

Charles Translator for Ukraine

Martin Popel
popel@ufal.mff.cuni.cz



Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics



unless otherwise stated

Charles Translator for Ukraine: Start

- March 1, 2022: *Hi folks,
Shouldn't we build a Czech-Ukrainian translator and put it on Lindat?*

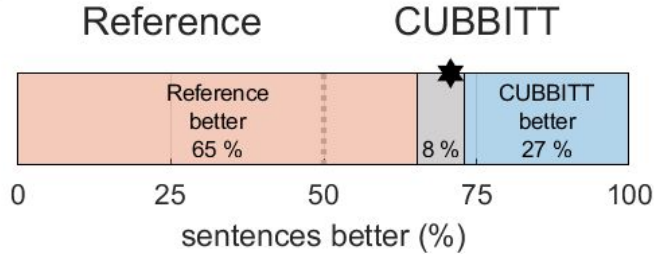
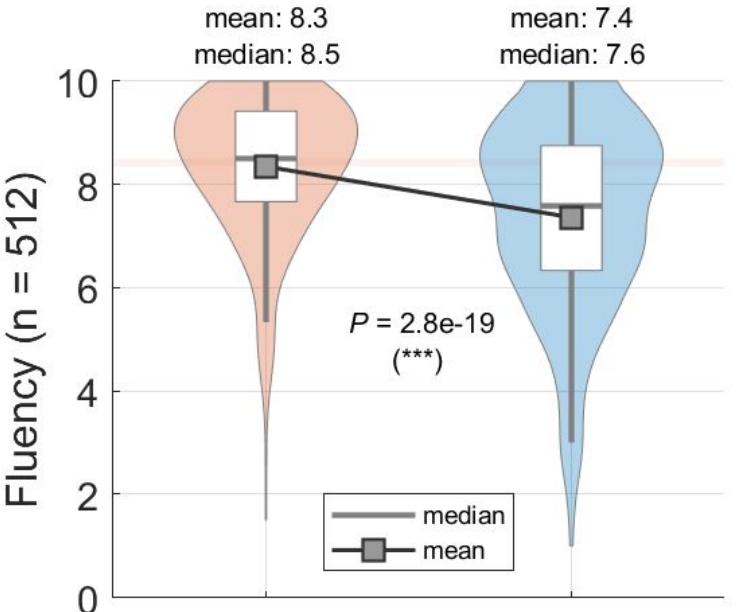
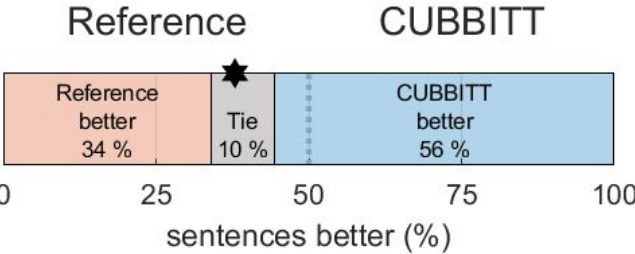
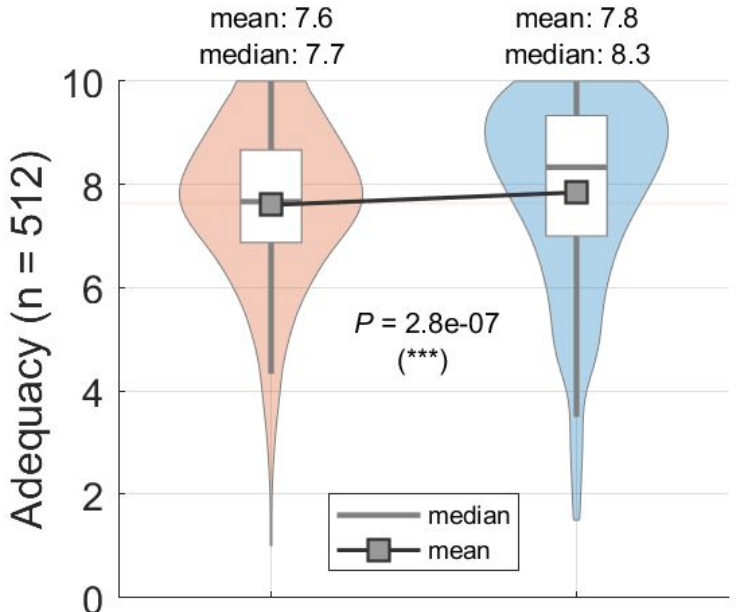
Many people use Google Translate that pivots through English and a direct translator would be better.

*Best,
Jindřich Libovický*



Experience with high-quality English-Czech NMT

Context-aware evaluation: 15 evaluators



Experience with high-quality English-Czech NMT

Context-aware evaluation: 15 evaluators



nature communications

nature > nature communications > articles > article

Article | [Open Access](#) | [Published: 01 September 2020](#)

Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals

Martin Popel [✉](#), Marketa Tomkova, Jakub Tomek, Łukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar & Zdeněk Žabokrtský

[Nature Communications](#) **11**, Article number: 4381 (2020) | [Cite this article](#)

6273 Accesses | **76** Altmetric | [Metrics](#)

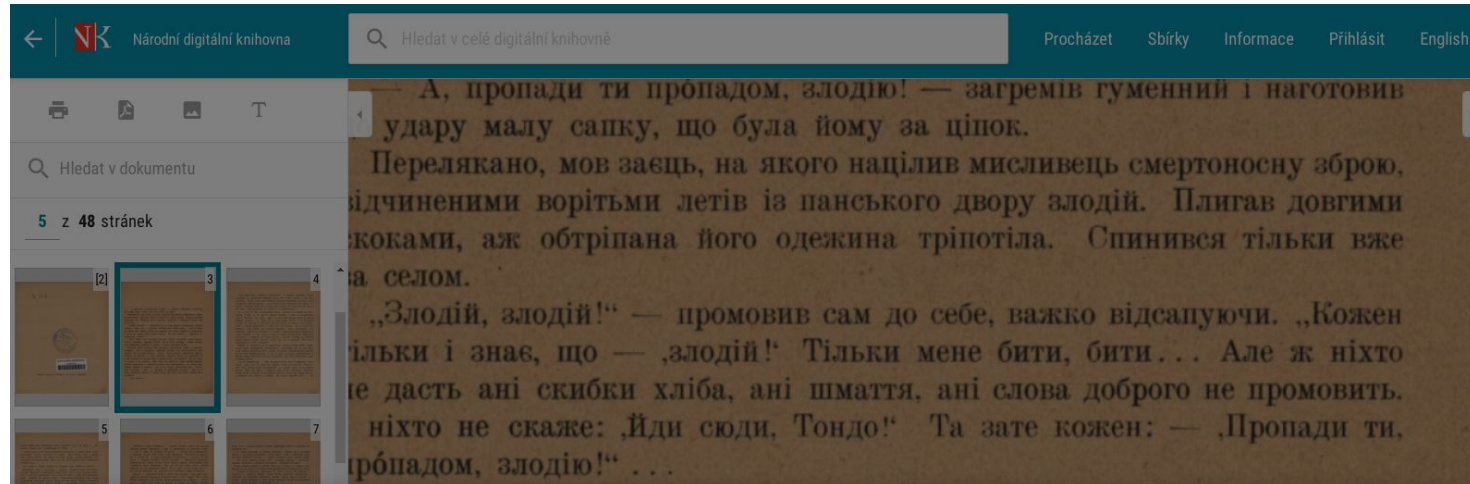
Referenc
better
34 %

Charles Translator for Ukraine: Timeline

- March 1: *Hi folks, Shouldn't we build a Czech-Ukrainian translator and put it on Lindat?*
- March 12: hackathon to develop a new frontend



Main work: collect and clean training data



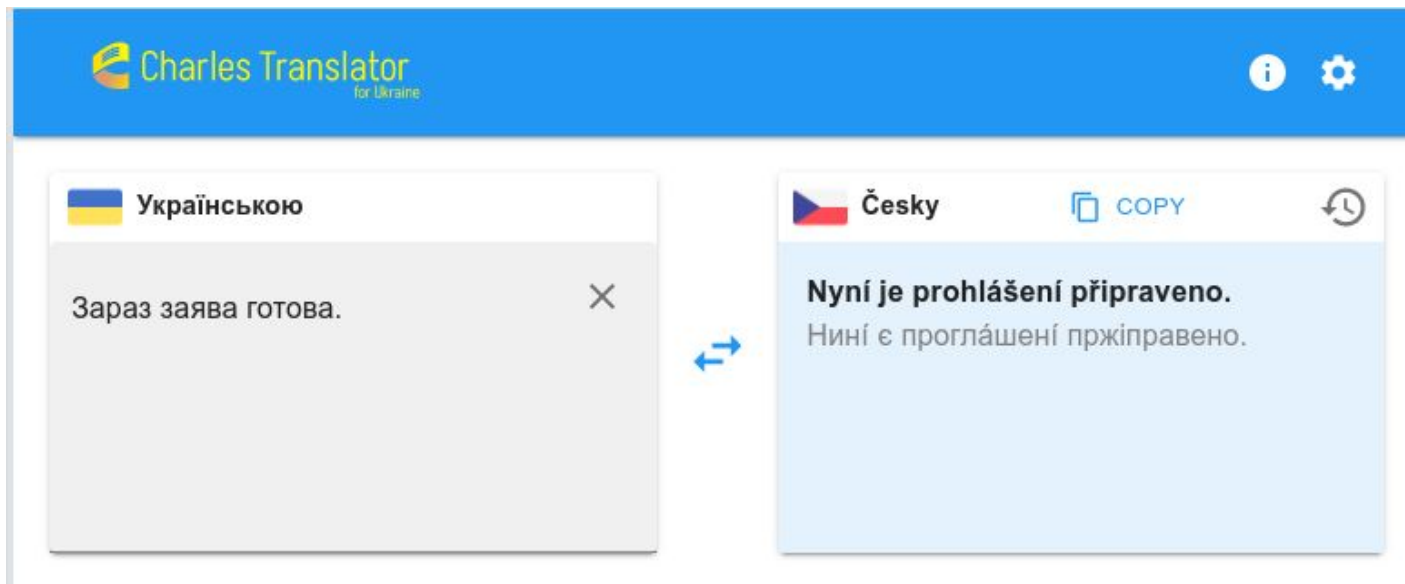
I. А, пропади ти пропадом, злодго! – затрем і н гуменнии і наготовив до удару налу сапку, що була йому за цшок. Перелякано, мов заєць, на якого надшив мисливець смертоносну зброю, вщчининими ворпьми лет і в іа панського двору злодш. Плигав довгими скоками, аж обтр!пана його одежина тріпотіла. Спинився тшки вже за селом. „Злодїй, злодш!“ – промовив сам до себе, важко відсапуючи. „Кожен тільки і знає, що – ,злодш!“ Тільки мене бити, бити... Але ж ніхто не дасть ані скибки хліба, ані шмаття, ані слова доброго не промовить. І ніхто не скаже: „Йди сюди, Тондо!“ Та зате кожен: – „Пропади ти, пропадом, злодш!“ ...

ndk.cz has thousands of Ukrainian books, but no mapping to Czech translations and Russian OCR software:

УКРАЇНА ⇒ УКРАЇНА or УКРАША or УКРА1НА or УКРА!НА

Charles Translator for Ukraine: Timeline

- March 1: *Hi folks, Shouldn't we build a Czech-Ukrainian translator and put it on Lindat?*
- March 12: hackathon to develop a new frontend
- March 14: first cs ↔ uk translator published at LINDAT/CLARIN

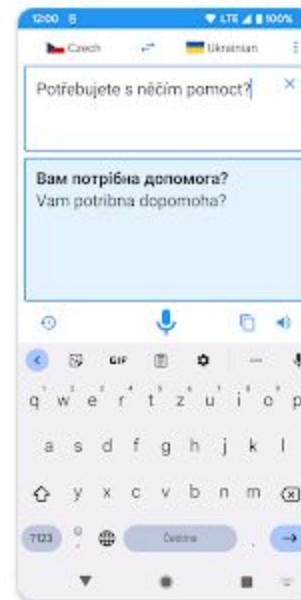
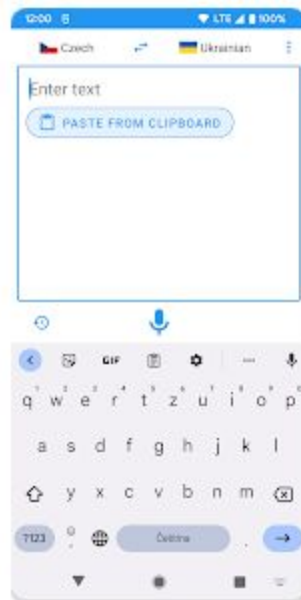


The screenshot displays the Charles Translator for Ukraine web interface. At the top, a blue header contains the logo and text "Charles Translator for Ukraine", along with information and settings icons. Below the header, two text input boxes are shown side-by-side, connected by a double-headed blue arrow indicating bidirectional translation. The left box is labeled "Українською" (Ukrainian) and contains the text "Зараз заява готова." (The statement is ready now.). The right box is labeled "Česky" (Czech) and contains the translated text "Nyní je prohlášení připraveno." (The statement is now ready.) and "Нині є проголошені пржіправено." (The statement is now ready.). The right box also features a "COPY" button and a refresh icon. At the bottom, a light blue footer contains the text: "THE LINDAT/CLARIAH-CZ PROJECT (LM2018101; formerly LM2010013, LM2015071) IS FULLY SUPPORTED BY THE MINISTRY OF EDUCATION, SPORTS AND YOUTH OF THE CZECH REPUBLIC UNDER THE PROGRAMME LM OF LARGE INFRASTRUCTURES".

THE LINDAT/CLARIAH-CZ PROJECT (LM2018101; formerly LM2010013, LM2015071) IS FULLY SUPPORTED BY THE MINISTRY OF EDUCATION, SPORTS AND YOUTH OF THE CZECH REPUBLIC UNDER THE PROGRAMME LM OF LARGE INFRASTRUCTURES

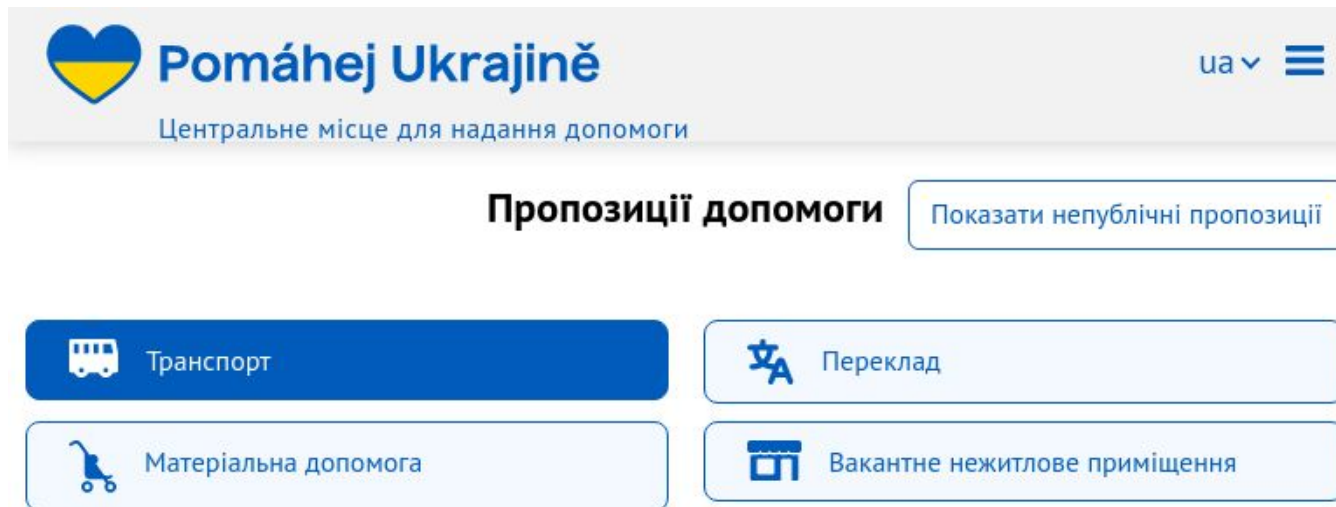
Charles Translator for Ukraine: Timeline

- March 1: *Hi folks, Shouldn't we build a Czech-Ukrainian translator and put it on Lindat?*
- March 12: hackathon to develop a new frontend
- March 14: first cs ↔ uk translator published
- April 11: improved cs ↔ uk models
- May 24: Android app



Why to develop our own MT?

- Quick solution until Big Tech companies improve their Ukrainian models
- Direct translation (not pivoted through English)
- Collect “in-domain” test set
- Focus on the needs of refugees and humanitarian organizations
- Offer free API (and on-premise for sensitive data)



The screenshot shows the website header for 'Pomáhej Ukrajině' with a heart logo and the text 'Центральне місце для надання допомоги'. Below the header is a section titled 'Пропозиції допомоги' with a button 'Показати неpubлічні пропозиції'. There are four category buttons: 'Транспорт' (highlighted in blue), 'Переклад', 'Матеріальна допомога', and 'Вакантне нежитлове приміщення'.

Pomáhej Ukrajině ua ▾ ☰
Центральне місце для надання допомоги


Пропозиції допомоги [Показати неpubлічні пропозиції](#)

- Транспорт**
- Переклад
- Матеріальна допомога
- Вакантне нежитлове приміщення

Why direct translation?



 Українською

Зараз заява готова. 



 Česky

 COPY



Nyní je prohlášení připraveno.
Нині є проголошені пржіправено.

Why direct translation?



заява ⇒ application ⇒ aplikace
declaration *computer app*

The screenshot shows the Google Translate web interface. At the top, the text 'заява ⇒ application ⇒ aplikace' is shown with the flags of Ukraine, the USA, and the Czech Republic. Below this, the Czech translation 'Nyní je aplikace připravena.' is displayed in a grey box. The interface includes the Google Translate logo, a 'Přihlásit se' button, and language selection options for 'UKRAJINŠTINA' and 'ČEŠTINA'. The input text 'Зараз заява готова.' is visible in the source language box, and the translated text 'Nyní je aplikace připravena.' is in the target language box. The interface also features a microphone icon, a speaker icon, and a character count '19 / 5 000'.

Why direct translation?

 Charles Translator
for Ukraine



 Česky

Jsem nemocná. A vy?
Jsem nemocný. A ty?
Jaké léky to jsou?



 Українською

 COPY



Я хвора. А ви?
Я хворий. А ти?
Які це ліки?
Ja chvora. A vy?
Ja chvoryj. A ty?
Jaki ce liky?

I am ill [fem]. And you [formal]?
I am ill [masc]. And you [informal]?
What kind of medicine is that?

Why direct translation?

☰ Google Překladač



Přihlásit se

ROZPOZNAT JAZYK

ČEŠTINA

ANGLICH



UKRAJINŠTINA

ČEŠTINA

ANGLICH

Jsem nemocná. A vy?
Jsem nemocný. A ty?
Jaké léky to jsou?



Я хворий. І ти?
Я хворію. І ти?
Які це наркотики?



YA khvoryy. I ty?
YA khvoriyu. I ty?
Yaki tse narkotyky?



58 / 5 000



Odeslat zpětnou vazbu

Why direct translation?



DeepL Překladač

DeepL Pro

Proč zrovna DeepL?

API

Přihlásit se



Čeština (detekováno) ▾



Ukrajinština ▾

Glosář

Jsem nemocná. A vy? ×

Jsem nemocný. A ty?

Jaké léky to jsou?

Я хворий. А ти?

Я хворий. А ти?

Що це за ліки?



Why direct translation?



Digital Europe
eTranslation

[About](#) | [Help](#) | [Liability](#)

Translate text

Translate documents

My translation requests

My settings

English

Jsem nemocná. A vy?
Jsem nemocný. A ty?
Jaké léky to jsou?

Я хворий. А як щодо тебе?
Я хворий. А ти?
Що це за ліки?

58 / 2500



From

Czech



To

Ukrain

Why direct translation?



Microsoft Bing

Prohledat web

Text

Překladatel

Konverzace

Aplikace

Pro firmy

Čeština

Jsem nemocná. A vy?
Jsem nemocný. A ty?
Jaké léky to jsou?



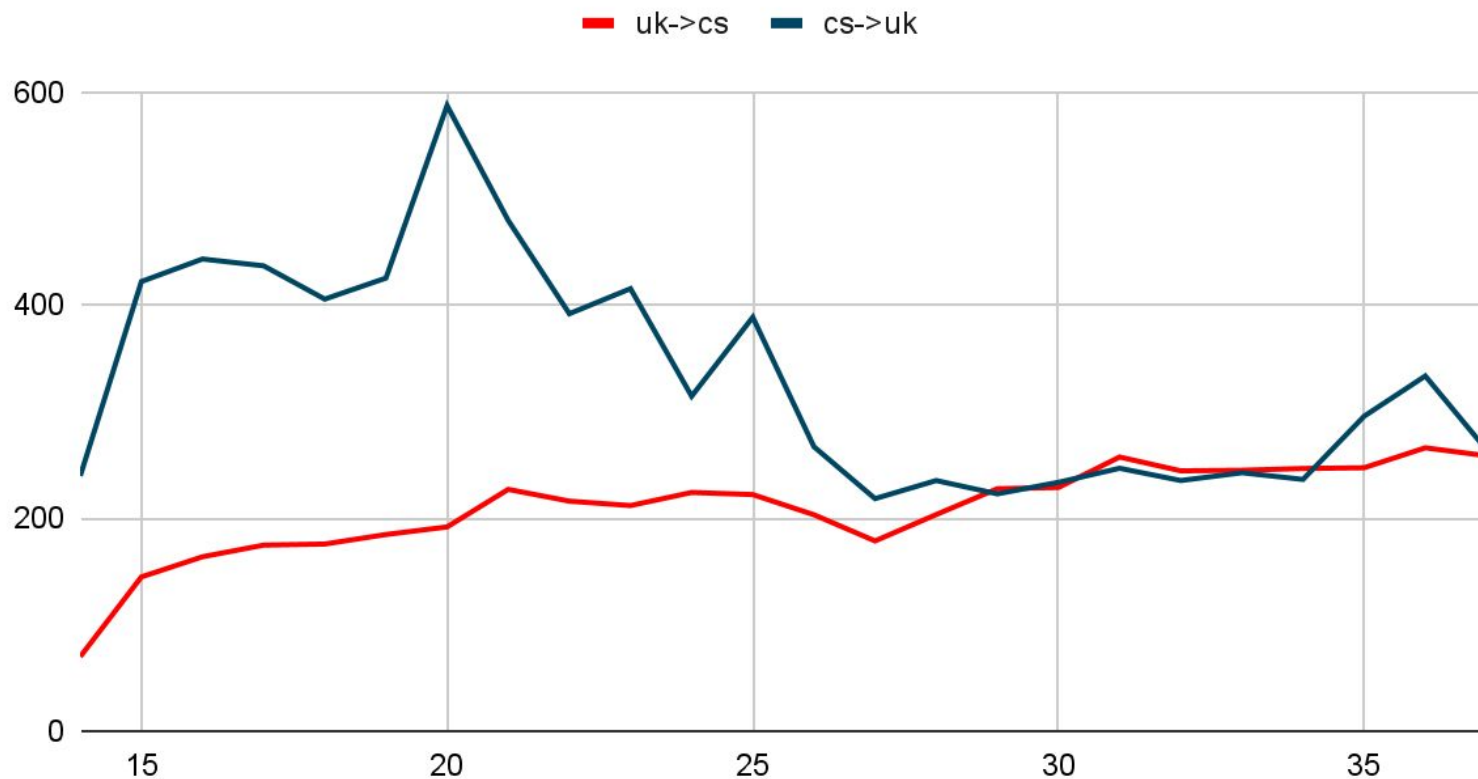
Ukrajnština

Я погано почуваюся. А ви?
Я погано почуваюся. А ви?
Що це за препарати?

yo pogano pochuvayusa. a vi?
yo pogano pochuvayusa. a vi?
shcho tse za preparati?

Thousands of translations per week (April-September)

Translation direction



Charles Translator for Ukraine: Timeline

- March 1: *Hi folks, Shouldn't we build a Czech-Ukrainian translator and put it on Lindat?*
- March 12: hackathon to develop a new frontend
- March 14: first cs ↔ uk translator published
- April 11: improved cs ↔ uk models
- May 24: Android app
- since September: evaluating 11 systems in WMT22 shared task

The screenshot shows the WMT22 shared task evaluation interface. At the top, there are tabs for 'Appraise' and 'Dashboard', and a user identifier 'ukrces1901'. The interface displays two translation examples, each with a rating scale from 0 to 6. The scale is labeled with '0: Nonsense/ No meaning preserved', '2: Some meaning preserved', '4: Most meaning preserved and few grammar mistakes', and '6: Perfect meaning and grammar'. Each example has a 'Reset' button on the left and a 'Submit' button on the right.

Example 1:

Ukrainian: У Бородянці під Києвом у ході розбирання завалів двох багатоповерхових житлових будинків виявлено тіла 7 цивільних.

Czech: v civilním oblečení, ale nepodařilo byla při demontáži sutin dvou vícepodlažních obytných budov nalezena těla 7 civilistů.

Rating: 0 (Nonsense/ No meaning preserved) to 6 (Perfect meaning and grammar). Submit button.

Example 2:

Ukrainian: Захисники Маріуполя заявили про розпилення над містом невідомої отруйної речовини з російського безпілотної.

Czech: Obránci Mariupolu oznámili postřik neznámé vaše země, sakra, ne dronu nad městem.

Rating: 0 (Nonsense/ No meaning preserved) to 6 (Perfect meaning and grammar). Submit button.

WMT 2022 General MT shared task evaluation

WMT22 Ukrainian→Czech

SQM	chrF	System
0.393	—	HUMAN-A
0.221	59.6	HuaweiTSC
0.181	60.7	AMU
0.137	60.4	Lan-Bridge
0.102	59.0	Charles
0.095	60.3	Online-B
...		
-0.036	57.5	Online-A
-0.157	56.3	Online-G
-0.276	55.3	Online-Y
—	53.7	eTranslation
-0.626	50.7	ALMAnaCH

WMT22 Czech→Ukrainian

SQM	chrF	System
0.295	—	HUMAN-A
0.225	64.0	Online-B
0.151	61.5	AMU
0.125	64.0	Lan-Bridge
0.065	62.6	HuaweiTSC
0.062	61.5	Charles
...		
-0.008	60.8	Online-G
-0.075	62.2	Online-A
-0.257	59.6	Online-Y
—	57.5	eTranslation
-0.669	54.5	ALMAnaCH

Слава Україні!

- Try Czech-English <https://cubbitt.lindat.cz>
- Try Czech-Ukrainian <https://translator.cuni.cz>

