ACL 2025 VIENNA

TRANSLATION**CORRECT**

A Unified Framework for Machine Translation Post-Editing with Predictive Error Assistance

Syed Mekael Wasti[†], Shou-Yi Hung[‡], Christopher Collins[†], En-Shiun Annie Lee[†,‡]

[†]Ontario Tech University [‡]University of Toronto

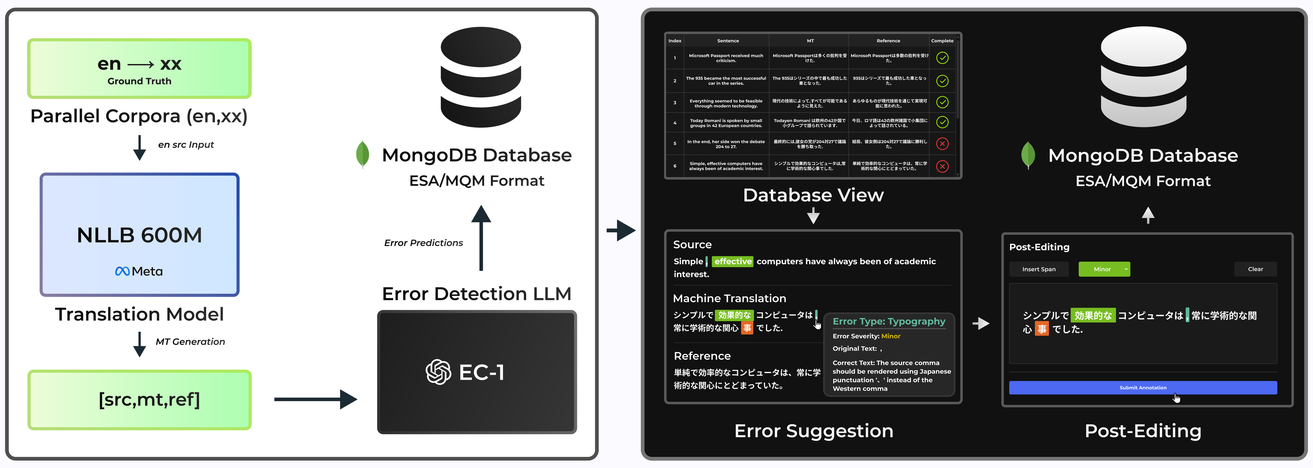
Motivation

Problem:

- Existing Neural Machine Translation (NMT) system workflows are **manual** and **disconnected**
- They do not offer **error prediction** or **ESA/MQM** integration
- Tools like **Appraise** help but lack detailed **annotation features** and post-editing **flexibility**

Our Solution:

- **MT generation** using models such as NLLB
- **Automated error prediction** using **XCOMET** or a custom **EC-1** assistant (GPT-4o)
- **Interactive post-editing UI** with real-time span highlighting and explanations
- **One-click export** in **MQM/ESA** formats for immediate use in research or training



Dataset Creation

Annotator Post-Editing

TRANSLATION**CORRECT** Workflow

Database View

Index	Sentence	MT	Reference	Complete
1	Microsoft Passport received much criticism.	Microsoft Passportは多くの批判を受けた。	Microsoft Passportは多数の批判を受けた。	✓
2	The 935 became the most successful car in the series.	The 935はシリーズの中で最も成功した車となった。	935はシリーズで最も成功した車となった。	✓
3	Everything seemed to be feasible through modern technology.	現代の技術によって,すべてが可能であるように見えた。	あらゆるものが現代技術を通じて実現可能に思われた。	✓
4	Today Romani is spoken by small groups in 42 European countries.	Todayen Romani は欧州の42か国で小グループで語られています。	今日、ロマ語は42の欧州諸国で小集団によって話されている。	✓
5	In the end, her side won the debate 204 to 27.	最終的には,彼女の党が204対27で議論を勝ち取った。	結局、彼女側は204対27で議論に勝利した。	✗
6	Simple, effective computers have always been of academic interest.	シンプルで効果的なコンピュータは,常に学術的な関心事でした。	単純で効率的なコンピュータは、常に学術的な関心にとどまっていた。	✗

Users begin with a database view that displays side-by-side source and MT outputs (generated using NMT), allowing easy loading, review, and status tracking of annotations stored in a MongoDB database

Error Suggestion

Source

Simple, effective computers have always been of academic interest.

Machine Translation

シンプルで 効果的な コンピュータは, 常に学術的な関心 事 でした。

Reference

単純で効率的なコンピュータは、常に学術的な関心にとどまっていた。

Error Type: Typography

Error Severity: Minor

Original Text: ,

Correct Text: The source comma should be rendered using Japanese punctuation '、' instead of the Western comma

Following the MT generation, TRANSLATION**CORRECT** integrates an error detection model of the user's choice to identify potential errors in the translated output automatically

*The feature display on the right showcases the flexible options provided to annotators by TRANSLATION**CORRECT***

Post-Editing

Insert Span | Minor | Clear

シンプルで 効果的な コンピュータは, 常に学術的な関心 事 でした。

Submit Annotation

The Post-Edit component allows users to make detailed, fine-grained error edits on top of the potential error spans generated by our error detection model.

Insert Span

You may insert a span by highlighting a segment of text using your mouse, and either using the pop-up insert shortcut button

シンプルで 効果的な コンピュータは, 常に学術的な関心 事 でした。 +

or by using the insert span button

Insert Span | Minor

Choose Error Category

Upon insertion, you can select the appropriate error category through the drop down menu that appears automatically. You can right click on a span to open this at any time

シンプルで 効果的な

Addition
Omission
Mistranslation
Untranslated
Grammar
Spelling
Typography
Unintelligible

Choose Error Severity

You can then assign an error severity to a span, choosing between Minor & Major options

シンプルで 効果的な

Minor
Minor
Major

Delete Individual Span

Deleting an individual span can be done by hovering over a span and clicking the pop-up delete button

x

シンプルで 効果的な コンピュータは

Clear All Spans

The clear button can be used to delete all spans and edits in the post-edit section. The button must be clicked twice to confirm

Clear | Clear?

Submit Your Annotations

Once your annotation is complete to your liking, you can click the Submit Annotation button to submit it to the backend database

Insert Span | Minor | Clear

シンプルで 効果的な コンピュータは, 常に学術的な関心 事 でした。

Submit Annotation

Evaluation

User Study:

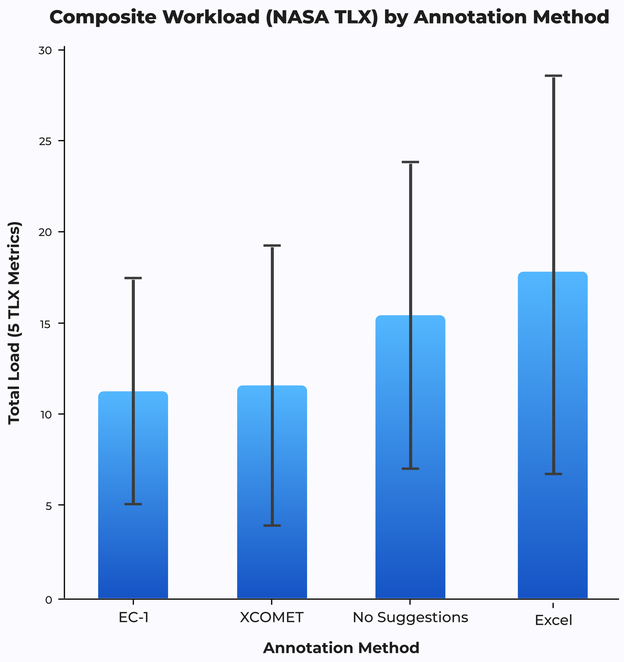
- To assess usability and cognitive load, we conducted a user study with **12 native speakers** across 6 languages using the NASA TLX framework:
  - **Manual (Excel)**
  - TRANSLATION**CORRECT: No Suggestions**
  - TRANSLATION**CORRECT: XComet Suggestions**
  - TRANSLATION**CORRECT: EC-1**

Key Results

- ***EC-1*** condition had the ***lowest workload across all metrics***
- TRANSLATION**CORRECT** (XCOMET or EC-1) consistently outperformed Excel in:

↓ Mental Demand ↓ Frustration

↓ Physical Effort ↑ Perceived Performance



Method	Mental (↓)	Physical (↓)	Temporal (↓)	Frustration (↓)	Effort (↓)	Performance (↑)
Excel	4.10 ± 2.51	3.40 ± 2.88	2.70 ± 2.26	3.50 ± 2.92	4.10 ± 2.38	7.80 ± 1.55
No Suggestions	4.17 ± 2.52	2.42 ± 2.57	3.58 ± 2.02	1.83 ± 2.41	3.42 ± 1.16	8.58 ± 1.16
XCOMET	2.92 ± 1.56	**1.58 ± 1.51**	2.50 ± 1.68	1.92 ± 2.31	**2.67 ± 1.07**	**8.67 ± 1.07**
EC-1	**2.67 ± 1.87**	**1.58 ± 1.08**	**2.17 ± 1.59**	**1.75 ± 2.26**	3.08 ± 1.00	8.50 ± 1.00

Table 1: Comparison of NASA TLX dimensions across annotation methods, including Excel spreadsheet-based annotations and TRANSLATION**CORRECT** with the different error detection settings available. Lower is better (↓) for all metrics except Performance (↑). Bold indicates the best score for each metric.

	Mental	Physical	Temporal	Effort	Frustration	Performance
Mental	1.00	0.44	0.46	0.47	0.52	-0.06
Physical	0.44	1.00	0.49	0.34	0.30	-0.44
Temporal	0.46	0.49	1.00	0.37	0.25	-0.20
Effort	0.47	0.34	0.37	1.00	0.60	-0.26
Frustration	0.52	0.30	0.25	0.60	1.00	-0.28
Performance	-0.06	-0.44	-0.20	-0.26	-0.28	1.00

Table 2: Correlation Matrix of the NASA TLX Metrics showcases that ***Perceived Performance*** was ***negatively correlated*** with most workload dimensions, suggesting that reducing user effort and fatigue may directly contribute to greater perceived success

Conclusion & Demo

We present TRANSLATION**CORRECT**, a unified framework that simplifies machine translation workflows by combining MT generation, error detection, and post-editing in a single interface. It supports ESA and MQM-standard exports, making it ideal for both translators and researchers. Our human study shows that TRANSLATION**CORRECT reduces workload, increases efficiency, and is more enjoyable than traditional spreadsheet-based workflows.** By prioritizing HCI design, TRANSLATION**CORRECT** improves the annotation experience while enabling the creation of high-quality, future-proof datasets for fine-tuning models like XCOMET, NLLB, and beyond.

TRANSLATION**CORRECT** is available here: *https://translation-correct-annotation-git-27a7e8-mekaelwastis-projects.vercel.app*

All project code is publicly available at: *https://github.com/MekaelWasti/TranslationCorrect*