

# Yet Another Formalism

Marek Grác  
NLP Centre, Faculty of Informatics  
Masaryk University, Brno

# Morfológia slovanských jazykov

- Veľmi flektívne jazyky
- Spoločný základ
- Rozdelenie
  - Tvarotvorba (konferencie, vidíme)
  - Slovo tvorba (bledomodrý, učiteľ)
- Vyriešený problém (?)

# Tradičné vzory

- 1 makroparadigma na slovný druh
- Základné (školské) vzory
- Formalizmus > prirodzený jazyk popisuje koncovky a alternácie
- Výnimky > zoznam výnimiek
- Vyžaduje > znalosť etymológie, významu slov, výslovnosti

# Koncovkové vzory

- Stovky vzorov pre češtinu/slovenčinu
- Jednoduchá transformácia na konečný automat/trie (ispell, ajka, ooffice)
- Formalizmus > pridávanie koncovky za nemennú časť
- Výnimky > vlastný vzor
- Vyžaduje >  $\emptyset$

# Morfologická databáza

- Obsahuje požadované údaje podľa formalizmu
- Opravovanie chýb
- Pridávanie nových slov
  - Ako určiť správny vzor ?
  - Výpis všetkých možných vzorov
  - Rozhodovací strom
  - Používanosť slovného tvaru

# Langusta

- Formalizmus na zápis tvarotvorby
- Grafémy (hlásky) x Znaky
- Podmienky na lemma (aj negatívne)
  - Končí na tvrdá hláska + -ý; posledná slabika v kmeni je krátka
- Niekoľko slovných základov
- Pripájanie predpôn a prípon

# Langusta

- Jednoznačné filtre
  - Platné pre celý jazyk: “t̃i” -> ti
  - Platné pre daný vzor: domáci -> domáci
- Grafémy
  - Zaradujeme do skupín
  - Vzťahy medzi nimi (černo<ch> -> černo<s>i)
- Rozdelenie slova na grafémy

# Langusta

- Jazykovo nezávislý formalizmus
  - Česky, slovensky, arabsky
- Možnosť metavzorov
  - Negácia (Nové lemma s rovnakým vzorom)
  - Stupňovanie (Nové lemma s iným vzorom)
  - Jednoduchá slovotvorba (slov. podst. mená)
- Nízky počet vzorov, zaujímavé dáta pre komparatívnu lingvistiku
- Vysoká efektivita zadávanie nových slov



Ďakujem, za pozornosť