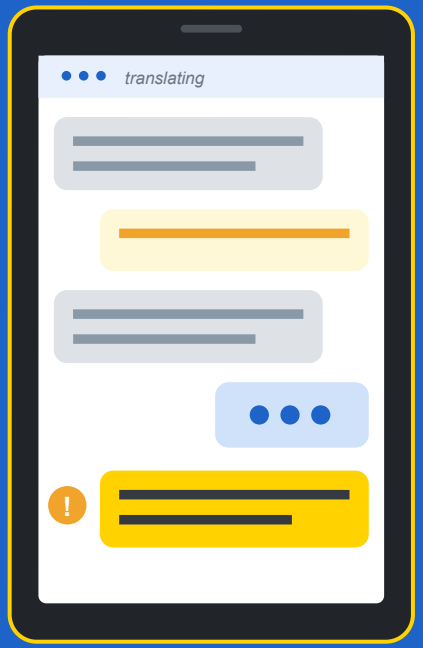
Multilingual Communication in the Asylum Context

Evaluating LLM-Based MT with Fuzzy-Match Augmentation and Adaptive NMT under Low-Data Constraints



Thomas Moerman · Arda Tezcan · Lieve Macken

Ghent University, LT³ Language and Translation Technology Team

EAMT 2026, Research / Technical Track

GHENT UNIVERSITY

language and translation technology team

ABOUT THE MATIAS PROJECT

Machine Translation to Inform Asylum Seekers. An interdisciplinary project at Ghent University.

PARTNERS AND FUNDING

GHENT UNIVERSITY

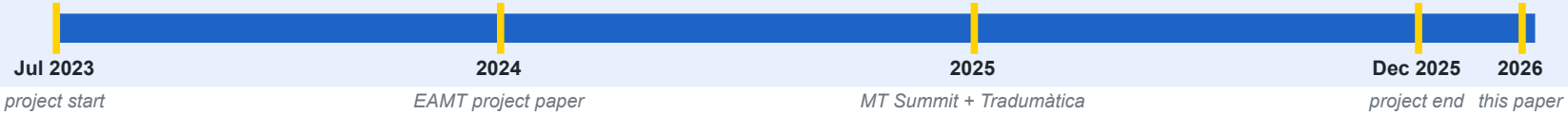
fedasil
FEDERAAL AGENTSCHAP VOOR DE OPVANG VAN ASIELZOEKERS

Federal asylum agency, Belgium

AMIF · EU Asylum, Migration and Integration Fund

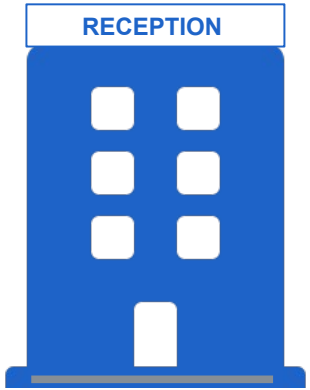
Grant AMIF 093-133. Co-funded by the European Union.

TIMELINE



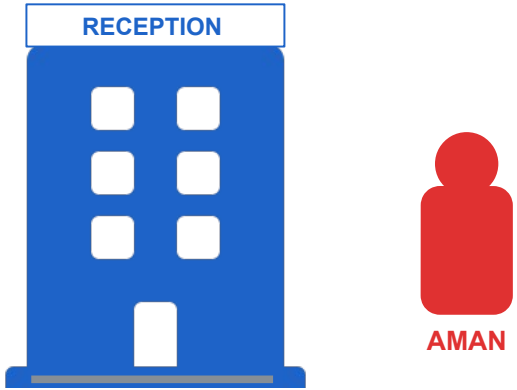
POSTER *Lieve Macken presents more on the broader MaTIAS project at the poster session tomorrow (Poster session 2, 11:25-12:25)*

THE SCENE (MaTIAS)



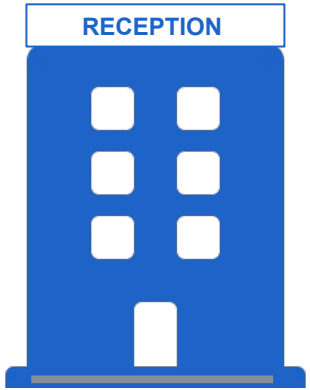
An asylum reception centre in Belgium

THE SCENE (MaTIAS)



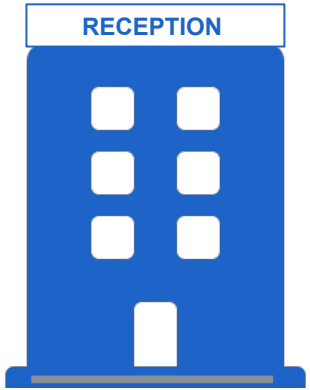
An asylum reception centre in Belgium

THE SCENE (MaTIAS)



An asylum reception centre in Belgium

THE SCENE (MaTIAS)

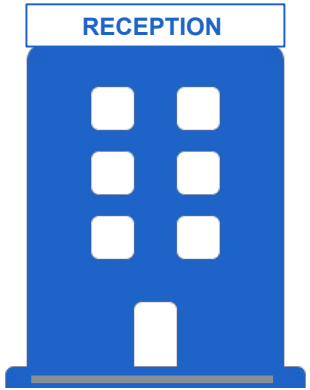


An asylum reception centre in Belgium

ትግርኛ

staff: **Dutch · French · English**

THE SCENE (MaTIAS)



An asylum reception centre in Belgium

ትግርኛ

staff: **Dutch · French · English**

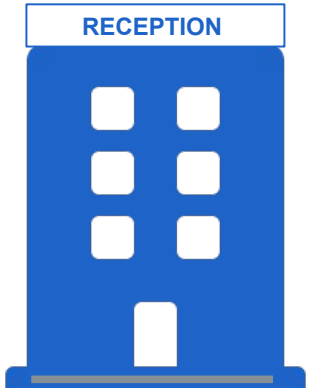
العربية | Deutsch | Español | فارسی | Português

Română | Русский | Türkçe | Հայերեն | Shqip

ქართული | پښتو | Soomaali | ትግርኛ

residents

THE SCENE (MaTIAS)



An asylum reception centre in Belgium

AMAN

ትግርኛ

staff: Dutch · French · English

العربية | Deutsch | Español | فارسی | Português

Română | Русский | Türkçe | Հայերեն | Shqip

ქართული | پښتو | Soomaali | ትግርኛ

residents

How does Aman communicate?

BEFORE MATIAS: HOW AMAN COMMUNICATES

Staff send messages every day. How does Aman, a Tigrinya speaker, read them?

Fellow resident as informal interpreter

Often a child or a more recently-arrived neighbour.

Google Translate on his phone

Works passably for high-resource languages. For Tigrinya, often unusable.

Posters in hallways

Static, not targeted, often out of date, not practical for all languages.

Or he misses the message

Missed class, missed meal, missed appointment.

GOOGLE TRANSLATE, EN → TI, TODAY

Input: "You have an appointment at the medical unit on 2 October at 10:00."

Output: ኣብ ሕክምናዊ ክፍሊት፣ ኣብ 2 ጥቅምት ሰዓት 10:00 ናይ መገብያ ቕጥዒት ኣለዎም። *← reads as something about a payment, not an appointment.*

Privacy: Google Translate routes sensitive content (medical, legal, personal) through external servers.

WITH MATIAS: WHATSAPP DELIVERY

Staff types one message; the platform translates into 14 languages and delivers via WhatsApp.

STAFF · Web composer

Source language	English
Recipient targeting	Centre X — all residents
Schedule	Send now

Message:

"You have an appointment at the medical unit on 2 October at 10:00."

→

WhatsApp · MaTIAS

EN *You have an appointment at the medical unit on 2 October at 10:00.*

TI ዕለት 2 ጥቅምቲ ሰዓት 10:00 ኣብ ኣሃዱ ሕክምና ቆጸራ ኣለኩም።

source + translation side-by-side
(from ethnographic fieldwork)

14 target languages. One-way delivery. Targeting by centre, group, floor, or resident.

[Macken et al. 2025a]

[Macken et al. 2024]

CURRENTLY DEPLOYED SYSTEM

Selected via separate multi-system comparison (Macken et al. 2025).

MMT+

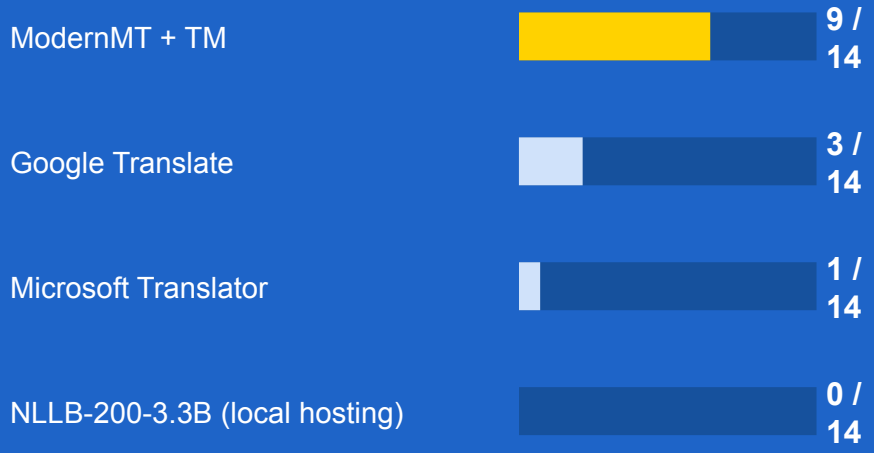
ModernMT + Translation Memory

Drag-and-drop TM upload for all 14 target languages was the decisive factor.

Same ModernMT instance + 358-sent TM is the baseline this paper measures against.

[Macken et al. 2025b]
[Bertoldi et al. 2018]

WIN COUNT ACROSS 14 LANGUAGES



Source: Macken et al. (2025) preliminary phase, ChrF on 200-msg test (Dec 2024 to Jan 2025).

HCAILT: THE FRAMEWORK BEHIND THE DESIGN

Briva-Iglesias & O'Brien (2026): a framework for high-stakes multilingual MT. Three pillars.

RELIABILITY	SAFETY	TRUSTWORTHINESS

HCAILT: THE FRAMEWORK BEHIND THE DESIGN

Briva-Iglesias & O'Brien (2026): a framework for high-stakes multilingual MT. Three pillars.

RELIABILITY	SAFETY	TRUSTWORTHINESS
Consistent, accurate outputs across users and conditions.	*Privacy, data sovereignty, compliance.*	*Users can justifiably rely on outputs.*

HCAILT: THE FRAMEWORK BEHIND THE DESIGN

Briva-Iglesias & O'Brien (2026): a framework for high-stakes multilingual MT. Three pillars.

RELIABILITY	SAFETY	TRUSTWORTHINESS
Consistent, accurate outputs across users and conditions.	*Privacy, data sovereignty, compliance.*	*Users can justifiably rely on outputs.*
HCAILT advocates RAG (retrieval-augmented generation). Here this is *fuzzy-match augmentation*.	**Open-source-vs-commercial is a safety trade-off.**	**Needs human evaluation, not only automatic metrics.**

The unresolved tension HCAILT identifies: data scarcity in less-spoken languages produces quality disparities.

[Briva-Iglesias & O'Brien 2026]

WHAT IS A FUZZY MATCH?

A **fuzzy match** is a sentence in the translation memory that is **similar but not identical** to the one you want to translate.

Give the model the **most similar** TM sentences as in-context examples (not random ones, not none).

WORKED EXAMPLE

TEST *You have an appointment at the medical unit on 2 October at 10:00.*

#1	Make **an appointment** for this through your social worker or **the medical unit**.
#2	You can get vaccinated **at the medical unit** of the reception centre.

← similarity …

source → **FM retrieval** → **LLM prompt** → **translation**

[Bulte & Tezcan 2019]
[Moslem et al. 2023]
[Bouthors et al. 2024]

RESEARCH QUESTIONS

Improve MT for low-resource languages in low-resource deployment settings, using FM augmentation.

1 **Can FM augmentation work at small TM scale?**

358 sentences, vs the hundreds of thousands in prior FM work.

2 **Does it help low-resource languages specifically?**

Where FM augmentation has not been shown to work before.

3 **How do adaptive NMT and retrieval-augmented LLMs compare?**

Across high, mid, and low resource tiers.

Source languages: Dutch, French, English. 14 target languages span high, mid, low resource.

DATASET - MaTIAS

Compiled via ethnographic fieldwork in 4 Belgian reception centres. Topics include house rules, hygiene, appointments, opening hours, classes.

400

Candidate messages

→

200

Curated, professionally translated into all 14 languages

→

100 even → 358 sentences

Translation Memory (TM)

100 odd → 423 sentences

Test set

358

sentences

[Macken et al. 2025a]
[Macken et al. 2025b]

WHY THIS IS CHALLENGING

14 TARGET LANGUAGES

HIGH RESOURCE

AR DE ES FA PT RO RU TR

MID RESOURCE

HY SQ KA

LOW RESOURCE

PS SO TI

[Joshi et al. 2020]
[NLLB Team 2022]

Pashto, Somali, Tigrinya : off-the-shelf MT often produces output residents can't use.

THE CONSTRAINTS

- Short, domain-specific messages
- Privacy: asylum-seeker data is sensitive
- Very little training data

EXPERIMENTAL OVERVIEW

Two paradigms compared. Open-source and commercial LLMs both tested.

ADAPTIVE NMT

MMT+ — *ModernMT with our 358-sentence TM. Currently deployed.*

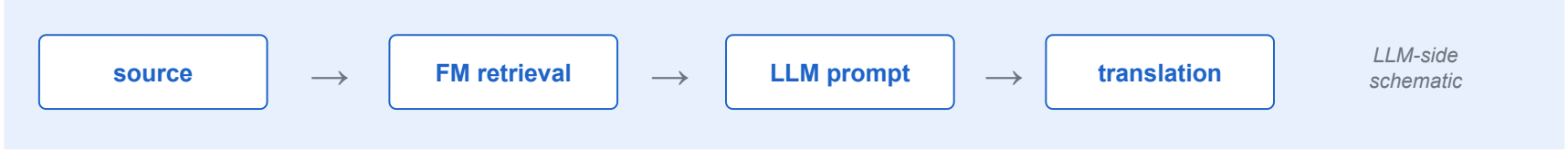
MMT — *ModernMT without the TM. Isolates the TM contribution.*

LLM + FM AUGMENTATION

TG-4B / 12B / 27B — *TranslateGemma. Open-source, local GPU.*

TowerPlus-9B — *Unbabel. Open-source, 5 of 14 langs only.*

Gemini Pro 2.5 — *Commercial, API call. Closed-source.*



LLM-side schematic

Adaptive NMT vs retrieval-augmented LLMs. Open-source (local) and commercial (API). Same TM, same test set.

THREE PROMPTS, ONE TEST SENTENCE

Same test sentence in all three conditions. The contrast is the methodological point.

ZERO-SHOT

Translate this message into Tigrinya:

(no in-context examples)

EN You have an appointment at the medical unit on 2 October at 10:00.

TI ____ ← *model fills this in*

RANDOM-15

Translate this message into Tigrinya:

EN Don't forget to apply sunscreen.

TI ሳንስክሪን ምግባር ኣይትረስዑ።

+ 14 more random examples (off-topic).

EN You have an appointment at the medical unit on 2 October at 10:00.

TI ____ ← *model fills this in*

FUZZY-15

Translate this message into Tigrinya:

EN Make an appointment through your social worker or the medical unit.

TI ብመገዲ ማሕበራዊ ሰራሕተኛኹም ወይ ኣሃዱ ሕክምና ኣቢልኩም ነዚ ቆጸራ ሓዙ።

+ 14 more most-similar examples.

EN You have an appointment at the medical unit on 2 October at 10:00.

TI ____ ← *model fills this in*

Output (TI): random-15 mistranslates 'appointment' as 'question'. Fuzzy-15 is character-identical to the human reference.

WHY CHRF?

We compute all three metrics. ChrF is the one that survives across all 14 languages of this study.

chrF — PRIMARY

Character-level overlap.

Works robustly across all 14 target languages, including non-Latin scripts (Ge'ez, Armenian).

BLEU — appendix

Word-level overlap.

Tokenisation issues for some scripts. Ge'ez punctuation is not properly stripped in standard preprocessing, so Tigrinya BLEU is unreliable.

COMET — appendix

Neural quality estimation (wmt22-comet-da).

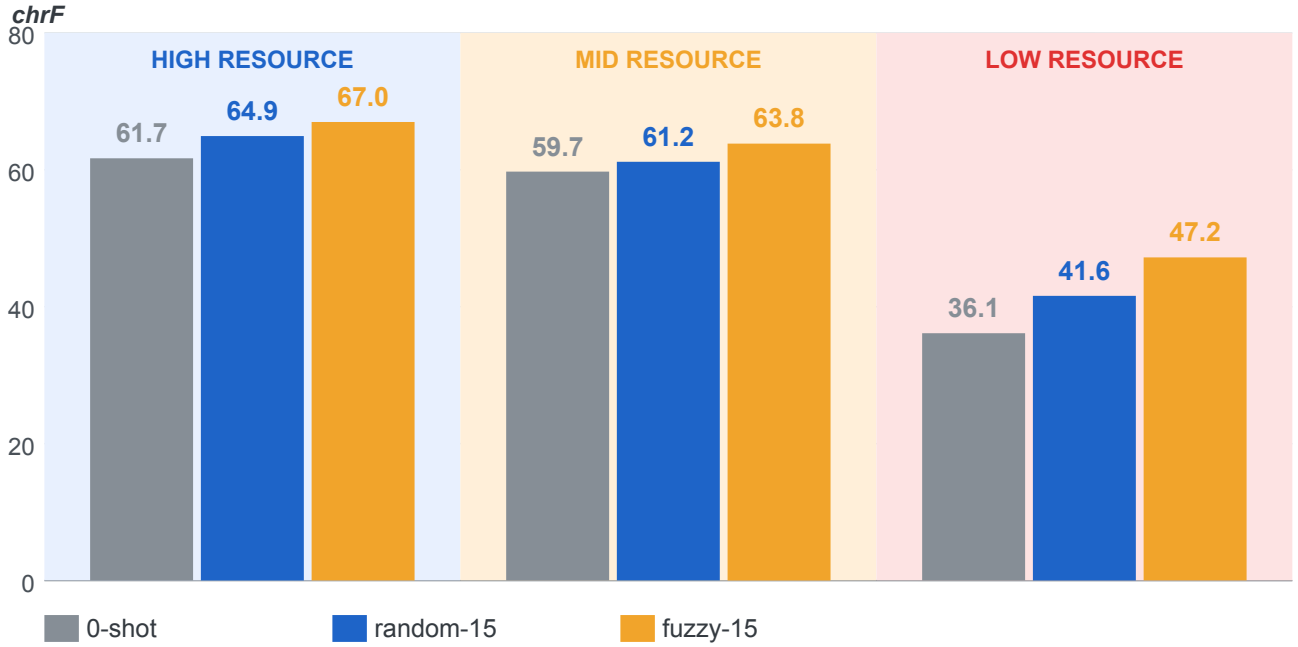
Unreliable for several low-resource languages.

ChrF is the only one of the three robust across all 14 languages. We report all three; the figures here use chrF.

[Popović 2015]
[Papineni et al. 2002]
[Rei et al. 2020]

RESULT 1: FM AUGMENTATION WORKS

TG-27B, chrF averaged per resource tier. Three conditions per tier.



RELIABILITY

evidence

Gain is consistent across tiers, biggest where reliability is hardest.

Low tier: +11 chrF

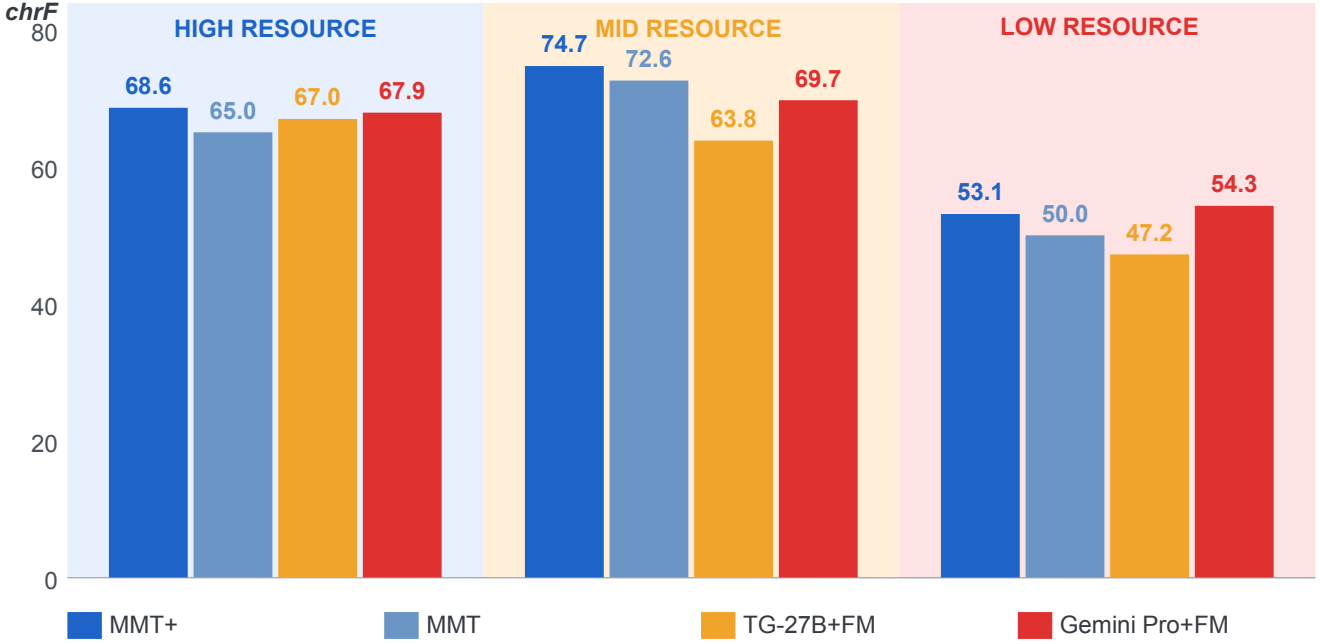
zero → fuzzy

11 of 14 langs:
fuzzy > random at p<0.01

[Paper Table 2 / §5.2]

RESULT 2: ADAPTIVE NMT VS LLM, BY TIER

Four systems × three tiers. ChrF, fuzzy-15 for LLMs.



AVG (14) MMT+ **66.6** · Gemini Pro **65.4** · MMT **63.4** · TG-27B **62.1**

[Paper Table 2]

SAFETY
trade-off

Mid tier: MMT+ dominates. Low tier: Gemini Pro overtakes MMT+ (54.3 vs 53.1).

TG-27B trails Gemini Pro by ~5 chrF: the cost of keeping data local.

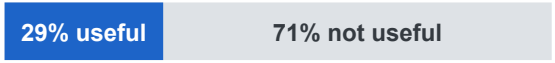
TIGRINYA RESULTS

Tigrinya is the language where the deployed system most clearly fails users.

HUMAN EVAL MMT+, TIGRINYA (Macken et al. 2025)

29%

of Tigrinya messages rated above 3 (out of 5) for meaning preservation



GHENT UNIVERSITY

[Macken et al. 2025b]
[Mager et al. 2023]

TIGRINYA RESULTS

Tigrinya is the language where the deployed system most clearly fails users.

HUMAN EVAL MMT+, TIGRINYA (Macken et al. 2025)

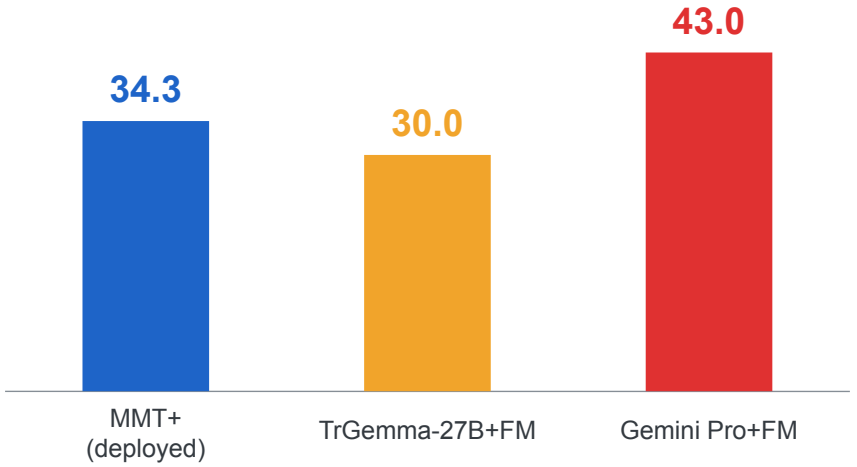
29%

of Tigrinya messages rated above 3 (out of 5) for meaning preservation

29% useful	71% not useful

CHRF, TIGRINYA (en → ti, fuzzy-15)

+8.7 over MMT+



+8.7 chrF over the deployed system, at the cost of routing sensitive data through a commercial API.

[Macken et al. 2025b]
[Mager et al. 2023]

GHENT UNIVERSITY

TAKEAWAYS AND LIMITATIONS

TAKEAWAYS

- FM augmentation beneficial even with very small TMs.
- Largest wins on **low-resource** languages.
- Adaptive NMT (MMT+) **still wins overall**.
- Commercial LLMs **now competitive**. Gemini Pro ≈ MMT+ on average.
- Low-resource gains concentrated on **Tigrinya**.

TAKEAWAYS AND LIMITATIONS

TAKEAWAYS

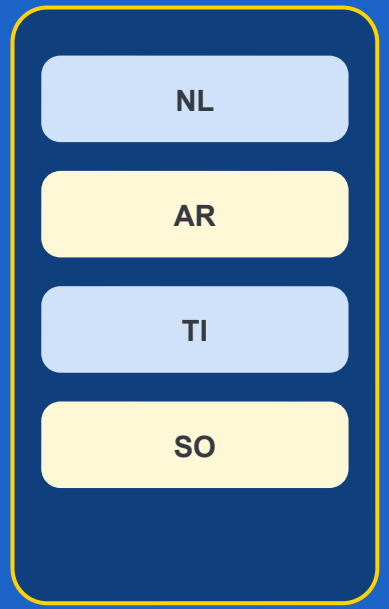
- FM augmentation beneficial even with very small TMs.
- Largest wins on **low-resource** languages.
- Adaptive NMT (MMT+) **still wins overall**.
- Commercial LLMs **now competitive**. Gemini Pro ≈ MMT+ on average.
- Low-resource gains concentrated on **Tigrinya**.

LIMITATIONS

- No human evaluation in this paper.
- Single domain (asylum reception).
- API systems are closed and non-reproducible.

Thank you.

Questions?



Thomas Moerman · thomas.moerman@ugent.be

Ghent University LT³ · EAMT 2026

GHENT UNIVERSITY

language and translation technology team

REFERENCES

MATIAS PROJECT

· Macken L., van Hest E., Tezcan A., Lumingu M., Maryns K., De Wilde J. (2024). MaTIAS: Machine Translation to Inform Asylum Seekers. EAMT.

· Macken L. et al. (2025a). MT to Inform Asylum Seekers: Intermediate Findings from the MaTIAS Project. MT Summit XX.

· Macken L., Fonteyne M., Tezcan A., van Hest E., Maryns K., De Wilde J. (2025b). MT in asylum reception centres: system selection and multilingual quality evaluation. Revista Tradumàtica 23.

FRAMING AND RESOURCEDNESS

· Briva-Iglesias V., O'Brien S. (2026). Human-Centred AI Language Technology (HCAILT). Empathetic design framework.

· Joshi P. et al. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. ACL.

METHODS AND EVALUATION

· Bulte B., Tezcan A. (2019). Neural Fuzzy Repair. ACL.

· Johnson J., Douze M., Jégou H. (2019). Billion-scale similarity search with GPUs (FAISS). IEEE Big Data.

· Koehn P. (2004). Statistical significance tests for MT evaluation. EMNLP.

· Moslem Y., Haque R., Kelleher J., Way A. (2023). Adaptive MT with large language models. EAMT.

· Popović M. (2015). chrF: character n-gram F-score for automatic MT evaluation. WMT.

· Rei R. et al. (2020). COMET: A neural framework for MT evaluation. EMNLP.

· Wang W. et al. (2021). MiniLMv2: multi-head self-attention relation distillation. ACL Findings.

SYSTEMS AND MODELS

· Bertoldi N. et al. (2018). The ModernMT project. EAMT.

· Gemini Team, Google (2025). Gemini 2.5 technical report.

· NLLB Team (2022). No Language Left Behind. arXiv:2207.04672.

· Rei R. et al. (2025). Tower+: open multilingual LLMs for translation.

· TranslateGemma Team, Google (2026). TranslateGemma technical report.

Entries shortened. Verify against latex/eamt26.bib before final distribution.

PER-LANGUAGE CHRF, FULL TABLE

BACKUP A

Lang	Name	MMT+	MMT	TG-27B+FM	Gemini Pro+FM
AR	Arabic	**66.76**	63.78	63.63	63.76
DE	German	73.67	69.26	70.97	**73.85**
ES	Spanish	72.64	69.77	69.98	**74.30**
FA	Farsi	54.01	51.36	**58.43**	55.03
PT	Portug.	**80.97**	75.48	75.65	76.54
RO	Romanian	**67.97**	63.10	63.80	67.22
RU	Russian	67.62	65.25	**67.95**	67.01
TR	Turkish	65.34	62.20	65.47	**65.60**
HY	Armen.	**74.26**	68.58	68.61	73.42
SQ	Albanian	70.01	65.46	61.15	**70.05**
KA	Georgian	79.87	**83.74**	61.70	65.69
PS	Pashto	**66.82**	64.86	57.10	62.53
SO	Somali	**58.19**	55.57	54.47	57.34
TI	Tigrinya	34.29	29.48	29.99	**43.00**

Yellow cells: per-row best system. Tier colours on Lang column match slides 10–12.

GHENT UNIVERSITY

FUZZY VS RANDOM Δ CHRF

BACKUP B

Lang	Name	Random-15	Fuzzy-15	Δ ChrF	Sig.
AR	Arabic	60.82	63.63	+2.81	†
DE	German	69.22	70.97	+1.75	†
ES	Spanish	70.35	69.98	-0.37	·
FA	Farsi	55.44	58.43	+2.99	‡
PT	Portug.	74.27	75.65	+1.38	‡
RO	Romanian	63.34	63.80	+0.46	·
RU	Russian	63.25	67.95	+4.70	‡
TR	Turkish	62.72	65.47	+2.75	‡
HY	Armen.	64.48	68.61	+4.13	‡
SQ	Albanian	61.08	61.15	+0.07	·
KA	Georgian	57.92	61.70	+3.78	‡
PS	Pashto	53.21	57.10	+3.89	‡
SO	Somali	50.39	54.47	+4.08	‡
TI	Tigrinya	21.16	29.99	+8.83	‡

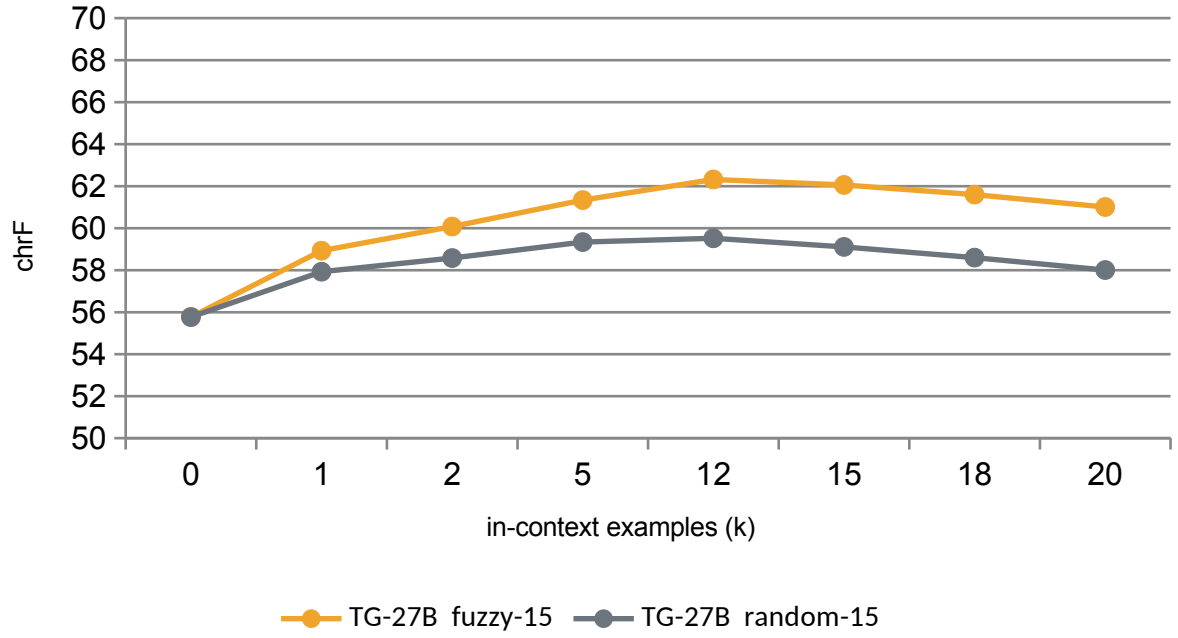
*‡ p<.001 · † p<.01 · * p<.05 · · not significant. Markers from MATEO bootstrap (n=1000).*

GHENT UNIVERSITY

SHOT-COUNT: FUZZY VS RANDOM OVERLAY

BACKUP C

yes, random examples help a little (any examples > none), but fuzzy match dominates at every k.



WHAT TO SAY

Random helps a bit (≈ 3 chrF).

Fuzzy adds another ≈ 3 chrF on top.

So roughly half the gain is from "prompt length" and half from "example quality".

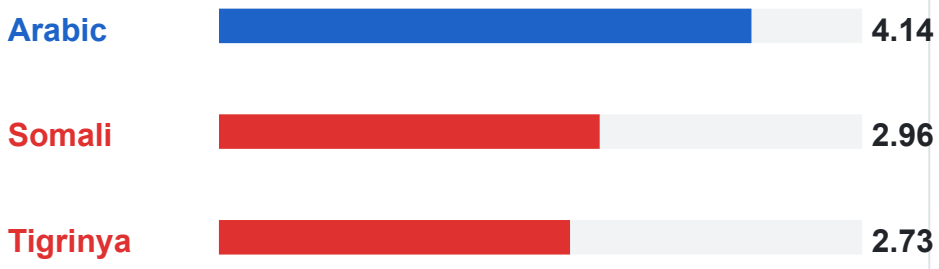
GHENT UNIVERSITY

PRIOR HUMAN EVALUATION (MACKEN 2025)

BACKUP D

Source: Macken et al. 2025, MaTIAS human evaluation (28 evaluators, XSTS 5-point scale).

MEAN XSTS RATING (1 = unrelated, 5 = perfect)



Tigrinya 2.73 / Somali 2.96 / Arabic 4.14

HOW DOES CHRF MAP TO PERCEIVED QUALITY?

ChrF is a **significant but modest** predictor of XSTS:

$p < .001$ · semi-partial $R^2 \approx 0.04$

Honest read: chrF tracks quality directionally but doesn't replace human evaluation, particularly on low-resource languages.

The +8.7 chrF gain on Tigrinya (slide 15) maps onto the language with the lowest prior XSTS rating.

GHENT UNIVERSITY

COMPUTE FOOTPRINT

BACKUP E

OPEN-SOURCE, LOCAL

Inference engine: vLLM
Hardware: 2 × NVIDIA A100 80 GB
Models: TG-4B / 12B / 27B, TP-9B
Privacy: data never leaves the cluster

Cost: *GPU-hours we can measure, energy attributable to our facility.*

CLOSED-SOURCE, API

Provider: Google (Gemini Pro 2.5)
Hardware: opaque
Energy: opaque
Privacy: messages traverse Google's servers

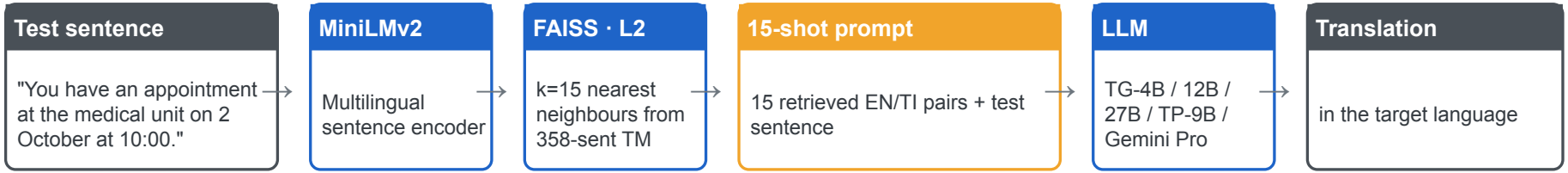
Cost: *per-token API spend; carbon footprint reported only at provider granularity.*

Comparing carbon footprints across local vs API systems on equal terms is itself an open question.

GHEN
UNIVERSITY

TECHNICAL PIPELINE: MINILMV2 + FAISS

BACKUP F



Conceptual summary: MiniLMv2 + FAISS = similarity search over the TM.

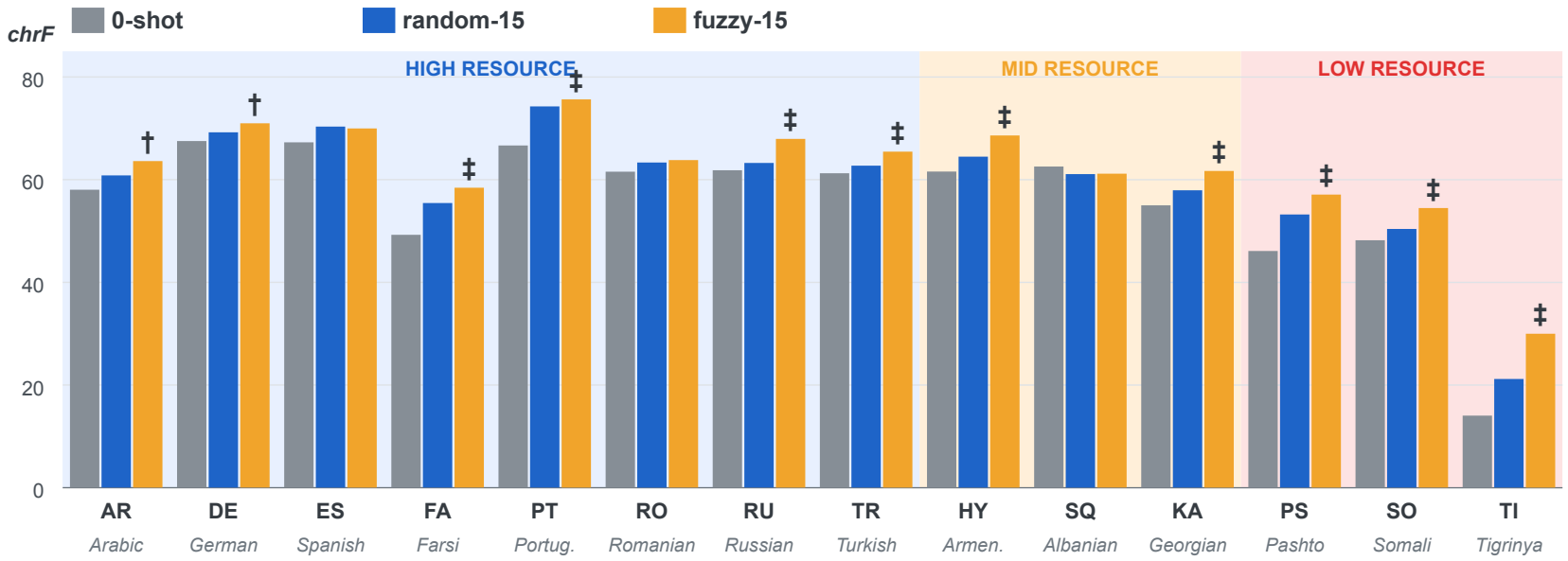
[Wang et al. 2021]
[Johnson et al. 2019]

GHENT UNIVERSITY

FM VS RANDOM VS ZERO-SHOT: FULL

BACKUP G

TG-27B, 14 languages, en → XX. ChrF for 0-shot, random-15, fuzzy-15. Significance markers vs the second-best.



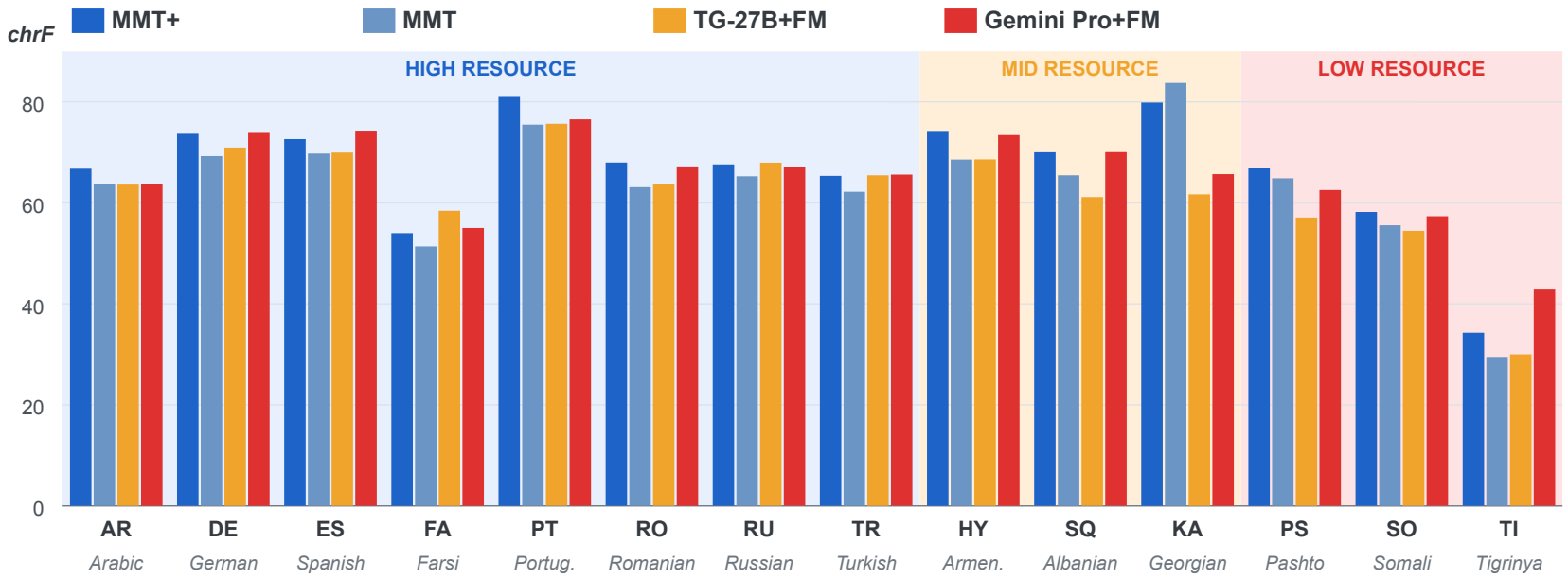
‡ p<.001 · † p<.01 · * p<.05 · 11 of 14 languages: fuzzy > random at p<.01.

GHENT UNIVERSITY

ADAPTIVE NMT VS LLM: FULL

BACKUP H

ChrF on en → XX, fuzzy-15 for LLMs. Tier-shaded backgrounds.



GHENT UNIVERSITY

PER-LANGUAGE WINNERS (FULL GRID)

BACKUP I

	MMT+	MMT	TG-27B	Gemini Pro
AR Arabic	**66.8**	63.8	63.6	63.8
DE German	73.7	69.3	71.0	**73.8**
ES Spanish	72.6	69.8	70.0	**74.3**
FA Farsi	54.0	51.4	**58.4**	55.0
PT Portug.	**81.0**	75.5	75.7	76.5
RO Romanian	**68.0**	63.1	63.8	67.2
RU Russian	67.6	65.3	**68.0**	67.0
TR Turkish	65.3	62.2	65.5	**65.6**
HY Armen.	**74.3**	68.6	68.6	73.4
SQ Albanian	70.0	65.5	61.1	**70.0**
KA Georgian	79.9	**83.7**	61.7	65.7
PS Pashto	**66.8**	64.9	57.1	62.5
SO Somali	**58.2**	55.6	54.5	57.3
TI Tigrinya	34.3	29.5	30.0	**43.0**

READING THE GRID

MMT+ dominates mid-resource (Georgian 84 chrF).

Gemini Pro wins 6/14, mostly low-resource.

TG-27B wins 3 cases (FA, RU, TR).

TowerPlus fails on coverage.

METRIC AGREEMENT

BACKUP J

Do chrF, BLEU, COMET tell the same story? Mostly yes. Not on low-resource.

AGREEMENT (overall)

chrF–BLEU ρ ≈ 0.90. chrF–COMET ρ ≈ 0.77. **Low-resource subset (ps, so, ti, ka): chrF–COMET ρ ≈ 0.**

WEAKNESSES

- chrF: character overlap, not meaning. Can reward fluent but wrong translations.
- BLEU: word overlap, tokenisation-sensitive.
- COMET: neural, but only as reliable as its training data. Does not cover Tigrinya.

HUMAN EVALUATION (Macken et al. 2025b)

28 evaluators, 90 messages per pair, XSTS 5-point scale. Messages rated above 3 of 5 for meaning preservation:

29% Tigrinya

40% Somali

57% Persian

ChrF and COMET are both significant but modest predictors of XSTS (p<.001, semi-partial R^2 ≈ 0.04-0.06).

[Macken et al. 2025b]
[Koehn 2004]

GHENT UNIVERSITY

REAL OUTPUTS

BACKUP K

Test sentence (EN): "You have an appointment at the medical unit on 2 October at 10:00."

Gold (human reference)	ዕለት 2 ጥቅምቲ ሰዓት 10፡00 ኣብ ኣሃዱ ሕክምና ቆጸራ ኣለኩም።
Zero-shot · TG-27B	ኣብ ሕክምናዊ ክፍሊት፣ ኣብ 2 ጥቅምት ሰዓት 10:00 ናይ መገብያ ቕጥዒት ኣለዎም።
Random-15 · TG-27B	ኣብ ክፍሊ ሕክምና 2 ጥቅምቲ ኣብ ሰዓት 10:00 ሕቶ ኣለዎም ← "question", not "appointment"
Fuzzy-15 · TG-27B	ዕለት 2 ጥቅምቲ ሰዓት 10:00 ኣብ ኣሃዱ ሕክምና ቆጸራ ኣለኩም። ← matches gold modulo one colon glyph
Fuzzy-15 · Gemini Pro 2.5	ብዕለት 2 ጥቅምቲ ሰዓት 10፡00 ኣብቲ ናይ ሕክምና ክፍል ቀጸራ ኣለኩም።

Arabic counterpoint (fuzzy-15 · TG-27B): 10:00 أكتوبر في تمام الساعة 2 يوم الطبية الوحدة في موعد لديك. ← essentially perfect.

Same script difficulty (non-Latin, RTL). Wildly different outcome. Training-data coverage is the bottleneck, not script.

GHENT UNIVERSITY

FULL 15 RANDOM TM ENTRIES

BACKUP L

1. We also don't know yet where you will be moving to.
2. Attention!
3. Get off at the Brussel Zuid / Bruxelles-Midi stop.
4. Bed linen is changed every Monday between 14:00 and 15:00 and between 19:00 and 19:30.
5. Mail for your fines.
6. Walk to the Jabbeke Dorp bus stop.
7. Bring your key in to the resident coordinator (Mark), office 539.
8. If you are in need of clothes, you can get second-hand clothes at the centre.
9. Do you like to write?
10. Don't forget to apply sunscreen.
11. From 10:00 to 11:00 on the first floor.
12. You have registered to volunteer in Herent.
13. We rehearse every Wednesday from 13.30 to 18.30 in the Yellow Room.
14. Location:
15. Parents or other reception centre residents who would like to help supervise activities can apply to Talent Twister.

GHENT UNIVERSITY

Off-topic: bus stops, bed linen, sunscreen, volunteering. None of them mention an appointment or the medical unit.

XSTS 5-POINT SCALE (MACKEN ET AL. 2025B)

BACKUP M

Cross-lingual Semantic Text Similarity. Scores shifted by a 20-item EN-EN calibration set per evaluator.

1	The messages are not equivalent, share very few details, and may be about different topics.
2	The messages share some details but are not equivalent. Some important information differs or is missing.
3	The messages are mostly equivalent but some unimportant details differ.
4	The messages are paraphrases of each other. They are equivalent in meaning but not in expression.
5	The messages are exactly and completely equivalent in meaning and expression.

'Above 3' = at least paraphrase-level equivalence. Tigrinya 29%, Somali 40%, Persian 57%. *[Macken et al. 2025b]*

GHENT UNIVERSITY