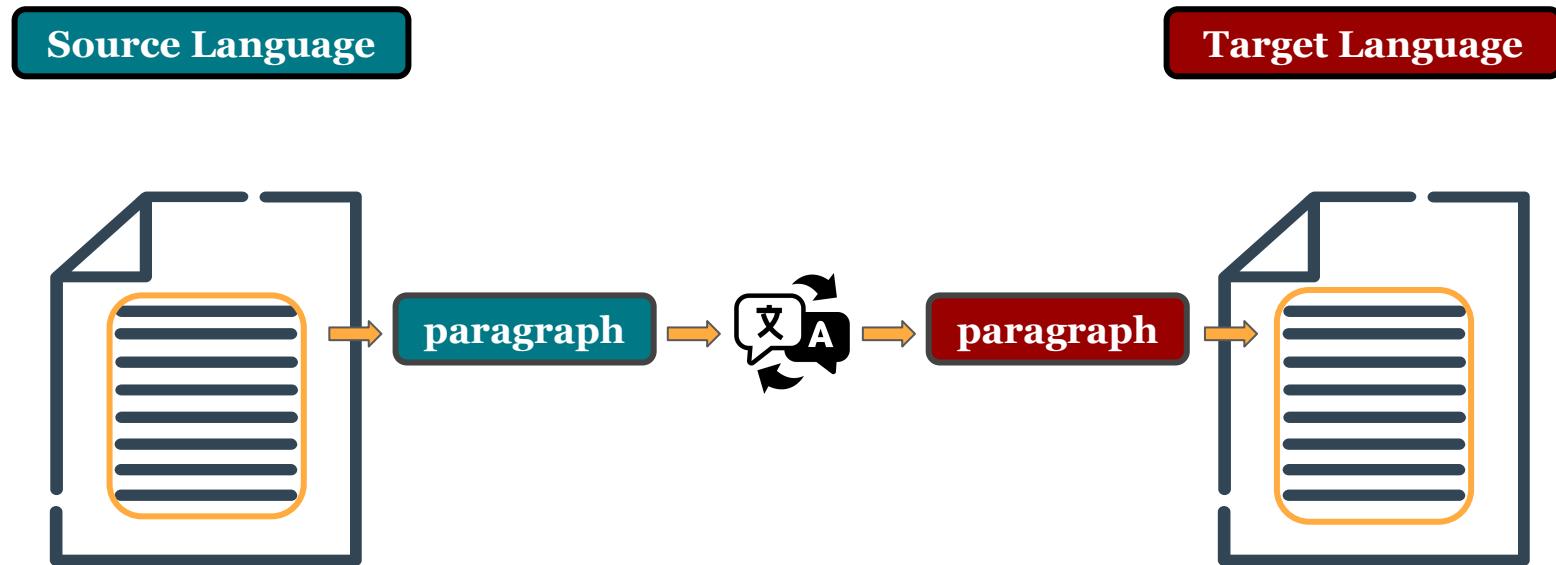
Original text in Japanese:

直子は立ちどまった。僕も立ちどまった。彼女は両手を僕の肩にあてて正面から、僕の目をじっとのぞきこんだ。彼女の瞳の奥の方ではまっ黒な重い液体が不思議な図形の渦を描いていた。そんな一対の美しい瞳が長いあいだ僕の中をのぞきこんでいた。それから彼女は背のびをして僕の頬にそっと頬をつけた。それは一瞬胸がつまってしまうくらいあたたかくて素敵な仕草だった。

Translation into Polish:

Naoko zatrzymała się. Ja też stanąłem. Położyła mi ręce na ramionach i zajrzała uważnie w oczy. W jej ciemnych jak atrament źrenicach tworzyły się przedziwne wirujące wzory. Te piękne oczy długo badały moje serce. Potem wyprostowała się i przytuliła policzek do mojego. To był cudowny ciepły gest, aż mi serce na chwilę zamarło.

(...)

translation →

Original text in Japanese:

店長は30歳の男性で、常にきびきびとしている。口は悪いが働き者の、この店で8人目の店長だ。
2人目の店長はサボり癖があり、4人目の店長は真面目で掃除好きで、6人目の店長は癖のある人で嫌われ、夕勤が全員一気に辞めるというトラブルになった。8人目の店長は比較的アルバイトからも好かれ、自分が体を動かして働くタイプなので、見ていて気持ちがいい。7人目の店長は気弱すぎて夜勤になかなか注意ができずに店がぼろぼろになってしまったので、少し口が悪くてもこれくらいのほうが働きやすいと、8人目の店長を見ると思う。

Translation into Polish:

🗨 Human evaluation

- Sampled **180 source paragraphs** (10 per language pair)

💬 Human evaluation

- Sampled **180 source paragraphs** (10 per language pair)

- **13 translators** hired on Upwork
  - Fluent in both source and target language, native speaker of target language

🗨 Human evaluation

- Sampled **180 source paragraphs** (10 per language pair)

- **13 translators** hired on Upwork
  - Fluent in both source and target language, native speaker of target language

- Give them a source paragraph, two candidate translations, and ask them to do the following:
  - Annotate and categorize span-level errors (MQM, Lommel et al., 2014)
  - Select the better translation
  - Justify their preference in a free-form comment

Multidimensional Quality Metrics (Lommel et al., 2014)
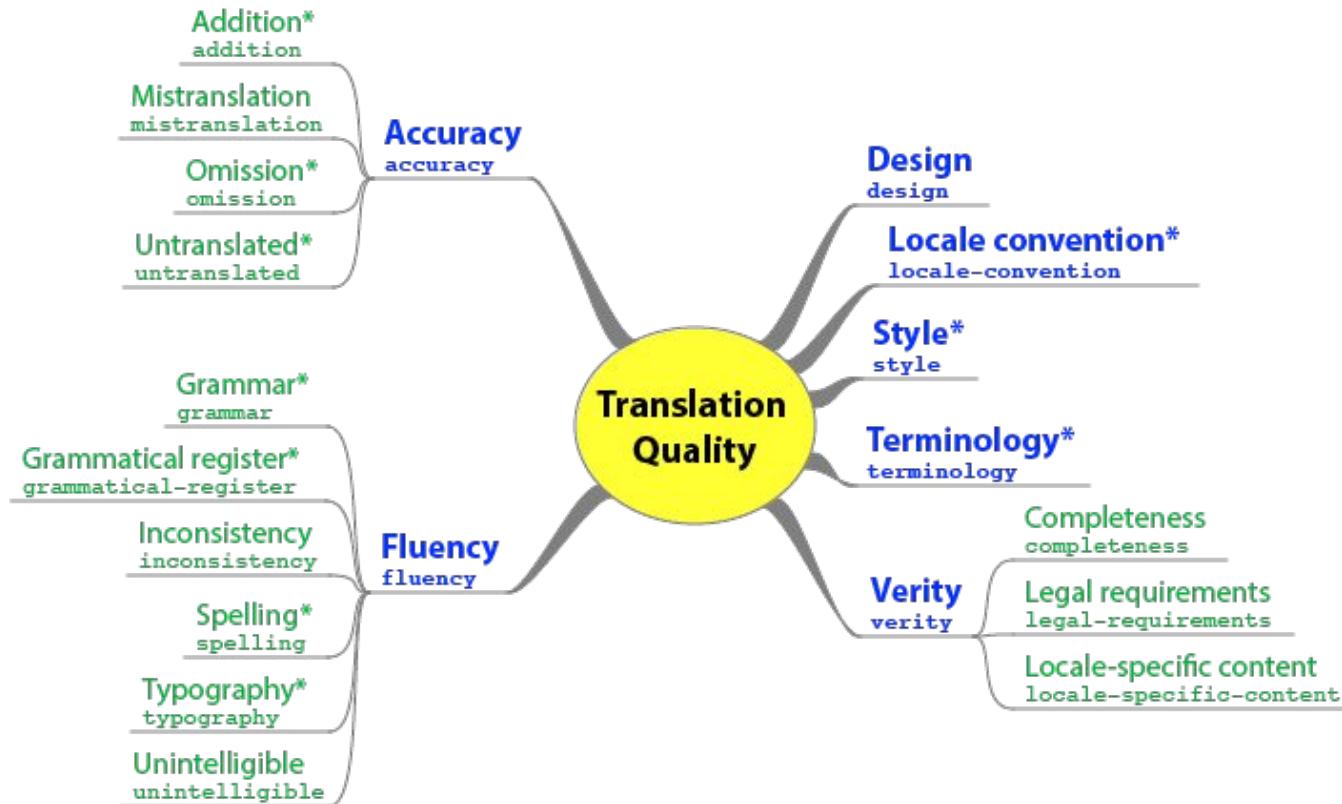
🗨 Human evaluation

Source Text: Une autre photo, signée du même photographe...

Read the source text.

💬 Human evaluation

Source Text: Une autre photo, signée du même photographe...

Read the source text.

👥 Human evaluation

Source Text: Une autre photo, signée du même photographe...

Read the source text.

Error Annotation

grammar   mistranslation   untranslated

inconsistency   register   format

Translation 1:

Inny zdjęcie, podpisane przez tego samego fotografa...

Translation 2:

Inne zdjęcie, podpisane tym samym fotografem...

Mark and categorize span-level errors

🗨 Human evaluation

Source Text: Une autre photo, signée du même photographe…

Read the source text.

Error Annotation

| grammar | mistranslation | untranslated |
| inconsistency | register | format |

Translation 1:
Inny zdjęcie, podpisane przez tego samego fotografa…

Translation 2:
Inne zdjęcie, podpisane tym samym fotografem…

Mark and categorize span-level errors

Any omissions or additions?

Binary choice (yes/no)

Which translation is better?

Binary choice (translation 1 vs 2)

🗨 Human evaluation

Source Text: Une autre photo, signée du même photographe...

Read the source text.

Error Annotation

grammar | mistranslation | untranslated
inconsistency | register | format

Mark and categorize span-level errors

Translation 1:
Inny zdjęcie, podpisane przez tego samego fotografa...

Translation 2:
Inne zdjęcie, podpisane tym samym fotografem...

Any omissions or additions?

Binary choice (yes/no)

Which translation is better?

Binary choice (translation 1 vs 2)

Why is it better? *I prefer T1 to T2 as it doesn't...*

Justify preference

Was it significantly better?

Binary choice (confident vs unsure)

💬 Human Evaluation

Source Text: Une autre photo, signée du même photographe...

→ Read the source text.

Error Annotation

- grammar
- mistranslation
- untranslated
- inconsistency
- register
- format

Translation 1:
Inny zdjęcie, podpisane przez tego samego fotografa...

Translation 2:
Inne zdjęcie, podpisane tym samym fotografem...

→ Mark and categorize span-level errors

Any omissions or additions? → Binary choice (yes/no)

Which translation is better? → Binary choice (translation 1 vs 2)

Why is it better? *I prefer T1 to T2 as it doesn't...* → Justify preference

Was it significantly better? → Binary choice (confident vs unsure)

Skip   Submit

Consider the following SOURCE PARAGRAPH and two competitive TRANSLATIONS:

Celui qui est sous ma joue est un manuscrit d'amour : c'est l'histoire d'un mec qui rencontre une fille mais il est marié et elle a un copain… J'en ai lu sept pages et je le connais déjà par cœur. Rien ne pourra me surprendre. Depuis des lunes, je ne lis plus, je relis. La même vieille bouillie dont on fait des « nouveautés », des saisons, des rentrées « littéraires », des succès, des bides, des bides. Du papier qu'on recycle, des camions qui partent le matin et qui rentrent le soir, bourrés de nouveautés déjà hors d'âge.

Annotate all mistakes in each translation. Refer to the gudelines for details.

GRAMMAR 1   MISTRANSLATION 2   UNTRANSLATED 3   INCONSISTENCY 4   REGISTER 5   FORMAT 6

TRANSLATION 1:

私の頬にあるのは、愛のための手稿だ。それは、男が女に出会う物語だが、彼は既婚者で、彼女は彼氏がいる……という物語である。7ページ読んだだけで、もう覚えてしまった。何も私を驚かせることはない。月日が経つにつれて、もう新しい本を読むことはなくなり、繰り返し読むことになった。それは、新しいものを作り出すために使われる古い粥のようなものであり、季節や新学期、文学的な成功や失敗、大失敗などを作り出すものだ。リサイクルした紙、朝出かけて夜帰ってくるトラックが、すでに古臭い新商品でいっぱいだ。

########

TRANSLATION 2:

私の下頬の下にあるのは、愛の物語の手書き原稿だ。それは、男が女に出会う物語だが、彼は既に結婚していて、彼女も彼氏がいる……。私はその七ページを読んだが、もう心から覚えている。何も驚かせるものはない。月日が経つうちに、私はもう新しいものを読むのではなく、古いものを改めて読んでいる。「新しいもの」として、季節や新学期、文学的な「成功」や「失敗」として、何度も何度も取り上げられる古いものだ。朝に出て、夜に帰ってくるトラックは、すでに時代遅れの「新しいもの」でいっぱいだ。

No Region selected

Regions   Labels

No Regions created yet

Relations (0)

No Relations added yet

LabelStudio

No Regions created yet

Relations (0)

No Relations added yet

Which of the following is true?

○ TRANSLATION 1 adds or omits information that significantly change the meaning of the text.[7]
○ TRANSLATION 2 adds or omits information that significantly change the meaning of the text.[8]
○ BOTH translations add or omit information that significantly change the meaning of the text.[9]
○ NEITHER of the translations add or omit information that significantly change the meaning of the text[o]

Which translation is better overall?

○ TRANSLATION 1[q]   ○ TRANSLATION 2[w]

Please motivate your choice in 2 to 5 sentences.

Add

Was it hard to decide between the two translations because of their quality (i.e., they are comparabily good or bad)?

○ YES, both translations are close in quality and it was hard to choose between them.[e]   ○ NO, the translations I have chosen is visibly better.[t]

## Press 'SUBMIT' to proceed.

Humans prefer PARA
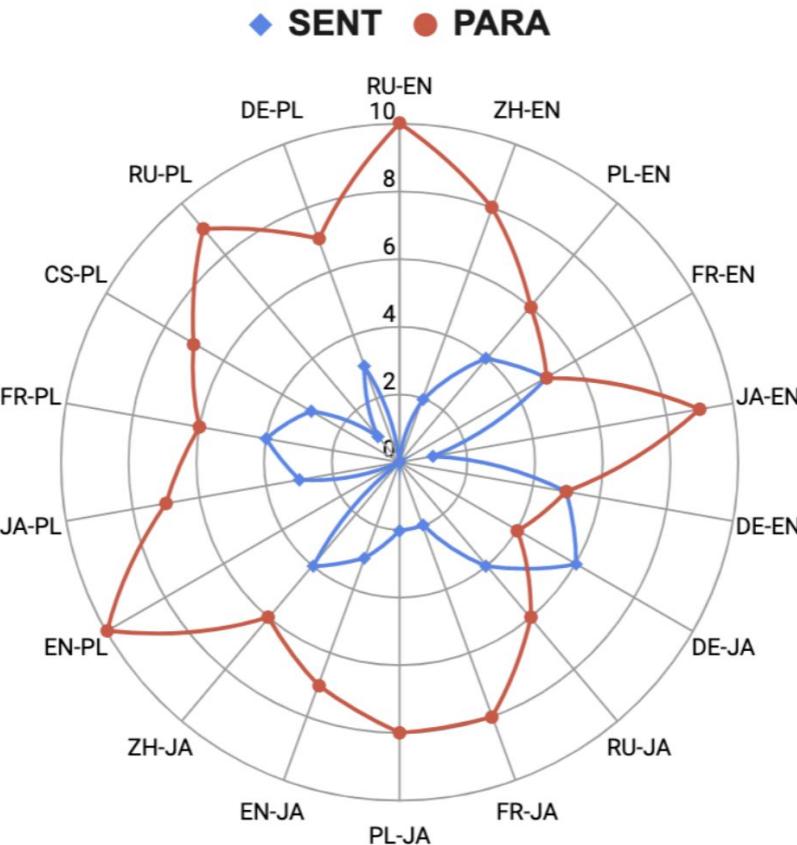
An Inventory of Losses by Judith Schalansky

Natürlich hatte ich schon davor andere bemerkenswerte Begräbnisstätten besucht: die Toteninsel San Michele etwa, wie sie mit hohen, roten Backsteinmauern aus dem blaugrünen Wasser der Lagune von Venedig emporragt gleich einer uneinnehmbaren Festung, oder das grelle Jahrmarktstreiben des Hollywood Forever Cemetery am alljährlich von der mexikanischen Bevölkerung begangenen Día de los Muertos mit den orange-gelb geschmückten Gräbern und den von der fortgeschrittenen Verwesung auf ewig zum Grinsen verdammten Totenschädeln aus bunt gefärbtem Zucker und Pappmaché. Doch keine hat mich so berührt wie der Friedhof jener Fischersiedlung, in dessen eigentümlichem Grundriss — einer Art Kompromiss aus Kreis und Quadrat ich nichts anderes als ein Sinnbild der ungeheuerlichen Utopie zu erkennen glaubte, die ich dort verwirklicht sah: mit dem Tod vor Augen zu leben. Lange Zeit war ich überzeugt, an diesem Ort, dessen dänischer Name »kleine Insel« oder »vom Wasser umgeben« bedeutet, sei man dem Leben näher, gerade weil seine Bewohner die Toten wortwörtlich in ihre Mitte geholt hatten, anstatt sie wie sonst in unseren Breitengraden üblich — aus dem Innersten der Gemeinden vor die Stadttore zu verbannen, auch wenn der urbane Raum sich die Gräberstätten durch sein ungehemmtes Anwachsen oft nur wenig später wieder einverleibt hat.

Humans prefer PARA

Legend: ■ confident ■ unsure

- SENT: unsure 55.77%, confident 44.23%
- PARA: unsure 34.38%, confident 65.63%
- PARA_SENT: unsure 54.10%, confident 45.90%
- PARA: unsure 50.42%, confident 49.58%
- GTr: unsure 48.39%, confident 51.61%
- PARA: unsure 21.48%, confident 78.52%
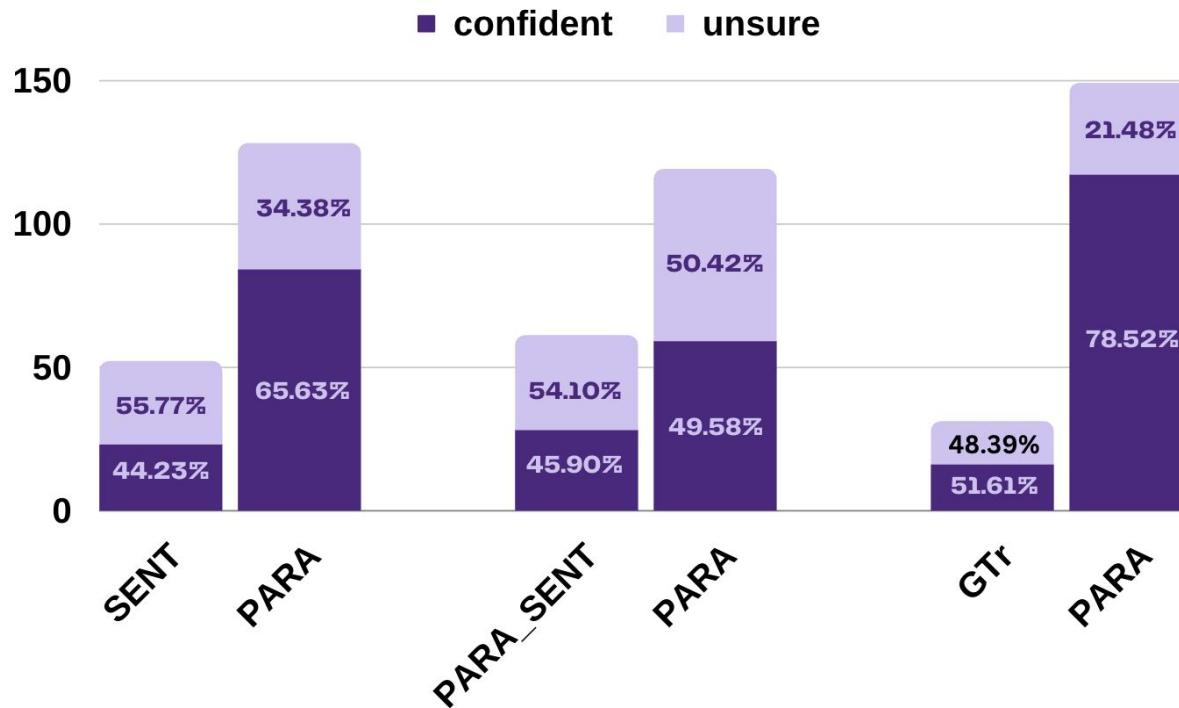
# PARA makes fewer errors
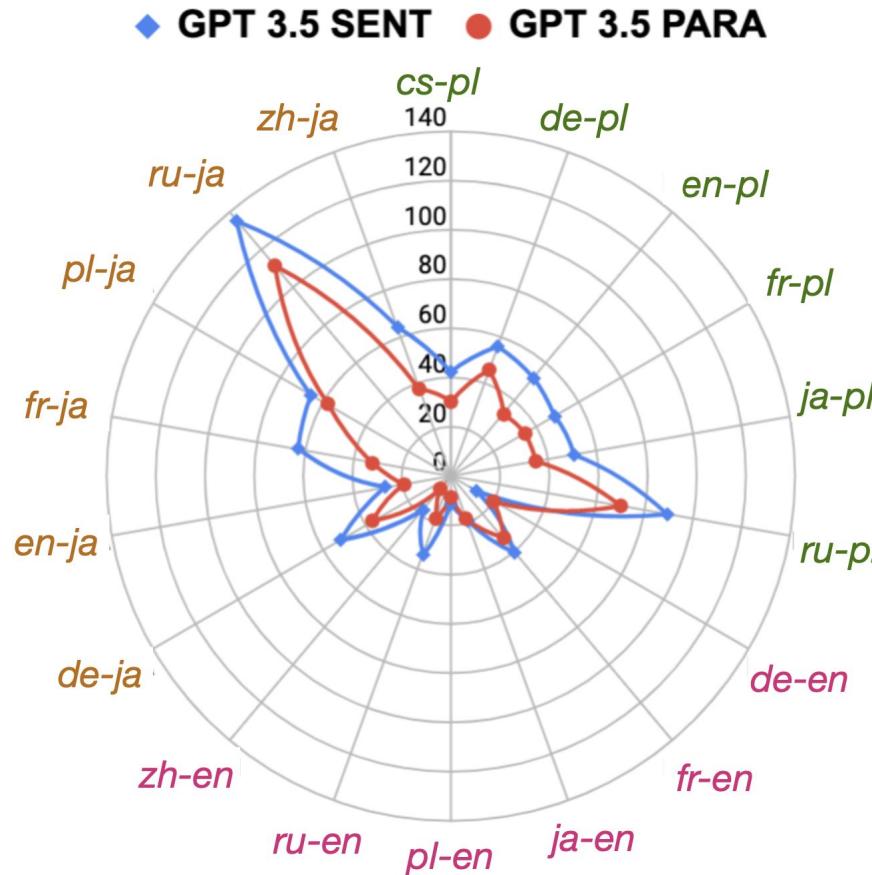
📋 PARA makes fewer discourse errors
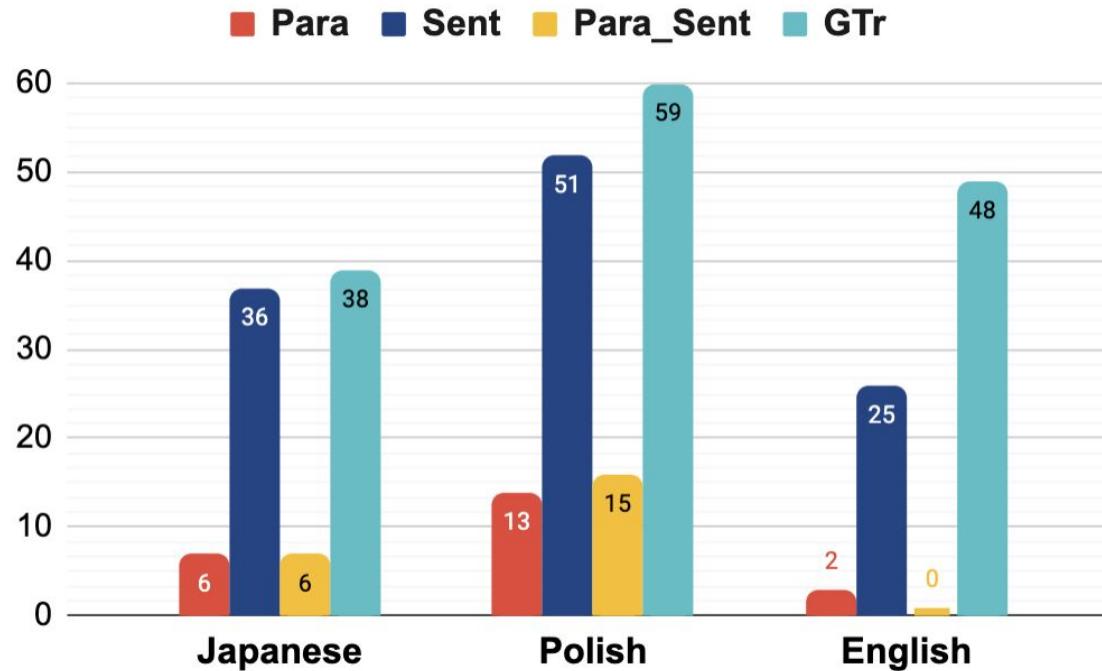
PARA makes fewer grammar errors
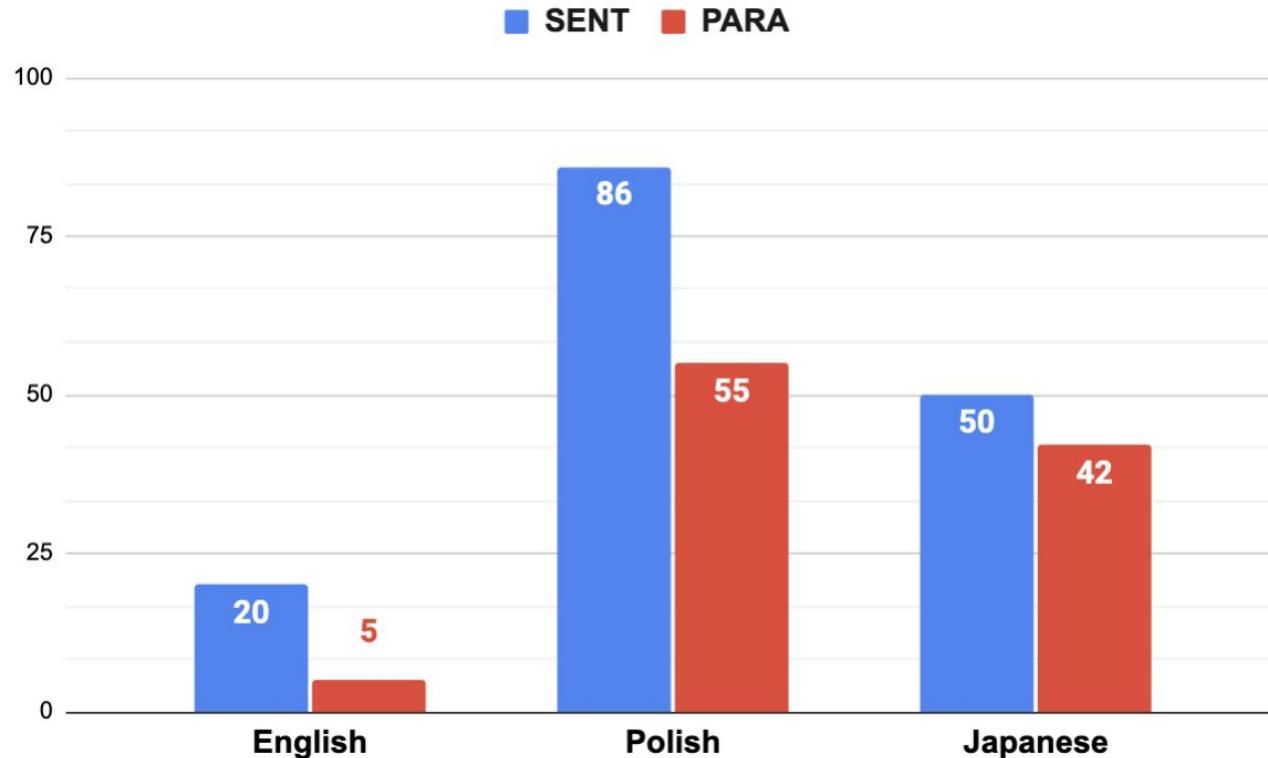
📋 Example of PARA and SENT translations

(...)
「あー、あと煙草の５番を一つ」
"Oh, and a one (pack) of cigarettes, number five."
「かしこまりました」
"Right away." (lit. (I) understood)
すばやくマルボロライトメンソールを抜き取り、レジでスキャンする。
(I) take out (a pack of) Marlboro Menthol Lights quickly, and scan it at the register.
「年齢確認のタッチをお願いします」
"Please confirm your age on the touch screen." (lit. "Age confirmation touch, please.")
画面をタッチしながら、男性の目線がファーストフードが並んだショーケースにすっと移ったのを見て、指の動きを止める。
As (he) touches the screen, (I) see that the man's gaze shifted to the showcase with the fast food, (and) (I) stop moving my finger(s).
—Japanese Source (from *Convenience Store Woman* by Sayaka Murata)

paragraph-level

(...)
"Ah, and one pack of cigarettes, number five."
"Right away."
I quickly pulled out a Marlboro Light Menthol and scanned it at the register.
"Please touch the screen for age verification."
His gaze shifted to the showcase with the fast food as he touched the screen, **and I stopped my finger's movement.**
—GPT 3.5 PARA (English)

sentence-level

(...)
"Ah, and one pack of cigarettes, number five."
"Understood."
Quickly remove the Marlboro Light Menthol and scan it at the register.
"Please confirm your age with a touch."
The man's gaze shifted quickly to the showcase where the fast food was lined up while he was touching the screen, **and he stopped his finger movement.**
—GPT 3.5 SENT (English)

Contextual Errors: Pronouns

Романы, как известно, печатались на разной **бумаге** **[*paper*]** . И гореть **она** **[*she*]** может по-разному.
                                                        –Russian Source (from *Maranaga*)

Paper is a **feminine** noun in Russian but a **masculine** noun in Polish.

Romany, jak wiadomo, drukowano na różnym **papierze** **[*paper*]** . I może **ona** **[*she*]** tęsknić na różne sposoby.
                                                        –GPT 3.5 SENT (Polish)

Jak wiadomo, powieści drukowano na różnym **papierze** **[*paper*]** . I może **on** **[*he*]** palić się na różne sposoby.
                                                        –GPT 3.5 PARA (Polish)

Contextual Errors: Subject Ellipsis

When we were done, the lipstick went back into some mother's Fendi handbag. We watched her apply it, **unaware** [(*she*)] .
> —English Source (from *A Children's Bible*)

Who is *unaware*, "her" or "we"?

Gdy skończyliśmy, szminka wróciła do jakiejś torebki Fendi należącej do matki. Patrzyliśmy, jak to robi, **nieświadomi** [*unaware (we)*] tego.
> —GPT 3.5 SENT (Polish)

PARA

Kiedy skończyliśmy, szminka wróciła do torebki Fendi jakiejś matki. Patrzyliśmy, jak ją nakłada, **nieświadoma** [*unaware (she)*] naszych działań.
> —GPT 3.5 PARA (Polish)

Contextual Errors: Cultural Nuances

「気が付かなくてすみません」
「いやいや、(…)。古倉さんは毎日勤務
 [lit. *Ms./Mrs./Mr. Furukura works every day*]
なのに手を抜かないからねー！」
—Japanese Source (from *Convenience Store Woman*)

In Japanese, it's common for the speaker to refer to the listener in the third person. In English, we use second-person pronouns ("you").

SENT

"I'm sorry I didn't notice."
"No, no, (…). **Furukura-san works hard every day** without taking any shortcuts!"
—GPT 3.5 SENT (English)

PARA

"I'm sorry I didn't notice."
"No, no, (…). **You work every day**, but you never slack off!"
—GPT 3.5 PARA (English)

Contextual Errors: Polisemy

Все прошло хорошо. Книга прочитана идеально – не быстро и не медленно, минимум дыма. Классика. Я был **в форме**. [*in shape*]

–Russian Source (from *Maranaga*)

This Russian noun can mean either "shape" or "uniform". From the context, it is clear that the former sense is preferred.

SENT

Wszystko poszło dobrze. Książka została przeczytana idealnie – nie szybko i nie wolno, minimalna ilość dymu. Klasyka. Byłem **w mundurze**. [*in uniform*]

–GPT 3.5 SENT (Polish)

PARA

Wszystko poszło dobrze. Książka przeczytana idealnie – nie szybko i nie wolno, minimalna ilość dymu. Klasyka. Byłem **w formie**. [*in shape*].

–GPT 3.5 PARA (Polish)

Contextual Errors: Consistency

Alles zu vergessen, ist gewiss **schlimm** [*bad*]. Noch **schlimmer** [*worse*] ist, nichts zu vergessen (...).
        –German Source (from *An Inventory of Losses*)

The German translates to "To forget everything is **bad**, certainly. **Worse** still is to forget nothing." The use of "tragic" instead of "bad" breaks stylistic consistency.

SENT

すべてを忘れることは確かに 悲惨な [*tragic*] ことです。さらに悪い [*worse*] のは、何も忘れないことです。
        –GPT 3.5 SENT (Japanese)

PARA

すべてを忘れることは確かに悪い [*bad*] ことです。もっと悪い [*worse*] ことは、何も忘れないことです。
        –GPT 3.5 PARA (Japanese)

Contextual Errors: Coherence

オープンまでの二週間、二人組になったり、社員を相手にしながら、架空の客に向かって、ひたすら練習が続いた。「お客様」の目を見て微笑んで一礼すること、生理用品は紙袋に入れること、温かい物は冷たい物と分けて入れること、ファーストフードを頼まれたら手をアルコールで消毒すること。

The two weeks until the opening we made pairs and practiced with other employees, we practice facing imaginary customers.

Look at the eyes of the "customer", smile, bow; put sanitary items in a paper bag; divide hot and cold items; if (the customer) asks you for fast food disinfect (your) hands.

–Japanese Source (from *Convenience Store Woman*)

SENT

For two weeks leading up to the opening, we practiced relentlessly with each other and with the staff, as if we were facing imaginary customers. Smiling and bowing when making eye contact with customers, putting sanitary products in paper bags, separating hot items from cold items, and disinfecting hands with alcohol when ordering fast food.

–GPT 3.5 SENT (Japanese)

PARA

For two weeks, we practiced relentlessly, in pairs and with our colleagues, pretending to be customers. We learned to smile and bow politely, to put sanitary items in paper bags, to separate hot and cold items, and to disinfect our hands with alcohol if we were asked for fast food.

–GPT 3.5 PARA (Japanese)

PARA adds a coherent opening to the second sentence resulting in better style

✂️ Limitations

- **PARA** still suffers from many errors, including *omissions* of content:

  It's a difficult choice because the translations are complementary in their flaws.
  —French to Polish Translator (Małgorzata Szymczak)

- Our study focused on **GPT 3.5**… what about **GPT-4**?
  - **Qualitatively:** at times better, but still makes similar errors

- Large language models, like other MT systems, are very literal:

  Both translations [SENT and PARA] *translate the words without the feeling; the original author's voice is lost.*
  —French to English Translator (Kristy Darling Finder)

Ethical Consideration

- **Toxicity and Bias in LLMs** (Blodgett et al., 2020; Bender et al., 2021)

- **Multiple stakeholders:**
  - the author – perception of author's work, transfer of ideas to the TL
  - the translator – underestimation of translator's work
  - the reader – impairing linguistic abilities

- **MT employed *responsibly* as an auxiliary tool can:**
  - alleviate the translator's cognitive burden (O'Brien, 2012)
  - make author's work accessible to a broader audience (Besacier, 2014)