

Uocaj... Gestaldi, 25-43

↳ DISTRIBUCIJSKA HIPOTEZA (danes)

↳ STRUKTURALISTIČNA HIPOTEZA (naslednjic)

✓ 2010ih ↗ NLP (natural lang. processing)

↳ velik in hiter porast

↳ pred tem neuspešni poskusi & neposrednim programiranjem jezikovnih pravil



mi se učimo drugače, a se lahko naučimo tudi preko pravil: knjižni/tuji...

stroji se tako ne nauč. jezika!

↳ od izvajanja pravil do ugibanja nast. besede (še prej - v 80ih PC, ki je znal pravilno uporab. nekaj nepravilne glagole)

↳ predhodniki neuron. mreži (NM)

↳ v grobem: PC se je sam naučil, a še daleč do GPTja



premalo podatkov in
premalo zmogljivi računalniki!

Danes... VJM [?] preht. telu. delovanje (jezika)

↳ spremeni naš pogled na jezik...?

Word Embedding = umesčanje besed!



mi & črkami, glasovi

VJM pa & številkami - vektorji!

↓
Kpostavljam jih besedam in
jim pustimo, da se sami učijo!

Kato rabimo

veliko podatkov! To pomeni, da beseda "medved"
je to samo beseda, brez pomena!

Preko izpost.

primerov! → Testiranje ... manjšanje napak!

↓
Po koncu učenja so besede lahko
predstavljene tudi kot več 100 dim prostori!
(na kaj. samo ena dim.)!

Analogija

x frekv. analizo

pojv. črk → Potem pa gledamo bližino med besedami!
v jeziku?? (spomni se: razd. ni najb. preprosta v N-D!)

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

tudi soc.
kontekst!?

↑
Ne v enem primeru, ampak v n ! $(\lim_{n \rightarrow \infty} \frac{P}{n})$

↑
+ co-occurrence(?) (= skupno pojavljanje)
pogostejše kot so skupaj ⇒ semantično so si blizu (= sintaksni!)

↓
dokaz je v pudingu?! ⇐ jezik prenamo iz konteksta
(če deluje, mora biti res...)

↓
Gastaldi: "Kateri puding?"

(strukturalistični ali distribucijski?)

↓
Posamezni protiprimeri, poudarek na posamezniki!?

