

KVANTITATÍVNA ANALÝZA POLYFUNKČNÉHO KOMPONENTU *sa* NA BÁZE SLOVENSKEHO NÁRODNÉHO KORPUSU

Radovan Garabík, Agáta Karčová, Adriána Žáková

Jazykovedný ústav Ľudovíta Štúra SAV

Panská 26, 811 01 Bratislava

e-mail: radovan.garabik@kassiopeia.juls.savba.sk

e-mail: agata.karcova@korpus.juls.savba.sk

e-mail: adriana.zakova@korpus.juls.savba.sk

GARABÍK, Radovan – KARČOVÁ, Agáta – ŽÁKOVÁ, Adriána: Quantitative Analysis of the Polyfunction Component *sa* on the Basis of Slovak National Corpus. *Slovak Language*, 2018, Vol. 83, No 1, pp. 23-34.

Abstract: We analyze reflexivity of Slovak verbs based on Slovak National Corpus. We derive a simple parameter describing the amount of reflexivity for a given verb, based on the distribution of the reflexive pronouns in the left and right context of the verb in the corpus, and apply the method to sort the verbs in the Slovak National Corpus according to the parameter. The method allows us to classify verbs automatically according to their reflexivity, given enough amount of their occurrences in the corpus.

Key words: Slovak language, verbs, reflexivity, corpus linguistics, Slovak National Corpus, statistical analysis

1. ÚVOD

Cieľom tohto príspevku je preskúmať zvrtnosť slovíes v slovenčine na základe reálnych jazykových dát. Snažíme sa nájsť taký parameter, ktorý by sa dal použiť na kvantitatívne určenie zvrtnosti (v najvšeobecnejšom chápaní) slovíes a vhodným spôsobom by eliminoval výskyty komponentu *sa* nesúvisiace so skúmanými slovesami. Pomocou exaktných štatistických metód určujeme relevantnosť výskytu komponentu *sa* v okolí slovíes, osobitne pre pozície napravo a naľavo od nich v rámci špecificky definovaných úsekov textov. V článku podrobne opisujeme použitý postup a jeho overenie na vzorke vybraných zvrtných a nezvrtných slovíes v Slovenskom národnom korpusu. Článok ukazuje, že taký parameter je v skutočnosti možné definovať a pri štatisticky dostatočnom množstve dát možno korigovať výskyty komponentu *sa* nesúvisiace so zvrtnosťou skúmaných slovíes.

2. KORPUS TEXTOV

Presnosť a relevantnosť kvantitatívneho výskumu zabezpečuje práca s rozsiahlymi a reprezentatívnymi textovými zdrojmi. Pre našu prácu sme zvolili korpus *prim-7.0-juls-all*, internú verziu Slovenského národného korpusu s veľkosťou 1 436 823 285 jednotiek textu (tokenov). Tvoria ho prevažne publicistické texty, zastúpená je v ňom aj umelecká a odborná literatúra. Korpus disponuje podrobnou bibliografickou a štýlovo-žánrovou anotáciou a všetky jednotky v texte sú automaticky lematizo-

vané a morfológicky anotované. Vďaka svojmu rozsahu, relevantnosti, spracovaniu a dostupnosti ide o zdroj spĺňajúci všetky požadované kritériá pre náš typ výskumu.

3. CIELE A METÓDY VÝSKUMU

3.1. Pracovné vymedzenie termínov

Veta sa tradične vymedzuje ako komplexná gramaticko-sémantická systémová jednotka, ktorá má povahu základnej komunikatívnej jednotky (Encyklopédia jazykovedy, s. 474), pričom jej primárnym konštitutívnym členom je sloveso. Na naše účely bolo potrebné v niektorých prípadoch pracovať s menšími úsekmi textu, ktoré sa však prevažne zhodujú s bežne vymedzenými vetami. Za samostatné úseky sme považovali aj polovetné konštrukcie oddelené čiarkami, úseky textu vyčlenené zátvorkami, ako aj jednotlivé časti priradovacích a podradovacích súvetí oddelené interpunkčnými znamienkami (čiarkou, bodkočiarkou, dvojbodkou, zátvorkami). Vznikli tak kratšie sémanticky relevantné a ucelené úseky. V ďalšom texte používame kvôli prehľadnosti pracovné označenie *veta* na pomenovanie týchto kratších úsekov.

Komponent *sa* môže byť formálnou súčasťou zvratných sloves, v odlišných prípadoch *sa* ako krátka podoba zvratného zámena (*seba/sebe*) slúži na vytvorenie zvratnej podoby nezvratných sloves, a napokon môže mať morféma *sa* čisto gramatickú funkciu, ak je súčasťou pasíva alebo neosobného zvratného tvaru. Vyskytuje sa, prirodzene, aj pri niektorých slovesných podstatných menách a okrajovo aj v iných prípadoch (porov. Morfológia slovenského jazyka, s. 376 – 388). Z morfeematického hľadiska je komponent *sa* pri reflexívach tantum prázdnu lexikálnu morfému, pri odvodených zvratných slovesách reflexívnou derivačnou morfému (porov. Morfeematický slovník slovenčiny, 1999, s. 507), keď ide o zámeno, je samostatnou koreňovou morfému. Ide teda o polyfunkčný komponent, ktorý môže mať v texte rozličné funkcie. R. Wagner (2011, s. 59) napríklad konštatuje, že aj české reflexíva môžu mať nielen funkciu zámena a derivačnej morfémy, ale aj častice. Klasifikácia konkrétneho *sa/si* (v češtine *se/si*) je v praxi často veľmi náročná (napr. K. Oliva, 2001, s. 206). Ako uvádza A. Jarošová (2003, s. 204), viacerí slovenskí aj českí lingvisti pri analýze konkrétnych prípadov (napr. pri slovese *umývať sa*) vysvetľujú mechanizmy reflexivizácie inak, aj keď ich závery môžu navzájom korešpondovať. Aktuálnym lingvistickým a zároveň i lexikografickým problémom je preto okrem iného logická a argumentačne zreteľná interpretácia reflexivity, resp. určenia, pri ktorých slovesách je reflexívny element *sa* krátkym tvarom zámena *seba* (porov. Jarošová, 2017, s. 103).

V našom výskume sa zameriavame na vždy zvratné slovesá, odvodené zvratné slovesá, ako aj na zvratné podoby nezvratných sloves a zvratné tvary nezvratných sloves (neosobný zvratný tvar, zvratný trpný rod). Do štatistiky preto vstupujú i tie nezvratné slovesá, pri ktorých komponent *sa* často figuruje ako súčasť zvratného pasíva. Ako sa ukáže v ďalšom opise, ich zastúpenie nie je natoľko vysoké, že by výrazne

ovplyvnilo štatistiky pri frekvenčnom výskume. V jednotlivých fázach skúmania sme do štatistiky zaradili všetky výskyty určitých plnovýznamových slovies s formantom *sa*. Na stanovenie vplyvu zvrtných pasív na spoľahlivosť výsledkov sme urobili kontrolnú analýzu, pri ktorej sme z materiálu vylúčili možné zvrtné tvary nezvrtných slovies (paušálne sme vylúčili všetky slovesá v tvare 3. osoby jednotného aj množného čísla). Vylúčenie tretej osoby výrazne neovplyvnilo kvalitatívne výsledky, ale prejavilo sa v miernom posune výsledných parametrov a hlavne v zníženej spoľahlivosti štatistiky vzhľadom na výrazne nižší počet analyzovaných slovies. Daný postup sa preto môže aplikovať aj na analýzu vzťahu zvrtných pasív k zvrtným a nezvrtným slovesám, podobný výskum však prekračuje rozsah tohto článku.

Pozícia zvrtných zámen v slovenčine je determinovaná rytmickým faktorom, a to v závislosti od toho, či vystupuje ako jednoslabičné neprízvučné slovo v pozícii napravo od slovesa vo funkcii enklitiky, alebo pred ním ako proklitika. Ako enklitiky sa komponenty *sa* realizujú v pozícii 1 alebo 2 napravo od slovesa (resp. v pozícii 3 v prípade kondicionálu). Naľavo od slovesa môže byť zvrtný komponent aj vo vzdialenejších pozíciách, lebo ako enklitika sa môže pripájať za prvé prízvukné slovo vety, jeho pozícia pred slovesom je teda do istej miery ľubovoľná (porov. Mistrík, 1966, s. 135 – 180). Situácia je zložitejšia, ak sa v jednej vete vyskytujú viaceré slovesá.

Vzhľadom na štatistický charakter výskytu slov v korpuse očakávame, že pozície komponentu *sa* v hrubých črtách budú zodpovedať uvedeným pravidlám.

3.2. Výber dát

Na začiatku sme ako skúšobnú vzorku vybrali 10 zvrtných slovies (*stať sa*, *čudovať sa*, *starať sa*, *zúčastniť sa*, *správať sa*, *postarať sa*, *zaoberať sa*, *dozvedieť sa*, *snažiť sa*, *podieľať sa*), 10 nezvrtných slovies (*dosiahnuť*, *pracovať*, *využiť*, *použiť*, *používať*, *bojovať*, *poskytnúť*, *služiť*, *očakávať*, *zabrániť*) a 10 takých slovies, ktoré sú nezvrtné, ale často sa vyskytujú aj s komponentom *sa* (*urobiť (sa)*, *vidieť (sa)*, *dostať (sa)*, *vrátiť (sa)*, *vyhnúť (sa)*, *stretnúť (sa)*, *zbaviť (sa)*, *rozhodnúť (sa)*, *hrať (sa)*, *porozprávať (sa)*). Prvotné hypotézy sme dôkladne testovali na týchto 30 slovesách, overovanie a koncipovanie záverov však prebiehalo na základe výsledkov štatistík realizovaných na všetkých plnovýznamových slovesách vyskytujúcich sa v prim-7.0-juls-all.

Z množiny slovies sme vopred vylúčili niektoré neplnovýznamové pomocné slovesá (*byť*¹ – toto sloveso je špecifické, vylúčili sme ho vo všetkých kontextoch ako plnovýznamové aj ako sponové sloveso; ďalej modálne slovesá *môcť*, *smieť*, *musieť*, ako aj ich negované podoby *nebyť*, *nemôcť*, *nesmieť*, *nemusieť*). Vo vetách, v ktorých sa tieto slovesá vyskytli v ktoromkoľvek tvare svojej paradigmy, sme ich nespracovávali, do úvahy sme brali iba ďalšie plnovýznamové slovesá vyskytujúce sa v daných vetách.

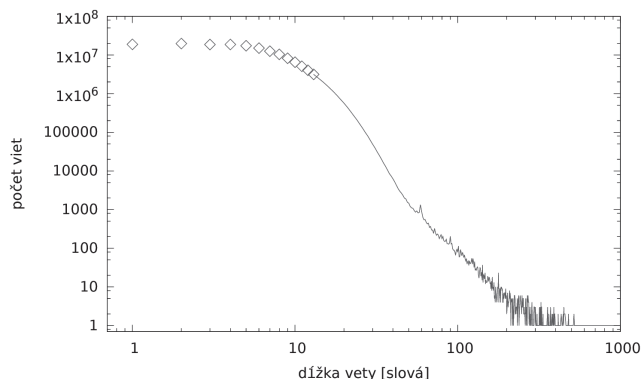
¹ Analýza „zvrtnosti“ slovesa *byť* v jeho rôznych funkciách si vyžaduje sémantický (alebo minimálne syntaktický) anotovaný korpus.

Počet výskytov plnovýznamových slovies v akomkoľvek tvare paradigmy predstavuje v korpuse *prim-7.0* celkovo 132 697 842 (92 355 na milión textových jednotiek). Zo všetkých tvarov sme vybrali tie, ktorých absolútny výskyt v afirmatívnom tvare je vyšší ako 400, čím sa počet tvarov zredukoval na 130 748 025 výskytov, ktorým zodpovedá 9 243 unikátnych lexém. Slovesá v negovanom tvare sme považovali za samostatné lexémy.

Pracovali sme s vetami, ktoré obsahovali maximálne jedno sloveso (okrem uvedených modálnych slovies a slovesa *byť*). Zostávajúcu interpunkciu vo vnútri viet sme dôsledne ignorovali, čo znamená, že nevstupovala do výpočtov dĺžok viet.

4. VSTUPNÉ DÁTA

Priložený graf názorne zobrazuje distribúciu dĺžky vety v slovenských textoch (ako sme ju definovali v úvode príspevku). Kvôli prehľadnosti sú hodnoty s dĺžkou vety nad 12 slov znázornené spojitou čiarou.



Obrázok 1: Distribúcia dĺžok viet v *prim-7.0-juls-all* v log-log mierke

Vzdialenosť slovesa od okraja vety sme definovali ako počet slov, ktoré sa nachádzajú medzi slovesom a koncom vety plus jeden (t. j. minimálna absolútna hodnota vzdialenosti, ak je sloveso na začiatku alebo na konci vety, bude 1). Zaviedli sme konvenciu, že vzdialenosti naľavo od slovesa sú záporné, vzdialenosti napravo kladné. Znamená to, že ak je skúmané sloveso posledným slovom vety, má vzdialenosť od okraja rovnú 1, a ak je sloveso prvým slovom vety, má vzdialenosť rovnú -1 .

Vzdialenosť zvrtného komponentu od slovesa sme definovali ako počet slov, ktoré sa nachádzajú medzi slovesom a zvrtným komponentom plus jeden. Analogicky k predchádzajúcemu prípadu budeme vzdialenosti naľavo od slovesa pokladať za záporné, vzdialenosti napravo za kladné.

Analýzou vybranej vzorky slovies sme zistili, že dĺžka vety nie je vhodný parameter pre analýzu umiestnení zvrtných komponentov *sa*, pretože má bezprostredný

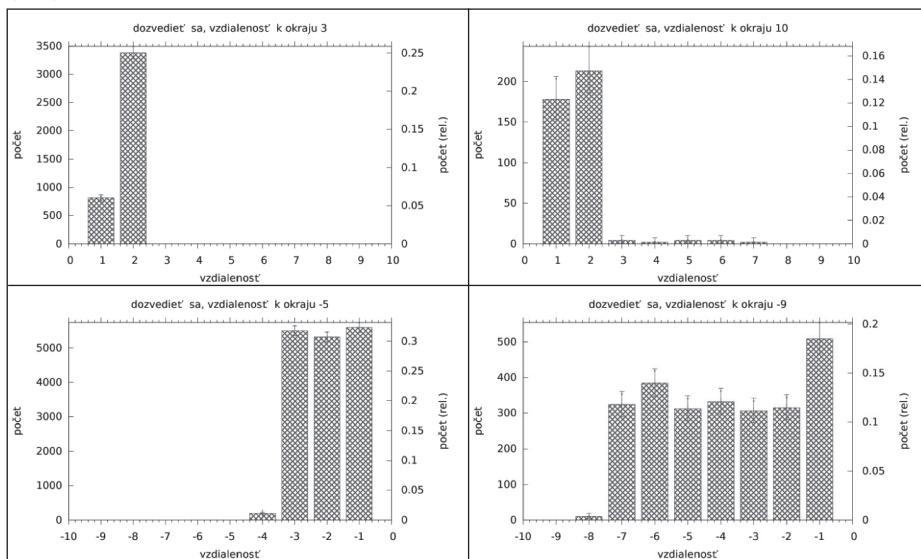
vplyv na vzdialenosť komponentu *sa* od slovesa. Ak napríklad vetu tvoria 3 slová, vzdialenosť *sa* od slovesa nemôže byť väčšia ako 2. Dĺžky viet sme preto po uskutočnení vstupných analýz nepovažovali za relevantné.

4.1. Spracovanie dát

Vety sme rozdelili do skupín podľa vzdialenosti slovesa od okraja. Do úvahy sme brali vzdialenosti od -10 do 10 vrátane. Viet s väčšími vzdialenosťami je zanedbateľný počet (porovnaj obrázok 1), ďalšie výsledky by preto ovplyvnili v minimálnej miere.

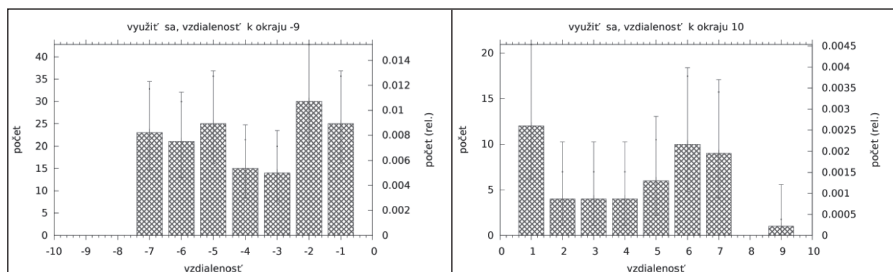
Pre každú skupinu následne sledujeme vzdialenosť komponentu *sa* od slovesa – na zjednodušenie pri skupine s kladnou vzdialenosťou slovesa od okraja sledujeme výskyty *sa* iba napravo od slovesa, pri skupine so zápornou vzdialenosťou slovesa skúmame výskyty *sa* len naľavo od slovesa (vzdialenosti *sa* od slovesa by na „opačnej“ strane nekorelovali so vzdialenosťou slovesa od okraja vety a podliehali by približne Poissonovmu rozdeleniu pravdepodobnosti (Poisson, 1837)).

Grafy na obrázku 2 ilustrujú, že pri zvrtných slovesách a okraji vety napravo od slovesa je veľký počet *sa* vo vzdialenosti 1 alebo 2 od slovesa, a to aj v prípade, ak je do konca vety k dispozícii viacero ďalších pozícií. Väčšie vzdialenosti *sa* majú mierne normálnu distribúciu, ale ich počet je výrazne nižší v porovnaní s počtom výskytov *sa* vo vzdialenosti 1 a 2.



Obrázok 2: Distribúcia počtu viet (zvislá os) podľa vzdialenosti *sa* od slovesa (vodorovná os) pre sloveso dozvedieť (*sa*) (ukážka niektorých skupín viet podľa vzdialenosti slovesa od okraja vety: 3, 10, -5, -9). Zvislé chybové intervaly zodpovedajú Clopperovým-Pearsonovým 95 % intervalom spoľahlivosti pre daný počet výskytov a sú uvedené na ilustráciu štatistickej spoľahlivosti danej analýzy.

Na druhej strane, pre okraj vety naľavo od slovesa je distribúcia počtu viet podľa vzdialenosti *sa* takmer uniformná. Následne sme vyslovili hypotézu, že takmer všetky komponenty *sa*, ktoré sa nachádzajú naľavo od slovesa, reálne súvisia so slovesom vety. Obrázok 2 naznačuje, že počet viet, ktoré sa začínajú na *sa* (prvý stĺpec naľavo, v zápornej vzdialenosti slovesa od okraja vety), je veľmi nízky (čo zodpovedá očakávanému výsledku). Ak je veta pred slovesom dostatočne dlhá, *sa* má miernu tendenciu nachádzať sa tesne pred slovesom. Tento predpoklad sme preverili na veľkom množstve slovies a vo všeobecnosti je okrem niekoľkých výnimiek platný.



Obrázok 3: Distribúcia počtu viet (zvislé osi) s danou vzdialenosťou *sa* od slovesa (vodorovné osi) pri slovese využiť pre dve skupiny viet podľa vzdialeností slovesa od okraja vety (-9, 10).

Na obrázku 3 vidíme, že počet výskytov *sa* pri nezvratných slovesách je v porovnaní s typicky zvratným slovesom veľmi nízky, čomu zodpovedajú aj podstatne väčšie chybové intervaly. Pri nezvratných slovesách je distribúcia počtu *sa* na pravej strane od slovesa výrazná, ale len relatívne vzhľadom na malý počet výskytov vo vzdialenosti 1 a 2. V prevažnej väčšine prípadov *sa* neprináleží k slovesu.

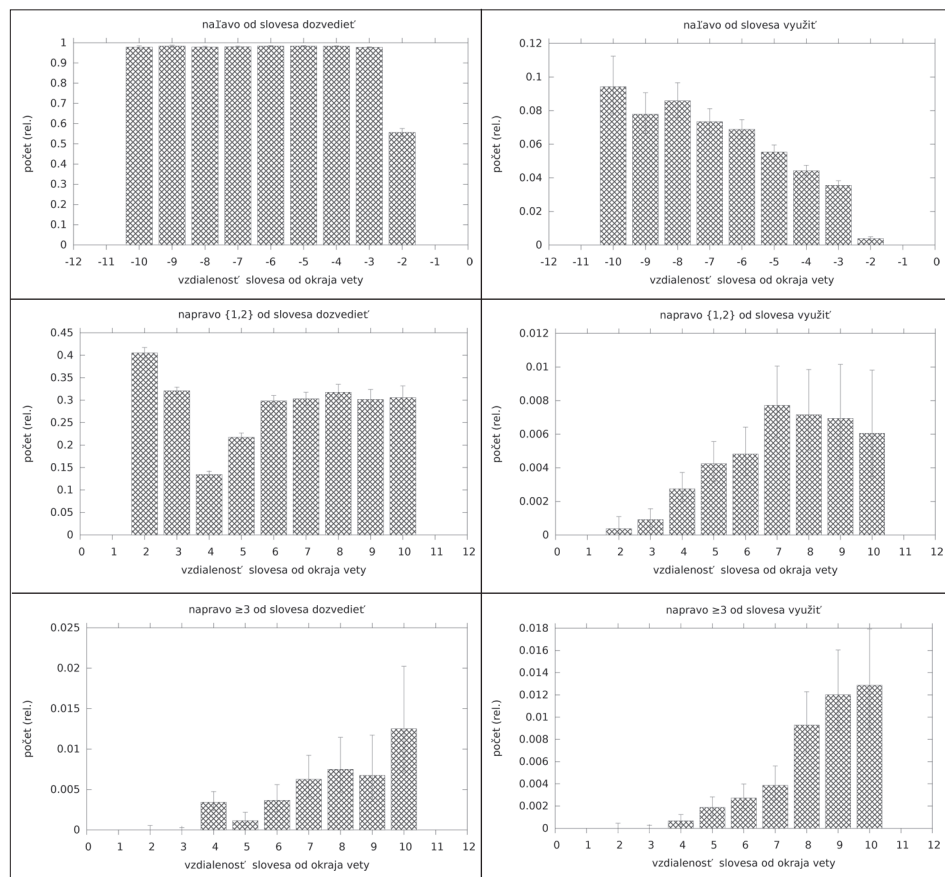
Pre komponenty *sa* naozaj prislúchajúce k slovesu sa tak nezávisle potvrdil vplyv rytmických faktorov na pozície výskytu komponentu – tieto sa nachádzajú (bez ohľadu na vzdialenosť slovesa k okraju vety) vo vzdialenosti 1 alebo 2. Vo vzdialenosti 3 a viac ide o *sa* nepatriace k slovesu – komponent *sa* zväčša patrí k tvarom (iných) slovies, ktoré sú relatívne samostatné a do štatistiky sme ich nezaradili – t. j. k slovesným podstatným menám či prídastiam vyskytujúcim sa v daných vetách. Tieto výskyty sme preverili na väčšom množstve slovies, a teda považujeme za všeobecne platné, že relevantná pozícia pre výskyt *sa* súvisiaceho so slovesom je napravo vzdialená 1 alebo 2 pozície, naľavo je rozptýlený vyššie, ale je tu tiež výrazná tendencia umiestnenia *sa* na čo najbližšej pozícii vzhľadom na súvisiace sloveso.

4.2. Parametre

Pre každé sloveso, osobitne pre každú vzdialenosť slovesa od okraja vety, zavedieme tri parametre (všetky sú relatívne, t. j. počet viet chápeme ako pomer daného počtu viet k celkovému počtu viet s daným slovesom):

- počet viet, keď je *sa* vzdialené od slovesa 1 alebo 2 (označíme $p_{1,2}$)
- počet viet, keď je *sa* naľavo, čiže má vzdialenosť menšiu ako 0 (označíme p_L)
- počet viet, keď je *sa* ďalej ako 2 (označíme $p_{\geq 3}$).

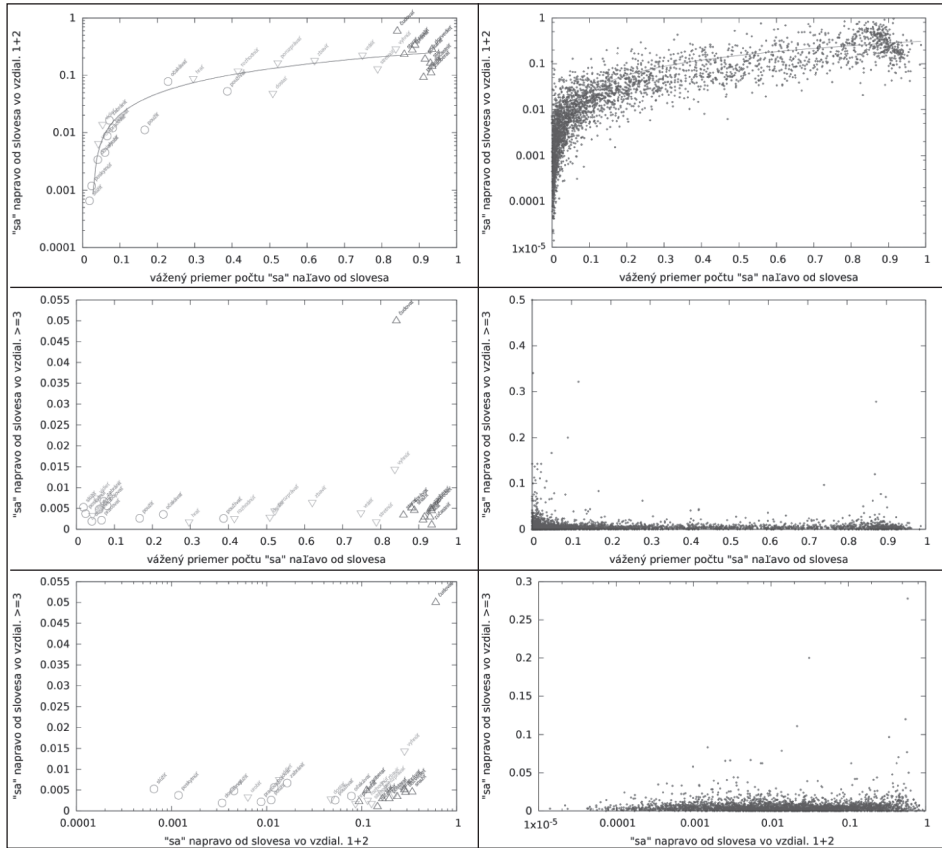
Berieme do úvahy relatívne počty vzhľadom na celkový počet výskytov slovesa pre danú vzdialenosť od okraja vety. To v štatistike vhodne odstráni závislosť od dĺžky vety (resp. od vzdialenosti slovesa od okraja vety).



Obrázok 4: Distribúcia parametrov podľa vzdialenosti slovesa od okraja vety pre slovesá *dozvedieť* a *využiť*.

Porovnanie parametrov plne potvrdzuje predpoklad, že $p_{\geq 3}$ zodpovedá komponentom *sa*, ktoré neprislúchajú k skúmanému slovesu. Na obrázku 4 je príklad distribúcie parametrov pre dve slovesá – *dozvedieť* a *využiť*.

4.3. KORELÁCIA VYBRANÝCH PARAMETROV



Na predchádzajúcich obrázkoch sú zachytené vzťahy medzi rôznymi dvojicami parametrov $p_{1,2}$, p_L a $p_{\geq 3}$ medzi sebou. V ľavom stĺpci je zachytená vzorka 30 skúmaných slovies, pričom sú slovesá zastúpené nasledujúcimi symbolmi:

- zvrtné slovesá – Δ
- nezvrtné slovesá – $\circ$
- nevyhranené slovesá – ∇ .

V pravom stĺpci je grafické zachytenie štatistiky všetkých slovies s počtom výskytov minimálne 400.

Uvedené grafy potvrdzujú náš predpoklad – parametre p_L a $p_{1,2}$ vykazujú zjavnú závislosť, pripomínajúcu lineárnu (hoci nie dokonalú) závislosť, znázornenú preloženou priamkou. Parametre p_L a $p_{\geq 3}$, ako aj $p_{1,2}$ a $p_{\geq 3}$ závislosť nepreukazujú.

Linearitu sme overili na všetkých slovesách Pearsonovým korelačným koeficientom (Pearson, 1895), ktorý má nasledovné hodnoty:

- p_L verzus $p_{1,2}$: 0.78
- p_L verzus $p_{\geq 3}$: -0.0244
- $p_{1,2}$ verzus $p_{\geq 3}$: 0.0266.

Korelačný koeficient v plnej miere potvrdzuje lineárnu závislosť parametrov p_L a $p_{1,2}$ a zároveň nezávislosť dvojíc ostatných parametrov.

Ide o lineárnu závislosť, preto môžeme znížiť počet parametrov (nie sú nezávislé) tak, že ich nahradíme vhodnou lineárnou kombináciou. Najjednoduchší spôsob, ako získať kombinácie, je získať parametre závislosti – dosiahneme to preložením priamky medzi parametrami p_L a $p_{1,2}$:

$$p_{1,2} = a \cdot p_L + b$$

Výsledky parametrov, ktoré sme získali, sú nasledujúce:

$$a = 0.320 \pm 0.004$$

$$b = -0.0085 \pm 0.0016$$

Intuitívne by sme očakávali, že parameter b by sa mohol rovnať nule (absolútne nezvratné slovesá by mali v ideálnom prípade $p_{1,2} = 0$ a $p_L = 0$, t. j. *sa* by sa nenachádzalo nikde naľavo ani napravo od slovesa). Získaná hodnota parametra b je naozaj takmer nulová.

Po získaní priamky môžeme normalizovať jeden z týchto parametrov vzhľadom na druhý. Ak má analyzované sloveso parametre p_L a $p_{1,2}$, môžeme položiť $p'_L = c \cdot p_L$ a $p'_{1,2} = c \cdot \frac{p_{1,2} - b}{a}$, kde c je neznáma multiplikatívna konštanta. Tým dosiahneme, že v ideálnom prípade, t. j. ak by parametre ležali na priamke, budú mať p'_L a $p'_{1,2}$ rovnakú hodnotu.

Hľadaný kvantitatívny parameter následne vyjadríme ako (Euklidovskú) vzdialenosť týchto transformovaných súradníc od začiatku súradnicovej osi:

$$p_{sa} = \sqrt{p'^2_L + p'^2_{1,2}}$$

Hodnotu konštanty c môžeme vo všeobecnosti určiť ľubovoľne, ale ak chceme smerovať k intuitívnym hodnotám, je podľa nás logické položiť pre „najviac zvrtné“ slovesá $p_{sa} = 1$ – to sú také, pre ktoré $p_L = 1$, potom $p'_L = p'_{1,2} = c$ a $p_{sa} = \sqrt{c^2 + c^2} = c\sqrt{2}$,

čiže $c = \frac{\sqrt{2}}{2}$. Znamená to, že sme urobili transformáciu súradníc, ktorá v ideálnom prípade zobrazí dvojicu bodov ($p_{1,2}$, p_L) na uhlopriečku štvorca s dĺžkou uhlopriečky 1 a stupeň zvrtnosti slovesa bude vyjadrený vzdialenosťou pozdĺž tejto uhlopriečky

od začiatku súradníc. Pre zrozumiteľnosť a úsporu vyjadrovania môžeme zaviesť pracovný termín „savost“, ktorým budeme označovať práve hodnotu tohto parametra (resp. menej formálne stupeň používania slovesa spolu so súvisiacim komponentom *sa*). Tento parameter teda matematicky možno vyjadriť nasledujúcim vzťahom:

$$p_{sa} = \frac{\sqrt{2}}{2} \cdot \sqrt{p_L^2 + \left(\frac{p_{1,2} - b}{a}\right)^2} \quad (1)$$

„Savost“ sa dá veľmi dobre interpretovať a môžeme pomocou nej porovnávať početnosť výskytu komponentu *sa* s ľubovoľným slovesom a ich vzájomný súvisiaci či nesúvisiaci vzťah. Ak sa hodnota „savosti“ rovná nule, sloveso je absolútne nezvratné. Ak je jeho hodnota blízka jednotke, sloveso je veľmi pravdepodobne iba zvratné (reflexívum tantum).

lema	p_L	$p_{1,2}$	savost'
čudovať	0,841	0,602	1,474
snažiť	0,889	0,341	0,995
dozvedieť	0,934	0,289	0,933
správať	0,881	0,281	0,894
vyhnúť	0,837	0,282	0,874
starat'	0,860	0,236	0,814
postarať	0,915	0,195	0,788
stať	0,930	0,165	0,762
zúčastniť	0,934	0,147	0,744
vrátiť	0,748	0,220	0,731
zaoberať	0,932	0,113	0,711
podieľať	0,911	0,094	0,683
stretnúť	0,789	0,127	0,633
zbaviť	0,620	0,177	0,600
porozprávať	0,521	0,162	0,526

lema	p_L	$p_{1,2}$	savost'
rozhodnúť	0,416	0,115	0,402
dostať	0,508	0,047	0,380
používať	0,387	0,053	0,305
hrať	0,296	0,086	0,296
očakávať	0,229	0,078	0,250
použiť	0,166	0,011	0,125
zabrániť	0,072	0,016	0,075
bojovať	0,080	0,012	0,073
vidieť	0,054	0,014	0,062
pracovať	0,066	0,009	0,060
využiť	0,060	0,005	0,051
urobiť	0,042	0,006	0,044
dosiahnuť	0,041	0,003	0,039
poskytnúť	0,024	0,001	0,027
slúžiť	0,019	0,001	0,024

Tabuľka 1: „Savost“ 30 lemm našej základnej vzorky

V tabuľke 1 sú uvedené výsledné hodnoty „savosti“ pre skupinu 30 lemm vzorových slovies (lemy uvádzame bez zvratného komponentu). Numerická hodnota veľmi spoľahlivo indikuje stupeň používania slovesa spolu s polyfunkčným komponentom *sa*.

„Savost“ môže nadobúdať hodnotu väčšiu ako jeden, a to v prípade anomálnych distribúcií komponentu pri slovese. Hodnota „savosti“ mierne uprednostňuje polohu *sa* napravo od slovesa, preto pri nadmernom výskyte (t. j. väčšom ako pri „bežných zvratných“ slovesách) *sa* naľavo od slovesa môže nadobúdať aj hodnoty väčšie ako jeden. Z korpusovej analýzy vyplýva, že anomálna „savost“ slovesa *čudovať sa* je spôsobená nadmerným výskytom fráz typu *nemožno sa čudovať*, *nemôžeme sa čudovať*, *netreba sa čudovať* a pod.

5. ZÁVER

Cieľom nášho výskumu bolo štatisticky overiť mieru výskytu polyfunkčného komponentu *sa* s prislúchajúcimi slovesami na rozsiahlom textovom materiáli, pričom sme využili exaktné postupy a metódy a overovali platnosť formulovaných premís. V súvislosti s tým sme vymedzili a analyzovali úseky textu vždy s jedným plnovýznamovým slovesom. Ďalej sme zisťovali, na ktorých pozíciách sa polyfunkčný komponent *sa* vyskytuje najčastejšie vzhľadom na sloveso, či jeho pozícia závisí od dĺžky vety, prípadne či ju výrazne ovplyvňujú aj iné parametre.

Postupovali sme tak, že sme stanovovali viaceré hypotézy, ktoré sme overovali na malej vzorke slov (30 sloviess), následne sme pracovali so všetkými slovesami s výskytom väčším než 400 nachádzajúcimi sa v korpuse *prim-7.0-juls-all*. Zo štatistickej analýzy korpusového materiálu vyplýva, že kvalitatívne vyjadrenie niektorých aspektov zvratnosti sloviess v slovenčine sa dá redukovat' na jeden parameter, ktorý sa ľahko interpretuje a zároveň umožňuje porovnávanie všetkých sloviess navzájom. Nazvali sme ho „savost“ a matematicky sme ho vyjadrili vzťahom (1). Pri hodnote parametra (v rozpätí 0 až 1) majú úplne nezvratné slovesá hodnotu 0, pri zvratných slovesách sa hodnota približuje k 1. Hodnoty vyššie ako 1 vyjadrujú preferenčný výskyt komponentu *sa* naľavo od slovesa. Hodnoty približne v strede tohto intervalu vyjadrujú iné možné funkcie komponentu *sa*.

V predkladanej práci sme pracovali predovšetkým so slovesami v pozitívnom tvare, ale vykonali sme aj samostatnú analýzu negatívnych tvarov týchto sloviess. Výsledky pre pozitívne aj negatívne tvary sloviess sú porovnateľné. Mierny posun nastal iba v hodnotách parametrov a v ich presnosti; keďže samotné negatívne tvary sa v korpuse vyskytujú zriedkavejšie ako pozitívne, štatistika pre ne je menej spoľahlivá.

Do analýzy sme zahrnuli aj zvratné pasívum a neosobný zvratný tvar nezvratných sloviess. Pri kontrolnej analýze po vylúčení zvratného pasíva a neosobného zvratného tvaru, t. j. po odstránení výskytov sloviess v tretej osobe jednotného i množného čísla, sa ukazuje, že výsledky sú aplikovateľné aj na vlastne zvratné slovesá, pričom sa hodnoty parametrov opäť posunú (a veľkosť tohto posunu by vyjadrovala pomer výskytov zvratného pasíva a neosobného zvratného tvaru k iným

funkciám komponentu *sa*). Analýza výskytu zvrtných tvarov nezvrtných slovies v porovnaní s výskytom zvrtných slovies však nie je predmetom tejto štúdie. Uvedenú metódu možno využiť predovšetkým v synchronnej, ale napríklad aj v diachrónnej analýze zvrtnosti v prípade relevantného množstva diachrónnych textov.

Literatúra

- DVONČ, Ladislav — HORÁK, Gejza — MIKO, František — MISTRÍK, Jozef — ORAVEC, Ján — RUŽIČKA, Jozef — URBANČOK, Jozef: Morfológia slovenského jazyka. Bratislava: Vydavateľstvo Slovenskej akadémie vied 1966. 896 s.
- JAROŠOVÁ, Alexandra: *Umývať sa*: predmetová syntagma alebo zvrtné sloveso? In: Tradícia a perspektívy gramatického výskumu na Slovensku. Ed. M. Šimková. Bratislava: Veda 2003, s. 204 – 208.
- JAROŠOVÁ, Alexandra: Problémy vymedzenia statusu komponentu *sa* v konštrukciách typu *umývať sa* a *vidieť sa*. In: Slovenská reč, 2017, roč. 82, č. 2, s. 103 – 115.
- MISTRÍK, Jozef: Slovosled a vetosled v slovenčine. Bratislava: SAV 1966. 276 s.
- MISTRÍK, Jozef et al.: Encyklopédia jazykovedy. Bratislava: Vydavateľstvo Obzor 1993. 515 s.
- OLIVA, Karel: Reflexe reflexivity reflexiv. In: Slovo a slovesnosť, 2001, roč. 62, č. 3, s. 200 – 207.
- PEARSON, Karl: Notes on regression and inheritance in the case of two parents. In: Proceedings of the Royal Society of London 1895, č. 58, s. 240 – 242.
- POISSON, Siméon Denis: Probabilité des jugements en matière criminelle et en matière civile, précédés des règles générales du calcul des probabilités. Paris: Bachelier 1837, s. 205 – 207.
- Slovenský národný korpus – prim-7.0-public-all. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2015. Dostupný na: <http://korpus.juls.savba.sk> [cit. 11. 9. 2017].
- SOKOLOVÁ, Miloslava — MOŠKO, Gustáv — ŠIMON, František — BENKO, Vladimír: Morfematický slovník slovenčiny. Prešov: Náuka 1999. 531 s.
- WAGNER, Roland: Reflexivní slovesa v češtině a prognóza jejich ekvivalentů v němčině. Disertační práce. Ústav českého jazyka Filozofické fakulty Masarykovy univerzity 2011. 388 s. Dostupný na: https://is.muni.cz/th/56270/ff_d/wagner_prognóza_reflex.pdf [cit. 11. 9. 2017].