

Presentation by local National Coordinator

CLARIN Annual Conference
October 10-12, 2022
Prague, Czech Republic

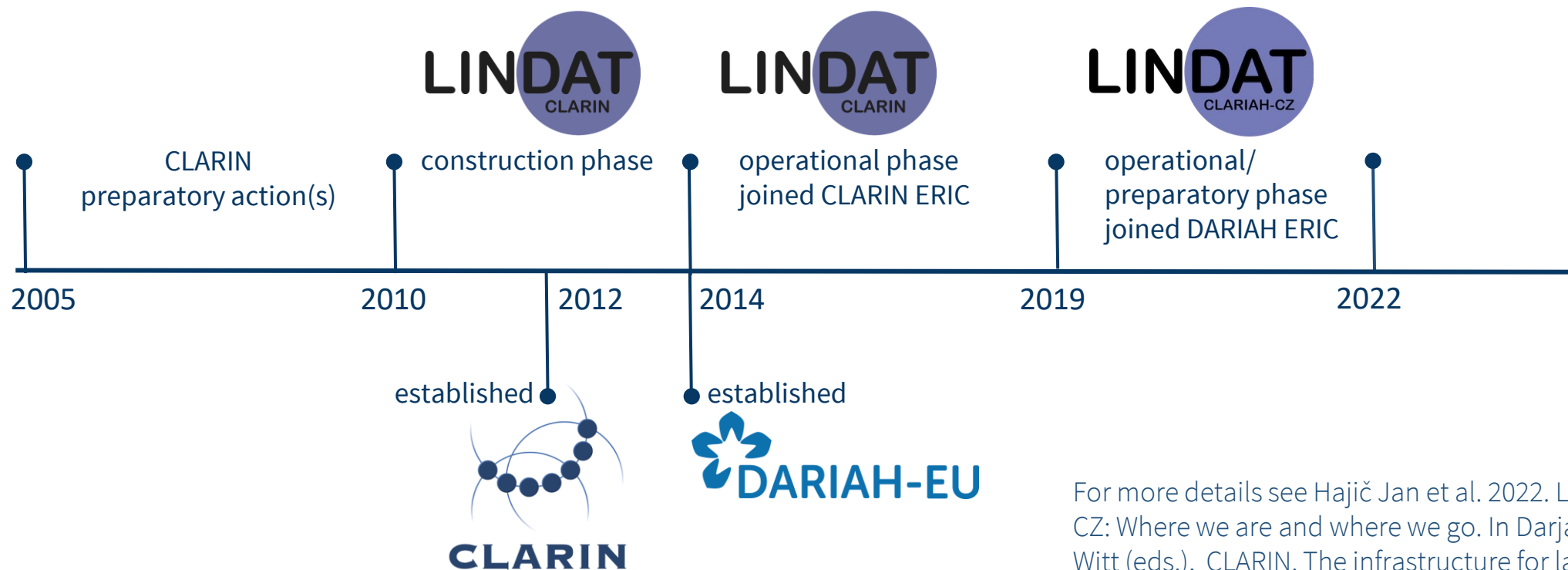
Eva Hajičová, Barbora Hladká, Martin Popel

Welcome to Prague



About LINDAT/CLARIAH-CZ (<http://lindat.cz>)

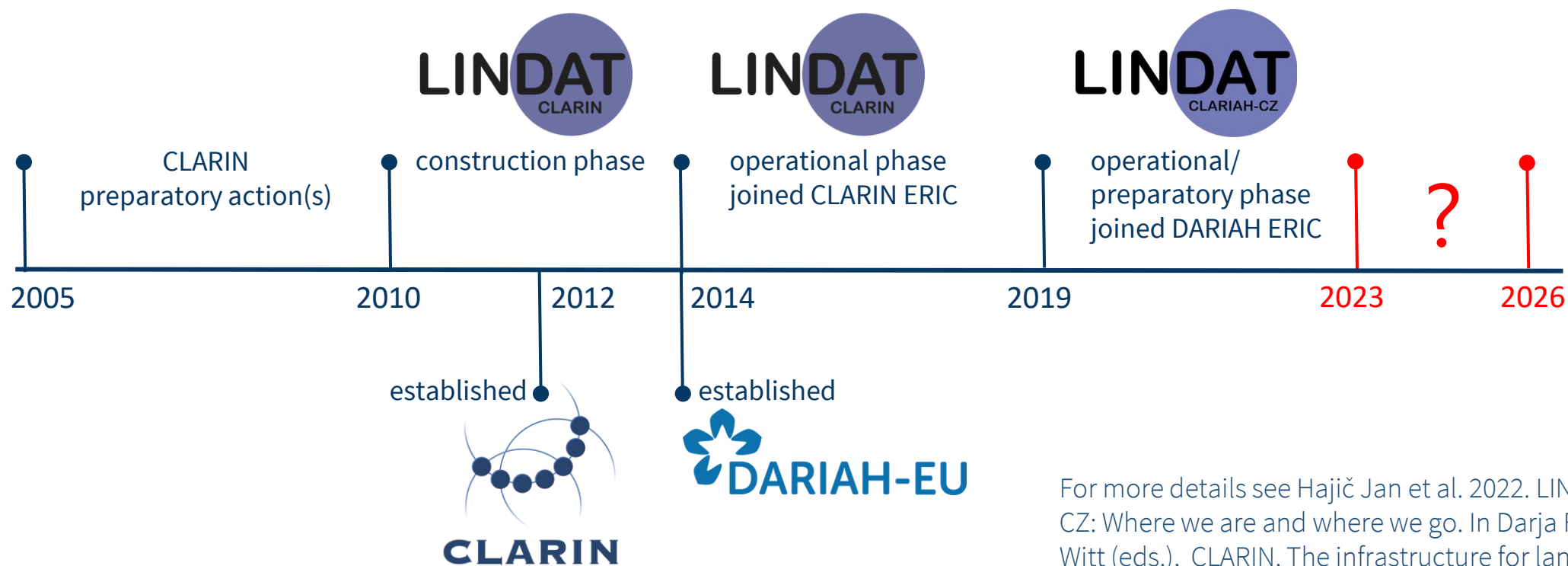
Digital Research Infrastructure for the Language Technologies, Arts and Humanities



For more details see Hajič Jan et al. 2022. LINDAT/CLARIAH-CZ: Where we are and where we go. In Darja Fišer & Andreas Witt (eds.), CLARIN. The infrastructure for language resources. Berlin: deGruyter.

About LINDAT/CLARIAH-CZ (<http://lindat.cz>)

Digital Research Infrastructure for the Language Technologies, Arts and Humanities



For more details see Hajič Jan et al. 2022. LINDAT/CLARIAH-CZ: Where we are and where we go. In Darja Fišer & Andreas Witt (eds.), CLARIN. The infrastructure for language resources. Berlin: deGruyter.

The LINDAT/CLARIAH-CZ PROJECT (LM2018101) is fully supported by the Ministry of Education, Sports and Youth of the Czech Republic under the programme LM of “Large infrastructures”.

About LINDAT/CLARIAH-CZ (<http://lindat.cz>)

<p>FACULTY OF MATHEMATICS AND PHYSICS Charles University</p> <p>Faculty of Mathematics and Physics, Charles University *</p>	<p>FACULTY OF ARTS Charles University</p> <p>Faculty of Arts, Charles University</p>	<p>MUNI ARTS Masaryk University Faculty of Arts</p> <p>Faculty of Arts, Masaryk University</p>
<p>FACULTY OF INFORMATICS Masaryk University</p> <p>Faculty of Informatics, Masaryk University</p>	<p>FAKULTA APLIKOVANÝCH VĚD ZÁPADOČESKÉ UNIVERZITY V PLZNI</p> <p>Faculty of Applied Sciences, University of West Bohemia in Pilsen</p>	<p>KNIHOVNA AKADEMIE VĚD ČR</p> <p>Library of the Czech Academy of Sciences</p>
<p>MORAVIAN LIBRARY</p> <p>Moravian Library in Brno</p>	<p>National Library of the Czech Republic</p> <p>National Library of the Czech Republic</p>	<p>Národní filmový archiv</p> <p>Národní filmový archiv</p>
<p>NATIONAL GALLERY PRAGUE</p> <p>National Gallery Prague</p>	<p>Ústav pro jazyk český Akademie věd České republiky</p> <p>Czech Language Institute of the Czech Academy of Sciences</p>	<p>Institute of History of the Czech Academy of Sciences</p> <p>Institute of History of the Czech Academy of Sciences</p>
<p>Filosofický ústav AVČR Institute of Philosophy</p> <p>Institute of Philosophy of the Czech Academy of Sciences</p>	<p>13 partners</p>	

* Center for Visual History Malach provides access to foreign oral history resources in the Holocaust and genocide research domains



- 4 new partners from European Holocaust Research Infrastructure (EHRI-CZ)
- oral + written resources

mia MASARYK INSTITUTE AND ARCHIVES OF THE CAS, v.v.i.

Terezín Initiative Institute



Data repository

As of Oct 1, 2022

- # items (metada + data) 1,395
- data volume 4.4 TB
- # unique downloads by unregistered users for fully open data 540k
- # downloads of items that require signing a license 1,200
- # registered users (only needed to submit data or sign a license) 1,218

Data repository

Non-linguistic content as well, e.g., collections created by the National Film Archive

Bohumil Veselý's Gallery of Famous Men and Women

This item is **Publicly Available** and licensed under:
Creative Commons - Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)



Name 3901475_12_hd-master.mov
Size 1.07 GB
Format QuickTime video
Description best quality
MD5 250b4401c3a2d9b98616321ff8c27e6f



[Download file](#)



Name 3901475_12_NAHLED.mp4
Size 54.9 MB
Format MPEG-4 video
Description lower quality
MD5 bf1b44128d276cea357bef6ec125038a

[Download file](#)

The Munich Agreement

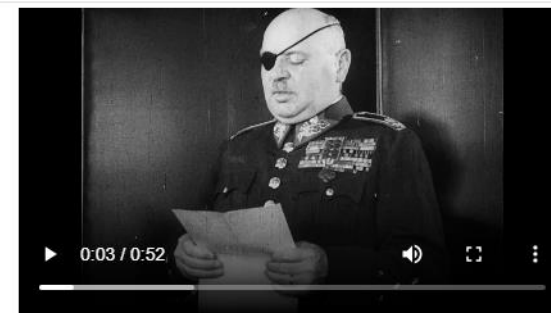
This item is **Publicly Available** and licensed under:
Creative Commons - Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)



Name ZF_0439_3901684_sot1_hd-master.mov
Size 1.07 GB
Format QuickTime video
Description best quality
MD5 47826f13ba8d8f7746f5be14518c243e



[Download file](#)



Name ZF_0439_3901684_sot1_NAHLED.mp4
Size 68.58 MB
Format MPEG-4 video
Description lower quality
MD5 3823be78de2394de300c12cb74bcfd61

[Download file](#)

Machine translation

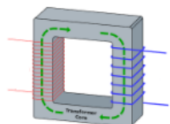
“ Please use the following text to cite this item or export to a predefined format: ”

BIBTEX CMDI

Košarko, Ondřej; Variš, Dušan and Popel, Martin, 2019, *LINDAT Translation service*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-2922>.



Share:



Authors:

Martin Popel, Dušan Variš, Ondřej Košarko

Description:

A neural networks based translation service provides a simple UI and API that lets you use Transformer models trained by our experts. Five models are currently provided with more to come.

Project home

Run



UDPipe

“ Please use the following text to cite this item or export to a predefined format: ”

BIBTEX CMDI

Straka, Milan and Straková, Jana, 2016, *UDPipe*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-1702>.



Terms of Use ▶

Most popular ▼

Machine translation

UDPipe

KonText

UWebASR

Text processing ▶

Search services ▶

Lexical resources ▶

Automatic speech recognition ▶

ToolService

LINDAT / CLARIAH-CZ

Universal Dependencies 2.10 models for UDPipe 2 (2022-07-11)



(Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL) / 2022-07-11)

Author(s):

Straka, Milan

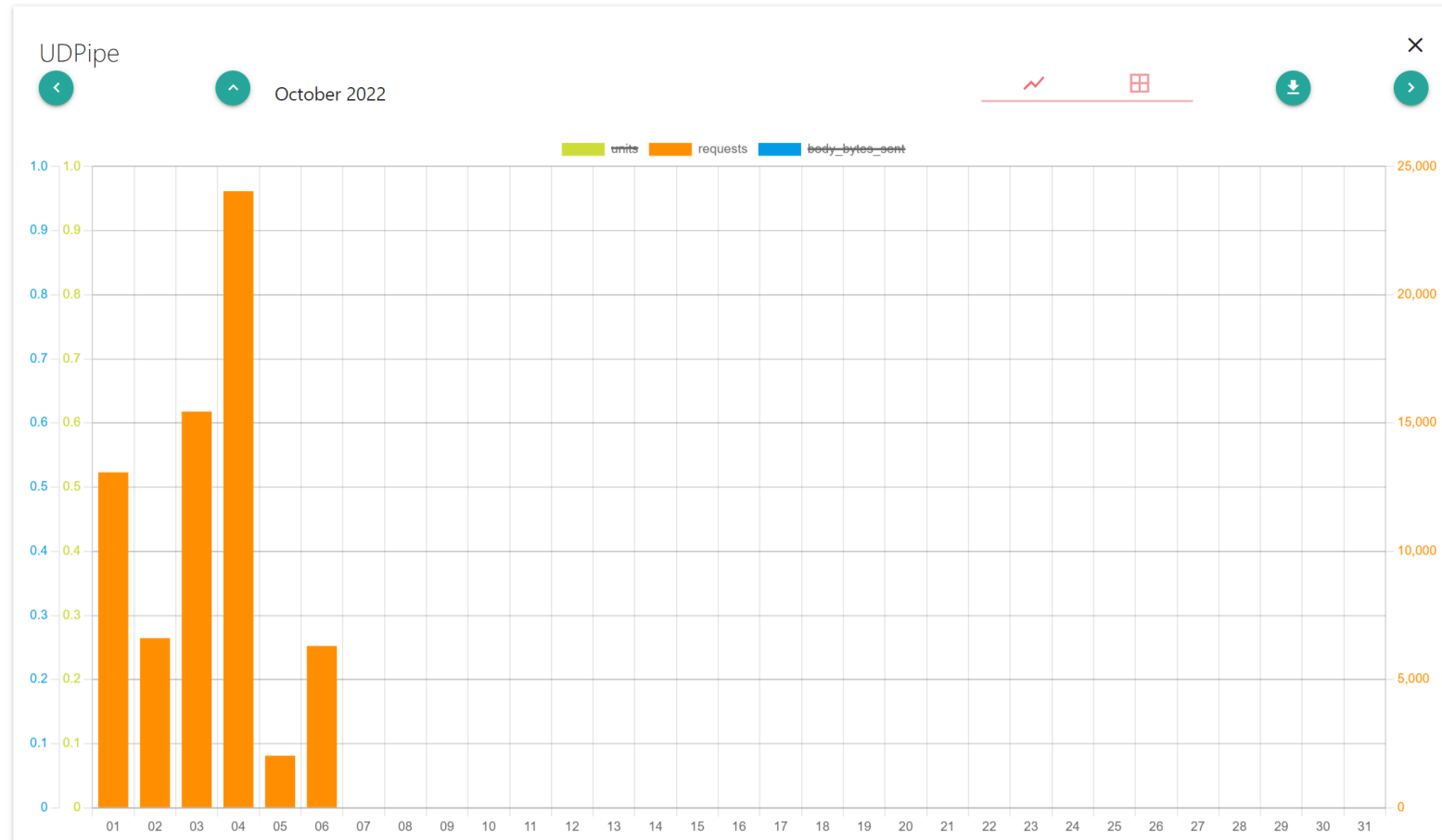
This item contains 1 file (6.88 GB).

Publicly Available

Language services

Lindat Billing

- User Panel
 - Overview
 - Endpoints
 - Tokens
- Admin Panel
 - Admin Overview**
 - Services
 - Users
 - IPs
 - Log Management



Machine translation

LINDAT/CLARIAH-CZ € Institute of Formal and Applied Linguistics € Charles University



Congratulations to Martin Popel and his u4u team on winning the Steven Krauwer Award

U4U (UFAL for Ukraine) Timeline

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

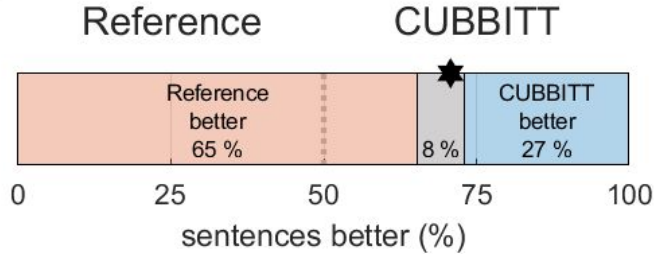
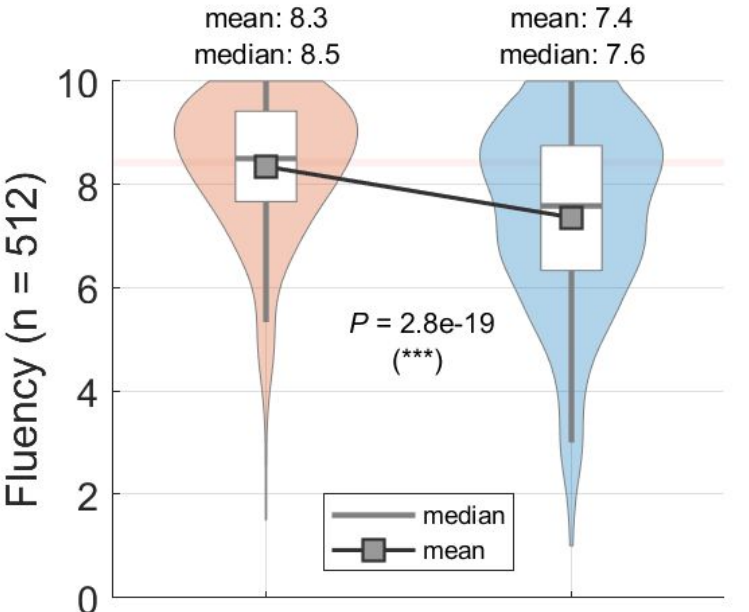
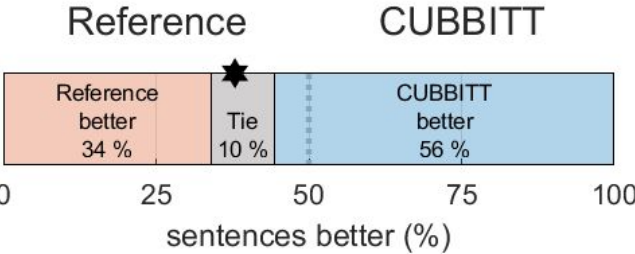
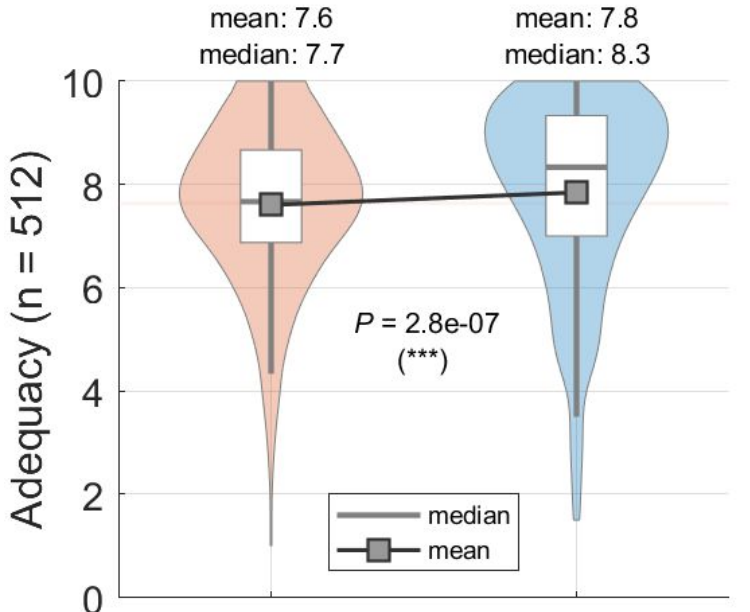
Many people use Google Translate that pivots through English and a direct translator would be better.

*Best,
Jindřich*



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](#)

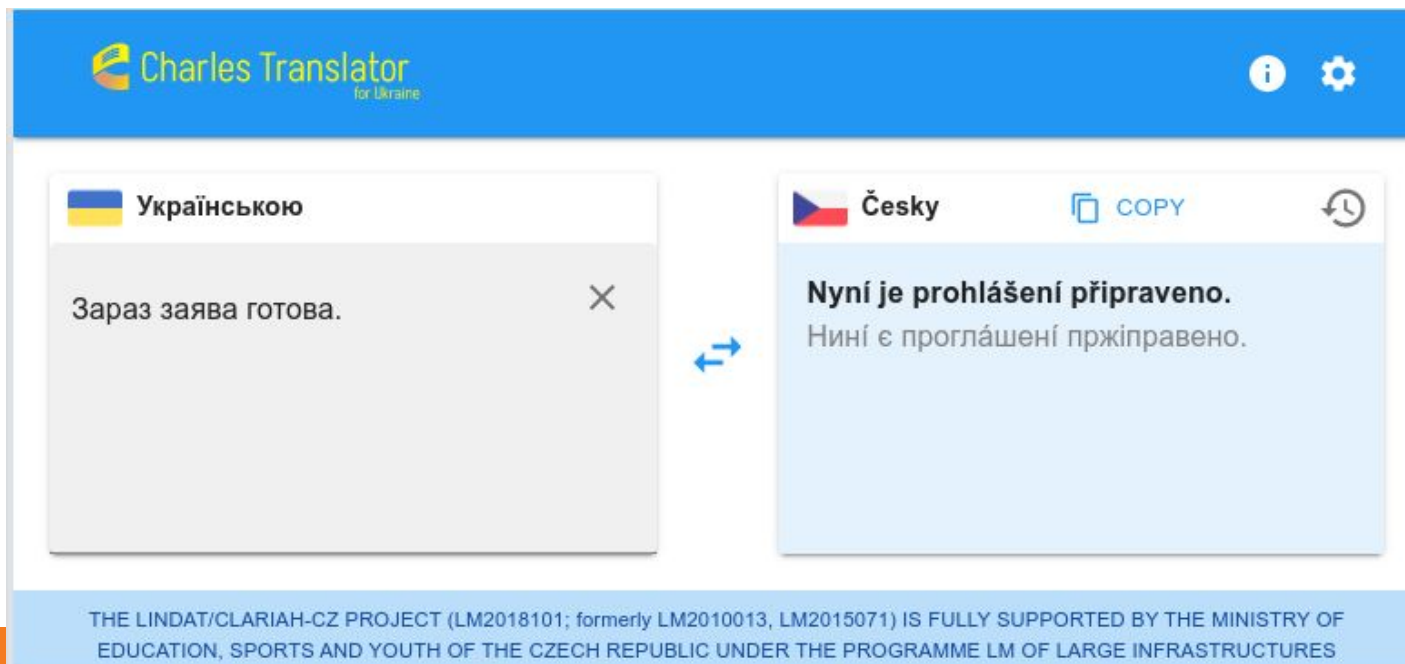
U4U (UFAL for Ukraine) Timeline

- 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



U4U (UFAL for Ukraine) Timeline

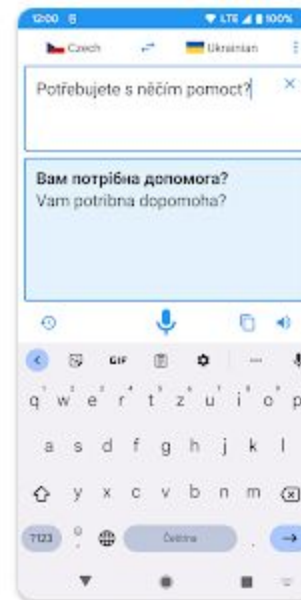
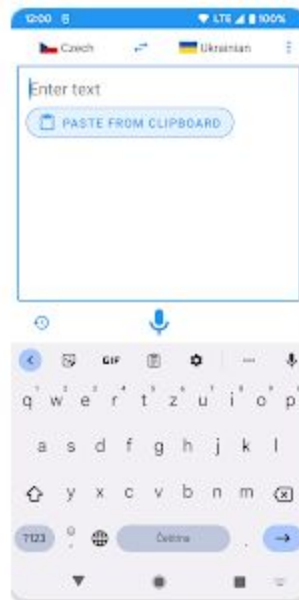
- March 1: *Hi folks, Shouldn't we build a Czech-Ukrainian translator and put it on Lindat?*
- March 12: hackathon to develop a new frontend
- March 14: first cs ↔ uk translator published at LINDAT/CLARIN



The screenshot displays the 'Charles Translator for Ukraine' web application. The interface is split into two main panels. The left panel, titled 'Українською' (Ukrainian) with a Ukrainian flag icon, contains the text 'Зараз заява готова.' (The statement is ready now). The right panel, titled 'Česky' (Czech) with a Czech flag icon, contains the translated text 'Nyní je prohlášení připraveno.' (The statement is now ready) and 'Нині є проголошені пржіправено.' (The statement is now ready). A double-headed blue arrow between the panels indicates the translation direction. The top blue header features the 'Charles Translator for Ukraine' logo and icons for information and settings. The bottom 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'.

U4U (UFAL 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



U4U (UFAL 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

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.

0 1 2 3 4 5 6

0: Nonsense/ No meaning preserved 2: Some meaning preserved 4: Most meaning preserved and few grammar mistakes 6: 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.

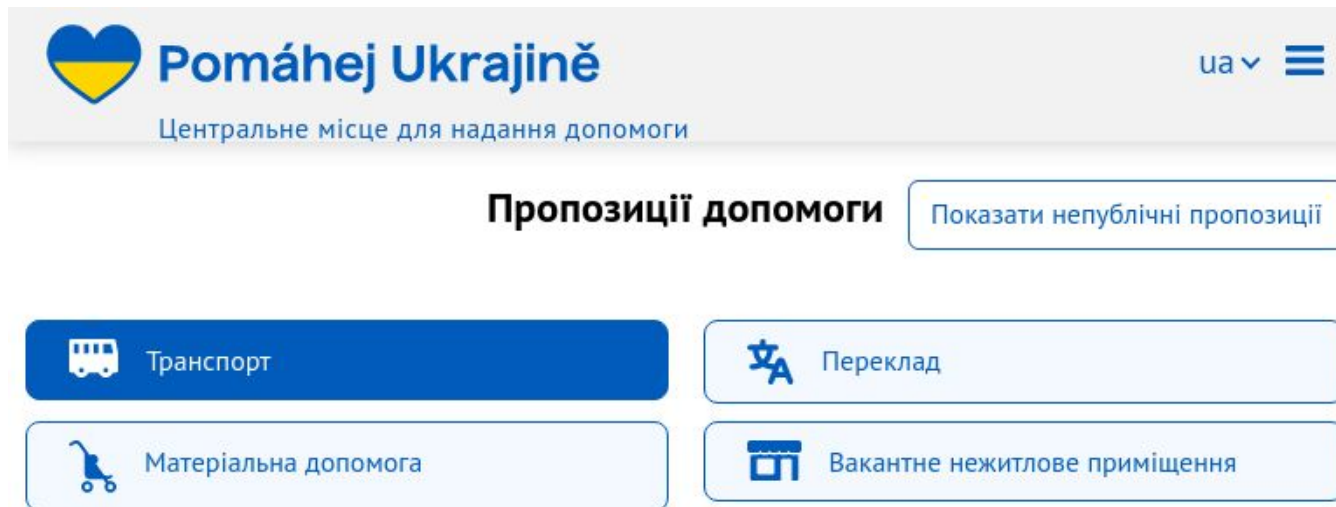
0 1 2 3 4 5 6

0: Nonsense/ No meaning preserved 2: Some meaning preserved 4: Most meaning preserved and few grammar mistakes 6: Perfect meaning and grammar

Reset Submit

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" (Help Ukraine). The header includes a logo with a heart containing the Ukrainian flag, the text "Pomáhej Ukrajině", and a subtitle "Центральне місце для надання допомоги". There is a language selector set to "ua" and a menu icon. Below the header, the main content area is titled "Пропозиції допомоги" (Offers of help) and includes a button to "Показати неpubлічні пропозиції" (Show non-public offers). The main content area features four buttons: "Транспорт" (Transport) with a bus icon, "Переклад" (Translation) with a translation icon, "Матеріальна допомога" (Material assistance) with a shopping cart icon, and "Вакантне нежитлове приміщення" (Vacant non-residential premises) with a building icon.

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 displayed with the flags of Ukraine, the USA, and the Czech Republic. Below this, the Google Translate interface is shown with 'UKRAJINŠTINA' selected as the source language and 'ČEŠTINA' as the target language. The input text is 'Зараз заява готова.' and the output is 'Nyní je aplikace připravena.' The interface includes a microphone icon, a speaker icon, a character count '19 / 5 000', and a dropdown menu for the source language.

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?



Microsoft Bing

Prohledat web

Text

Překladatel

Konverzace

Aplikace

Pro firmy

Čeština

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



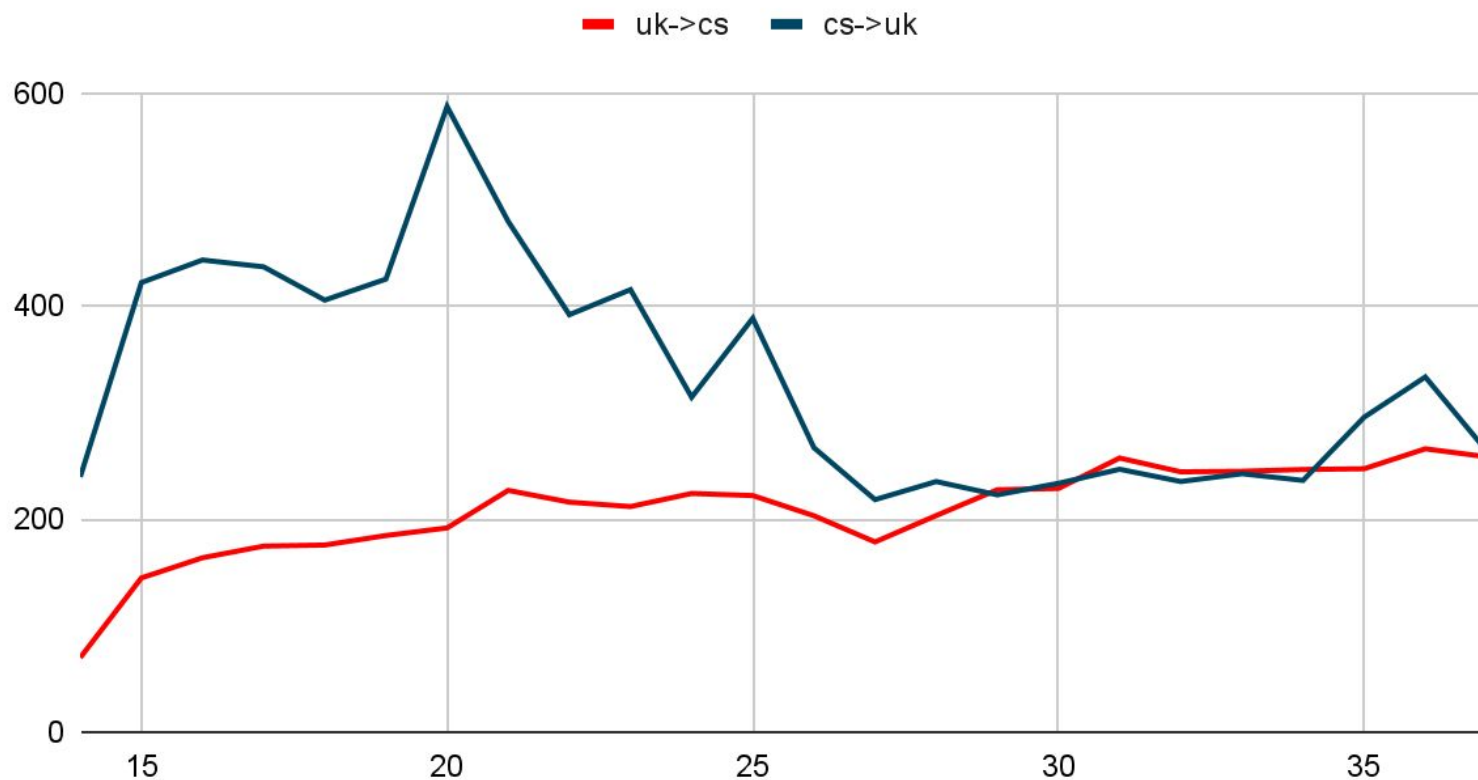
Ukrajnština

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



Слава Україні!

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

