

MultiGEC-2025

A Multilingual Grammatical Error Correction shared task

Arianna Masciolini and Ricardo Muñoz Sánchez
Språkanken Text, University of Gothenburg

What is GEC?



Sequence-to-sequence task where:

- ▶ **input**: an (ungrammatical) text, typically written by a learner
- ▶ **output**: a normalized version of the same text, aka *correction hypothesis*, which can be
 - ▶ *minimal* or
 - ▶ *fluency-edited*

Example

original	normalized (minimal)	normalized (fluency)
My moter became very sad, no food.	My <i>mother</i> became very sad, <i>and ate</i> no food.	My <i>mother was very sad and refused to eat.</i>
Min mama bliv väldigt ledsen, ingen mat.	Min <i>mamma</i> blev väldigt ledsen, <i>och åt</i> <i>ingen mat.</i>	Min <i>mamma</i> blev väldigt ledsen <i>och slutade äta.</i>
Mia mama era tanto triste, mangiava niente.	Mia <i>mamma</i> era tanto triste <i>e non mangiava</i> niente.	Mia <i>madre</i> era molto triste <i>e aveva smesso di mangiare.</i>
Mi mama era tan triste, no comia.	Mi <i>mamá</i> <i>estaba muy triste, no comía.</i>	Mi <i>mamá</i> <i>estaba muy triste y no comía nada.</i>

The shared task



SPRÅKBANKENTEXT
En forskningsinfrastruktur för språkliga data
och en språkteknologisk forskningsenhet

English

Språkbanken Text är en avdelning inom [Språkbanken](#).

Aktuellt **Forskning** Verktyg Data Frågelåda Om oss Kontakt

[Hem](#) / [Computational SLA](#) / MultiGEC-2025

Computational SLA

Mini-workshop on GEC 2024

MultiGEC-2025

MultiGED-2023

MultiGEC-2025



Shared task on Multilingual Grammatical Error Correction

For the shared task **results**, click [here \(minimal edits track\)](#) and [here \(fluency edits track\)](#).

The [Computational SLA](#) working group invites you to participate in the shared task on text-level Multilingual Grammatical Error Correction, **MultiGEC**, covering 12 languages: Czech, English, Estonian, German, Greek, Icelandic, Italian, Latvian, Russian, Slovene, Swedish and Ukrainian (see also the [call for participation on the ACL portal](#)).

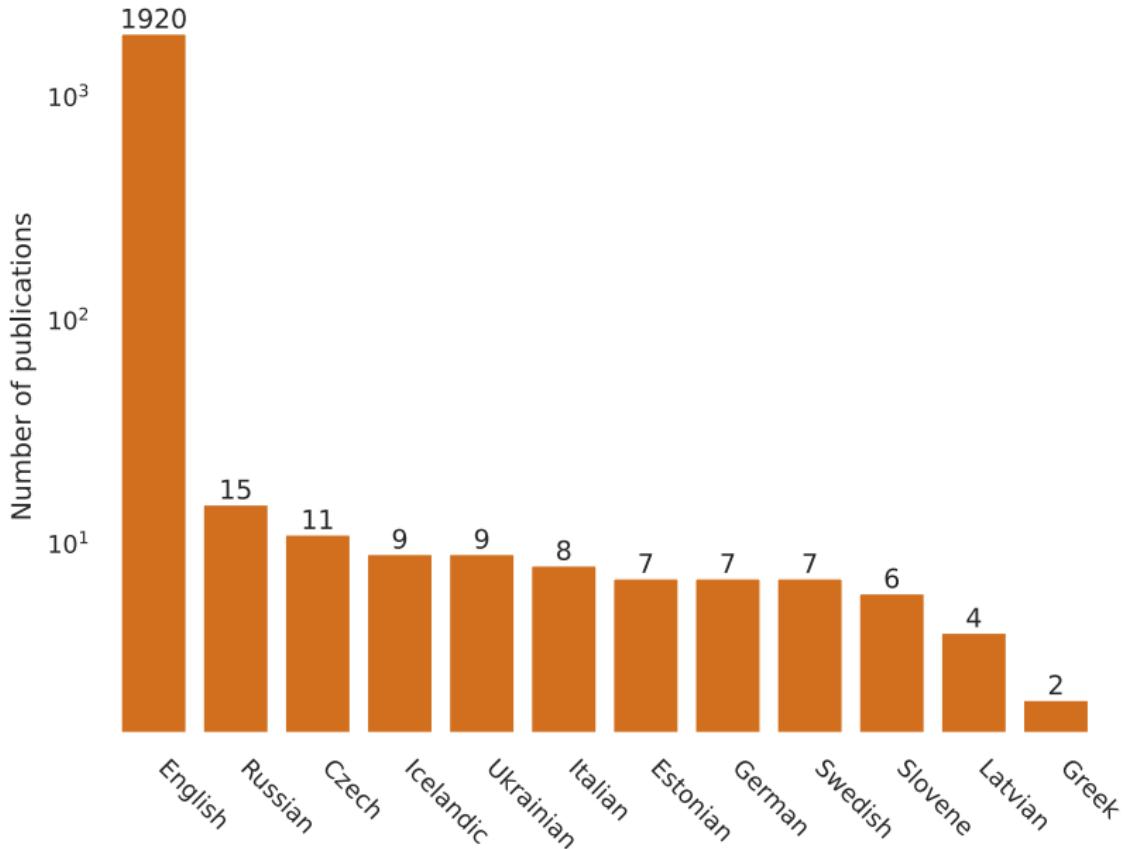
Automatic system evaluation will be carried out on [CodaLab](#), but the official leaderboard will be hosted on this website.

The results will be presented on March 5, 2025, at the [NLP4CALL workshop](#), colocated with the [NoDaLiDa conference](#) to be held in Estonia, Tallinn, on 2–5 March 2025. The publication venue for system descriptions will be the proceedings of the NLP4CALL workshop, co-published in ACL anthology.

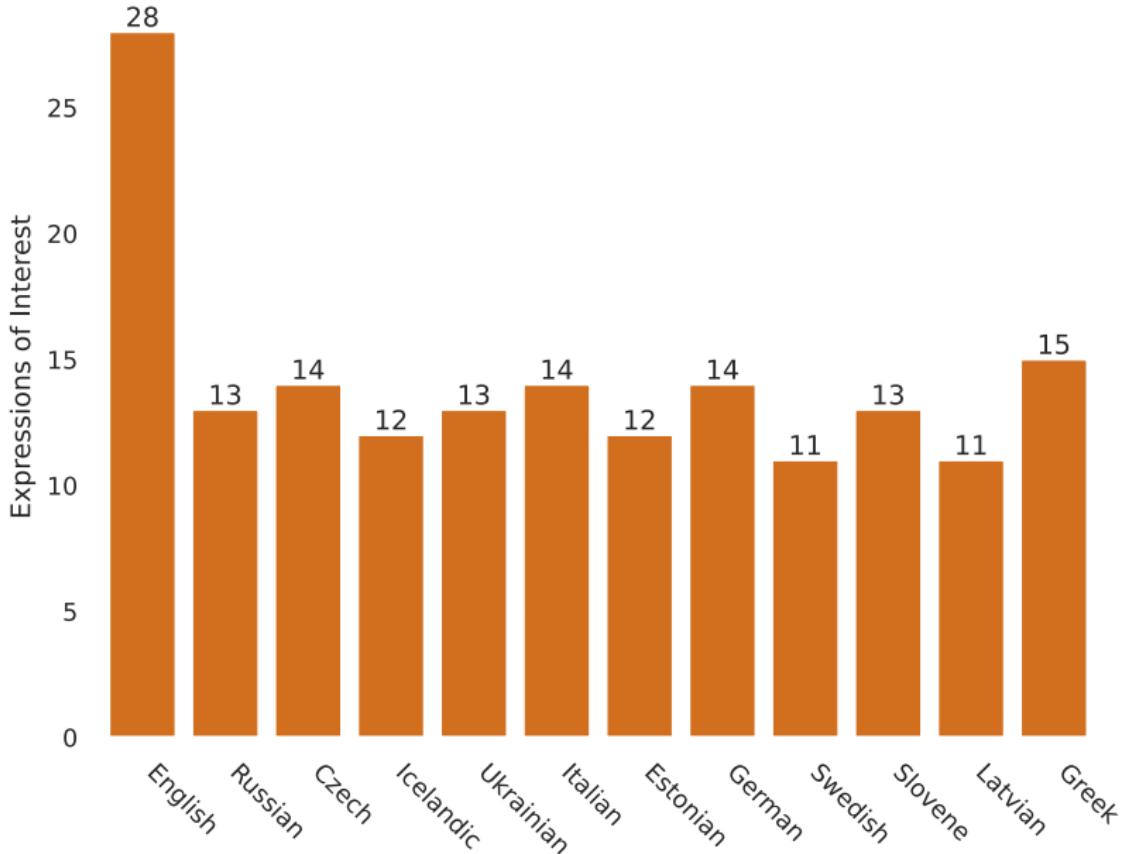
To register for/express interest in the shared task, please fill in [this form](#).

To get important information and updates about the shared task, please join the [MultiGEC-2025 Google Group](#).

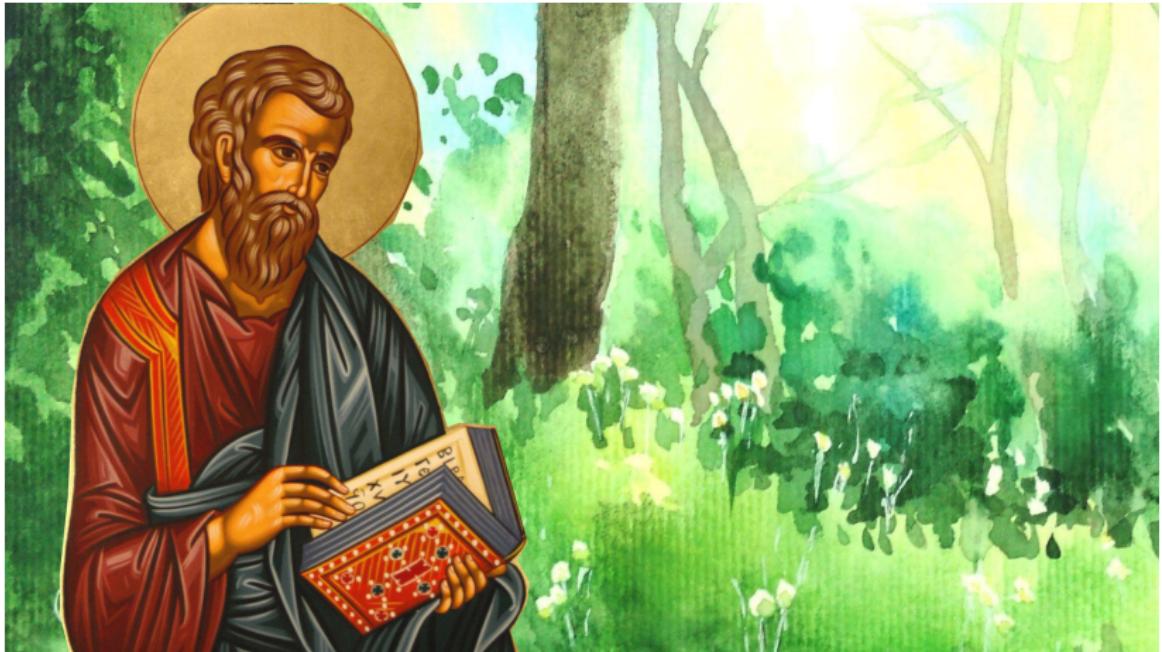
Why are we doing this?



Why are we doing this?



Why are we doing this?



Why are we doing this?



The shared task in numbers



- ▶ 12 languages and 18 subcorpora
- ▶ 8 organizers and 28 data providers
- ▶ 2 tracks (minimal + fluency)
- ▶ 3 evaluation metrics
- ▶ 1 multilingual baseline

Dataset

	train			dev			test			total		
	orig	ref1	ref2	orig	ref1	ref2	orig	ref1	ref2	orig	ref1	ref2
Czech – NatWebInf	3620	3620	-	1291	1291	687	1256	1256	1216	6167	6167	1903
Czech – Romani	3247	3247	-	179	179	84	173	173	163	3599	3599	247
Czech – SecLearn	2057	2057	183	173	173	97	177	177	170	2407	2407	450
Czech – NatForm	227	227	-	88	88	47	76	76	74	391	391	121
English – Write & Improve	4040	4040	-	506	506	-	504	504	-	5050	5050	-
Estonian – EIC	206	206	206	26	26	26	26	26	26	258	258	258
Estonian – EKIL2	1202	1202	1202	150	150	150	151	151	151	1503	1503	1503
German – Merlin	827	827	-	103	103	-	103	103	-	1033	1033	-
Greek – GLCII	1031	1031	-	129	129	-	129	129	-	1289	1289	-
Icelandic – IceEC	140	140	-	18	18	-	18	18	-	176	176	-
Icelandic – IceL2EC	155	155	-	19	19	-	19	19	-	193	193	-
Italian – Merlin	651	651	-	81	81	-	81	81	-	813	813	-
Latvian – LaVA	813	813	-	101	101	-	101	101	-	1015	1015	-
Russian – RULEC-GEC	2539	2539	-	1969	1969	-	1535	1535	1535	6043	6043	1535
Slovene – Solar-Eval	10	10	-	50	50	-	49	49	-	109	109	-
Swedish – SweLL-gold	402	402	-	50	50	-	50	50	-	502	502	-
Ukrainian – UA-GEC	1706	1706	-	87	87	87	79	79	79	1872	1872	166
total	18833	18833	2523	4514	4514	1178	4023	4023	3414	27370	27370	7115

TODO: this is just a temporary placeholder, the table should (only?) contain info about learners and domains

Evaluation

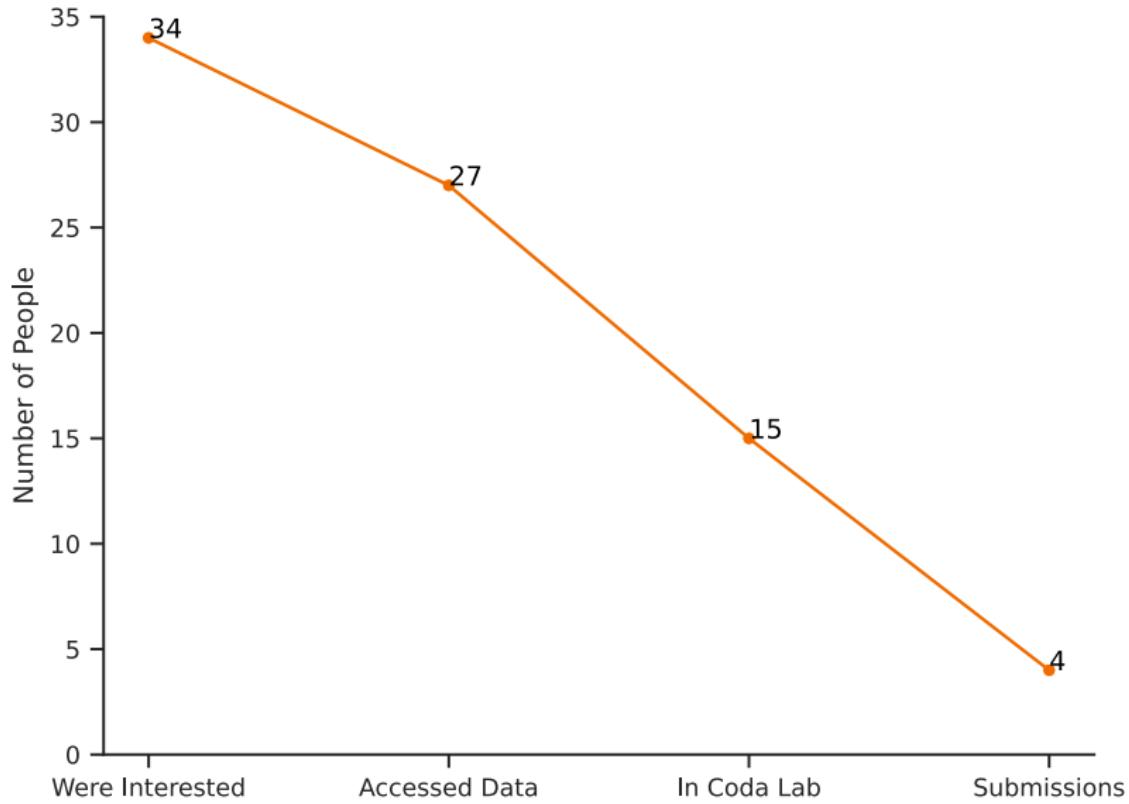


- ▶ 2 **reference-based** metrics (better for minimal edits):
 - ▶ $F_{0.5}$
 - ▶ GLEU
- ▶ Scribendi score (**referenceless** and LM-based, better for fluency edits)

Baseline



Participation



Results (non-random sample)



Ukrainian (UA-GEC), “minimal edits” track

Rank	Team name	GLEU	P	R	F0.5	Scribendi
1	UAM-CSI	79.55	74.31	54.11	69.15	0.89
2	Lattice	74.0	58.55	34.28	51.29	0.1
3	baseline	68.03	26.1	14.82	22.66	0.41
4	Grammaticks	62.93	16.53	13.48	15.81	-0.1
5	Rum-Cull	65.38	3.15	1.18	2.36	0.62

Results (overview)



TODO: Up for discussion how to best visually summarize them (we will need something like this in the paper as well), I have some ideas we can talk about one of these days but not sure they're any good.

What's next?



- ▶ open phase of the shared task (ongoing)
- ▶ presentation of the results at the NLP4CALL workshop
- ▶ manual evaluation of system output