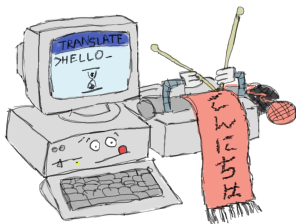


# Strojový překlad a umělá inteligence

**Mgr. Martin Popel, Ph.D.**

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

<http://ufal.cz/martin-popel>



source	Great talkers are little doers.
Yandex	Velké talkers jsou trochu činitelé.
Bing	Velcí vysíláčky jsou malí činitelé.
Google	Velcí mluvčí jsou malí lidé.
TectoMT	Velcí řečníci jsou malí vrazi.
CUBBITT	Velcí mluvkové jsou malí dřiči.



Chlapci šly.

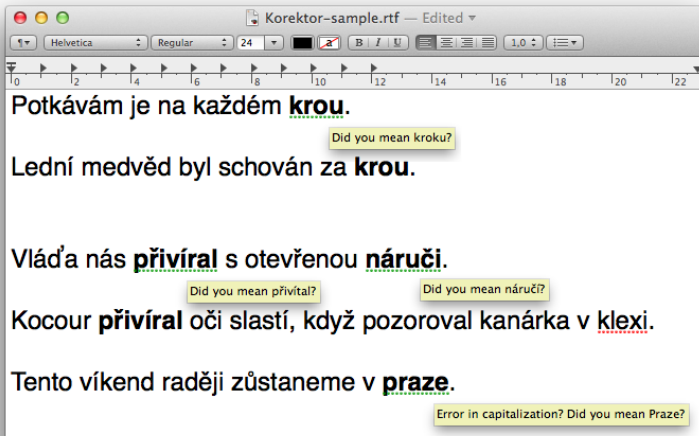
Chlapec šli do školy.

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

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

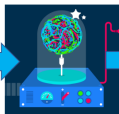


<http://ufal.cz/korektor>

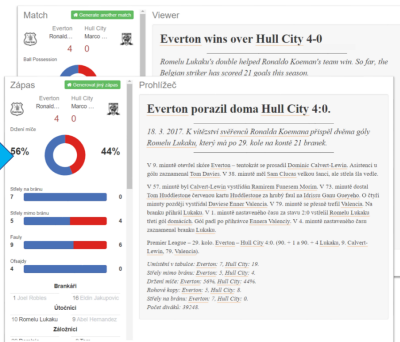


zpráv přibývá (FB, Twitter), novinářů ubývá, zisky klesají, čtenáři chtějí zprávy personalizované a hned

Time (UTC)	Category	Player	Team	Goal	Goalkeeper
2017-03-18 11:00:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 11:20:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 11:30:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 11:40:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 11:50:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 12:00:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 12:10:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 12:20:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 12:30:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 12:40:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 12:50:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 13:00:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 13:10:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 13:20:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 13:30:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 13:40:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 13:50:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 14:00:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 14:10:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 14:20:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 14:30:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 14:40:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 14:50:00	GOAL	James	Everton	Goal	Simon Mignolet
2017-03-18 15:00:00	GOAL	James	Everton	Goal	Simon Mignolet



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 plánována na leden 2021 (100 let po R.U.R.)



<https://www.theaitre.com>

## 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 MAFRIA – 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 záso-

**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 poradenství v projektech týkajících se oblasti nestrukturovaných a velkých dat. Nyní se zaměřuje zejména na projekty z oboru forenzní analytiky a řízení rizik.

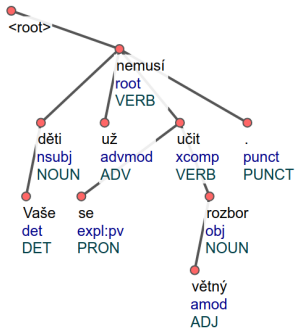
**Jaroslava Kračinová**

■ 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.

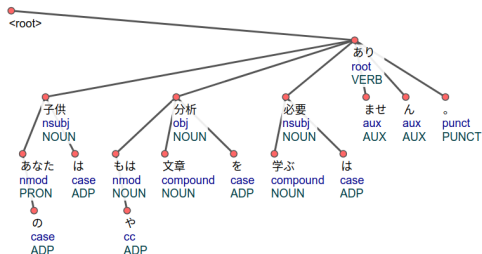
rozmolněné hranice toho, co je v této oblasti už za hranou a co je ještě v pořádku. V některých zemích je to ale něco naprosto nepřehledného,“ vzpomíná Kračinová. Byl tím specialistů prokázal, že se

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。



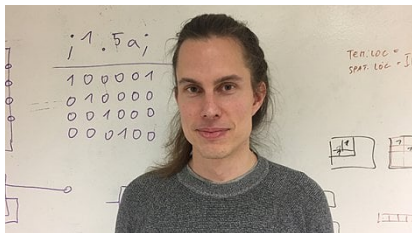
<https://lindat.cz/services/udpipe/>

Umíte sčítat a odčítat čísla?  
A co slova a obrázky?

král - muž + žena = ?

král - muž + žena = **královna**

Tomáš Mikolov, 2012, word2vec



<https://projector.tensorflow.org/>



## Country and Capital Vectors Projected by PCA

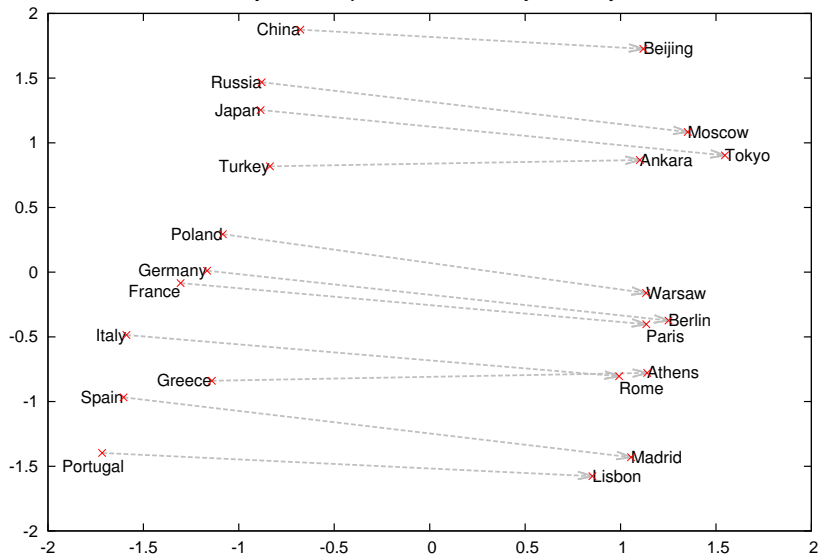
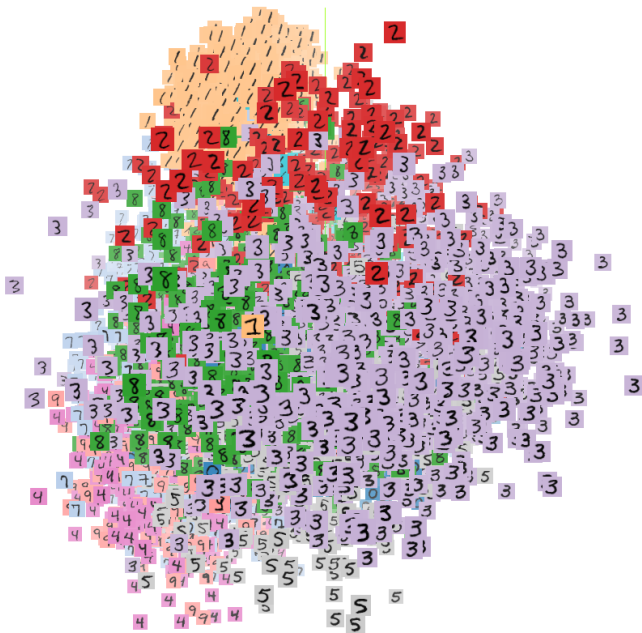
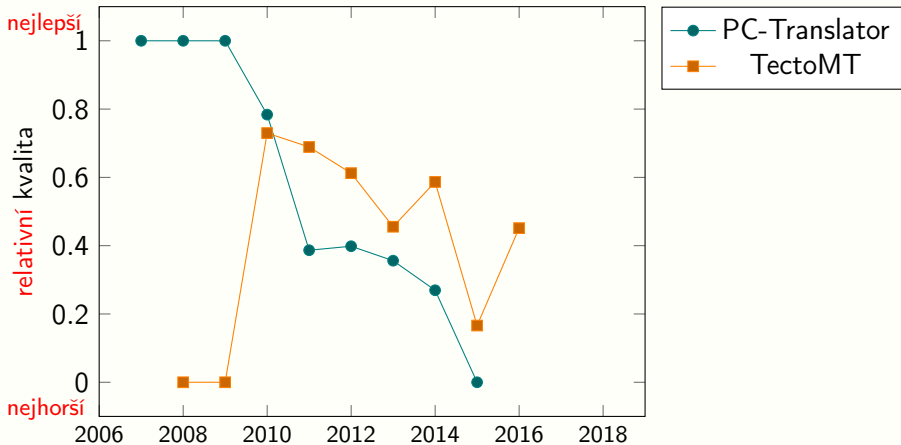
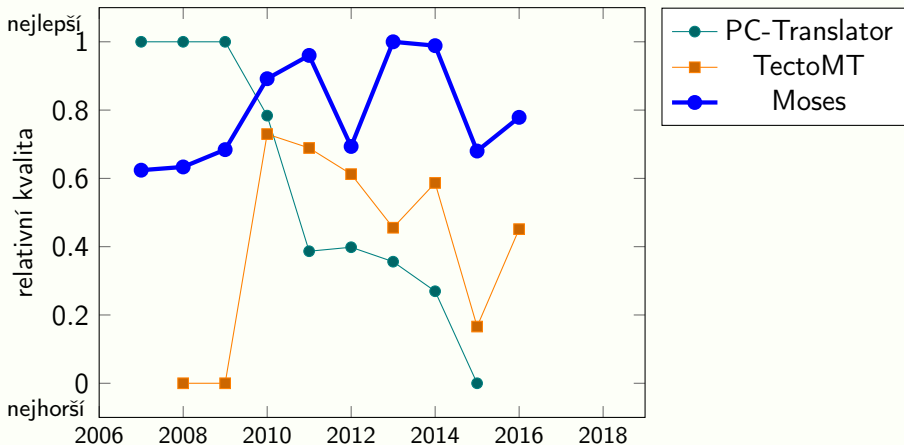


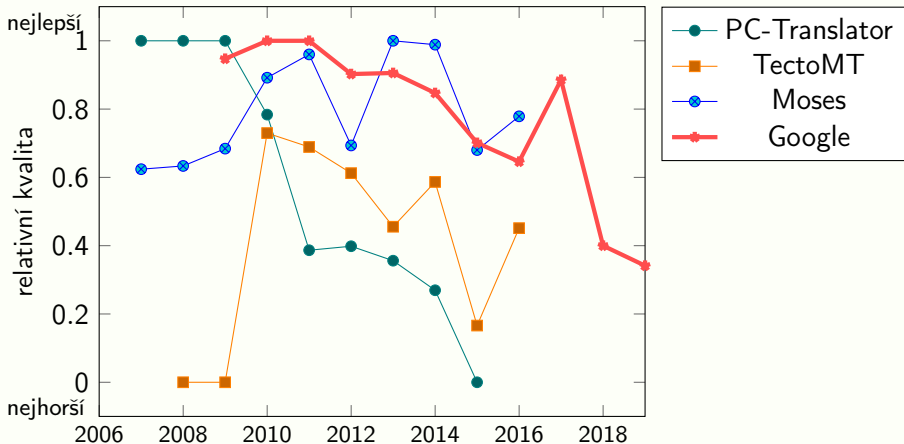
Table 8: *Examples of the word pair relationships, using the best word vectors from Table 4 (Skip-gram model trained on 783M words with 300 dimensionality).*

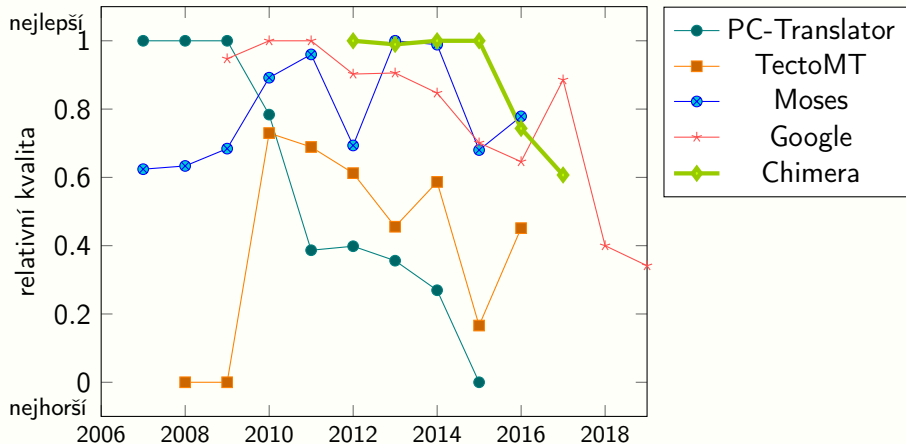
Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

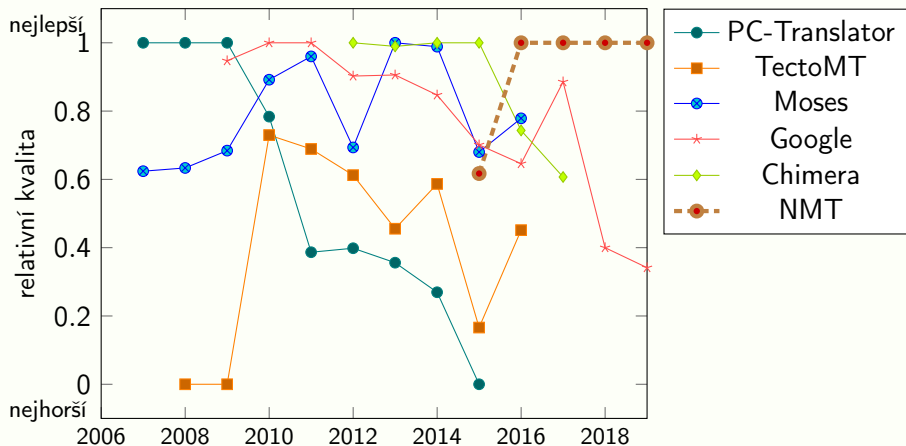




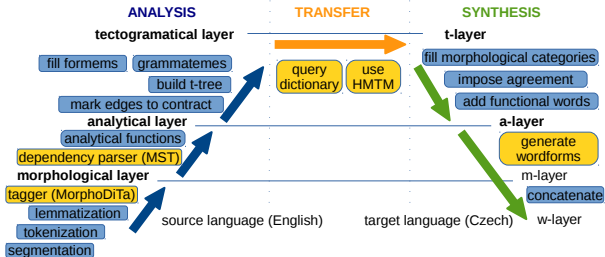
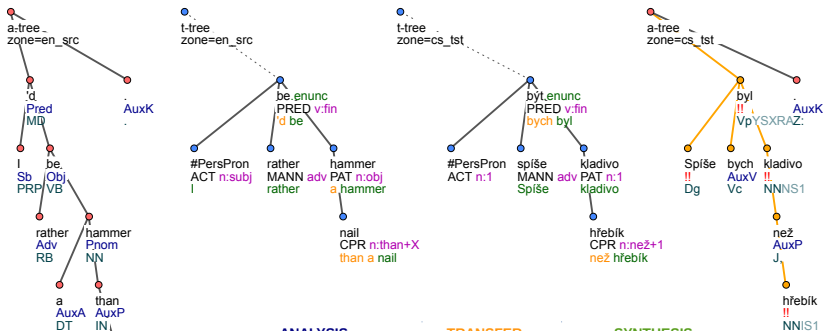












I'd rather be a hammer than a nail.

Spíše bych byl kladivo než hřebík/nehet.

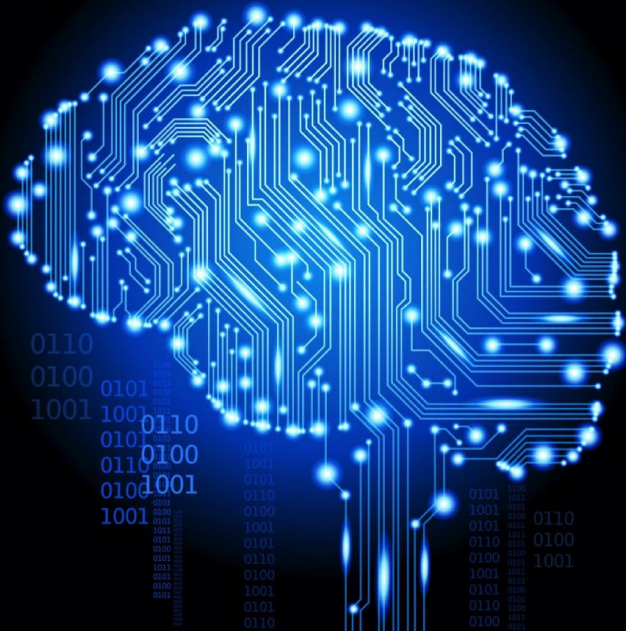




umělá inteligence  
~1950

strojové učení  
~1980

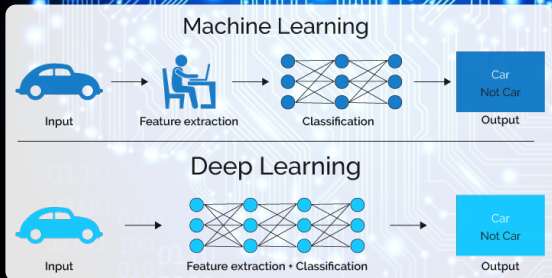
hluboké učení  
~2010



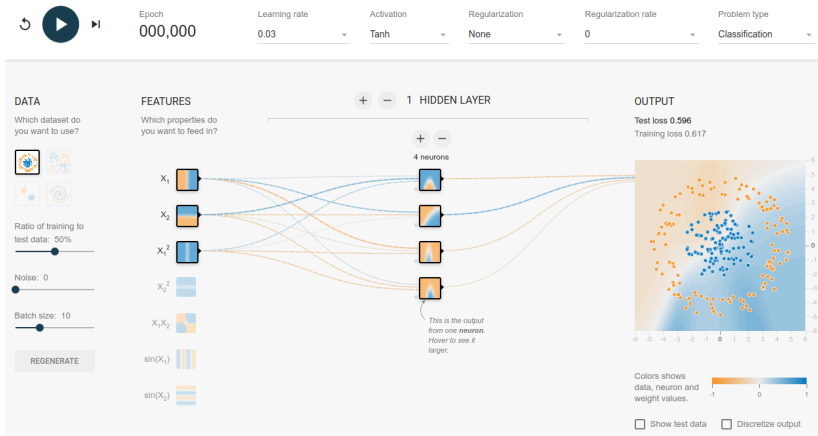
umělá inteligence  
~1950

strojové učení  
~1980

hluboké učení  
~2010

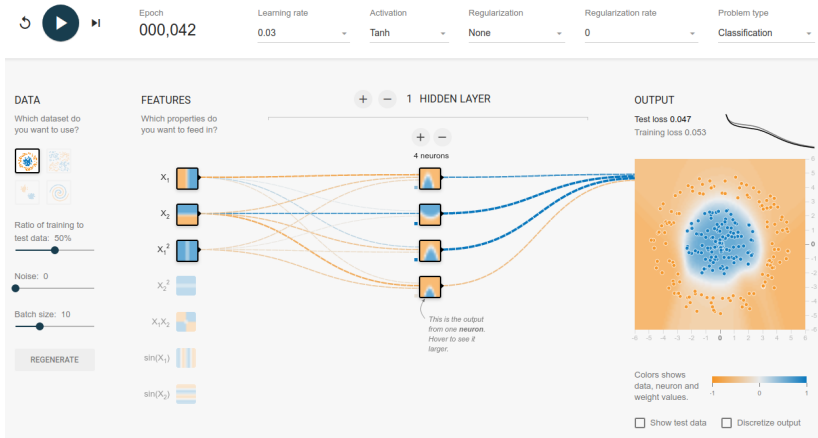


## jednoduchá architektura (16 parametrů)

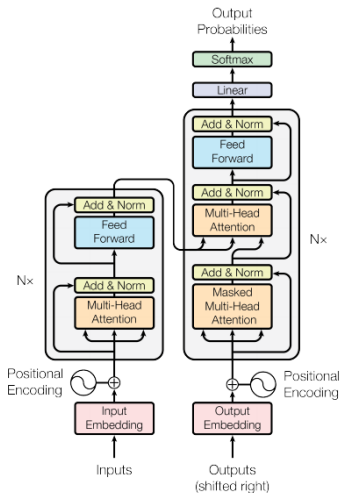


<https://playground.tensorflow.org>

## jednoduchá architektura (16 parametrů)

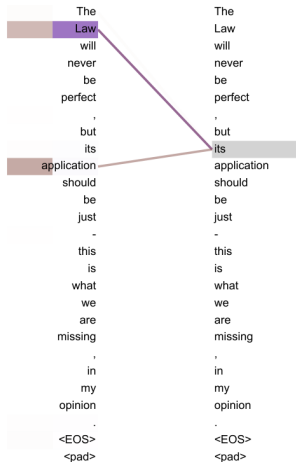


architektura Transformer (213 milionů parametrů)

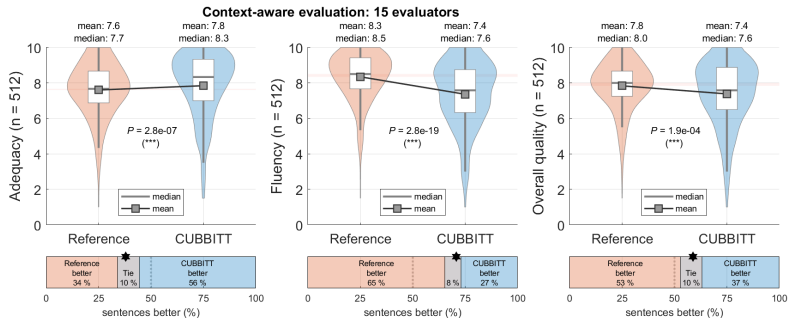




síť se sama učí důležité vztahy mezi slovy (díky self-attention)



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



## 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, Lukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar & Zdeněk Zábokrtský

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

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

Zkuste si CUBBITT sami:

<https://lindat.cz/cubbitt>

Support The Guardian    Subscribe    Find a job    Sign in

# The Guardian

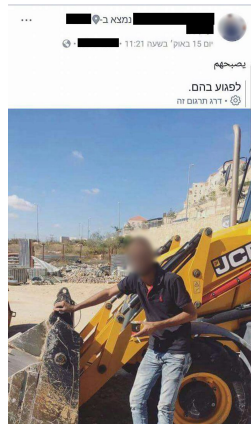
[News](#)   [Opinion](#)   [Sport](#)   [Culture](#)   [Lifestyle](#)

**World** ▶ Europe US Americas Asia Australia **Middle East** Africa Inequality

## 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



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>