

Nocoj... Gašaldi, 25-43

- ↳ DISTRIBUCIJSKA HIPOTEZA (danes)
- ↳ STRUKTURALISTIČNA HIPOTEZA (nastavljač)

↑ N 2010ih → NLP (natural lang. processing)

↳ velik in hiter porast

↳ pred tem neuspešni poskusi k neposrednemu programiranju jezikovnih pravil



↳ stroji se tako ne nauč.  
jezik!

mi se năimo drugače, a ne lahko naučimo  
tudi preko pravil: knjizeni / triji ...

→ od izvajanja pravil do učivanja masl. besede  
(če prej - v doih PC, ki jeenal pravilno uporab.  
nemike nepravilne glagole)

↳ predhodniki neuron. mrež (NN)

↳ v grobenem: PC se je sam naučil, a če  
daleč do GPT-ja



premalo podatkov in

premalo temognjivi računalniški!

Danes... VJM prekt. tehnič. delovanje (jezika)

↳ spremeni naš pogled na jezik...?

Word Embedding = umesčanje besed!



ni s říkami, glasovi



VJM pa s řetvikami - vektorji!

↓  
skupostavljamo jih besedam in  
jim pustimo, da se same učijo!

Zato rabimo

↓

veliko podatkov! To pomeni, da beseda „medved“

↓

je to samo beseda, brez pomena!

Pretko izpost.

↓

primerek! → Testiranje ... manjšanje napak!

↓

To koncu učenja so besede lahko  
predstavljene tudi kot več 100 dim prostori!  
(na nač. samo ena dim.)!

Analogija

iz frezro. analizo

pojav. črk → Potem pa gledamo bližino med besedami!  
v jekaku?? (spomni se: matkd. ni majb. preprosta v N-D!)

Morda ne: DH - Pomen besede določamo in bližine  
drugih besed!

↑

(lim.?)

Ne v enem primeru, ampak v n!

tudi soc.

kontekst!!?

+ co-occurrence(s) ( = skupno pojavljanje )

pogosteje kot so skupaj ⇒ semantično so si blizu (= sintaksi!)

↓

dokaz je v pudingu?! <= jekak poznano iz konteksta  
(če deluje, mora biti res...)



Gastaldi: „Hateri puding?“

(strukturalistični ali distribuciji?)

↓

Poznejši protiprimeri, podprtih na posamezniki!?