

Analýza slovenských adjektív

Radovan Garabík, Agáta Karčová

Jazykovedný ústav Ľ. Štúra SAV, v. v. i.

Inšpirácie Ľudovíta Štúra a 21. storočie
16. celoslovenské stretnutie jazykovedcov
29. október 2025, Univerzitná knižnica v Bratislave

Prídavné mená v slovenčine

Štúdie a štatistiky na báze korpusov slovenčiny

Prídavné mená v slovenčine

- ▶ slovník slovenských adjektív
- ▶ dôraz na flexiu a gradáciu
- ▶ corpus driven
- ▶ príklady použitia
- ▶ Skloňovanie podstatných mien v slovenčine s korpusovými príkladmi, 2015
- ▶ *Časovanie slovies v slovenčine, 2022*
- ▶ bez príspevia LLM

Radovan Garabík, Agáta Karčová

Štúdie a štatistiky na báze korpusov slovenčiny



Prídavné mená v slovenčine

- ▶ Zdrojové dáta:
 - ▶ Morfológická databáza SNK – 114 636 lem, 27 897 adjektív, 6 958 participií
 - ▶ KSSJ (60 000 lexém, 13 658 – 57 adjektív), SSSJ, OGS
 - ▶ Araneum Slovacum VII Maximum (5 273 781 496 tokenov)
 - ▶ prim-10.0-public-sane (1 649 561 653 tokenov)
- ▶ participiá, adjektíva, substantivizované participiá (obe)
- ▶ adjektív 27 785, participií 2 554
- ▶ pomerne veľa informácií
- ▶ Retrográdny slovník súčasnej slovenčiny, 2018
- ▶ inkluzívny spôsob poskytovania informácií

Štruktúra slovníka

- ▶ Skloňovanie
 - ▶ Skloňovanie adjektív v pozitíve
 - ▶ Skloňovanie adjektív v komparatíve (a superlatíve)
- ▶ Stupňovanie
- ▶ Najčastejšie adjektíva – štatistické údaje a kolokácie
- ▶ Anomálie (s príkladmi)
- ▶ Pozitivita a negativita

Niektoré princípy

- ▶ formálny prístup k morfológii
- ▶ založený na ortografii
- ▶ osobitné paradigmy pre pozitív a komparatív (*d'alší* je pozitív)
- ▶ neberieme do úvahy: analytická gradácia, elatív, potenciálny komparatív, vokatív, sémantika
- ▶ berieme do úvahy: nesklonné adjektíva, niektoré „cudzie“ adjektíva, fosílné krátke tvary
- ▶ početnosť v korpuse je veľmi dôležitá, ale nie absolútne
- ▶ 4 rody (3 rody + životnosť), 2 čísla, 6 pádov, 3 stupne
- ▶ lema (lexikálna) je pozitív alebo komparatív, nominatív, singulár, „generický“ rod

Paradigmy

- ▶ skupiny identických Levenštejnových editovacích operácií
 - ▶ paradigma zahŕňa rody, čísla, pády
 - ▶ dvojtvary patria k paradigme
- ▶ superlatív je 100 % pravidelný (okrem tých, čo nemajú komparatív)
- ▶ počet spracovaných prídavných mien: 30 339
- ▶ počet paradigm (POS): 13
- ▶ počet paradigm (COMP): 4
- ▶ počet paradigm s gradáciou: 9 + 13
- ▶ ukážka

Najčastejšie adjektíva

- ▶ 546 najčastejších adjektív, usporiadané abecedne
- ▶ lema, rank, ipm/arfpm, pozitivita, fic/prf/inf/web
- ▶ kolokácie (logDice, 1. slovo napravo)
- ▶ ukážka

Anomálie

- ▶ anomália – počet výskytov hodnoty nejakej gramatickej kategórie je iný, než by sme očakávali
- ▶ počet hodnôt gramatickej kategórie – zložka vektora v n -rozmernom priestore
- ▶ vektor pre každé adjektívum
- ▶ vektor pre všetky adjektíva
- ▶ normalizujeme lineárnou transformáciou (diagonálna matica)

Anomálie – príklady

- ▶ $\vec{\hat{G}} = (\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$
- ▶ $\vec{\hat{G}} = T\vec{G}$
- ▶ $\vec{g}_w \rightarrow \vec{\hat{g}}_w = T\vec{g}_w$ pre každé adjektívum w
- ▶ uhol vektorov $\cos \angle(\vec{\hat{g}}_w, \vec{\hat{G}})$, čím menší, tým typickejšie adjektívum
- ▶ $\cos \angle = \frac{\vec{\hat{g}}_w \cdot \vec{\hat{G}}}{|\vec{\hat{g}}_w| \cdot |\vec{\hat{G}}|}$
- ▶ $\Delta_i = \vec{\hat{g}}_w^i - \vec{\hat{G}}^i$
- ▶ $\overline{\Delta}_i = \frac{\Delta_i}{1 - \frac{1}{\sqrt{n}}}$

Anomálie – kategórie

- ▶ rod: mifn
- ▶ číslo: sp
- ▶ pád: [NV]GDALI
- ▶ kombinácie: ms is fs ns mp [ifn]p
- ▶ ukážka

Negácia

- ▶ iba *ne*-
- ▶ pozitivita: $\mathcal{P}(w) = \mathcal{P}(\neg w) = \frac{|w| - |\neg w|}{|w| + |\neg w|}$
- ▶ ukážka

Ďakujem(e) za pozornosť