



Charles Translator for Ukraine

Martin Popel
popel@ufal.mff.cuni.cz



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



March 2022: over 200 000 Ukrainian refugees in Czechia



Photo: René Volfík, irozhlas.cz

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



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*





ÚFAL 4 Ukraine Team:

Martin Popel, Jindřich Libovický,
Jindřich Helcl, Ondřej Košarko, David
Nápravník, Tomáš Krabač, Michal
Novák, Rudolf Rosa, Zdeněk Kasner,
Lucie Poláková, Jaroslava Hlaváčová,
Mariia Anisimova, Martin Majliš, Denys
Bojko, Tomáš Musil, Tereza Chlaňová,
Oxana Čmelíková and other volunteers

Lindat (CLARIN/CLARIAH-CZ) translation service

← → ↻ 🏠 🔒 lindat.mff.cuni.cz/services/translation/ ⚙️ ⭐ 🗂️ 📄 ⚙️ 🖨️ 👤 ⋮

LINDAT CLARIN/CLARIAH-CZ Catalog Repository Education Projects Tools Services About ▾  

LINDAT Translation

Translate Docs

The translation service is available for **personal and non-commercial use** (see [terms of use](#) for more details).

Source Czech ▾ **Target** English ▾

☐ advanced


Input sentences

Translation

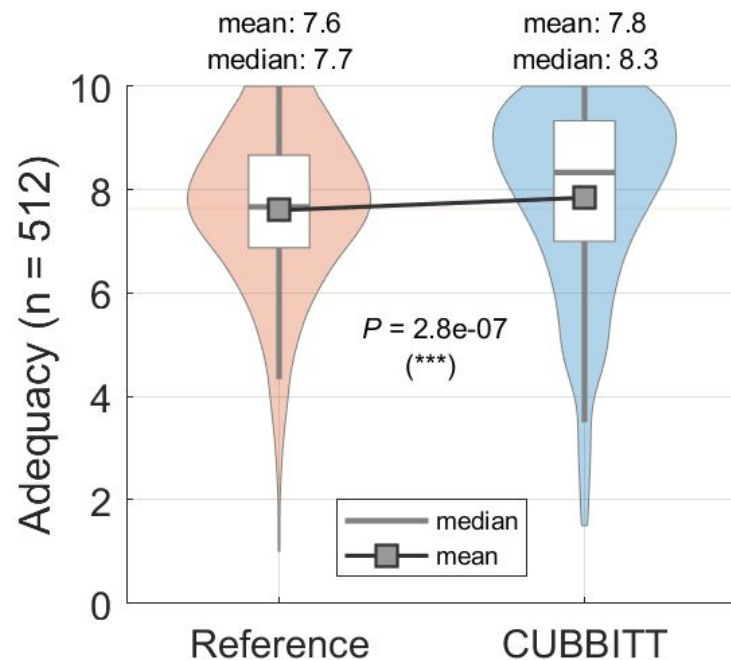
Translate Choose file

Credits: The service runs systems trained by:

Martin Popel
CUBBITT models, en↔cs, en↔fr, en↔pl as described in
Popel, M., Tomkova, M., Tomek, J. et al. *Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals*. Nat Commun 11, 4381 (2020). <https://doi.org/10.1038/s41467-020-18073-9>

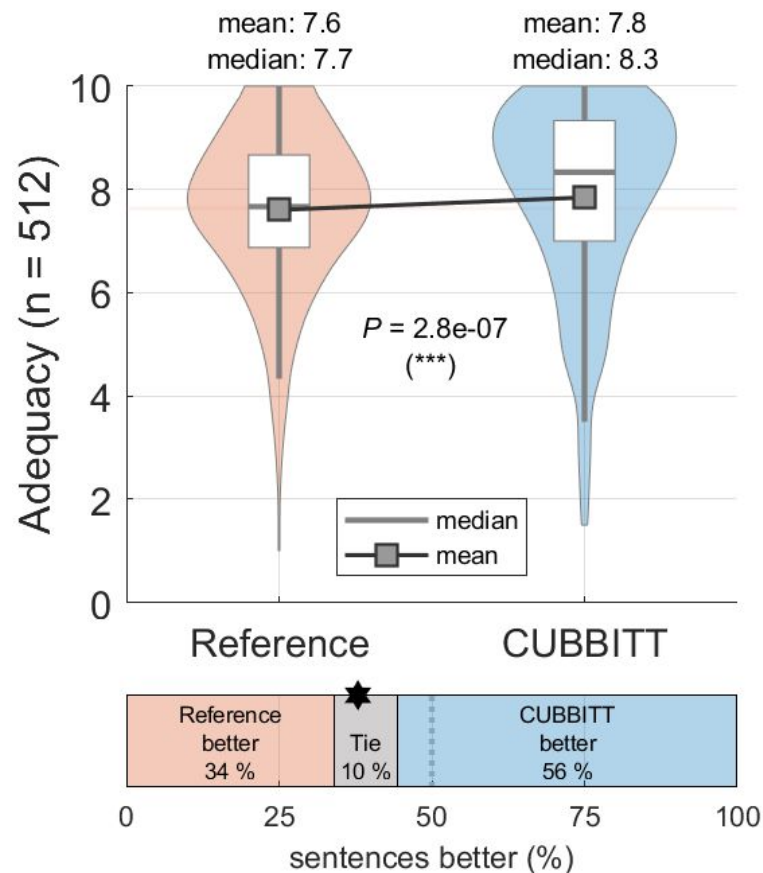


Experience with high-quality English-Czech NMT



In 2020, a human evaluation showed our NMT system called CUBBITT is **more adequate than a human** reference translation (by a translation agency).

Experience with high-quality English-Czech NMT

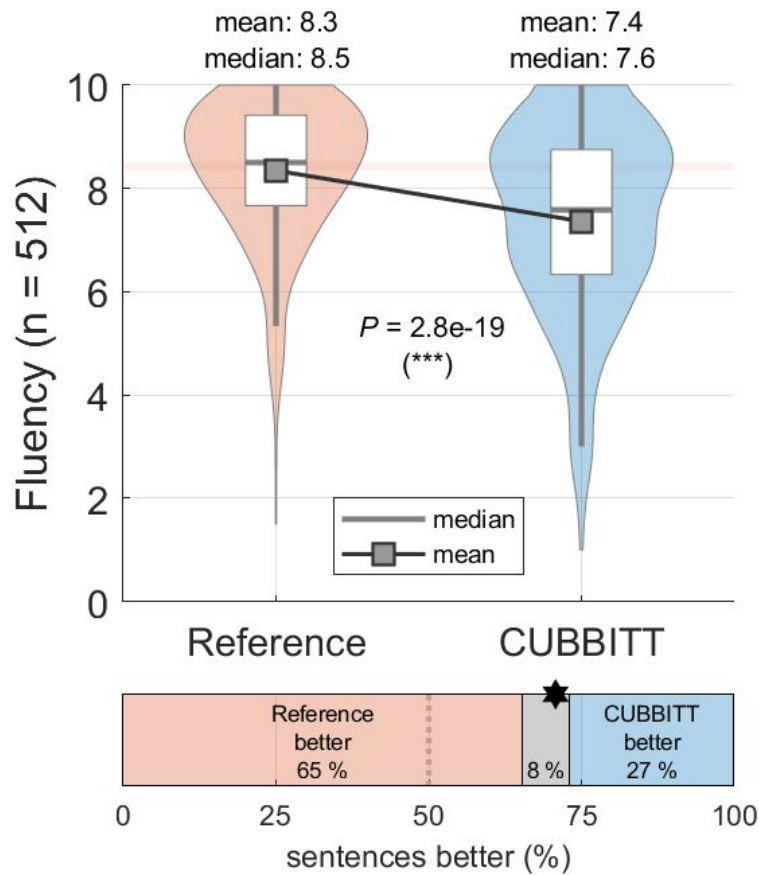
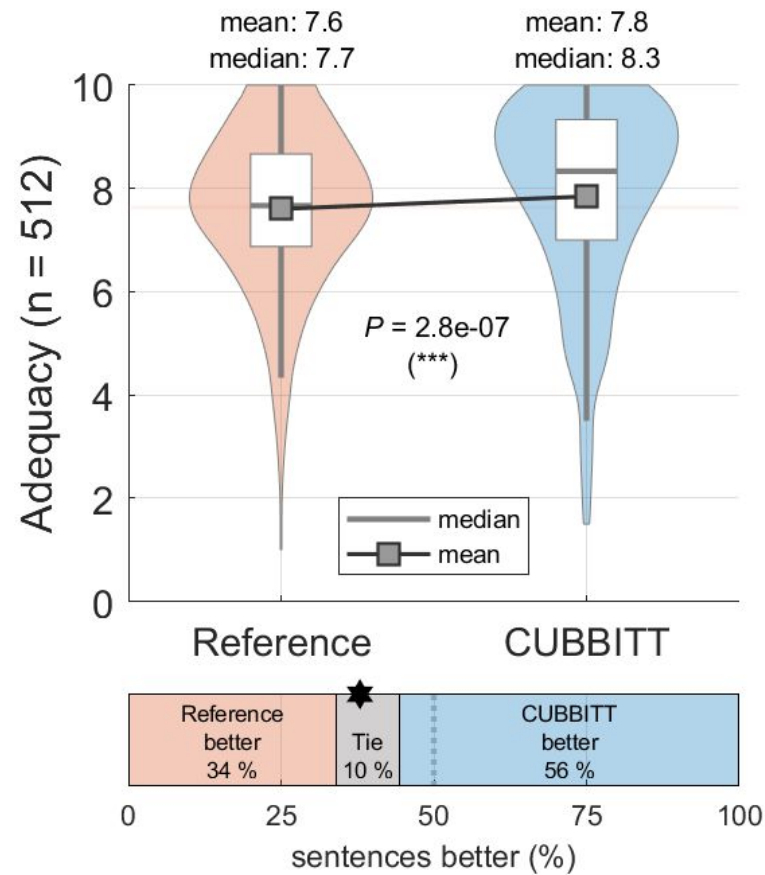


In 2020, a human evaluation showed our NMT system called CUBBITT is **more adequate than a human** reference translation (by a translation agency).

34% sentences: human more adequate
56% sentences: NMT more adequate

Experience with high-quality English-Czech NMT

Context-aware evaluation: 15 evaluators



Experience with high-quality English-Czech NMT



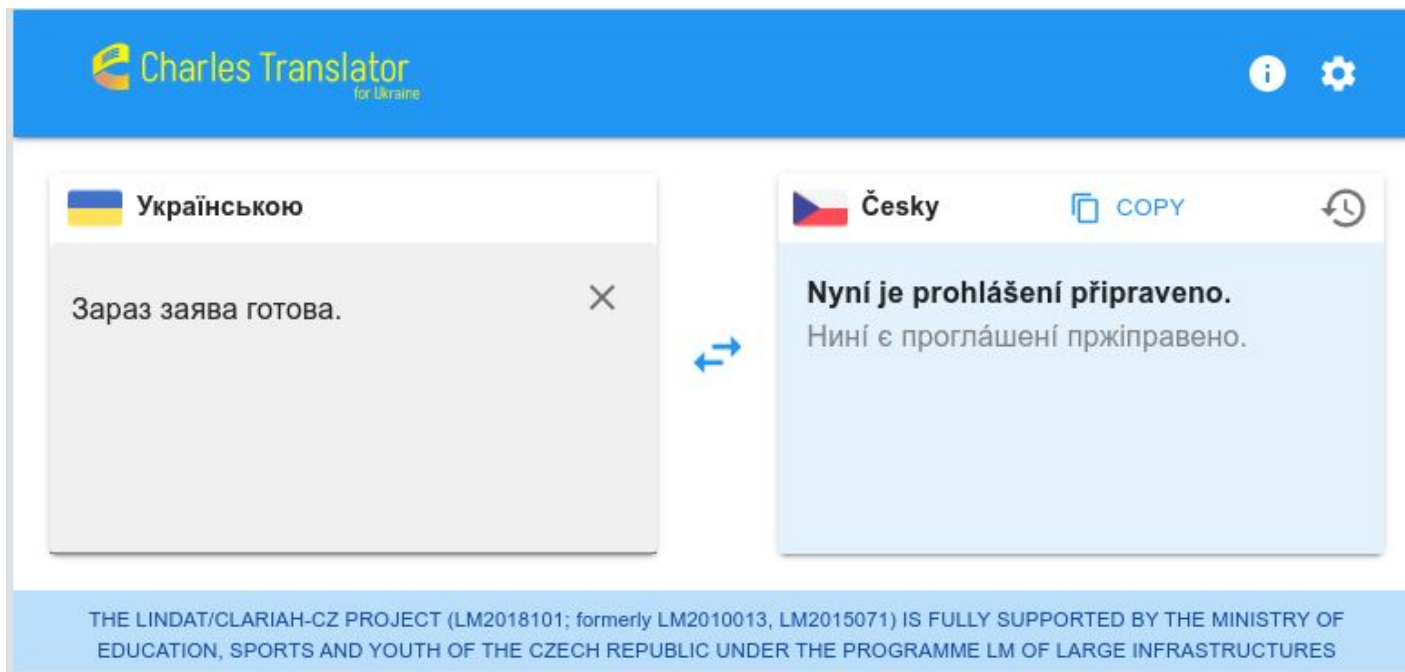
Charles Translator for Ukraine: Timeline 2022

- 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



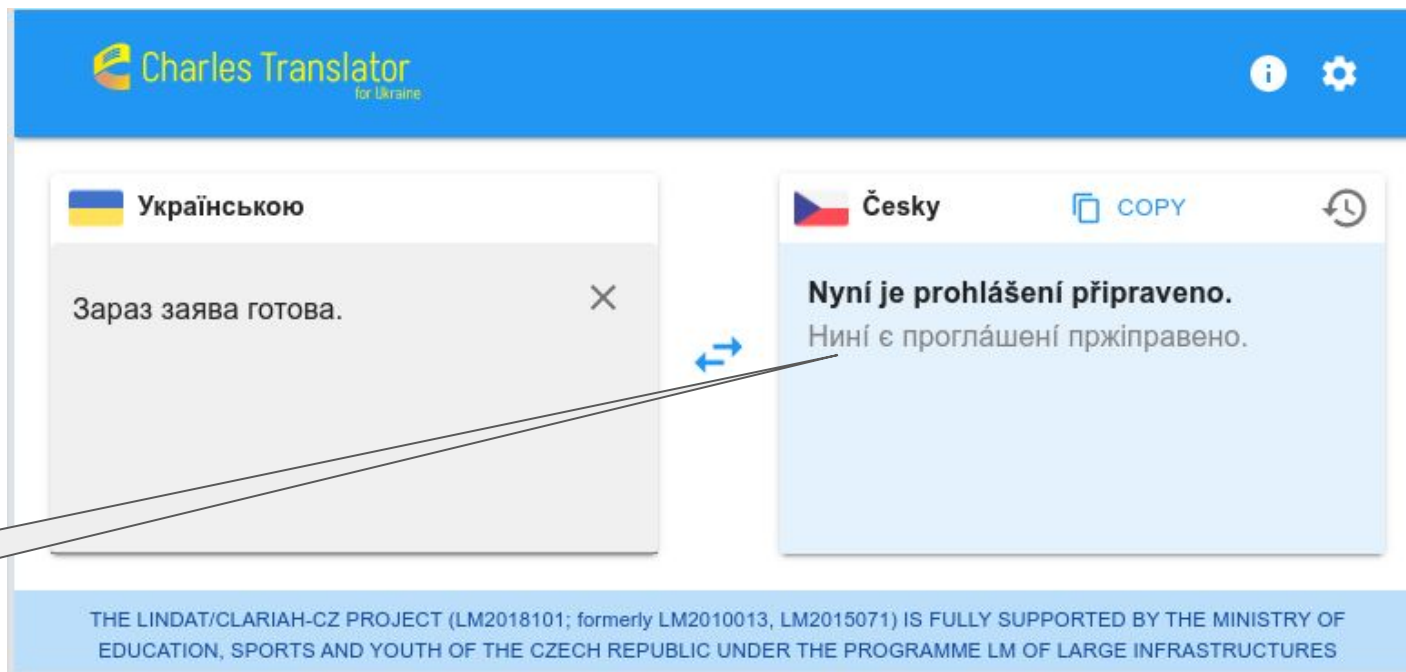
Charles Translator for Ukraine: Timeline 2022

- 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



Charles Translator for Ukraine: Timeline 2022

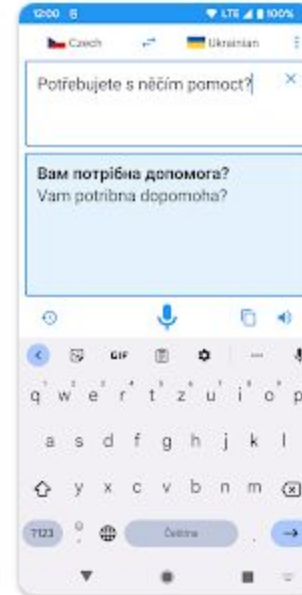
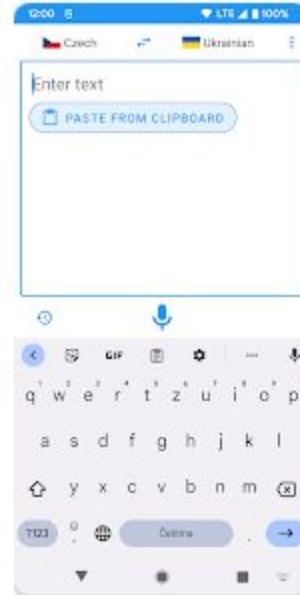
- March 1: *Hi folks, Shouldn't we build a Czech-Ukrainian translator and put it on Lindat?*
- March 12: hackathon to develop a new frontend
- March 14: first cs ↔ uk translator published at LINDAT/CLARIN



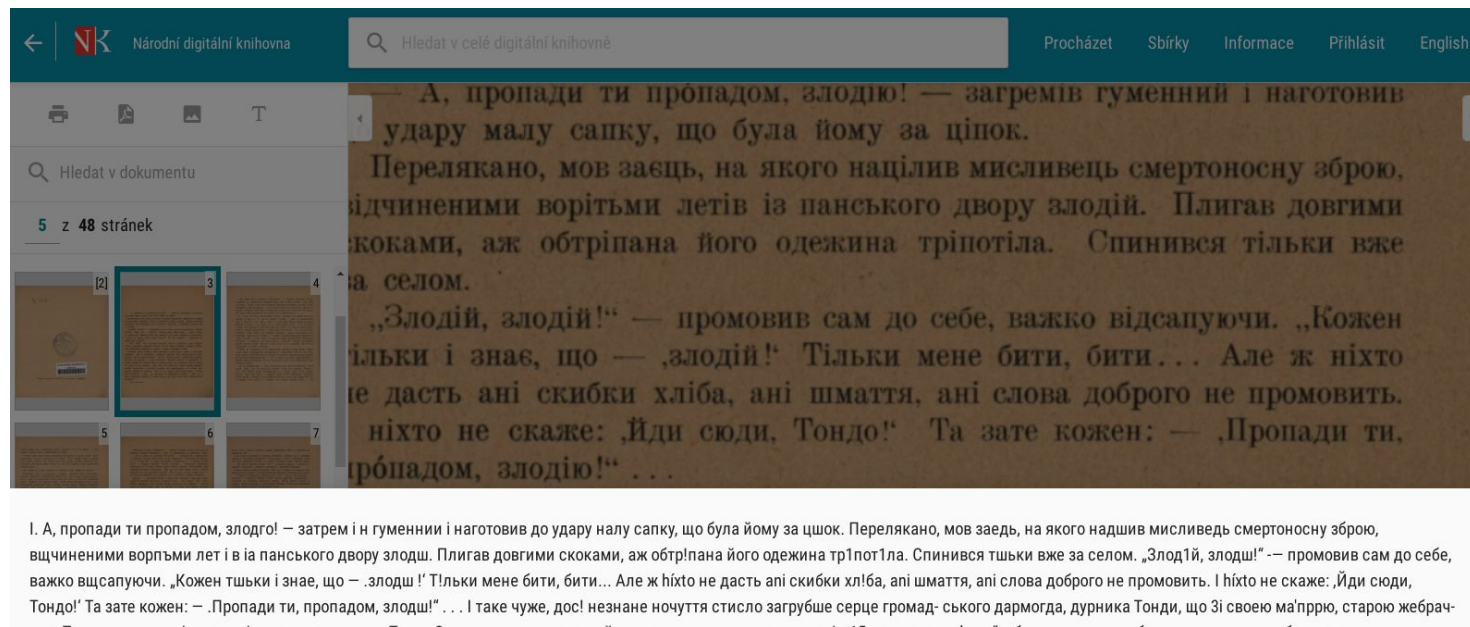
phonetic
transcription

Charles Translator for Ukraine: Timeline 2022

- 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



Main work: collect and clean training data

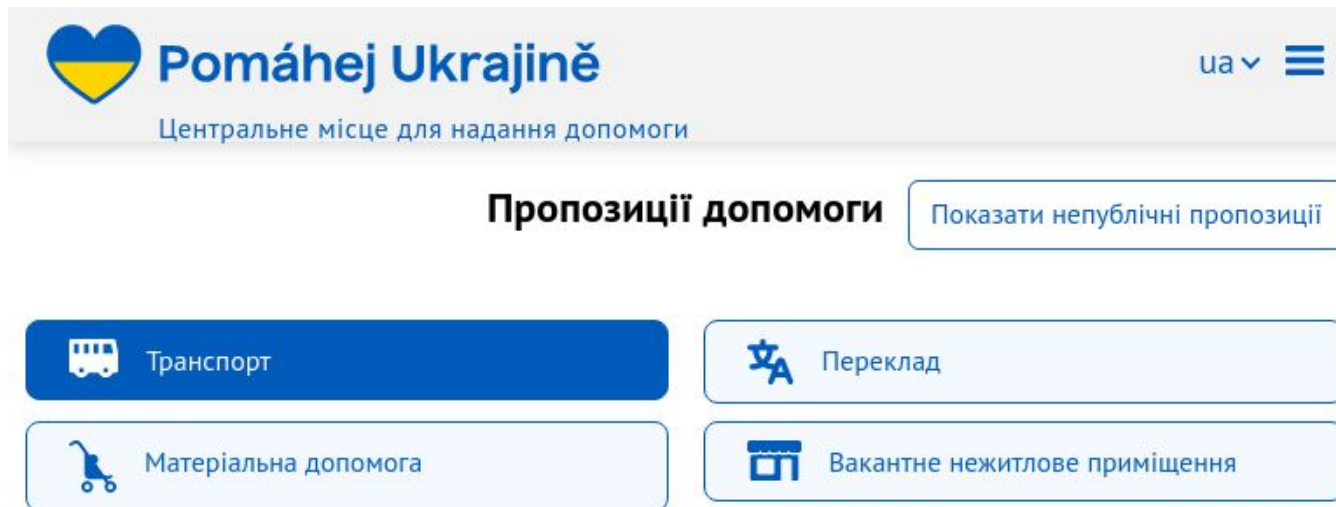


ndk.cz has thousands of Ukrainian books, but no mapping to Czech translations and Russian OCR software:

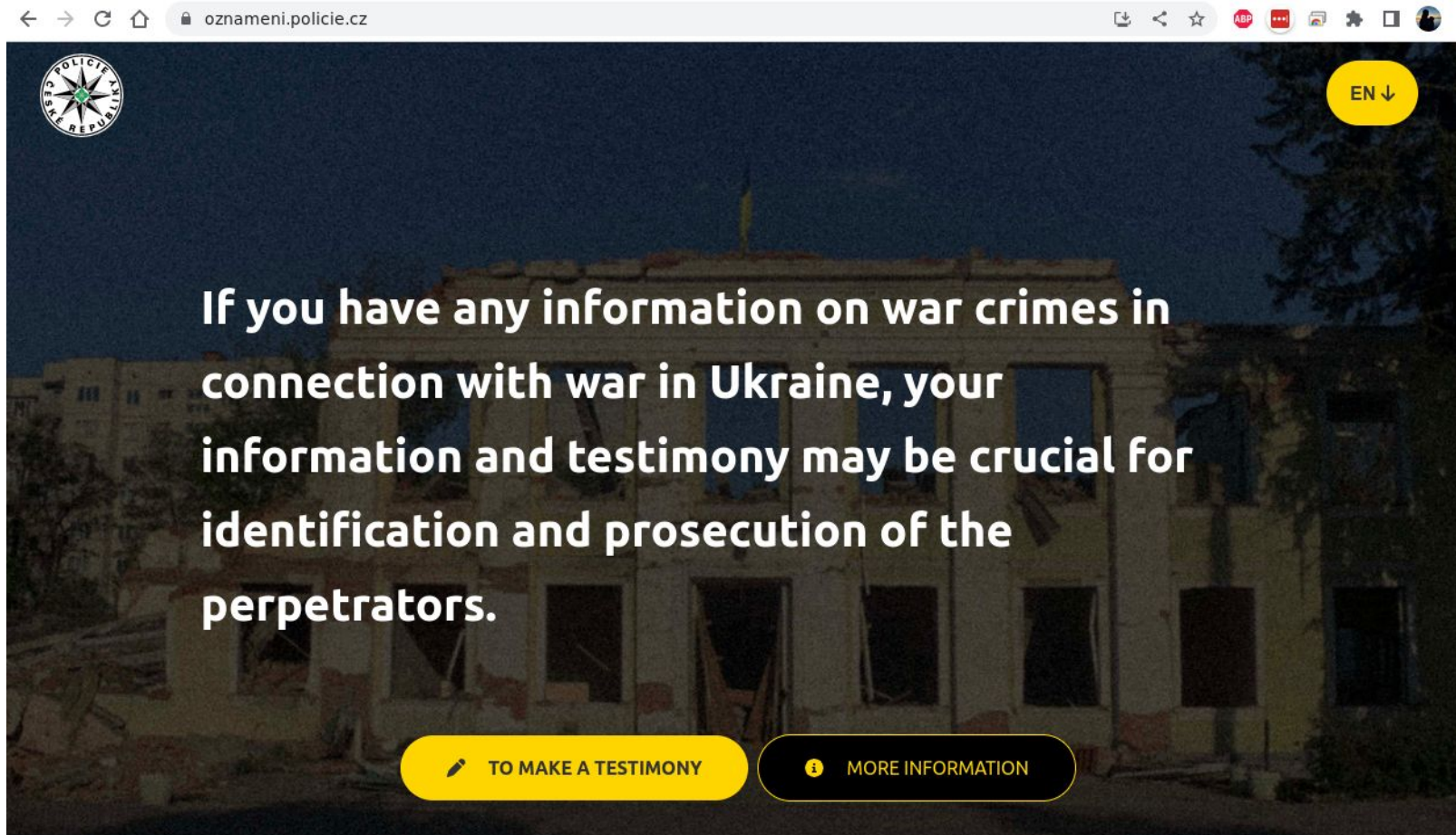
УКРАЇНА ⇒ УКРАЇНА or УКРАША or УКРА1НА or УКРА!НА

Why develop our own MT?

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



Czech Police uses Charles Translator





← → ↻ 🏠 oznameni.policie.cz

📄 🔍 ⭐ ABP 📱 ⚙️ 🖨️ 👤

 EN ↓

If you have any information on war crimes in connection with war in Ukraine, your information and testimony may be crucial for identification and prosecution of the perpetrators.

 **TO MAKE A TESTIMONY**  **MORE INFORMATION**

Czech Police uses Charles Translator

Preferred Language

- ☐ Čeština
- ☐ English
- ☐ Український
- ☐ Русский

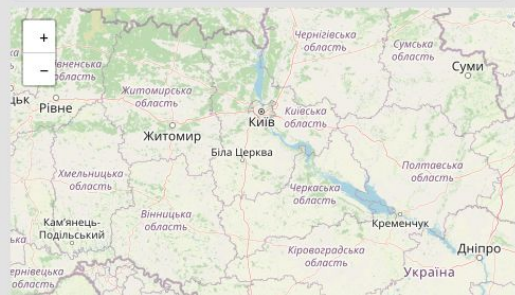
3 / 4 Incident questions Circumstances

Did you witness an attack of armed forces of any side of the conflict against civilians, infrastructure and civil objects in Ukrainian territory? *

- ☐ Yes
- ☐ No

Place *

Please write most accurate address: county, city, street, number



Event Type *

- ☐ Damage to life and health of civilians
- ☐ Attacks against unarmed civilians and residential areas
- ☐ Attacks against civil infrastructure (educational, cultural or medical facilities)
- ☐ Attacks against industrial infrastructure (factories, power plants, food storages, agriculture)
- ☐ Looting, pillaging, theft
- ☐ Kidnappings, hostage taking
- ☐ Sexual violence, torture and abuse
- ☐ Executions
- ☐ Occurrence of mass graves
- ☐ Conducting of combat operations and usage of prohibited means of combat in residential areas (area bombing, mine planting)
- ☐ Prohibited usage of special ammunition (e.g. cluster bombs, vacuum bombs, phosphorus munition, chemical and biological weapons)
- ☐ Other

Information on the Offender *

- ☐ Russian armed forces and allied armed groups under united command

- Prosecution of **war crimes in Ukraine**
- On-premise installation (sensitive data)
- Ukrainian ⇌ Czech, Russian ⇌ Czech

Another information that may be useful

Another information on the incident, offender, witnesses etc.



ADD ANOTHER INCIDENT

Why direct translation?

Зараз заява готова.
(The declaration is now ready.)



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*

Google Překladač

ROZPOZNAT JAZYK UKRAJINŠTINA ↕ ČEŠTINA ANGLIČTINA NĚMČINA

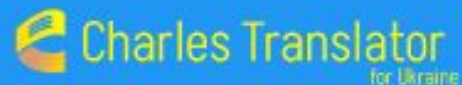
Зараз заява готова. ×

Nyní je aplikace připravena. ☆

Zaraz zayava hotova.

19 / 5 000

Why direct translation?



 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?

Google Překladač

Přihlásit se

ROZPOZNAT JAZYK ČEŠTINA ANGLIČTINA

UKRAJINŠTINA ČEŠTINA ANGLIČTINA

feminine formal masculine informal

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) ▾



Ukrajínština ▾

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?



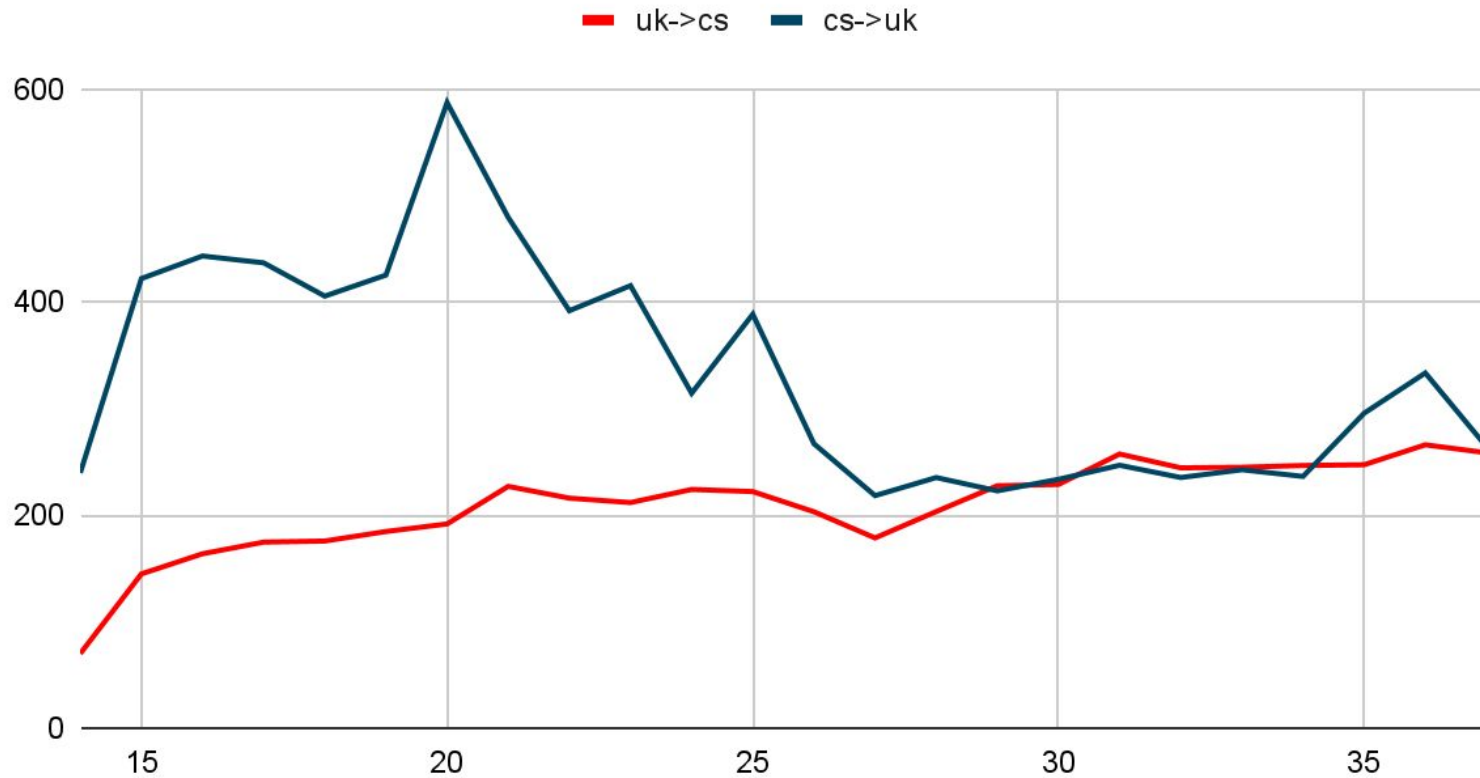
Ukrajínština

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 2022

Translation direction



Charles Translator for Ukraine: Timeline

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

[Appraise](#) [Dashboard](#) ukrces1901 ▾

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

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

0123456

0: Nonsense/ No meaning preserved2: Some meaning preserved4: Most meaning preserved and few grammar mistakes6: Perfect meaning and grammar

Reset

Submit

^ ☐ Захисники Маріуполя заявили про розпилення над містом невідомої отруйної речовини з російського безпілотнока.

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

0123456

0: Nonsense/ No meaning preserved2: Some meaning preserved4: Most meaning preserved and few grammar mistakes6: Perfect meaning and grammar

Reset

Submit

28

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	ALMAAnaCH

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	ALMAAnaCH

Слава Україні!

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

