

LANGUAGE AND TRANSLATION TECHNOLOGY TEAM (LT3)

Bram Vanroy, Arda Tezcan, Lieve Macken



# MATEO (2022-2023): MACHINE TRANSLATION EVALUATION ONLINE

## Project goals

- Accessible website to evaluate machine translation for both experts (MT analysts or system builders) and non-experts (SSH, students, others)
- Python command-line interface for expert users
- Established and state-of-the-art metrics
- Advanced, research-focused features: custom metric options, batch processing, exporting results to Excel/LaTeX
- Integration in CLARIN B (INT-NL) and open-source access

## Funding

Starting grant: European Association for Machine Translation  
Full grant: CLARIN.eu *Bridging Gaps*

## Current demo (first, alpha version)

- <https://vtctools.ugent.be/mateo/evaluate>
- Translate your data automatically with Facebook's "many-to-many" model (M2M)
- Easily add DeepL, Google Translate or other translations
- Select evaluation metrics
- Evaluate! Save visualizations and results

## 1. Translate automatically

Source language: English | Target language: Dutch; Flemish | Model size: 1.2B

Source text: As investors took fright, gilt yields surged, prompting the Bank of England to close early on September 28th.

Reference translation: Omdat beleggers schrokken, stegen de rendementen op obligaties waardoor de Bank of England op 28 september vroeger sloot.

Machine translation: Terwijl beleggers angst kregen, stegen geleegde opbrengsten, waardoor de Bank of England vroeg op 28 september moest sluiten.

## 2. Add other translations (optional)

Other translation #1 (Google): Terwijl beleggers bang werden, stegen de rendementen op staatsobligaties, wat de Bank of England ertoe aanzette om op 28 september vervroegd te sluiten.

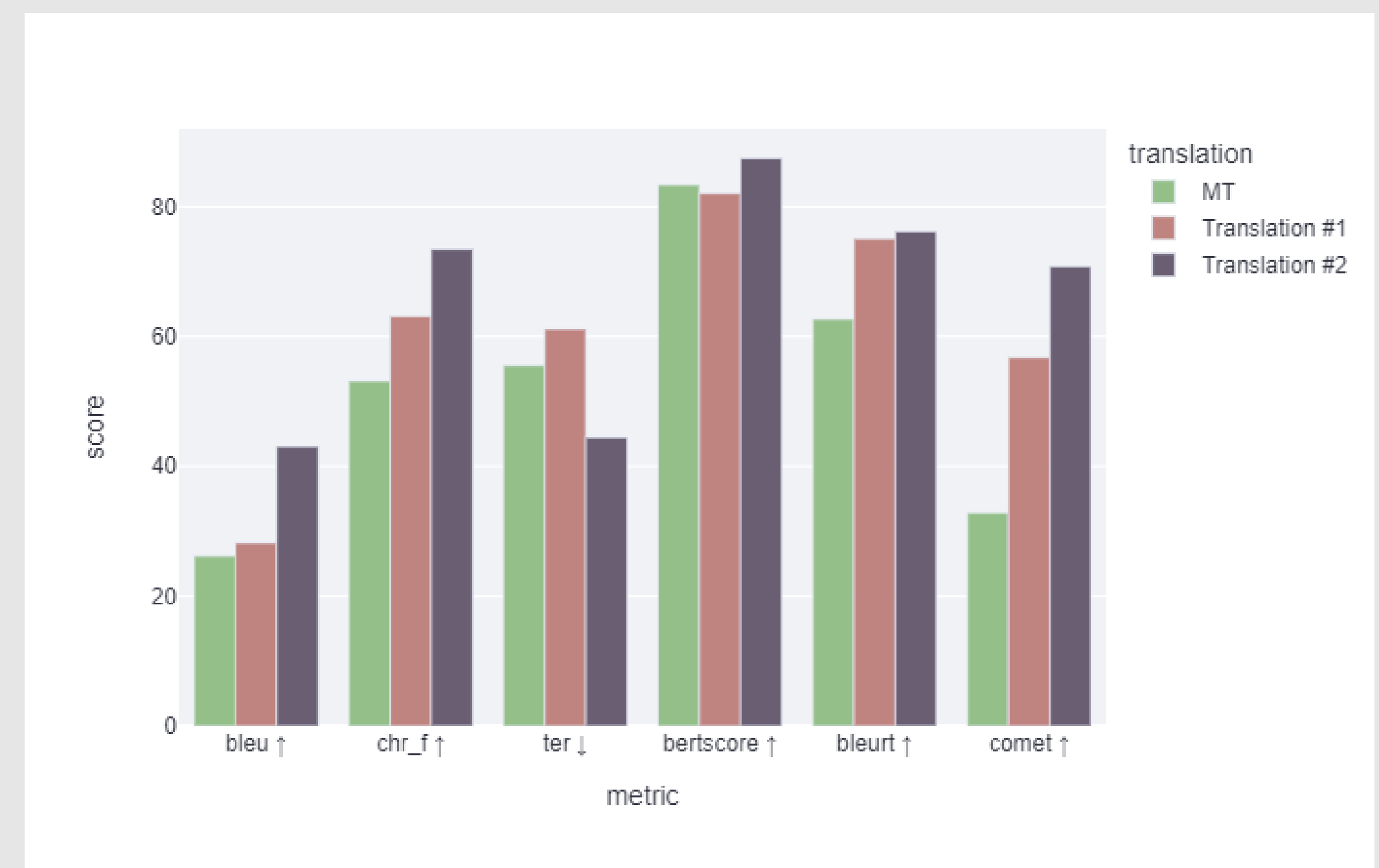
Other translation #2 (DeepL): Omdat beleggers bang werden, stegen de obligatierendementen, waardoor de Bank of England op 28 september vervroegd moest sluiten.

## 3. Select evaluation metrics

Which metrics do you wish to use to evaluate the given MT output?

Metrics to calculate: bleu, chr\_f, ter, comet, bleurt, bertscore

## 4. Evaluate, visualize, save results



What do you need? What would make your life easier in terms of MT evaluation? Let us know!

@ [bram.vanroy@ugent.be](mailto:bram.vanroy@ugent.be)  
@bramvanroy / @lt3ugent  
Bram Vanroy

		bleu	chr_f	ter	bertscore	bleurt	comet
Source	close early on September						
Ref.	de Bank of England op 28						
MT	nk of England vroeg op 28	26.153118	53.090791	55.555556	83.347470	62.613690	32.840598
Translation #1	vat de Bank of England er	28.196659	63.053753	61.111111	82.002395	74.994314	56.754404
Translation #2	de Bank of England op 28	42.976414	73.387787	44.444444	87.370849	76.107210	70.776457