silňuje „milou“ (usmievavou) kinetikou, jej jazykový prejav je falošný, tak trocha sentimentálny, až mierne gýčovitý. A to je škoda, lebo ako moderátorka má nesporné predpoklady na to, aby kultúrne a kultivovane vysielala aktuálne, obsahovo-tematicky diferencované spravodajské texty.

Pravdaže, v tejto chvíli si netrúfam povedať, či si tento falošný intonačný pôdorys moderátorka vybudovala sama, alebo či za ním stojí nejaký odborný („odborný“?) znalec jazyka a jeho ducha i duše, ktorú ústnemu jazykovému prejavu vdychujú predovšetkým jazykovo-intonačné prostriedky. V tomto druhom prípade by to bolo na zamyslenie... Tam vlastne mieri aj táto moja úvaha.

Najčastejšie a najmenej časté slová

MÁRIA ŠIMKOVÁ

Používatelia jazyka i záujemcovia o jeho hlbšie poznanie nám neraz kladú otázku, koľko slov má slovenčina a ktoré sú najviac a najmenej frekventované slová. V našom časopise sme sa už tejto témy dotkli (v príspevku M. Šimkovej *Slovenský národný korpus v druhej etape*, ktorý bol publikovaný v Kultúre slova v r. 2007, č. 4, s. 202 – 211), ale vzhľadom na opakovaný záujem a nové informácie sa k nej vraciame ešte raz.

Vymedzenie jadra slovnej zásoby, na ktorom je postavený napr. aj Krátky slovník slovenského jazyka (4. vyd., 2003; ďalej KSSJ), patrí od konca 19. storočia k základným poznatkom o každom jazyku a zaoberajú sa ním nielen jazykovedci, ale zužitkúvajú ho tiež odborníci z oblasti pedagogických a didaktických disciplín, logopédie, neurológie, počítačového spracovania prirodzeného jazyka a pod. V neposlednom rade výsledky frekvenčných analýz slúžia na porovnávacie výskumy jazykových systémov rôznych, typologicky odlišných i blízkych jazykov. Donedávna sa frekvenčné analýzy robili ručne – takto vznikli aj prvé frekvenčné slovníky slovenčiny, ktorých autorom je J. Mistrík: *Frekvencia slov v slovenčine,* 1969; *Retrográdny slovník slovenčiny,* 1976; *Frekvencia tvarov a konštrukcií v slovenčine*, 1985. Ide o lingvistické, presnejšie kvantitatívno-lingvistické spracovanie frekvenč-

ných ukazovateľov, ktoré sa získali z proporčne presne vybraných textov rôznych štýlov a žánrov v rozsahu 1 milión slov. V súčasnosti existujú už pre mnohé jazyky vrátane slovenčiny rozsiahle elektronické databázy textov v podobe národných jazykových korpusov a frekvenčné analýzy sa robia automatizovane na materiáli v rozsahu niekoľko stoviek miliónov až niekoľko miliárd jednotiek v prípade písaných textov a niekoľko miliónov v prípade hovorených korpusov (textových prepisov rečových prehovorov).

Na Slovensku existuje takáto verejne prístupná databáza od r. 2002 v podobe ***Slovenského národného korpusu*** (http://korpus.juls.savba.sk; ďalej SNK). Najnovšia verzia všeobecného, základného korpusu písaných textov ***prim-4.0*** bola sprístupnená začiatkom r. 2009 a obsahuje okolo 550 miliónov jednotiek. To však neznamená, že sú v nej už obsiahnuté všetky jazykové jednotky a že rozsah korpusu zodpovedá počtu slov v slovenčine. V prípade korpusu totiž uvádzame jeho rozsah v (textových) jednotkách (angl. *token*, mn. č. *tokens*), čo je jednak pomenovanie širšie ako slovo, jednak pri tomto rozsahu nejde o ich jednotlivé výskyty, ale o množiny výskytov všetkých jednotiek. Z korpusu však vieme získať aj ďalšie, hoci v niektorých prípadoch zatiaľ iba orientačné informácie.

Textové jednotky, tokeny, definované ako reťazce znakov medzi dvoma medzerami zahŕňajú v korpusovej lingvistike nielen slová (*slovo* podľa KSSJ = základná jazyková jednotka s ustálenou formou i významom; doplnili by sme: v slovenčine spravidla zložená z písmen alebo v reči z hlások), ale toto pomenovanie zahŕňa aj interpunkciu a neslovné vyjadrenia v podobe emotikonov, rôznych znakov, číslic a pod. Pri počítačovom spracovaní sa musia brať do úvahy všetky súčasti textu – programom, ktoré slúžia na jeho spracovanie a následne na vyhľadávanie, nemožno povedať, že to a to si nemajú všímať. Teda, naprogramovať sa dá aj to, len potom by už nešlo o úplné spracovanie textu, ale o nejakú (a ako rozhodnúť akú?) selekciu toho, čo sa pri počítačovom spracovaní bude brať do úvahy a čo nie – bez ohľadu na to, aké požiadavky na výskum sa perspektívne môžu vynoriť. Vzhľadom na tieto podmienky sa v jazykových korpusoch napr. umelo, technicky pridáva medzera pred interpunkčné znamienka, čo je síce v rozpore s pravopisnými pravidlami (a na začiatku používania korpusov si to viacerí všimli a takisto pátrali po príčinách), ale umožňuje presné štatistické výpočty: ak by sme v predchádzajúcej časti vety nedali v korpusovej databáze pred dvojbodku

medzeru, počítačový nástroj by považoval toto znamienko za súčasť slova *výpočty* a vo frekvenčných tabuľkách by sme mali raz slovo *výpočty* samostatne, inokedy s dvojbodkou, prípadne čiarkou alebo bodkou, ak by stálo na konci vety v súvetí alebo na úplnom konci vety, alebo ešte s inými znamienkami. V lingvistike však ide predovšetkým o slová a vo všetkých využitiach korpusu o presnosť, a tak interpunkciu technicky oddeľujeme a rátame osobitne.

Ak uvádzame, že najnovšia verzia SNK obsahuje 550 miliónov jednotiek, tak ide o súčet množín výskytov všetkých slov vo všetkých tvaroch a všetkých ďalších znakových súčastí textov zhromaždených v korpuse. V tomto počte je napr. čiarka ako najfrekventovanejšie interpunkčné znamienko zahrnutá takmer 35-miliónkrát. Čiarka je len jedno jednotlivé znamienko, ale tvarov pomocného slovesa *byť* je vyše 20 a my vieme povedať, že toto sloveso sa so všetkými svojimi tvarmi vyskytuje v celom korpuse vyše 12-miliónkrát, pričom výskyt jeho najfrekventovanejšieho tvaru *je* predstavuje z tohto počtu takmer tretinu. Všetky tieto údaje sa zverejňujú vždy spolu s najnovšou verziou korpusu a sú voľne dostupné (http://korpus.juls.savba.sk/stats/index.sk.html). Registrovaní používatelia, ktorí pracujú s databázou SNK prostredníctvom korpusového manažéra Manatee s klientom Bonito, si v ňom v položke Korpus/Súhrnné informácie môžu všimnúť ďalšie údaje.

Pri výbere podkorpusu ***prim-4.0-public-all*** (celý verejne prístupný primárny korpus z verzie 4.0) zisťujeme v danej položke, že jeho celková veľkosť je presne 526 082 640 jednotiek. Počet jednotlivých výskytov jednotlivých tvarov je však „len“ 3 969 719 jednotiek – tento údaj zahŕňa čiarku ako 1 jednotku, tvar *byť* ako 1 jednotku, tvar *je* ako 1 jednotku atď. V poradí tretí údaj, ktorý sa nachádza v položke Korpus/Súhrnné informácie, uvádza počet lem, teda základných tvarov slov ako zástupcov celého súboru vyskloňovaných, vyčasovaných alebo vystupňovaných tvarov, ktoré sa v textoch reálne vyskytujú. V podkorpuse prim-4.0-public-all je jednotlivých výskytov základných tvarov 2 467 595 jednotiek – slovo *byť* je tu započítané ako 1 jednotka, v ktorej sú zahrnuté tvary *je, bola, budeme* atď.

Ani tento údaj však nepredstavuje počet slov v slovenčine. Po prvé preto, lebo v korpuse nie sú všetky texty, ktoré boli kedy vydané/napísané v slovenskom jazyku, a teda ani všetky slová. Po druhé preto, lebo okrem jednotlivých slov existujú aj iné jazykové jednotky s ustálenou formou a význa-

mom – dvojice alebo trojice slov v podobe kolokácií, t. j. ustálených a lexikalizovaných spojení, ktoré predstavujú akési prefabrikáty a pomáhajú nám rýchlejšie a presnejšie konštruovať potrebné výpovede. Vďaka nim rodený hovoriaci vie, že napr. *bystrý krok* je rýchly, ale *bystrý človek* je vnímavý, *jasný deň* je slnečný, ale *jasná odpoveď* je jednoznačná, s čím majú cudzinci pri učení sa slovenčine problémy tak, ako ich máme my, keď sa učíme iný cudzí jazyk. V rámci korpusovolingvistických výskumov registrujeme pri každom z frekventovaných podstatných mien okolo 120 – 150 takýchto formálno-významových spojení, ktoré sa z istého hľadiska dajú považovať za jazykové (alebo aspoň textové) jednotky a ktoré v tom prípade zvyšujú hľadaný počet slov v slovenčine.

Jednotlivé výskyty základných tvarov v rozsahu 2 467 599 jednotiek nepredstavujú počet slov v slovenčine po tretie preto, lebo oproti najfrekventovanejším jednotkám s miliónovými alebo aspoň tisícovými výskytmi (spomínané čiarka, *byť*) sú vo frekvenčnej tabuľke aj slová s jedným alebo dvoma výskytmi. Tých je v podkorpuse prim-4.0-public-all dohromady 1 639 623 jednotlivých výskytov, čo sú približne dve tretiny výskytov jednotlivých základných tvarov. Medzi nimi sa nachádzajú rôzne neslovné znaky, chybné zápisy (ide nielen o pravopisné chyby alebo preklepy, ktoré sa v tlačených textoch často vyskytujú a v korpusových databázach sa evidujú tak, ako v daných textoch vyšli, ale sú to aj rôzne skomoleniny, jazykové hry a pod.), ale i citátové slová z cudzích jazykov. Táto skupina môže predstavovať aj polovicu z uvedeného počtu – azda príde perspektívne čas na ich presnejšiu analýzu. Navyše, chybné zápisy slov sa vyskytujú aj s frekvenciou vyššou ako 1 alebo 2, pretože niektoré druhy preklepov sú akoby všeobecne rozšírené, často sa zamieňajú napr. písmená *m* a *n*, na začiatku slova sa ponechávajú dve veľké písmená a pod. Preto by sme z počtu reálne existujúcich správnych slov v podobe základných tvarov mali túto skupinu vylúčiť. Druhú časť predstavujú regulárne slová zvyčajne už zastarané až archaické alebo naopak úplne nové, zriedkavé, s neustáleným spôsobom písania, úzko špecializované termíny, nárečové, slangové výrazy, vlastné mená a pod. Vzhľadom na predpoklad, že s výskytom 1 môže byť takýchto reálnych slov niekoľko stotisíc (ak by sme počítali polovicu z počtu jednotiek s výskytom 1, tak by to bolo 650 000), nevieme zodpovedne (logicky) odpovedať na otázku, ktoré slovo je ***najmenej frekventované***.

Keď sa však opäť vrátime na vrchol frekvenčnej tabuľky, vieme presne povedať, že v podkorpuse prim-4.0-public-all sú ***najčastejšie*** tieto jednotky (podľa základného tvaru): čiarka (1. v poradí, 34 749 635 výskytov), bodka (2., 34 570 215), *byť* (3., 12 339 227), *a* (4., 12 020 904), *v* (5., 10 874 817), *sa* (6., 9 024 381), *na* (7., 8 196 340), úvodzovky (8., 6 392 116), pomlčka (9., 6 187 349), *to* (10., 4 208 172). Z podstatných mien sú to (v každom z nasledujúcich výpisov podľa slovného druhu uvádzame príslušné jednotky z prvej stovky najfrekventovanejších slov): *rok* (26., 1 655 462), *človek* (59., 740 659), *Slovensko* (86., 517 865), *čas* (92., 481 088), *strana* (96., 453 741). V prvej stovke je 5 podstatných mien. Zo slovies sú to: *mať* (23., 2 414 658), *môcť* (40., 1 089 966), *povedať* (66., 657 189), *musieť* (78., 584 054), *chcieť* (80., 545 389), *nebyť* (82., 526 940), *ísť* (83., 521 257). Slovies je v prvej stovke 7 + pomocné sloveso *byť* z prvej desiatky. Z prídavných mien sú to: *veľký* (56., 781 375), *nový* (63., 713 173), *slovenský* (71., 631 342), *dobrý* (90., 485 079), *ďalší* (91., 481 761), teda 5. Ostatné jednotky v prvej stovke sú najmä spojky, predložky, zámená a častice. Ale napr. v podkorpuse pôvodných slovenských umeleckých textov ***prim-4.0-public-skimg*** v rozsahu 26 462 144 jednotiek sú najviac frekventované podstatné mená *človek* (47., 61 153), *ruka* (66., 38 425), *rok* (83., 33 268), *deň* (85., 32 366), *oko* (86., 32 244), *život* (87., 31 701), *žena* (96., 29 065), teda 7; slovesá *byť* (3., 802 727), *mať* (25., 134 244), *povedať* (44., 63 427), *môcť* (51., 48 688), *chcieť* (53., 47 192), *ísť* (54., 46 620), *vedieť* (55., 46 375), *nebyť* (58., 44 044), *musieť* (68., 37 777), *prísť* (74., 34 229), *dať* (76., 33 892), *vidieť* (84., 33 031), *nemať* (97., 28 685), *nevedieť* (98., 28 272), *začať* (100., 28 046), teda 15, a iba dve prídavné mená: *veľký* (80., 33 647), *celý* (92., 30 315).

Záver. Počet slov fungujúcich alebo aspoň raz použitých v písaných textoch súčasnej slovenčiny nachádzajúcich sa v Slovenskom národnom korpuse môžeme odhadovať na viac ako 3 milióny (podľa jednotlivých tvarov) alebo približne 2 milióny (podľa počtu základných tvarov, v ktorých sú obsiahnuté všetky ich tvary nachádzajúce sa v textoch korpusu). Najmenej frekventovaných slov (s výskytom 1) je asi 650 tisíc a sú to spravidla slová staršie až archaické, úplne nové, špeciálne termíny, slangové výrazy a pod. Najviac frekventované slovné jednotky (*a, v, na, sa, byť*) majú spolu so základnou interpunkciou (čiarka, bodka) stabilné postavenie na čele frek-

venčnej tabuľky bez ohľadu na štýlovú príslušnosť textov v korpuse. Na prvých miestach sa tieto jednotky nachádzajú aj vo frekvenčnom slovníku J. Mistríka z r. 1969. Na ďalších miestach sa jednotlivé slová nachádzajú na rôznych pozíciách v závislosti od rozsahu a štýlovej príslušnosti textov v korpuse. Zoznamy najfrekventovanejších slov a tvarov sa zverejňujú vždy s novou verziou korpusu (http://korpus.juls.savba.sk/stats/index.sk.html).

Motivácia nomenklatúry psích plemien

ANDREA EIBENOVÁ

Kynológia je zjednodušene veda o psoch (gréc. *kyón, kynos* = pes a lat. *logos* = veda; *canis* je rodové označenie psovitých šeliem v latinčine). Kynológia sa zaoberá psom ako najstarším a najbližším spoločníkom človeka všeobecne, čiže spája vedomosti o pôvode a genetike psa, šľachtiteľstvo psích plemien, výcvik psy a prácu so psami, psie športy, chov a odchov psov, ošetrovanie psov a starostlivosť o psy a okrajovo aj časti veterinárneho lekárstva a psiu psychológiu. Na Slovensku kynológiu zastrešuje Slovenská kynologická jednota (SKJ), ktorá je členom Medzinárodnej kynologickej federácie (FCI) so sídlom v hlavnom meste Belgicka Bruseli. Medzi základné úlohy tejto organizácie patrí vytvárať a upravovať štandardy jednotlivých plemien psov a tvoriť právne predpisy na chov.

Kynologická terminológia v sebe zahŕňa podskupinu, ktorú tvorí nomenklatúra plemien psov.

Motivácia nomenklatúrnych názvov môže zohľadňovať inherentné znaky (farbu, veľkosť, vzhľad, tvar, zloženie atď.), adherentné znaky (účel – činnosť, funkciu, použiteľnosť) a môžeme vyčleniť aj onymickú motiváciu, t. j. tzv. onymické názvy, ktoré by sme rozdelili do dvoch podskupín: podskupina obsahujúca geografické názvy (názvy domovskej krajiny, oblasti, územia, ostrova, polostrova, pohoria, mesta) a eponymá – dedikačné názvy.

Nomenklatúrne názvy psích plemien môžeme rozdeliť do dvoch základných skupín: