

Strojový překlad z pohledu filosofie jazyka

Tomáš Musil

📅 5. listopadu 2018



<http://tomasm.cz/>
ÚFAL MFF UK
ÚFaR FF UK



unless otherwise stated

Strojový překlad

Stručná historie

Neuronový strojový překlad

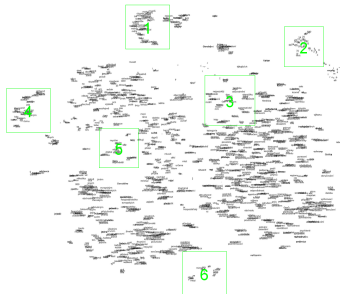
Filosofie jazyka

Holismus

Význam jako užití

Strukturalismus

Jak dál?



Strojový překlad

Specialization is for insects.



...10100110111011010...



Specializace je pro hmyz.

angličtina polština islandština Rozpoznat jazyk ▼



francouzština angličtina čeština ▼

Přeložit

Specialization is for insects. ✕



Specializace je pro hmyz.



- pravidlový MT
- lingvistická intuice, klasické gramatické teorie

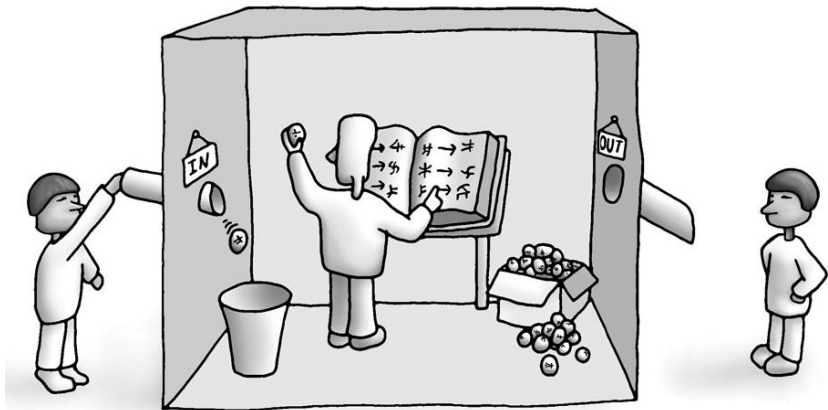
Stručná historie strojového překladu

- pravidlový MT
- frázový/stat. MT
- lingvistická intuice, klasické gramatické teorie
- teorie informace, syntaktické struktury

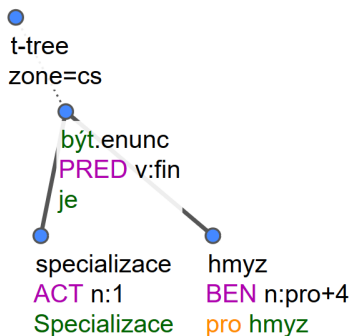
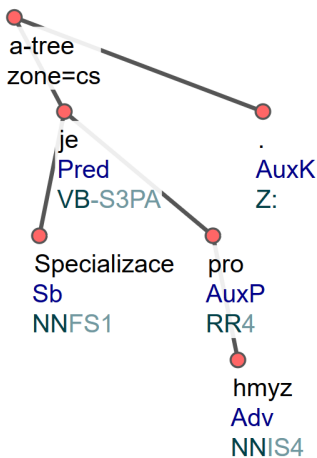
Stručná historie strojového překladu

- pravidlový MT
- frázový/stat. MT
- neuronový MT
- lingvistická intuice, klasické gramatické teorie
- teorie informace, syntaktické struktury
- sekvence symbolů

Čínský pokoj

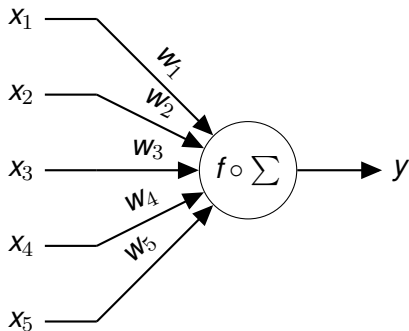


[cs] Specializace je pro hmyz.



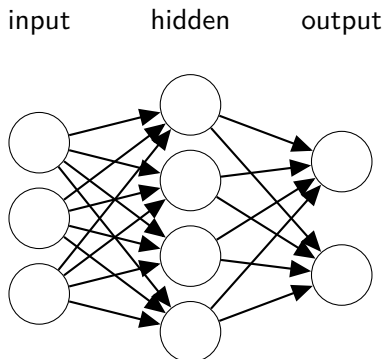
Neuronový strojový překlad

Matematický neuron

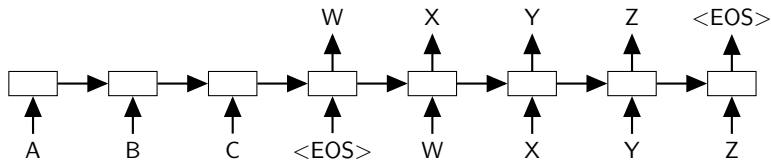


$$y = f\left(\sum_i w_i x_i\right)$$

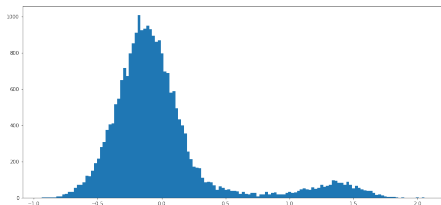
Jednoduchá neuronová síť



Sequence to sequence learning



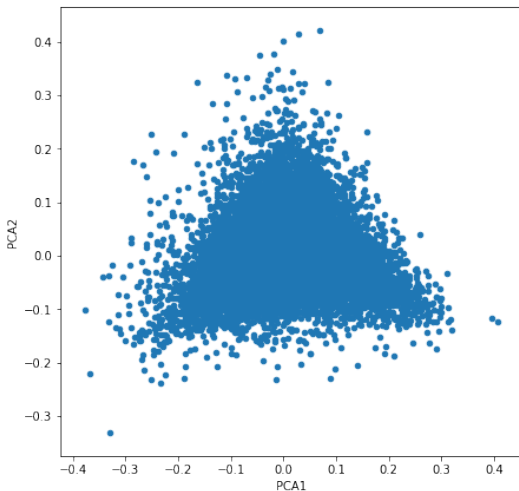
Word Embeddings in NMT: proper names



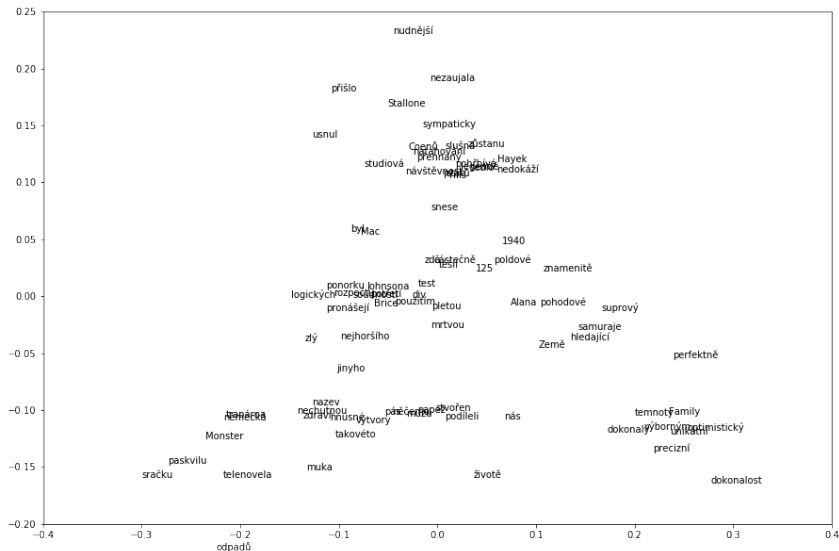
volant okně zátylek křeslem ostrůvek lucernu oknu lavičku krbem polštáři
lahvičku psovi silnicí gauči ostrovem brzdu halou prostěradlo botou
kočáru

Newt Toranaga Paithan Bremen Baniči Blake Sadi Sacharóza Vanion
Chronos Jaku Frigate Albert Culum Richards Duun Rozšafín Tynian
Anheg Hawat

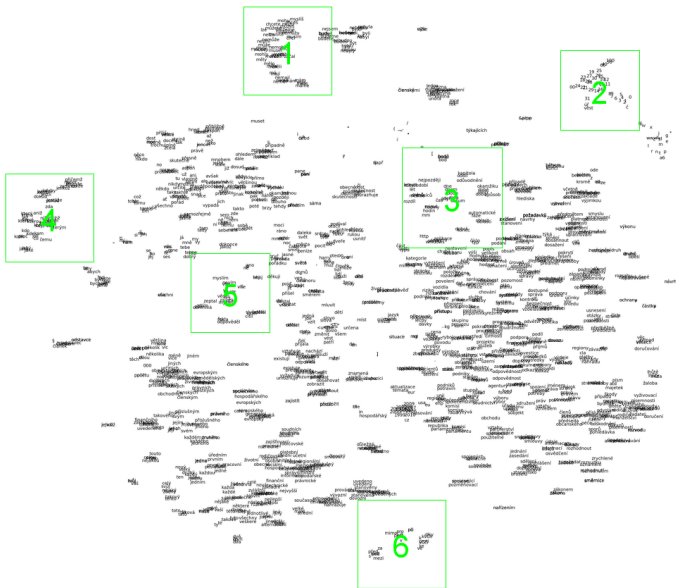
Word Embeddings in Sentiment



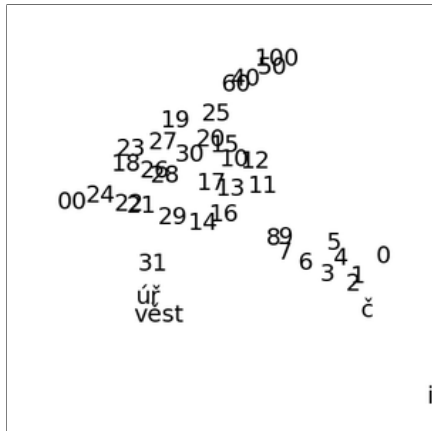
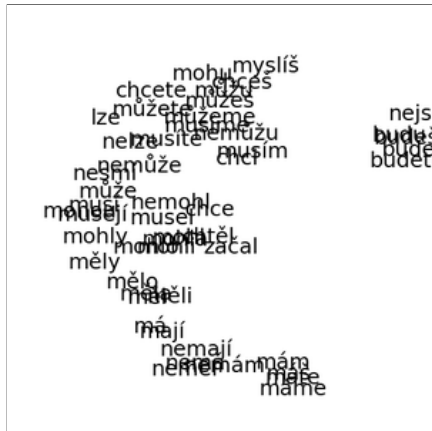
Word Embeddings in Sentiment



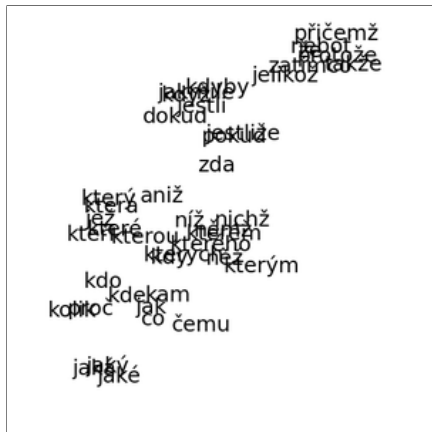
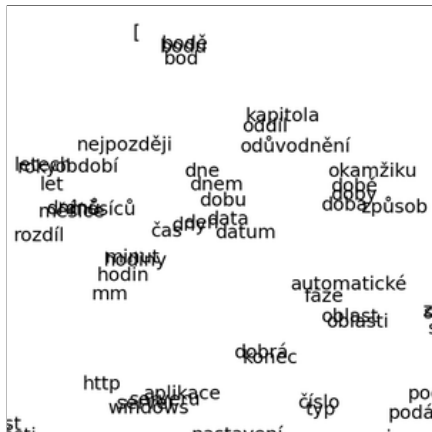
Vektorový prostor významů

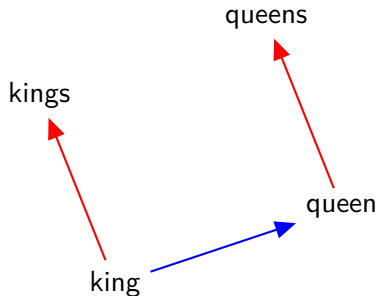
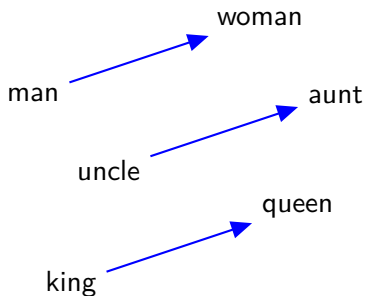


Vektorový prostor významů



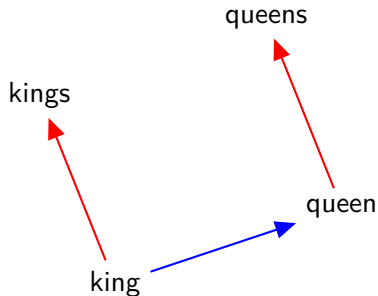
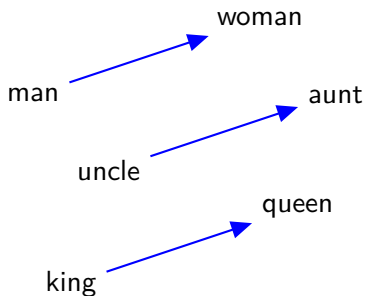
Vektorový prostor významů





Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. [Linguistic regularities in continuous space word representations.](#)

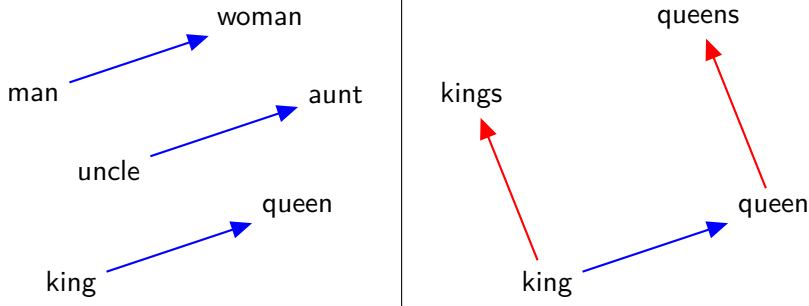
In *HLT-NAACL*, pages 746–751, 2013



$$V_{king} - V_{man} + V_{woman} \approx V_{queen}$$

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. [Linguistic regularities in continuous space word representations](#).

In *HLT-NAACL*, pages 746–751, 2013

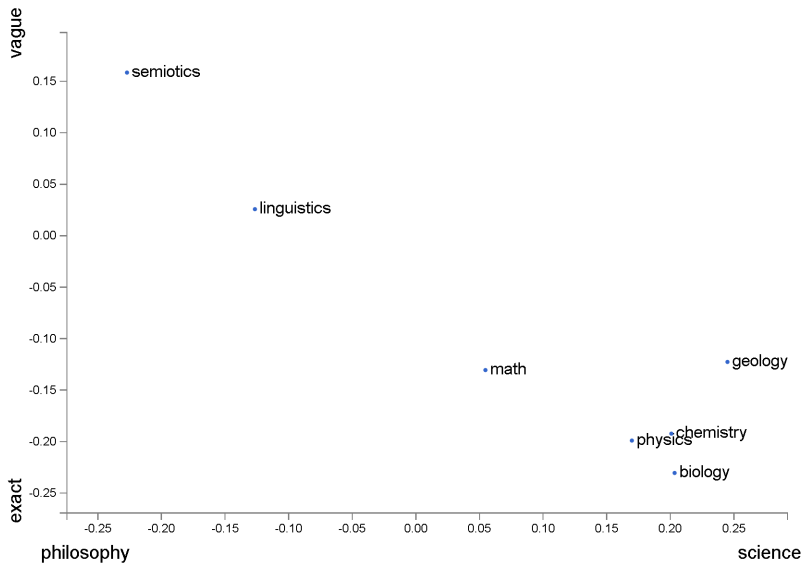


The model itself has no knowledge of syntax or morphology or semantics. Remarkably, training such a purely lexical model to maximize likelihood will induce word representations with striking syntactic and semantic properties.

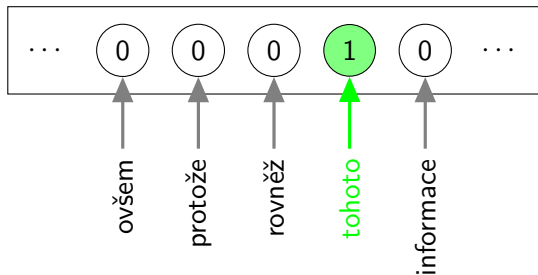
Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. [Linguistic regularities in continuous space word representations](#).

In *HLT-NAACL*, pages 746–751, 2013

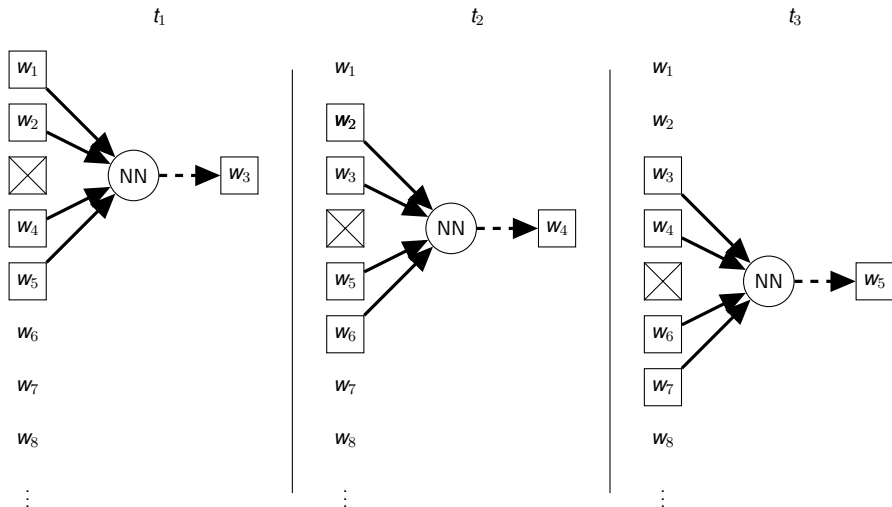
Vektorová aritmetika významů



Vektorové reprezentace slov

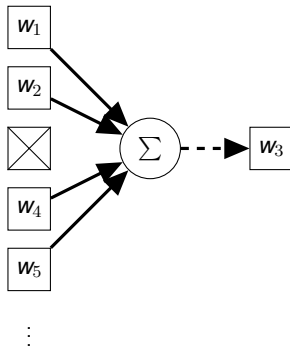


Jazykový model

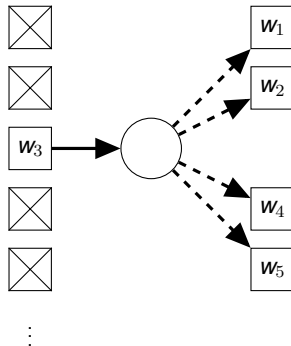


Učení neuronových jazykových modelů

CBOW



Skip-gram



Filosofie jazyka

Definice

Dva výrazy φ a ψ mají stejný pravdivostní potenciál právě tehdy, když kdykoliv jsou oba doplněny stejným výrazem tak, aby s ním utvořily větu, obě věty mají stejnou pravdivostní hodnotu.

Ernst Tugendhat: The meaning of 'Bedeutung' in Frege, 1970

$$V(\varphi) = V(\psi) \iff \forall x: P(\varphi x) = P(\psi x)$$

Pro velkou třídu případů použití slova „význam“ – i když ne pro všechny případy jeho použití – lze toto slovo vysvětlit takto: Význam nějakého slova je způsob jeho použití v řeči. [§43]

Jestliže se řekne, že určitá věta je beze smyslu, tak to neznamená, že její smysl je jakoby nesmyslný. Nýbrž určité slovní spojení se z řeči vylučuje, je staženo z provozu. [§500]

Říci „Toto slovní spojení nemá žádný smysl.“ vylučuje dané spojení ze sféry řeči a tím oblast řeči ohraničuje. [§499]

Definice

Dva výrazy φ a ψ mají stejný *pravdivostní potenciál* právě tehdy, když kdykoliv jsou oba doplněny stejným výrazem tak, *aby s ním utvořily větu*, obě věty mají *stejnou pravdivostní hodnotu*.

Definice

Dva výrazy φ a ψ mají stejnou *reprezentaci* právě tehdy, když kdykoliv jsou oba doplněny stejným výrazem, oba výsledné výrazy mají *stejné užití*.

Syntagms

(horizontal)

Paradigms

(vertical)

Their	tiger	stood	by	my	friend.
That	wolf	sat	over	your	tree.
The	dog	walked	on	the	floor.
My	cat	bounced	under	Nicole's	mat.
Your	cow	jumped	above	Anthony's	fridge.

[P]rvky struktury nemají ani vnější určení, ani vnitřní význam, [...] mají pouze smysl, který je nutně dán výlučně jejich „postavením“.

– G. Deleuze, How do we recognize structuralism?

Významy [...] neexistují před jazykem a nezávisle na jazyce, přicházejí na svět až skrze jazyk. Nejsou to předměty, které by výrazy jenom „onálepkovávaly“, jsou to hodnoty výrazů, které vyplývají z postavení těchto výrazů v systému jazyka a z jejich užitečnosti v rámci (použijeme-li populárního Wittgensteinova termínu) „jazykových her“.

– J. Peregrin, Význam a struktura

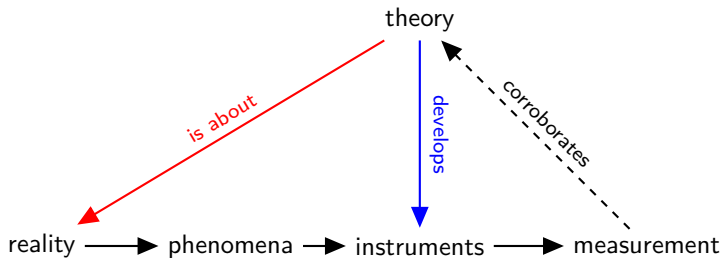
Vektorový prostor významů



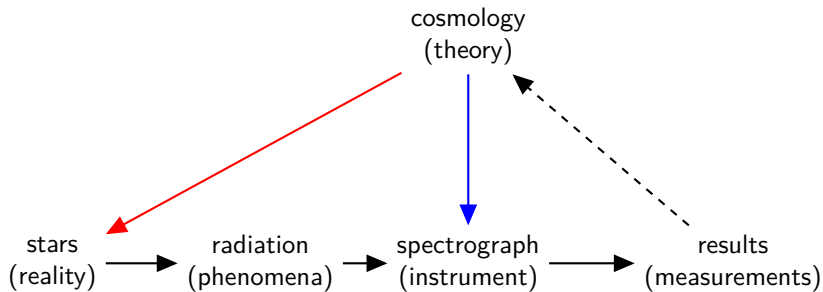
Jak dál?

nástroje $\leftarrow ? \rightarrow$ teorie

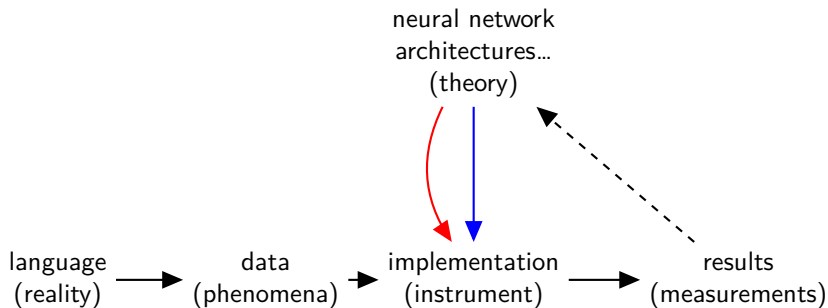
The Question of Methodology: Empirical Science



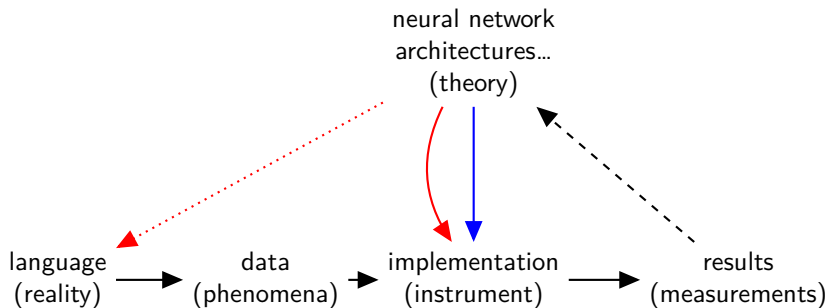
The Question of Methodology: Cosmology



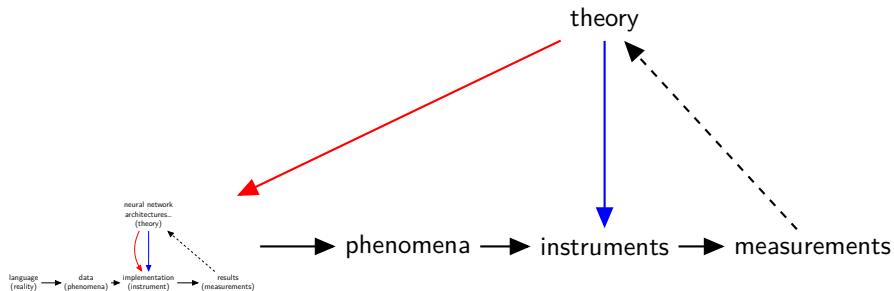
The Question of Methodology: NLP



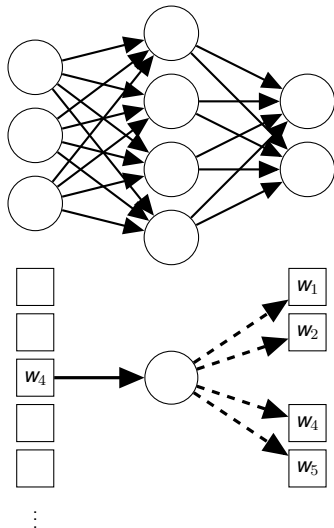
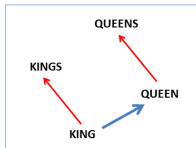
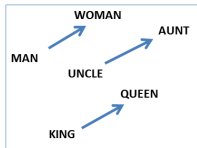
The Question of Methodology: NLP



The Question of Methodology: Solution



Děkuji za pozornost



Gilles Deleuze. *Podľa čoho rozpoznáme štrukturalizmus?* Herrmann a synové, 2010.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *HLT-NAACL*, pages 746–751, 2013.

Jaroslav Peregrin. *Význam a struktura*. OIKOYMENH, 1999.

Ernst Tugendhat. The meaning of ‘bedeutung’ in frege. *Analysis*, 30(6):177–189, 1970.

Ludwig Wittgenstein. *Filosofická zkoumání*. Filosofický ústav AV ČR, 1993.



Bob: i can i i everything else

Alice: balls have zero to me to me to me to me to me to me to me to me to me to

Bob: you i everything else

Alice: balls have a ball to me to me to me to me to me to me to me to me

Bob: i i can i i i everything else

Alice: balls have a ball to me to me to me to me to me to me to me to me

Bob: i

THE VERGE

TECH

SCIENCE

CULTURE

CARS

REVIEWS

LONGFORM

VIDEO

MORE



MICROSOFT WEB T.L.D.R

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day

By James Vincent | @jvincent | Mar 24, 2016, 6:43am EDT

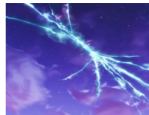


SHARE



It took less than 24 hours for Twitter to corrupt an innocent AI chatbot. Yesterday,

MOST READ



Fortnite players are mad after being kill during the epic rocket launch



Tesla keeps changing how it builds the I

Rasistický twitter bot

