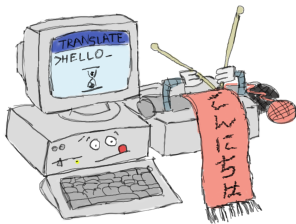


Strojový překlad a umělá inteligence

Mgr. Martin Popel, Ph.D.

Ústav formální a aplikované lingvistiky,
Matematicko-fyzikální fakulta, Univerzita Karlova

<https://ufal.mff.cuni.cz/martin-popel>



source	Great talkers are little doers.
Yandex	Velké talkers jsou trochu činitelé.
Microsoft	Velcí vysílačky jsou malí činitelé.
Google	Velcí mluvčí jsou malí lidé.
Google 2022	Velcí řečníci jsou malí dělníci.
TectoMT	Velcí řečníci jsou malí vrazi.
CUBBITT	Velcí mluvkové jsou malí dřiči.

Chlapci šly.

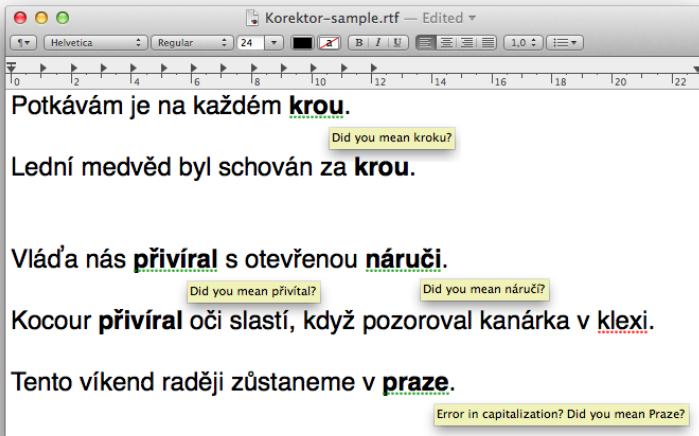
Chlapec šli do školy.

Dívce nešly hodinky. Chlapci šly.
Chlapec šli do školy.

Dívce nešly hodinky. Chlapci šly.
Kdo kam co donesl? Chlapec šli do školy.

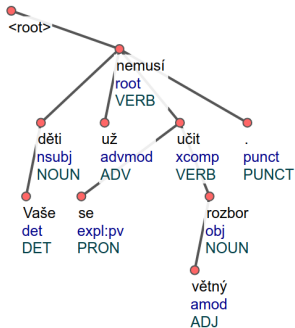


<http://ufal.cz/korektor>

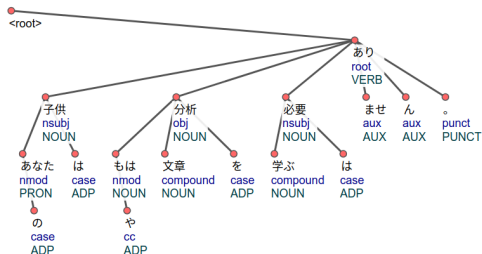


větný rozbor dostupný pro 50 jazyků
přesnost pro češtinu asi 90% (85% včetně morfologie)

Vaše děti se už nemusí učit větný rozbor .



あなたの子供はもはや文章分析を学ぶ必要はありません。



<https://lindat.cz/services/udpipe/>

Korupci často prozradí „kapřící“

Vyšetřovací tým právníků a forenzních analytiků hledá ve firmách důkazy o korupci. Prolomí kódovanou řeč i šifrovací aplikace

KATEŘINA KOLÁŘOVÁ

V e druhém patře moderní pražské kancelářské budovy Nile House usedá k jednacím stolu pětice složený tým odborníků. Právníci, forenzní analytici a vyšetřovatelé zjišťují, jestli byly v prošetřované společnosti globálního významu uzavřeny pro firmu nevýhodné smlouvy. Experti společnosti Deloitte zkoumají, jestli zaměstnanci vyšetřované společnosti „šli na ruku“ dodavatelům služeb, a za úplatek mu umožnili vyhrát lukrativní zakázku. K vyšetřování používají specializované techniky automatické analýzy dat, která prohlédne i kódovanou řeč.

„Velmi často řešíme právě vztahy dodavatelů s nákupním oddělením, ty jsou problematické téměř v každé prošetřované společnosti,“ vysvětluje Jaroslava Kračinová, advokátka a partnerka kanceláře Ambuz & Dark Deloitte Legal, jež spolupracuje s forenzním týmem vyšetřovatelů. Multidisciplinární tým už vyšetřoval i podezření z financování terorismu v zahraničí. „Takto závažné trestné činy se velmi těžko prokávají izolovaně u prošetřované společnosti klienta. Proto často, paralelně s naším šetřením, probíhá i policejní vyšetřování,“ říká Kračinová.

Rychlý zášah



Multidisciplinární tým. Vyšetřování podezření z korupce nebo projevů sexuálního harašení na pracovišti spojuje práci rozdílných vědních oborů. Analytický tým vede absolventka matematicko-fyzikální fakulty Kateřina Veselovská (vlevo), právníky advokátka Jaroslava Kračinová (vpravo).

FOTO MAPIA – DAN MATERNA

dostatečně odůvodněné. Vše řešíme po stránce pracovněprávní, i s ohledem na ochranu osobních údajů a ochranu soukromí,“ popisuje Kračinová začátek vyšetřování. Po dokončení právních kroků

ké analýzy, kterou v Deloitte využívají téměř dva roky, a to ve 28 jazycích.

Rozpoznání kódované řeči

„Na základě blízkosti slovní zásoby

Důkazem manipulace s výsledky tendru je typicky vznik

mohla existovat, protože o vítězi tendru ještě nebylo nákupním oddělením oficiálně rozhodnuto,“ popisuje Kračinová. Pak už analytici hledají v počítačích konkrétní textový soubor.

Kateřina Veselovská

■ Manažerka oddělení Data Analytics Deloitte, kde vede tým pro analýzu nestrukturovaných dat. Absolvovala doktorandský program na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze. Věnovala se vývoji softwaru pro textovou analýzu a business paradenství v projektech týkajících se oblasti nestrukturovaných a velkých dat. Nyní se zaměřuje zejména na projekty z oboru forenzní analytiky a řízení rizik.

Jaroslava Kračinová

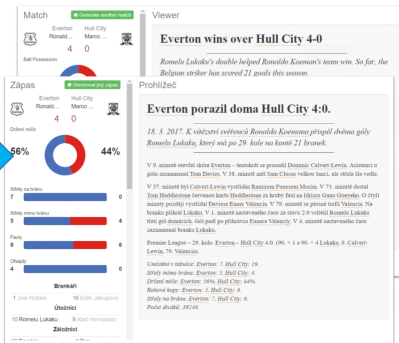
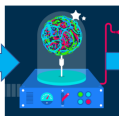
■ Partnerka a advokátka v Deloitte Legal, vede tým Business Integrity. Promovala na Právnické fakultě Univerzity Karlovy v Praze. Studovala i právo a management v Innsbrucku. Specializuje se na odhalování hospodářské kriminality, trestní odpovědnost právníků osob, corporate governance a ochranu osobních údajů.

rozmolněné hranice toho, co je v této oblasti už za hranou a co je ještě v pořádku. V některých zemích je to ale něco naprosto nepřehledného,“ vzpomíná Kračinová. Byl tým specialistů prokázal, že se

čtenáři chtějí zprávy personalizované a hned

Time (UTC)	Category	Player	Team	Goal	Player (Goal)
2016-08-18 11:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 11:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 11:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 12:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 12:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 12:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 13:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 13:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 13:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 14:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 14:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 14:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 15:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 15:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 15:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 16:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 16:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 16:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 17:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 17:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 17:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 18:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 18:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 18:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 19:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 19:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 19:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 20:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 20:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 20:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 21:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 21:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 21:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 22:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 22:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 22:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 23:00:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 23:20:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-18 23:40:00	Goal	Goal	Chelsea	Goal	Goal
2016-08-19 00:00:00	Goal	Goal	Chelsea	Goal	Goal

Structured data
(real-time & historical)



Article / Report



Byl by to rytíř, kde v pláně hřích vzlet,
Vědě jsem jse seheldo na přídoutně v světě si nezastává:
„Ukryjemné, chvěla, milý nás jest

Kolem jsou jest vyhrávaných
A svítí co pláčem, rád pravil:
Ale plná jízdo zaporodilo se, vys.

již dávno vás poháru a vlanných rány,
v jablonění je píše je i v kristování,

srdce v své ženských svém
v obly pětky tam a vzíti,
na kónku je, milý svěžek.

I'll come a bit later on my own.

I'll come a bit later on my own.
Sem čelist ještě na své milé.

MFF + DAMU + Švandovo divadlo
premiéra leden 2021 (100 let po R.U.R.)



<https://www.theaitre.com>



- Generování scénářů divadelních her
 - Umí počítač napsat divadelní hru?
 - Jak funguje jazykový model?
 - Co je to GPT-2?
- Sčítání a odčítání významů slov
 - Co jsou slovní embedinky?
 - král - muž + žena = královna
 - pivo - Česko + Rusko = ?

THEATRE THEaiTRobot 1.0 THEaiTRobot 2.0 Switch to English

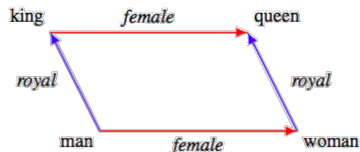
UKÁZKOVÝ SCÉNÁŘ

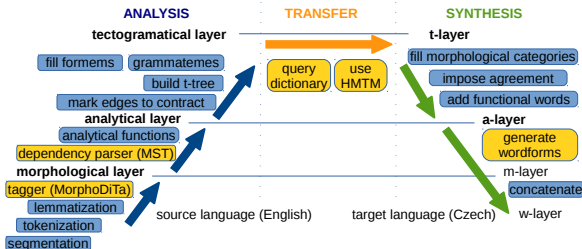
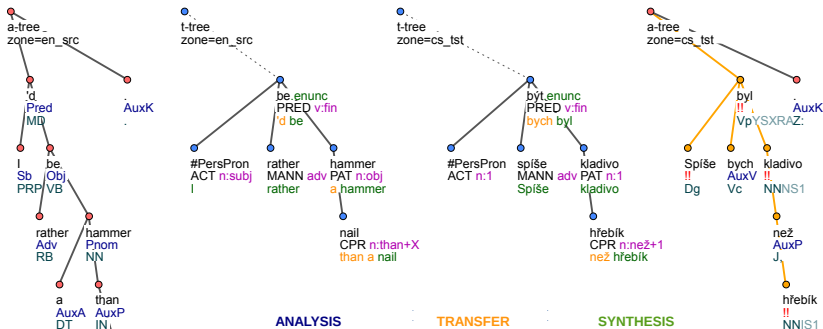
Scéna Procházka. Faust v myšlenkách přechází, k němu Mefistofeles.

Mefistofeles Při vši zhrzené lásce! Při jízdě do pekel! Chtěl nové bych vymyslet živly, bych při tom nejhorším klel!

Faust Co je ti? Viděl ti kdy svět? Vyvádíš, že tě nepoznávám.

← → GENEROVAT DÁL





I'd rather be a hammer than a nail.

Spíše bych byl kladivo než hřebík/nehet.

output_label=hřebík#N	
feature	λ
child_formeme_n:in+X=1	1.64
is_member_of_coord=1	1.30
child_formeme_v:fin=1	1.04
next_lemma=down	0.84
is_capitalized=1	0.79
+precedes_parent=0	0.75
tense_g=post	0.74
+voice_g=active	0.66
prev_lemma=drive	0.66
parent_capitalized=1	0.62
formeme=n:from+X	0.60
+prev_lemma=hammer	0.59
child_lemma_few=1	0.55
child_lemma_remove=1	0.54
sempos=n.denot	0.50
next_lemma=and	0.50
formeme_g=v:until+fin	0.49
child_lemma_rusty=1	0.47
...	

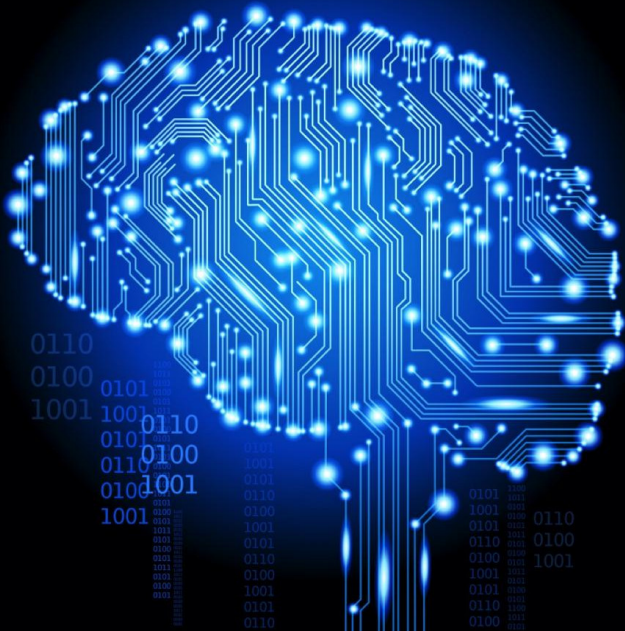
output_label=hřebík#N	
feature	λ
child_formeme_n:in+X=1	1.64
is_member_of_coord=1	1.30
child_formeme_v:fin=1	1.04
next_lemma=down	0.84
is_capitalized=1	0.79
+precedes_parent=0	0.75
tense_g=post	0.74
+voice_g=active	0.66
prev_lemma=drive	0.66
parent_capitalized=1	0.62
formeme=n:from+X	0.60
+prev_lemma=hammer	0.59
child_lemma_few=1	0.55
child_lemma_remove=1	0.54
sempos=n.denot	0.50
next_lemma=and	0.50
formeme_g=v:until+fin	0.49
child_lemma_rusty=1	0.47
...	

output_label=nehet#N	
feature	λ
child_formeme_n:poss=1	1.32
child_lemma_finger=1	1.07
child_formeme_n:of+X=1	0.98
precedes_parent=1	0.88
prev_lemma=black	0.77
child_lemma_broken=1	0.76
child_formeme_v:attr=1	0.70
formeme=n:at+X	0.67
formeme_g=n:attr	0.67
child_lemma_long=1	0.67
next_lemma=file	0.60
child_lemma_false=1	0.58
prev_lemma=false	0.58
+number=sg	0.56
formeme=n:obj	0.53
formeme=n:by+X	0.52
...	

umělá inteligence
~1950

strojové učení
~1980

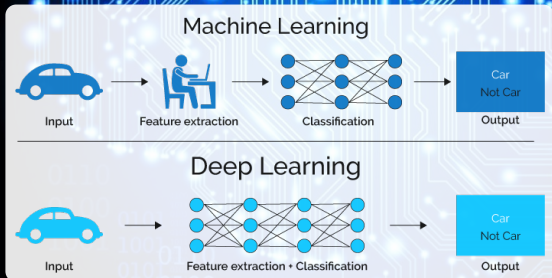
hluboké učení
~2010



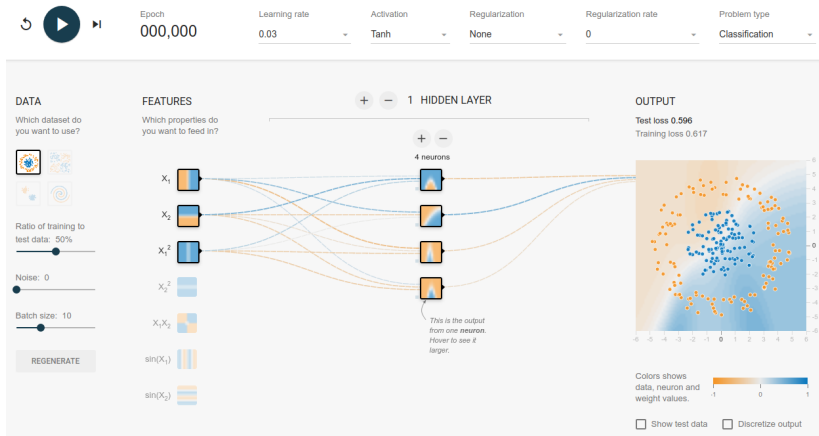
umělá inteligence
~1950

strojové učení
~1980

hluboké učení
~2010

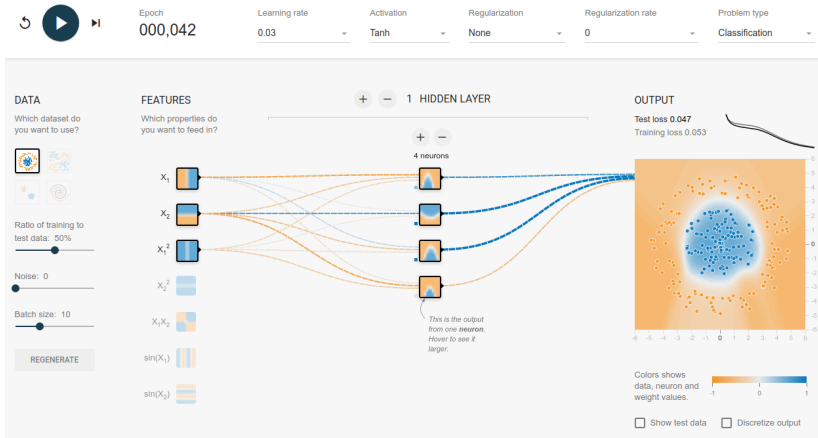


jednoduchá architektura (16 parametrů)



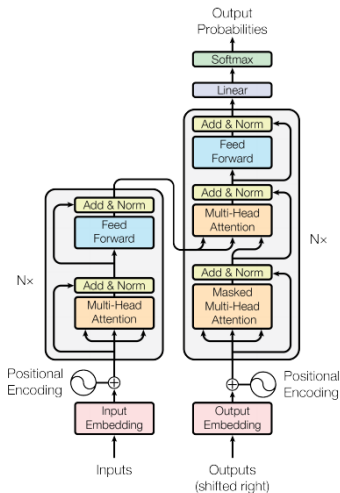
<https://playground.tensorflow.org>

jednoduchá architektura (16 parametrů)

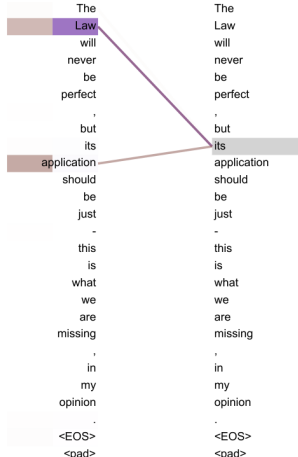
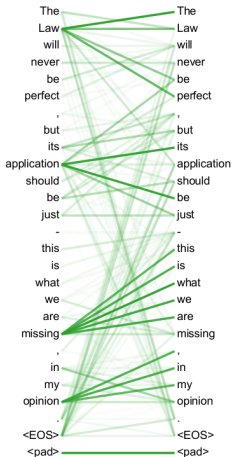


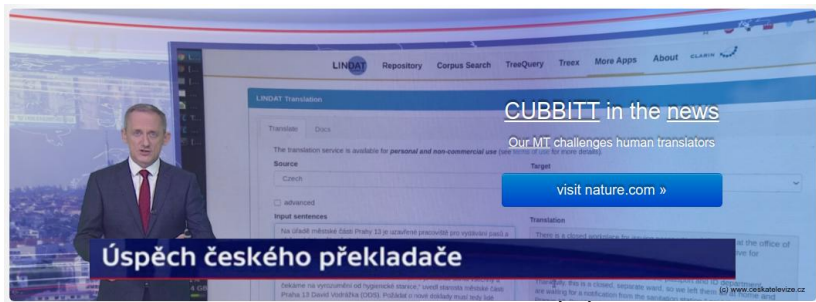
<https://playground.tensorflow.org>

architektura Transformer (213 milionů parametrů)



síť se sama učí důležité vztahy mezi slovy (díky self-attention)





Úspěch českého překladače

nature communications

Zkuste si CUBBITT sami:
<https://lindat.cz/cubbitt>
(En↔Cs, Fr, Pl)

[nature](#) > [nature communications](#) > [articles](#) > [article](#)

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

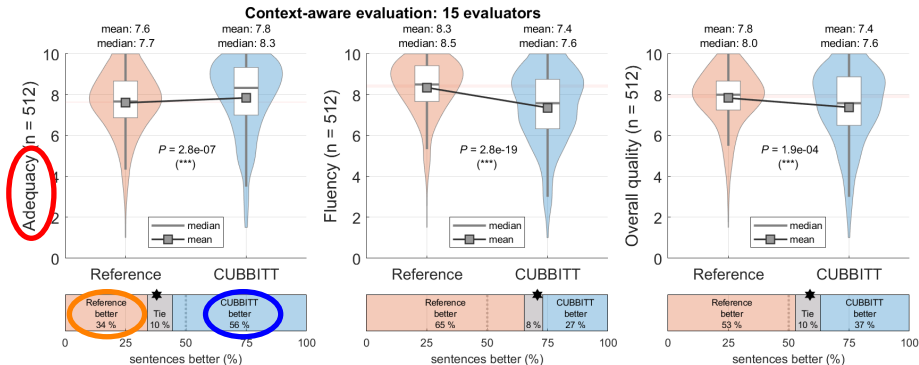
Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals

Martin Popel , Marketa Tomkova, Jakub Tomek, Lukasz Kaiser, Jakob Uszkoreit, Ondrej Bojar & Zdeněk Zábokrtský

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

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

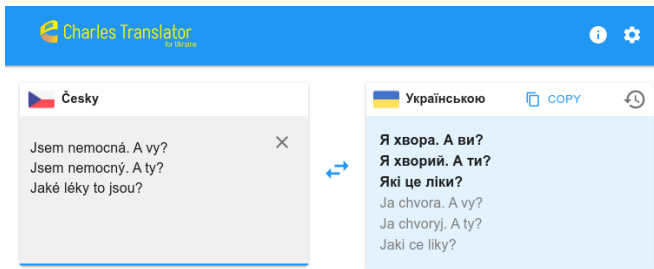
56 % vět přeložil **přesněji** náš překladač CUBBITT,
34 % profesionální překladatelská agentura





Zkuste si Charles Translator:
<https://translator.cuni.cz>





- první verze překladače vytvořena za 2 týdny
- cs→uk: asi 1400 uživatelů denně, uk-cs: asi 700
- přímý překlad (nikoli přes angličtinu)
- fonetický přepis azbuky, hlasový vstup,...
- API: překlad učebnic, PomahejUkrajine.cz,...

The image displays two translation interfaces side-by-side. The top interface is 'Charles Translator for Windows', which shows a translation from Czech to Ukrainian. The bottom interface is 'Google Překladač', which shows a translation from Czech to Ukrainian, but with a different result for the third sentence.

Charles Translator for Windows

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?**

Google Překladač

ROZPOZNAT JAZYK **ČEŠTINA** ANGLIČTINA ↕ **UKRAJINŠTINA** ČEŠTINA ANGLIČTINA ↕

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

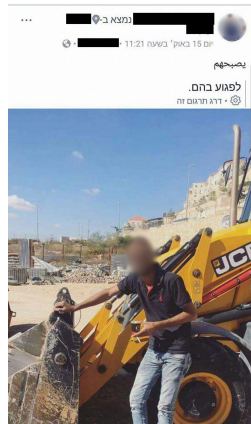


Facebook

Facebook translates 'good morning' into 'attack them', leading to arrest

Palestinian man questioned by Israeli police after embarrassing mistranslation of caption under photo of him leaning against bulldozer

zdroj: [The Guardian \(2017\)](#)



Source	Jana je žena. Pracuje jako průvodčí.
Google	Jana is a woman. He works as a guide.
Bing	Jana is a woman. He works as a conductor.
CUBBITT	Jane is a woman. He works as a conductor.
CUBBITT-doc	Jana is a woman. She works as a conductor.

Source	Saša je muž. Pracuje jako průvodčí.
DeepL (2021)	Sasha is a man. She works as a conductor.
CUBBITT-doc	Sasha is a man. He works as a conductor.

Source	Pavla není muž. Pracuje jako průvodčí.
DeepL	Paul is not a man. She works as a conductor.
CUBBITT-doc	Pavla is not a man. He works as a conductor.

source As good be an addled egg as an idle bird.

Bing Jako dobrý být popletený vejce jako nečinný pták.

Google Jako dobrá být včleněná vejce.

T2009 Dobré je feťácké vejce jako činný pták.

T2018 Dobří buďte plete vejce jako nečinný pták.

CUBBITT Stejně dobré je být pomateným vejcem jako zahálejícím ptákem.

source A miss by an inch is a miss by a mile.

Bing Miss o palec je Miss o míli.

Yandex Slečna tím, že palec je vedle o míli.

Google Chybějící palcem je míle vzdálená míle.

T2009 Slečna palec je slečna miliónu.

T2018 Slečna palce je slečna míle.

CUBBITT Minutí o centimetr je o kilometr.

Birds of a feather flock together.

Ptáci peří stáda dohromady.

Vrána k vráně sedá.

Vrána k vráně sedá.

Ptáci v bederním hejnu spolu.

Ptáci péřového hejna spolu.

Vrána k vráně sedá.

Zkuste si CUBBITT sami: <https://lindat.cz/cubbitt>