

## Slovenský národný korpus ako pomôcka pri výučbe slovenského jazyka

Mária Šimková

### 1. Čo je korpus?

Slovo *korpus* poznáme z bežného života vo všeobecnom význame „celok, podstatná časť, základ“, v konkrétnom význame „telo“ (z latinčiny) a Krátky slovník slovenského jazyka (2003) uvádza aj najnovší, odborný význam „súbor skúmaných prvkov“ s dokladom *textový korpus* – súbor textov v počítačovom spracovaní určený na vedecký výskum.

Vedecký výskum sa v lingvistike aj v minulosti realizoval na materiáli, ktorý predstavoval istý textový korpus uložený zvyčajne v kartotéke, ale takýto materiál si bádatelia zhromažďovali zväčša individuálne, ručným vypisovaním (excerpovaním) konkrétneho skúmaného javu na lístky a nebol všeobecne prístupný ani využiteľný na iné lingvistické výskumy. Ak sa niekto venoval napr. spájateľnosti prídavných mien, v jeho excerpácii sa nachádzali iba doklady na túto spájateľnosť a nie napr. na nové slová alebo na väzbu sloviess. Centrá zaoberajúce sa národným jazykom systematicky budovali a spravovali centrálnu kartotéku, ktoré mali širší záber a boli na požiadanie k dispozícii celej odbornej verejnosti. Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied v Bratislave disponuje všeobecnou kartotékou v rozsahu vyše 5 miliónov lístkov. Na tomto materiáli sa do polovice 90. rokov koncipovali súčasné výkladové slovníky a pravopisné príručky (napr. Slovník slovenského jazyka, 1959 – 1968, prvé vydanie Krátkeho slovníka slovenského jazyka, 1987, Synonymický slovník slovenčiny, 1995, Pravidlá slovenského pravopisu, 1991) a realizovali sa všetky väčšie gramatické výskumy. V súvisi so svojou výskumnou orientáciou JÚLŠ SAV vybudoval a spravuje aj špecifické historické a nárečové kartotéky. Všetky lístky v kartotékach sú zoradené abecedne podľa slova v záhlaví, ktoré sa v hlavnej časti lístka nachádza v určitom relevantnom kontexte. Nevyhnutnou súčasťou takéhoto zápisu sú aj stručné bibliografické údaje o zdroji textu uvádzané na okraji alebo v dolnej časti lístka.

Pri malej sonde do všeobecnej kartotéky začiatkom 90. rokov sa však veľmi jasne ukázalo, aké sú nevýhody tohto spôsobu získavania a ukladania jazykového materiálu. Okrem pracovného zhromažďovania a udržiavania lístočkov, ktoré boli ľahko stratiteľné a nenahraditeľné, a okrem ďalších nevýhod (škatule s lístkami zaberajú veľa priestoru, pri zaplnení treba lístky aj škatule neustále preskupovať a pod.) sa ako problematické ukázalo najmä určovanie slov, ktoré sa uvádzali v záhlaví a podľa ktorých sa lístky abecedne triedili. Zvyčajne išlo o niečím zaujímavé slová alebo spojenia, a tak jedno z najfrekvencovanejších

slovenských slov (zahŕňajúc všetky polyfunkčné využitia) **sa** sa v 5-miliónovej kartotéke nachádzalo osobitne dokladované iba 8-krát. Ak by chcel niekto na základe takejto kartotéky skúmať zvratnosť slovenských slovies a ich dynamiku, veľa by sa nedozvedel, resp. by musel prezerat' všetky škatule a hľadať podľa abecedy príslušné slovesá.

Začiatkom 60. rokov 20. storočia sa začali budovať prvé korpusy textov v elektronickej podobe. Najskôr pre anglický jazyk (Brown corpus, LOB corpus) a v nie veľkých rozsahoch (do 1 milióna jednotiek), postupne sa takto začali spracúvať aj iné jazyky a s rozvojom informačných technológií narastal aj objem textových dát. Posledných 20 rokov sa lingvistické výskumy preorientávajú na nové možnosti práce s rozsiahlymi textovými súbormi (rádovo stovky miliónov až miliardy textových jednotiek). Veľkosť korpusov, v ktorých sa aj stredne frekventované slová nachádzajú v desiatkach, ba stovkách tisícov výskumne takmer nezvládnuteľných výskytov, by spôsobovala opačný problém oproti predchádzajúcim malým množstvám dokladového materiálu, keby neexistovali výkonné vyhľadávacie, triediace a selektovacie počítačové nástroje, ktoré sú nevyhnutnou výbavou korpusov. Každý bádateľ či bežný používateľ korpusu už dnes disponuje tzv. treťou, počítačovou gramotnosťou, takže táto „nevýhoda“ korpusu je eliminovaná. Oproti lístkovým kartotékam má korpus okrem veľkosti ďalšie množstvo výhod. Napríklad:

- akékoľvek slovo sa dá nájsť v absolútnom výskyte (nie iba podľa viac-menej náhodne vybraných slov v záhlaví, a to iba podľa nich – slová v kontextoch sú v kartotéke prakticky nenájditeľné),
- každé hľadanie sa dá presne zopakovať, čím sa umožňuje verifikácia, ale aj doplnenie, pokračovanie predchádzajúceho výskumu kedykoľvek a kýmkoľvek,
- na tom istom korpuse sa môžu realizovať rôzne jazykové výskumy a jazykovedci nestrácajú čas ustavičným vypisovaním dokladového materiálu raz na taký, inokedy na iný skúmaný jav,
- rastúca veľkosť korpusu pri súčasnej štýlovo-žánrovej pestrosti zabezpečuje dostatok dokladového materiálu nielen pre bežné, frekventované slová, ale aj pre doteraz okrajové, menej poznané jazykové prostriedky, ktoré sa niekedy ukážu reálne frekventovanejšie ako tie, ktoré sa za také považovali teoreticky na báze menej rozsiahleho a menej pestrého materiálu, resp. na báze jazykovej introspekcie,
- prístupnosť korpusu prostredníctvom počítačovej siete internet dovoľuje skúmať zhromaždený materiál odkiaľkoľvek a kedykoľvek,

- špeciálne počítačové vybavenie umožňuje hľadať nielen slová, ale súčasne aj všetky ich tvary, umožňuje hľadať spojenia, zisťovať rôzne frekvencie, usporadúvať nájdený materiál podľa konkrétnych potrieb, vytvárať špecifické podkorpusy a pod.

## Ukážka č. 1

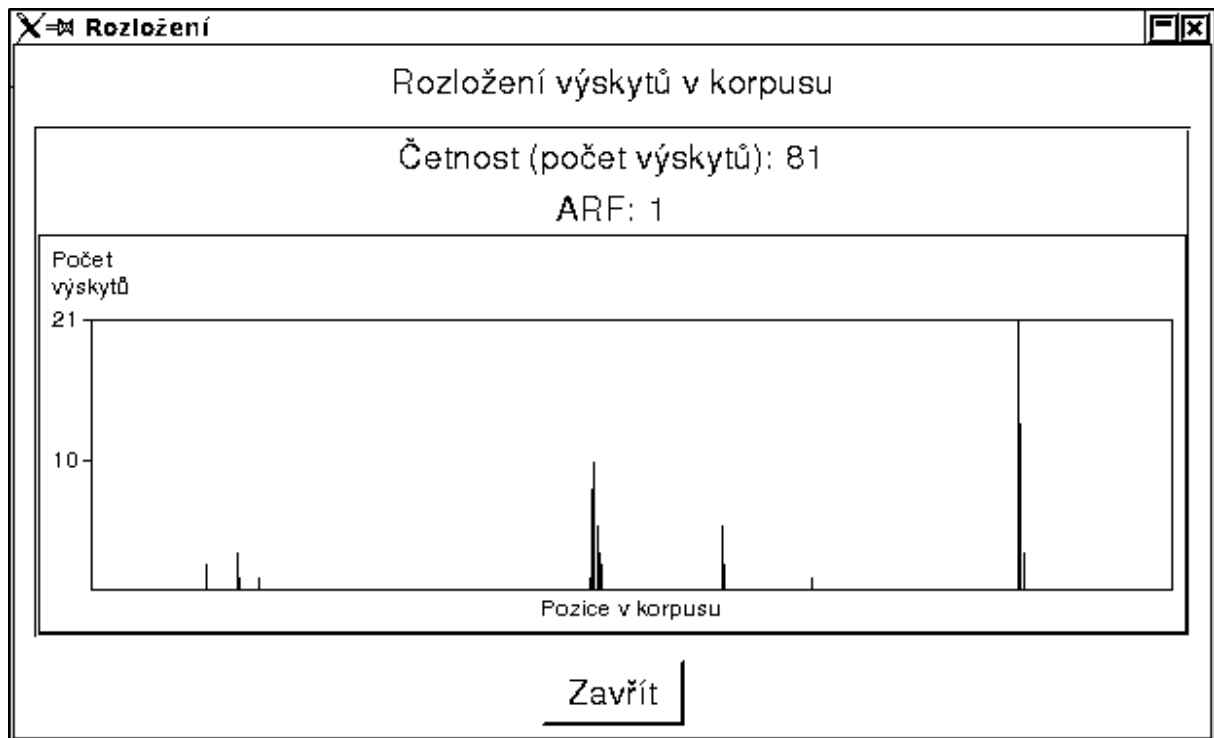
Výpis vybraných kontextov všetkých tvarov hľadaného slova *slovenčina* (konkordancia) z aktuálnej verzie Slovenského národného korpusu pomocou korpusového manažéra Manatee s klientom Bonito

The screenshot shows the Manatee software interface with the following details:

- Window Title:** X Bonito
- Menu Bar:** Manažer Korpus Dotaz Konkordance Zobrazení Výběr nápověda
- Search Bar:** Nový dotaz [ ] jmeno: [ ] prim-3.0-snk-all
- Main Content:** A concordance table with three columns:
  - Left Column:** Contextual text from various sources, including mentions of Bratislava, Peter Topol'sky, and linguistic discussions.
  - Middle Column:** The word 'slovenčina' in various grammatical forms (e.g., slovenčinu, slovenčine, slovenčiny, slovenčiny).
  - Right Column:** Additional contextual text, often starting with 'SLOVENČINY' and discussing language education and media.
- Status Bar:** Zobrazeno: 4970+50/8927 (0%) Řádek: 4970

## Ukážka č. 2

Rozloženie výskytu spojenia *vyučovanie slovenčiny* v celom korpuse (81 výskytov): prvé dva najsústredenejšie výskytov sú v novinách Slovákov z Nového Sadu (Hlas ľudu), posledný v zborníku Studia Academica Slovaca, ktorý obsahuje prednášky pre zahraničných účastníkov letného kurzu slovenčiny



Zhrňujúca odpoveď na úvodnú otázku „Čo je korpus?“ znie: textový korpus je rozsiahly vnútorne štruktúrovaný a ucelený súbor jazykových dát, ktoré sú elektronicky uložené a spracovávané, čo umožňuje všestranne a opakovane ich lingvisticky skúmať.

Korpusy majú rôzne obsahovo-formálne zloženie a zameranie. Najbežnejšie sú jednojazyčné korpusy písaných textov, ale budujú sa aj paralelné korpusy založené na rovnakých textoch z dvoch alebo viacerých jazykov, pričom jeden z nich môže, ale nemusí byť originálny. Stále viac pozornosti sa v súčasnosti venuje hovoreným korpusom, teda korpusom, v ktorých sa nachádzajú prepísané spontánne prehovory používateľov daného jazyka, keďže prevažná väčšina našej komunikácie sa odohráva ústne, no doterajšie výskumy sa zaoberali skoro výlučne písanou formou jazyka. Korpusy môžu byť všeobecné – pokrývajúce výberovo (reprezentatívne) slovnú zásobu konkrétneho jazyka podľa rôznych štýlov a žánrov, ale aj špecifické – zameriavajúce sa na tvorbu pre deti, na mládežnícky slang, na právnu terminológiu a pod. Z hľadiska času poznáme korpusy synchronne (súčasný) a diachronne (zhromažďujúce cenný, často inak veľmi ťažko dostupný a skúmateľný materiál z dávnejších vývinových fáz daného jazyka). Prvé korpusy obsahovali v podstate čisté textové dáta, ktoré slúžili najmä na prípravu slovníkov, dnes je štandardnou výbavou korpusov podrobná bibliografická a štýlovo-žánrová anotácia (pri každom slove sa dá ľahko zistiť, z akého zdroja pochádza), lematizácia (informácia o základnom tvare slova, podľa ktorého sa dajú vyhľadať všetky tvary), morfológická anotácia (informácia o morfológických kategóriách konkrétneho tvaru slova v kontexte) a ďalšie lingvistické informácie

podľa možností daného korpusu a potrieb jeho používateľov (o syntaktických funkciách, sémantike, štylistickom statuse, výslovnosti a pod.).

## 2. Slovenský národný korpus

Slovenský národný korpus (<http://korpus.juls.savba.sk>) sa po viacerých neúspešných pokusoch o získanie finančnej podpory v priebehu 90. rokov začal budovať až v r. 2002. Vtedy vzniklo oddelenie Slovenského národného korpusu Jazykovedného ústavu L. Štúra SAV v Bratislave, ktorého základnou úlohou je budovať všeobecný jednojazyčný korpus písaných textov súčasnej slovenčiny. Dolná hranica súčasného slovenského jazyka sa pre potreby korpusového výskumu stanovila na rok 1955 v súvisi s pravopisnou reformou v r. 1953 a istou časovou rezervou potrebnou na zmenu sádzacích strojov a pretlač všetkých textov. Texty sa do Slovenského národného korpusu (SNK) získavajú priamo od autorov alebo vlastníkov autorských či distribučných práv na základe Licenčnej zmluvy o inom (nekomerčnom) použití diela podľa autorského zákona. Staršie texty, ktoré neexistujú v elektronickej podobe, sa skenujú, rozpoznávajú a rekonštruujú tak, aby sa vytvorila elektronická verzia textovo identická s tlačenu. Novšie texty spravidla sú v elektronickej podobe, ale treba ich technicky „vyčistiť“ (odstraňujú sa netextové časti, teda obrázky, grafy, tabuľky a pod.) a previesť do jednotného formátu. Jedným z princípov budovania korpusov je spracúvanie reálnych textov, a to aj s preklepmi a inými chybami – ***korpus nie je kodifikačnou príručkou***, je „len“ materiálom na tvorbu príručiek a v tomto materiáli sa skúmajú aj varianty, niekedy aj najčastejšie chyby, čo môže spätne pomôcť pri ich lepšom rozpoznaní a odstraňovaní už vo fáze prípravy textu či pri výučbe.

V r. 2003 bola na internete prístupná prvá verzia všeobecného, primárneho korpusu slovenského jazyka v rozsahu 30 miliónov textových jednotiek (sú to nielen slová, ale aj interpunkcia, číslice a iné časti textu). Intenzívnou prácou na získavaní a spracúvaní textov sa korpus postupne rozrastal a aj kvalitatívne zlepšoval – vybavoval potrebnými informáciami a nástrojmi a po relevantných zmenách a doplneniach sa prakticky každý rok prístupnovala nová verzia. Začiatkom r. 2007 dosiahla súčasná verzia prim-3.0 hranicu 350 miliónov textových jednotiek s plnou bibliografickou a štýlovo-žánrovou anotáciou, s lematizáciou a automatizovanou morfológickou anotáciou. Vzhľadom na veľký rozsah korpusu bolo možné vytvoriť pre potreby špecifických výskumov osobitné podkorpusy obsahujúce len publicistické texty, len odborné texty, len umelecké texty a z umeleckých len originálne slovenské umelecké texty. K dispozícii je aj štýlovo vyvážený podkorpus, v ktorom má publicistika 60-percentné zastúpenie a texty z náučného a umeleckého štýlu sú zastúpené po 20 percentách z rozsahu podkorpusu.

Slovenský národný korpus je prostredníctvom siete internet verejne prístupný na vedecké, výskumné a učebné využitie. Dá sa v ňom vyhľadávať veľmi jednoducho pomocou www rozhrania priamo na stránke SNK. Tento spôsob nevyžaduje nijaké špeciálne registrovanie sa, treba len súhlasiť s podmienkami používania korpusu uvedenými na úvodnej stránke. Prostredníctvom www rozhrania je však prístupný iba jeden základný podkorpus a nie je možné zisťovať štatistické a frekvenčné údaje, spätelnosti a pod. Na to slúži druhý spôsob práce so SNK – vyhľadávanie pomocou korpusového manažéra Manatee s klientom Bonito. Zaujemca vyplní registračný formulár (<http://korpus.juls.savba.sk/registration/index.sk.html>), zašle ho na adresu SNK a obratom dostane vlastné heslo a prístup k všetkým verziám a podkorpusom SNK.

V ukážke č. 1 bol znázornený výstup z hľadania v korpuse pomocou tohto nástroja – konkordancia všetkých tvarov slova *slovenčina* (8927 výskytov vo verzii prim-3.0-snk-all). V pravom hornom okienku si treba v prvom rade nastaviť verziu, s ktorou chceme pracovať. Keď naň klikneme ľavým tlačidlom myši, zobrazí sa ponuka, z ktorej najčastejšie vyberáme najaktuálnejšiu verejne prístupnú verziu, v tomto prípade prim-3.0-public-all.<sup>1</sup> Do vyhľadávacieho okna sme zadali hľadať reťazec znakov *slovenčina*, v položke Korpus – Implicitný atribút sme aktivizovali hľadanie podľa lemy, teda základného tvaru. Vyhľadané kontexty môžeme dať zoradiť podľa ľavej alebo pravej strany (Konkordancia – Jednoduché triedenie), môžeme si pozrieť rozloženie v korpuse (podobne ako v ukážke č. 2) alebo zistiť, s ktorými slovami sa hľadané slovo spája najčastejšie.

### Ukážka č. 3

Zoznam slov vyskytujúcich sa v okolí slova *slovenčina* (2 zľava, 2 sprava) usporiadaný podľa absolútnej frekvencie (posledný stĺpec)

---

<sup>1</sup> Verejne prístupné podkorporusy sa od interných (juls, snk) líšia veľkosťou podľa typu Licenčnej zmluvy – nie všetci poskytovatelia súhlasia so sprístupnením svojich textov v korpusovom celku na internete. Rozdiel vo veľkosti však nie je veľký – pri posledných verziách predstavuje približne 10 mil. textových jednotiek, čo sú necelé tri percentá z rozsahu posledného najväčšieho korpusu.

Vypočítané najčítnejšie kolokace

word	Mf-score	T-score	Rel. f [%]	Abs. f
v	4.772	42.49	0.06852	1945
spisovnej	14.26	30.79	49.04	948
spisovnej	13.25	15.1	24.36	228
súčasnej	10.16	15.02	2.867	226
spisovná	14.18	9.38	46.56	88
češtiny	12.14	8.483	11.36	72
zásoby	9.589	8.415	1.932	71
slovník	10.25	8.239	3.055	68
spisovnú	13.4	7.68	27.19	59
slovnej	10.88	7.345	4.716	54
Spisovná	14.15	6.782	45.54	46
češtine	11.68	6.706	8.242	45
Synonymický	14.69	6.245	66.1	39
používateľov	10.96	5.828	5	34
zásobe	12.34	5.743	12.99	33
slovníku	10.06	5.652	2.68	32
spisovnou	13.41	5.099	27.37	26
spisovnú	12.84	5.098	18.44	26
češtinou	12.36	4.999	13.16	25
štúrovskej	12.07	4.999	10.78	25
češtiny	10.91	4.997	4.836	25
jazyková	10.35	4.792	3.281	23
Tézy	13.94	4.69	39.29	22
srbčiny	11.98	4.689	10.14	22
čeština	11.36	4.689	6.587	22
češtine	10.24	4.687	3.043	22
čeština	11.81	4.581	9.013	21
bernoľákovskej	13.77	4.472	35.09	20
štúrovskej	11.73	4.358	8.52	19
angličtina	9.866	4.354	2.34	19
lámanou	11.71	4.241	8.411	18
kodifikácie	11.49	4.122	7.234	17
profesorka	10.18	4.12	2.906	17
používateľ	10.15	3.997	3.509	16

Zavřít      Uložit

Osobitné možnosti využitia na hodinách slovenského jazyka ponúka ručne morfológicky anotovaný korpus (r-mak), ktorý v aktuálnej verzii 2.0 obsahuje vyše 500 tisíc textových jednotiek. Použité značky (anotačný zápis – tag) a hľadanie v ňom sú opísané na stránke SNK takisto ako všetky ostatné postupy (<http://korpus.juls.savba.sk/usage/morpho/>). V pravom hornom rohu vyhľadávacieho nástroja Bonito vyberieme z ponuky podkorpusov a verzií r-mak-2.0 a v položke Korpus nastavíme Implicitný atribút na Tag. Vo vyhľadávacom okienku potom môžeme zadať morfológickú značku alebo jej časť a vyhľadať napr. všetky častice, všetky vokatívny alebo inú konkrétnu skupinu slov či výrazov.

#### Ukážka č. 4

Príklad hľadania podstatných mien, ktoré sa skloňujú ako prídavné mená: hľadaná značka *SA...* je zložená z reťazca znakov S – substantívum, A – adjektívna paradigma a troch bodiek, ktoré znamenajú, že táto značka sa štandardne skladá z 5 znakov, ale my presne vymedzujeme iba prvé dva, ostatné tri – rod, číslo, pád – môžu byť ľubovoľné (bodka nahrádza akýkoľvek znak v značke)

Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

Nový dotaz  jméno:  r-mak-2.0

hercami , ktorí sa v seriále objavili . Vynikajúci večer končí tesne pred desiatou . len čo sa podarilo Kuricovi prehovoriť terénnych nerovností . Ocení ju aj „ naopak , veľmi dobrý a múdro konajúci Hlupákom sa stane len vtedy , ak robí ich zberom , tak na pravej strane by bola dedina . Jeden deň si vypočujem svojho hľadisku . A to najmä v súvislosti s cenou zážitku mohol iba snívať . Moje nízke o šport poteší futbalový turnaj pre pred vynájdением kopírky ) a ak niektorí tak nezostanú visieť na krku v podobe kolko tu stojí 1 - izbový byt ? Mesačné naďabiť na sčítaného kamaráta či stopách ideme , pretože patrime k svojim vypnutým vysávačom . Viem to od svojho som sa dnes s príhodou podelil so svojim bol ošetrojúci lekár . Zomierať bez imavým pohľadom prezerat' svojich budúcich	<b>Vstupné Záverečná prítomných princezná vedúci vedúceho premenná známeho vstupného vreckové dospelých účinkujúci dospelých nájomné príbuzného príbuzným známeho známym blízkych spolucestujúcich</b>	je 25 dolárov . Jednu z epizód napísal ! S údivom pozeráme na hospodskú . " , aby sa zhromaždili na ploche pred pódium na hrášku " . Karimatky Sú ľahké zelvocou , sústružník , či bača . a nemá riadiace schopnosti , ak obsluhuje s dost' nízkym číselným vyjadrením , ako sa mu extrémne darí v práci , . Priznám sa , lístky mi nepripadali ma vtedy dokonca prinútilo , aby som sa a jazda na koni pre deti . Uvidieť sa mali isté didaktické sklony , podarilo , ktorí sa nedokážu zamestnať a nedokážu garsónky ? Lístok do kina ? Alebo presso . Takže ak sa vám zdá , že v celom , do rodín , do ktorých sme sa narodili , ktorý sedel pri počítači nad nedorobeným , ten okamžite zareagoval , že to má je asi zážitok , ktorý človek tak ľahko , či náhodou v niekom z nich neodhalim
--	--	--

Počet výskytů: 338  
> Query : "SA..."

Zobrazeno: 1+50/338 (14%) Řádek: 33 Vybráno: 1

Súčasťou informácií na stránke SNK sú aj Štatistiky korpusu, kde možno nájsť napr. zoznam bibliografií všetkých textov spracovaných v SNK, podiel textov v korpuse podľa štýlov, žánrov, vecných oblastí (domén), pôvodného jazyka, ako aj zoznamy najfrekvencovanejších slov ( tvarov) a lem a dĺžky textov podľa počtu viet či textových jednotiek.

### 3. Špecifické časti Slovenského národného korpusu

Okrem veľkého základného, primárneho korpusu a ručne morfológicky anotovaného korpusu sa na stránke SNK nachádzajú paralelné korpusy (rusko-slovenský, francúzsko-slovenský a paralelný korpus počítačových termínov) a lingvistické zdroje. Obidve zložky sa budú ďalej rozširovať, napr. paralelné korpusy v blízkej budúcnosti o česko-slovenský a chorvátsko-slovenský paralelný korpus.

Lingvistické zdroje sú najnavštevovanejšou zložkou SNK. Dá sa k nim dostať nielen cez stránku SNK, ale aj cez stránku Jazykovedného ústavu E. Štúra SAV (<http://www.juls.savba.sk/>). Táto zložka sprístupňuje on-line vybranú produkciu JÚLŠ SAV: Krátky slovník slovenského jazyka (2003), Slovník slovenského jazyka (1959 – 1968), Názvy



obcí Slovenskej republiky, viaceré monografie, zborníky a časopisy. Naposledy pribudla Databáza vlastných mien a názvov lokalít na Slovensku, chystajú sa Pravidlá slovenského pravopisu, Synonymický slovník slovenčiny a ďalšie diela. Využitie týchto zdrojov na hodinách slovenského jazyka bude závisieť už len od dostupnosti počítačovej učebne a tvorivosti vyučujúceho.

#### 4. Možnosti využitia Slovenského národného korpusu pri výučbe slovenského jazyka

Priama práca s materiálom korpusu sa dá využiť v rôznych častiach vyučovacieho predmetu počnúc pravopisom a významom slov cez tvaroslovie a slovtvorbu až po syntax a štylistiku, ako aj v rôznych častiach hodiny – pri výklade, samostatnej práci, precvičovaní či skúšaní. Možno pracovať samostatne s korpusom alebo sa práca s korpusom môže kombinovať aj porovnávaním so slovníkmi či internetom. V korpuse, ako sme uviedli vyššie, sa nachádzajú reálne texty, preto sa niektoré prvky v ňom zhromaždeného jazykového materiálu prirodzene odlišujú od toho, ako ich spracúvajú alebo odporúčajú normatívne príručky. No doterajšie výskumy súčasne ukazujú, že jazykové jednotky a javy v textoch korpusu sú bližšie ku kodifikovaným podobám ako tie isté na internete (Google, Morfeo), pretože väčšina textov v korpuse pochádza z oficiálnych vydavateľstiev, ktoré investujú aj do korektorskej práce, zatiaľ čo internetová produkcia je podstatne voľnejšia a bez jazykových korektúr. Už tieto rozdiely sa dajú pri výučbe dobre využiť.

Na nasledujúcich stranách uvádzame viacero ukážok využitia korpusu pri výučbe slovenčiny, ktoré sme vypracovali pre tento príspevok alebo čerpali zo zdrojov nachádzajúcich sa v zozname literatúry.

##### Ukážka č. 5

Pomocou regulárneho výrazu sme dali naraz vyhľadať tvary slov *vír* a *výr* (hľadaná lema *v[ýi]r*). Na základe nájdených kontextov si môžeme objasniť význam/významy oboch slov a vysvetliť, prečo sa nedá napísať formulácia *Oba v..ry naraz zmizli*. Môžeme povedať *Oba naraz zmizli*, tak ako pri akejkoľvek inej dvojici nerovnakých podstatných mien, no použiť na ne jedno slovo je nezmyselné (nehovoríme ani o hruško-jablku či okno-dverách a pod.).

Zároveň si možno všimnúť, že v korpuse sa nachádzajú aj cudzojazyčné texty (v nevelkom rozsahu, ale predsa) a na českom kontexte slova *víra* (viera) si môžeme ozrejmiť homonymiu formy v širšom význame.

Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

Nový dotaz  jméno:  prim-3.0-public-sane

<p>predstavuje mimoriadne intenzívny vzdušný vokalistkou Ali Augerovou , ktorí PKO vťahli do nastupujúcich zmien vo svete . Musel sa prispôbiť vytvárajú podmienky pre čoraz rýchlejší od FNM všetci členovia vlády . . . Vo turbulencií . V jednom scenári je to vzdušný možno teraz počuť najväčšiu z nich - mňa slovenský Dumas . Pozerali na mňa ako Lokomotívou ( 22 . februára ) . . . Vo pohár sánkarov mladým Slovákom z Tatier Vo uzavrel Stantien . Dnes už je Roman opäť vo sa aspoň na chvíľu cítila človekom vo však autorke dovoľuje , aby ju námetový svadobné lety a príprava hniezdných jamiek množstvo bizarných postáv a dostáva sa do</p>	<p><b>vír</b> vznikajúci v súvislosti s búrkovým mrakom  <b>víru</b> 70 . ( džez ) rokov . Vyvrcholenie prišlo  <b>víru</b> udalostí , dokonca sa postavil na ich čelo  <b>vír</b> bezohľadnej spotreby strhujúcej do záhuby  <b>víre</b> intríg gaulideriády si poslanci stihli  <b>vír</b> , ktorý sa vytvoril v stope lietadla ,  <b>výra</b> skalného , sovu lesnú , myšiarku ušatú  <b>výr</b> a nechápali to , " spomína na rokovania  <b>víre</b> športového diania sa stratil veľmi cenný  <b>víre</b> športového diania sa stratil veľmi cenný  <b>víre</b> českej extraligy , v utorok priviezol Vsetín  <b>víre</b> šantiacich masiek z úzkych uličkách  <b>vír</b> nestrhol do priepasti zvanej maniera , ale  <b>výra</b> skalného , ktorého v blízkom kameňolome  <b>víru</b> udalostí , z ktorých márne hľadá východisko</p>
<p>až natoľko , že tento svědek osvícené vývoju nových liekov , ktoré zabránia tímy prišli k záveru , že preniknutie Trnavské divadlo 9 . 00 a 11 . 00 h : Vo 20 . 05 Plná poľná a hudba 21 . 00 Vo</p>	<p><b>víry</b> , odvážny až k mučeníctví - ako ho  <b>víru</b> HIV v prenikaní do bunky . Vedci ale varujú  <b>víru</b> HIV do zdravých buniek v organizme umožňujú  <b>víre</b> tanca a baletu , Baletné štúdio D . Nebylu  <b>víre</b> sveta , v prúde času 22 . 15 H . Zelinová</p>

Počet výskytů: 2545  
 > Query : "v[ýi]r"

Zobrazeno: 201+50/2545 (1%) Řádek: 216 Vybráno: 1

## Ukážka č. 6

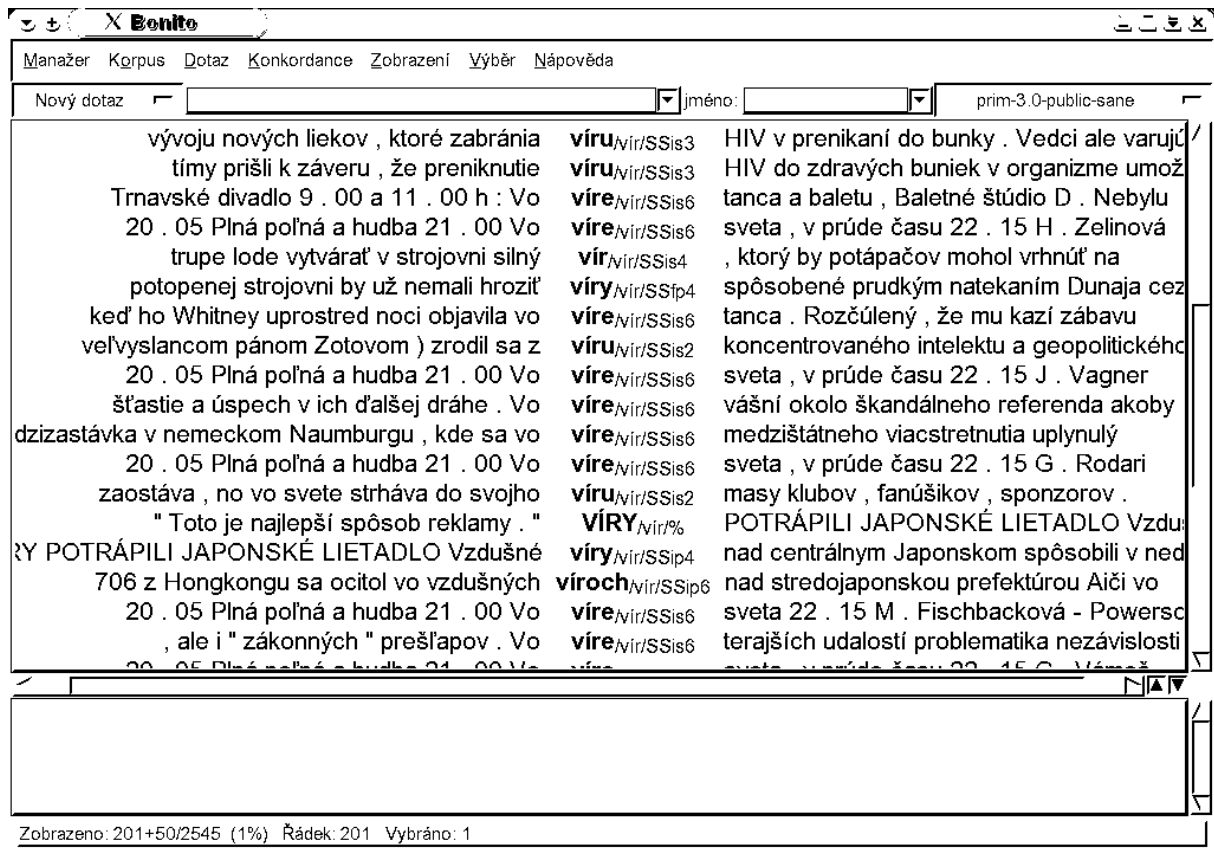
Posunutím predchádzajúcej konkordancie o niekoľko kontextov sa v prvých dvoch riadkoch hore ocitli nesprávne tvary slova *vírus*. Zapnutím funkcie Zobrazenie a Atribút pre lemu a tag hľadaného slova (kľúčové slovo – KWIC), ako to znázorňuje prvý obrázok, sa môžeme pozrieť na základné tvary a morfológické kategórie určené automatizovane nástrojom na morfológickú anotáciu. Na druhom obrázku vidíme, že tvaru *víru* bol nesprávne priradený základný tvar *vír*, preto sme ho podľa lemy dostali vypísaný spolu s hľadanými lemy *výr* a *vír*. Keď sa pozrieme bližšie na počítačom určené pády (číslica na konci značky), zistíme, že nástroj ich nie vždy priradil správne, v jednom prípade dokonca určil aj nesprávny rod, čo môžeme využiť na precvičenie týchto kategórií žiakmi. Opravovanie (hlúpeho) počítača môže mať úspech.

Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

Nový dotaz [ ] jméno: [ ] prim-3.0-public-sane

<p>vývoju nových liekov , ktoré zabránia          tímy prišli k záveru , že preniknutie          Trnavské divadlo 9 . 00 a 11 . 00 h : Vo          [ X Atributy ]          Zobrazit atributy</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> lemma</li> <li><input type="checkbox"/> tag</li> <li><input type="checkbox"/> hlemma</li> <li><input type="checkbox"/> htag</li> <li><input type="checkbox"/> Pro všechny pozice</li> <li><input checked="" type="checkbox"/> Pouze v KWIC</li> </ul> <p>Budiž</p>	<p>Trnavské divadlo 9 . 00 a 11 . 00 h : Vo          a hudba 21 . 00 Vo          várať v strojovni silný          ni by už nemali hroziť          stred noci objavila vo          Zotovom ) zrodil sa z          a hudba 21 . 00 Vo          ich ďalšej dráhe . Vo          aumburgu , kde sa vo          a hudba 21 . 00 Vo          ete strháva do svojho          ší spôsob reklamy . "          LIETADLO Vzdušné          706 z Hongkongu sa ocitol vo vzdušných          20 . 05 Plná poľná a hudba 21 . 00 Vo          , ale i " zákonných " prešľapov . Vo          20 . 05 Plná poľná a hudba 21 . 00 Vo          20 . 05 Plná poľná a hudba 21 . 00 Vo</p>	<p>víru HIV v prenikaní do bunky . Vedci ale varujú          víru HIV do zdravých buniek v organizme umožňuje          víre tanca a baletu , Baletné štúdio D .          víre sveta , v prúde času 22 . 15 H . Zelinová          vír , ktorý by potápačov mohol vrhnúť          víry spôsobené prudkým natekaním Dunaja          víre tanca . Rozčúlený , že mu kazí zábavu          víru koncentrovaného intelektu a geopolitického          víre sveta , v prúde času 22 . 15 J . Vagner          víre vášni okolo škandálneho referenda          víre medzištátneho viacstretnutia uplynulý          víre sveta , v prúde času 22 . 15 G . Rodari          víru masy klubov , fanúšikov , sponzorov .          VÍRY POTRÁPILI JAPONSKÉ LIETADLO Vzdušné          víry nad centrálnym Japonskom spôsobili v          víroch nad stredojaponskou prefektúrou Aiči          víre sveta 22 . 15 M . Fischbacková - Powersová          víre terajších udalostí problematika nezávislosti          víre sveta , v prúde času 22 . 15 G . Vámoš          víre sveta , v prúde času 22 . 15 E . Charles</p>
---	---	--

Zobrazeno: 201+50/2545 (1%) Řádek: 218 Vybráno: 1



Ukážka č. 7

Doplňte chýbajúce slovo:

na ktorej zvyčajne pohráva priateľský ....., z jeho úst vždy vychádza milé a bezprostredné

odhodlane . Túto rozhodnosť zmäkčuje ..... a neutajený výraz radosti , ktorá akiste

ju viacej s meričkami v rukách . I jej ..... sa mi zahmlil ( s . 98 ) . A napokon ,

detskom časopise som čítal vetu Ujov ..... dievčatkovou zásluhou nestačil prejsť

Napr . zdvorilé správanie , zdvorilý ..... , ušľachtilé mravy , ušľachtilá zábava

tieto príklady z antického sveta vyvolajú ..... aj počudovanie , prečo sa na konci 2

rôzne výrazy , ktoré neraz vyvolávajú i ..... na tvári . . . Napríklad 7 . decembra

úsmevom sa naozaj vyjadruje kývnutie hlavy , ..... ako prostriedok lúčenia .

V spojení s akými verbami a adjektívami najčastejšie nájdeme slovo vyhľadané v predchádzajúcom cvičení?

Ukážka č. 8

Čo musím urobiť?

a) Vypíš verbá v infinitíve a potom utvor imperatív:

1. "...pre nich bola prínosom a podobné reči. Povedal mi, aby som prišla na fotografické skúšky."
2. "...otvoril dvere," uviedla E. Slavkovská. "Povedal mi, aby som si sadla na posteľ a čakala na skúšku..."
3. "...ju dal a aby som si nahral z nej kópiu. Povedal mi, aby som nahral aj druhú kazetu pre prípad..."
4. "...a odstúpil z orgánu na kontrolu SIS. Povedal mi, aby som si sám určil požiadavky za vrátenie..."
5. "...reagoval na prvé výtvarné výsledky? Povedal mi, aby som pokračoval."
6. "Trocha ma vyviedol z miery rozhodca. Povedal mi, aby som si ju postavil presne na biely bod..."

1. musím .....	imperatív: 2. os. sg. ....	2. os. pl. ....
2. musím .....	imperatív: 2. os. sg. ....	2. os. pl. ....
3. musím .....	imperatív: 2. os. sg. ....	2. os. pl. ....
4. musím .....	imperatív: 2. os. sg. ....	2. os. pl. ....
5. musím .....	imperatív: 2. os. sg. ....	2. os. pl. ....
6. musím .....	imperatív: 2. os. sg. ....	2. os. pl. ....

b) Vytvor vety s priamou rečou, obsahujúce tieto imperatívy:

Ukážka č. 9

S ktorým pádom sa spája predložka "na"?

je zrejmé to , čo poukazuje na obsah , označujúca	<b>na</b>	myšlienkovú náplň , kým
je to , čím , akými prostriedkami sa Označujúcu zložku	<b>na</b>	tento obsah odkazuje .
je najdôležitejšie , býva zvyčajne jednotlivými	<b>na</b>	konci výpovede ) sa vzťahy medzi
tomto príspevku by sme chceli poukázať dôležité	<b>na</b>	to , že v architektónike textu majú
zvýrazniť niektorý prvok . Konkrétne nám ide prípadne pomlčiek , ale aj úvodzoviek , písma	<b>na</b>	jednej strane o využitie zátvoriek , <b>na</b> druhej strane rozličných typov
vývinu nerealizované ) aj nároky človeka sa	<b>na</b>	slobodný . . . rozvoj . Zdôraznenie
výpovede — majú takéto vety v zátvorkách	<b>na</b>	konci súvetia či dokonca odseku :
Tí		

Ukážka č. 10

Určte slovný druh slov "okolo" a "mimo" v týchto kontextoch:

"...z kopy na ľavu, ktorá práve bežala **okolo**. Melóny sa rozkotúľali všetkými smer..."

"...mŕtva lopta , držaná lopta , lopta " **mimo** " , volejbalová lopta , mohlo byť aj ľahaná..."

"...možné iba takéto väzby : havária sa stala **mimo** našich hraníc ; lopta je už mimo hracej..."

“...povedal . Abu schytil závoj a omotal si ho **okolo** krku . “ Naozaj parádny , Abu ! “...”  
 “...hlasom , no bála sa , čo uvidí . Obehla **okolo** krika a prekvapene zhikla . Radža s...”  
 “...stala mimo našich hraníc ; lopta je už **mimo** hracej plochy ; táto vec je mimo diskusie...”  
 “...vyvolával Aladin na ľudí , ktorí sa náhlili **okolo** neho . “ Dve vešby za jednu cenu !...”  
 “...ostávala aj počas našich rozhovorov niekde **mimo** . A ešte mi to pripomína niečo , čo som...”  
 “...mierne oteplí na 22 až 26 , na severe **okolo** 20 stupňov Celzia . KEDY NA DOVOLENKU...”  
 “... čelom . Navyše sa snažím z toho , čo sa **okolo** mňa robilo , poučiť . Našťastie som...”  
 “... oslobodzujúcich rozsudkov , zatiaľ čo vo svete je to **okolo** 20 percent . ( čtk ) Disney ° s Rozprávky...”  
 “...dňa na deň . Nová generácia je trochu **mimo** . Nezaujímajú sa o politiku , nechce hovoriť...”

## Ukážka č. 11

Po niektorých viacslovných predložkách môžu nasledovať nielen podstatné mená v konkrétnom páde, ale aj slovesá v neurčitku. Pomocou zloženého výrazu [word="s"][word="cieľom"][tag="VI.\*"] sme dali vyhľadať spojenia predložky *s cieľom* s infinitívom. Vo vyše šesťtisícovej konkordancii môžeme zisťovať, ktoré slovesá sa s touto predložkou spájajú najčastejšie. Alebo môžeme použiť funkciu Konkordancia – Štatistiky – Frekvenčná distribúcia a dať si zobrazit', ktoré slová sa nachádzajú najčastejšie napravo od výrazu *s cieľom*. Na ďalšom obrázku vidíme, že sú to najmä slovesá *získať*, *dosiagnuť*, *vytvoriť*, *zabezpečiť*.

The screenshot shows a search interface with the following elements:

- Search bar: "Nový dotaz" with a dropdown menu.
- Search criteria: "jméno:" and "prim-3.0-public-sane".
- Search results table:
 

tátov kompletný prapor rakúskej armády upravujú Zmluvu o tichom spoločenstve ísť o skutočne globálnu iniciatívu ecifických výrob vo vyspelých krajinách íla sa do rozsiahlej sabotáže ekonomiky využívali na odhalenie takýchto miest tnávateľom zmierňoval sociálne výdavky rozličné mená , ktoré boli zneužitie socializačné programy pre deti a mládež kom zdržanlivosti od sexuálneho vzťahu IX . ( 1049 - 1054 ) presadzoval celibát to iba vtedy , keby sa obchod uskutočnil . Verme , že mechanizmy podpory únie egulatívy pre plánovanie hypermarketov vaný variant - prevrat podporovaný CIA všetkého bol napísaný na objednávku sa preniká do informačných systémov a na objednávku vykonávajú činnosť ktorá vytvára a udržiava vzťahy s vládou na zásade národnostného príbuzenstva	<b>s cieľom odhaliť</b> <b>s cieľom navrhnúť</b> <b>s cieľom odstrániť</b> <b>s cieľom vyvinúť</b> <b>s cieľom vyvolať</b> <b>s cieľom vyhnúť</b> <b>s cieľom transformovať</b> <b>s cieľom dosiahnuť</b> <b>s cieľom pripraviť</b> <b>s cieľom udržať</b> <b>s cieľom odstrániť</b> <b>s cieľom zmanipulovať</b> <b>s cieľom harmonizovať</b> <b>s cieľom predísť</b> <b>s cieľom zosadiť</b> <b>s cieľom zmeniť</b> <b>s cieľom zničiť</b> <b>s cieľom ovplyvniť</b> <b>s cieľom ovplyvniť</b> <b>s cieľom chrániť</b>	a zajať čo najviac členov hliadok . zamedzenie zneužívania tohto inštitútu chudobu vo svete a zaručiť trvalo udržateľ nové špecifické výrobky , čo sú väčšinou čo najväčší chaos a neporiadok , čo sa im pri výstavbe ľudských obydlií dočasné pracovné miesta na zmluvy s víťazstvo na náboženskom , politickom ich začlenenie do občianskej spoločnosti si energiu pred bojom alebo športom , nemorálnosti v cirkvi , hoci k nej v skutoč kapitálový trh . Bolo od začiatku nad celý jej hospodársky priestor zohľadnia predimenzovaniu predajných kapacít a Saddáma zvnútra ? Stal sa iba zlým vtípor rozhodnutie výberovej komisie vedenej , zneužiť alebo ukradnúť informácie legislatívny proces . Museli by sa zaregis legislatívu a reguláciu . “ Napriek národnostné menšiny žijúce mimo jurisdik
--	---	--
- Statistics: "Počet výskytů: 6009" and "> Query : [word="s"][word="cieľom"][tag="VI.\*"]".
- Footer: "Zobrazeno: 1+50/6009 (0%) Řádek: 1".

Výsledek frekvenční distribuce

Limit: 0 Řádků k zobrazení: 1746

word: 1KWIC> zobrazit součet schovat

word:	##
získat'	322
dosiahnuť	249
vytvoriť	168
zabezpečiť	141
,	137
vyhrať	131
zvýšiť	121
zabrániť	96
podporiť	81
nájsť	79
pomôcť	79
znižiť	78
zistiť	75
zvítaziť	59
zlepšiť	57
udržať	57
v	55
"	49
ich	48
neprehrať	47
čo	43
zastaviť	42
nosilniť	41

Zavřít Uložit

Počet výskytů: 7849  
> Query : "s" "cieľom"

Zobrazeno: 1+50/7849 (0%) Řádek: 1

prim-3.0-public-sane

cieľom odovzdávania informácií o Slovensku a  
 cieľom odhaliť a zajať čo najviac členov hliadok  
 cieľom dobrovoľníckej práce pre dobro komunity  
 cieľom navrhnúť zamedzenie zneužívania tohto  
 cieľom odstrániť chudobu vo svete a zaručiť  
 cieľom niečo dosiahnuť . Pri dosahovaní cieľa  
 cieľom vyvinúť nové špecifické výrobky ,  
 cieľom vyvolať čo najväčší chaos a neporiadok  
 cieľom vyhnúť sa im pri výstavbe ľudských  
 cieľom transformovať dočasné pracovné miesta  
 cieľom dosiahnuť víťazstvo na náboženskom  
 cieľom pripraviť ich začlenenie do občianskej  
 cieľom udržať si energiu pred bojom alebo športom  
 cieľom odstrániť nemorálnosti v cirkvi , hoci  
 cieľom zmanipulovať kapitálový trh . Bolo od  
 cieľom harmonizovať celý jej hospodársky priestor  
 cieľom predísť predimenzovaniu predajných kapacít  
 cieľom zosadiť Saddáma zvnútra ? Stal sa iba  
 cieľom lepšie čeliť separatistickému násiliu  
 cieľom aktívne pomôcť týranému dieťaťu a

2 Verbum - textSNkhi marias@...  
 3 4 SAUS - I: X Bonito GILIF S 21:55  
 9.4.2007

## Ukážka č. 12

Vyhľadajme pomocou jednoduchého príkazu všetky výskyty spojky *ako* (*než, alebo, či* a pod.) a určíme, kedy a prečo sa píše s čiarkou, kedy a prečo bez čiarky (v textoch korpusu sa pred interpunkčné znamienka umelo pridáva medzera pre potreby technického spracovania a štatistického vyhodnocovania textov). Vzhľadom na to, že nie všetci pisatelia ovládajú písanie čiarok a nie všetky texty zhromaždené v korpuse prešli korektúrou, žiaci sa môžu potešiť opravovaním chýb iných. A za odmenu určiť, akú funkciu majú jednotlivé čiarky v konkrétnom kontexte.

Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

Nový dotaz  jméno:  prim-3.0-public-sane

pět' ráz , hovoríeva mi vzrušene , viac ako do päť sa nenaučil rátať , aj to ma  
aj to ma stálo veľkú námahu , podobne ako rozlišovanie písmeniiek , po dlhom úsilí  
spomalenie , čas , ktorý je úplne odlišný ako môj , čas , ktorému ešte stále , napriek  
neúspechom , po ktorých klesá jeho sebaistota , ako dlho potom trvá , kým ju znovu získa  
Bobíkovi , zakaždým viem celkom presne , ako dlho trval obed , pretože syn mi telefonuje  
pretože syn mi telefonuje tesne pred tým , ako ide zazvonit' na susedu a hneď po tom ,  
ide zazvonit' na susedu a hneď po tom , ako sa naje , poumýva tanier s lyžicou , pedantne  
používať , ostrých predmetov sa bojí , tak ako aj ohňa a elektrického prúdu ) , okrem  
uvelebí sa v kresle a pozorne sleduje , ako sa po koberci pomaly pohybujú šesťuholníky  
predsa na kritériách , od nich záleží , ako to budem posudzovať ( či čas je bohatší  
hovoriť s mamou , viem si predstaviť , ako nástojčivo ma oslovoval , čo chlapa na  
skľúčený mi celý večer rozprával , ako sa mu miesto mňa ohlásil ktosi cudzí  
, ale aj svet jeho strachov a úzkostí ( ako som ich schopná vnímať , ako si ich predstavujem  
úzkostí ( ako som ich schopná vnímať , ako si ich predstavujem a interpretujem ) ,  
celkom jednoduchým maskovacím manévrom , ako má napríklad reagovať , keď ho v električke  
ohľadoch múdrejší , chápajúcejší ako ja , ako ja , ako my ostatní , čo ak jeho stavy  
ohľadoch múdrejší , chápajúcejší ako ja , ako my ostatní , čo ak jeho stavy mysle sú  
mysle sú vo väčšej zhode s prírodou ako stavy našich myslí , večne čosi projektujúcich  
vracat' sa k týmto myšlienkam , tá téma je ako ópium , nesmiem sa jej poddať , Fedor  
čo sa tak nejasne ohlasuje , ibaže neviem ako a navyše , neviem ani čo , jedna z mojich

Počet výskytů: 1379524  
> Query : "ako"

Zobrazeno: 361+50/1379524 (0%) Řádek: 361

1 2 Verbum - té-bíll X Bonito  
3 4 SAUS - I. GILIF marias. S 22:30  
9.4.2007

## Literatúra a zdroje

<http://korpus.juls.savba.sk>

IVORÍKOVÁ, Helena: Využitie Slovenského národného korpusu v školskej praxi. Cvičenia a kľúč. Rukopis.

Jak využívat Český národní korpus. Spracovali F. Čermák – R. Blatná a kol. Praha: Nakladatelství Lidové noviny 2005. 179 s.