

USTREZNOST STROJNO IZLUŠČENIH ŠTIRIPRIPONSKIH IZPRIDEVNIŠKIH BESEDOTVORNIH NIZOV IZ UČNE MNOŽICE BSSJ

V sodobnem slovenskem jezikoslovju se do preloma stoletja besedotvorne raziskave niso načrtno ukvarjale s sinhrono večmorfemsko obrazilno kombinatoriko, ampak je le-ta bila omenjana obrobno (Vidovič Muha 1988, Toporišič 2000). V zadnjih dveh desetletjih se je tej problematiki namenila konkretnjša pozornost z besednodružinskim oz. stopenjskim pristopom, ki je bil preizkušen na gradivu besednih družin (BSSJ, Stramljič Breznik 2004) in na primerih glagolov čutnega zaznavanja (Kern 2017).

Povod za sistematične morfotaktične raziskave slovenskega jezika je interdisciplinarni projekt *Kombinatorika besedotvornih obrazil v slovenščini* (2021–2025), ki združuje jezikoslovni in jezikovnotehnološki pristop. Raziskave se osredinjajo na analizo in opis družljivosti besedotvornih priponskih obrazil v okviru večstopenjske tvorbe, pri tem se upoštevajo tovrstna novejša spoznanja poljskega besedotvorja (Burkacka 2012).

Izhodiščno bazo besedotvornih podatkov predstavlja BSSJ (*Besednodružinski slovar slovenskega jezika na b*), ki ob nemotivirani iztočnici po stopnjah tvorbe prikazuje vse istokorenske tvorjenke, členjene na podstavni in obrazilni del. Izdelan je bil s posebnim računalniškim programom za slovarje in izdan v knjižni obliki. Obsega 666 iztočnic na b in njihovih besednih družin, v katerih je razvrščenih 11 136 enot. Slovar je bil ob nastanku pri vključevanju gradiva zavestno in hote izčrpen z namenom, da bi bilo zajetih čim več tvorbenih podatkov. Temeljni gradivni viri so bili: *Slovar slovenskega knjižnega jezika* (SSKJ), *Besedišče slovenskega jezika* (BSJ), *Slovenski pravopis 2001* (SP) ter takrat edino razpoložljivi korpusu *Fida*, po katerem so bile dodatno paberkovalno zbrane nekatere besede črke b (Stramljič Breznik 2004: 9).

Raziskovalci jezikovnotehnološkega dela so na podlagi izvoženih podatkov iz BSSJ oblikovali obsežno bazo realiziranih besedotvornih nizov, primernih za nadaljnjo strojno analizo (Erjavec, Pranjić idr. 2023: 449–465). Vsebovani so podatki o stopnji tvorbe, besednovrstnostnem izhodišču tvorjenk, zaporedju obrazil v nizu in dodan konkretni primer tako uresničene tvorjenke ali več njih iz BSSJ. Na podlagi avtomatskega luščenja podatkov so bile iz SSKJ in baze Sloleks dodane še druge tvorjenke, ki so bile prepoznane kot ustrezne tvorbenemu nizu iz BSSJ.

Določen delež strojno najdenih primerov iz obeh dodatnih virov je bil popolnoma pravilen. Npr. izluščena tvorbena shema četrstopenjskih tvorjenk s pridevniškim izhodiščem: ADJ:X → NOUN:X-ec → ADJ:X-ev → ADJ:X-ski → ADV:X-o, izdelana na podlagi tvorbenega vzorca pridevnika *bel* iz BSSJ (*bel* → *bel-ec* → *belč-ev* → *belčev-ski* → *belčevsk-o*), je bila ustrezno potrjena z dodatnima nizoma tvorjenk s pridevniškima izhodiščema *nor* in *star*, ki sta bila najdena v SSKJ in Sloleks.

Precejšen delež tako avtomatsko izluščenih primerov pa je bilo treba izločiti. Naj konkretiziramo. Po tvorbeni shemi iz BSSJ s samostalniškim izhodiščem NOUN:X → ADJ:X-en → NOUN:X-ik → NOUN:X-ica → ADJ:X-in je bil iz SSKJ in baze Sloleks avtomatsko izluščen tudi neustrezen niz: **pesa* → **pes-en* → *pesn-ik* → *pesn-ica* → *pesnič-*

¹ irena.stramljic@um.si

in, v katerem se mešajo tvorjenke izhodiščnega samostalnika *pesa* 'kulturna rastlina' in glagola *pesniti* 'pisati pesmi'.

V nadaljevanju smo predstavili pet iz BSSJ izluščenih štiripriponskih nizov s pridevniškim izhodiščem in zanje preverili uspešnost avtomatsko prepoznanih primerov iz širše leksikalne baze, ki jo predstavljata SSKJ in Sloleks, da bi lahko na konkretnih primerih ovrednotili, v kolikšni meri sta pri novih avtomatsko najdenih kandidatkah skladni zunanja formalno (morfemsko) izražena struktura in notranja (besedotvornopomenska) logična tvorbeno pot. Na podlagi predstavljenih primerov lahko izpostavimo tako prednosti kot slabosti avtomatskega luščenja štiripriponskih izpridevniških nizov.

Med prednostmi lahko izpostavimo naslednje: (1) Za strojno učenje so uporabni na vseh tvorbenih stopnjah zastopani morfološki nizi, ki se pojavljajo znotraj besednih družin posameznih slovarskih iztočnic na b. (2) Izluščen morfološki niz iz učne množice je lahko v celoti uspešno apliciran na prepoznavo še drugih tvorjenk s pridevniškim izhodiščem v večji leksikalni bazi, kot sta SSKJ in Sloleks (prim. niza *nor* in *star* po izhodiščnem vzorcu *bel*). (3) Na določenih mestih so ti nizi lahko celo morfološko in pomensko prekrivni, čeprav je njihovo besednovrstno izhodišče različno (prim. niz *norec* s pridevniškim in niz *borec* z glagolskim izhodiščem: svojilni pridevnik iz obeh na 2. stopnji (*borčev* – *norčev*), vrstni pridevnik na 3. stopnji (*borčevski* – *norčevski*) in prislov na 4. stopnji tvorbe (*borčevsko* – *norčevsko*). (4) Z dodatnim pregledom in ročno narejenimi korekcijami v izluščenih nizih lahko odstotek ustreznih še povečamo (prim. niz *drevesen* → *drevesnik* → *drevesnica* → *drevesničar* → *drevesničarski*).

Na slabostmi, ker gre le za delno ustreznost dodatno najdenih primerov tvorjenk, avtomatsko izluščenih po naučenih vzorcih, pa lahko vpliva več dejavnikov, med katerimi so: (1) Morfološko zaporedje v nekem delu niza ne podpira logičnega besedotvornopomenskega vzorca (*žičen* → *žičnik* → *žičnica* → *žičničar* → *žičničarski*), ker je bilo napačno posplošeno, da vsakemu samostalniku na *-ik* sledi samostalnik na *-ica*, kar velja le s pogojem, da je samostalnik na *-ik* živo, človeško in da se obrazilo moškega spola krni in zamenja z ženskospolskim (zamenjevalni tip feminativne tvorbe). (2) Strojno neprepoznana enakopisnost, a ne tudi enakozvočnost vpliva na napačno posplošitev vzorca (prim. *bólnica* proti *bolnica*). (3) Sinhrono predpostavljeno besedotvorno razmerje med motivirajočo in motivirano besedo (*bolan* → *bolnik* → *bólnica*), ki etimološko ni podprto, ker naj bi šlo za diahrono gledano istostopenjski tvorjenki (*bolan* → *bolnik* in *bolan* → *bólnica*), lahko vodi do napačnega avtomatsko posplošenega vzorca. (4) Strojno napačno prepoznano tvorbeno razmerje med deležnikom in glagolom tako na ravni motivacijske smeri, v kateri je deležnik izhodišče ter glagol njegova tvojenka (**ojačan* → **ojačati*) in ne obratno, kot je ustrezno (*ojačati* → *ojačan*), in še dodatne obrazilne neustreznosti pri izpeljavi deležnikov iz glagolov z različnima glagolskima priponama (ne *ojačati* → **ojačen*, ampak *ojačati* → *ojačan* in *ojačiti* → *ojačen*) zahtevajo korekcijo. Dodatna rekonstrukcija celotnega tvorbenega niza pa pokaže, da je le-ta mnogo daljši (petostopenjski), saj ima v izhodišču korenski pridevnik *jak* → *jačati* → *ojačati* → *ojačevati* → *ojačevalen* → *ojačevalnica*.

Iz predstavljenih primerov izhaja, da bo za izboljšanje strojne prepoznave tvorjenk po naučenih modelih vsaj v tej fazi potrebnega še veliko dodatnega ročnega pregleda in odstranjevanja napačno posplošenih vmesnih faz tvorbe in s tem obrazilnega zaporedja v priponskih nizih, s čimer se navadno ne samo skrajša dolžina priponskega niza, ampak tudi stopnja tvorbe.

Ključne besede: slovenščina, besedotvorje, izpridevniška tvorba, štiripriponski nizi, avtomatsko prepoznavanje tvorjenk

LITERATURA

- Burkacka I. *Kombinatoryka sufiksalska w polskiej derywacji odrzeczownikowej*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2012.
- Erjavec T., Pranjić M., Pelicon A., Kern B., Stramljič Breznik I., Pollak S. «Automating derivational morphology for Slovenian» *eLex 2023: electronic lexicography in the 21st century: proceedings of the eLex 2023 conference. Lexical Computing CZ*. Ur. M. Medved', M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubiček, S. Krek. 449–465.
<https://elex.link/elex2023/wp-content/uploads/92.pdf> 15. 11. 2024.
- Kern B. *Stopenjisko besedotvorje (Na primeru glagolov čutnega zaznavanja)*. Ljubljana: Založba ZRC SAZU, ZRC SAZU, 2017.
- Kern, B. «Potencjał słowotwórczy wybranych polskich i słoweńskich czasowników ruchu. V: Gawlak, M. (ur.). *Polsko-słoweński dialog międzykulturowy = Slovensko-poljski medkulturni dialog*. 1. mednarodna znanstvena slovenistična konferenca Slovensko-poljski medkulturni dialog, Katowice: Wydawnictwo Uniwersytetu Śląskiego, 2023. Str. 185–201.
<https://monograph.us.edu.pl/index.php/wydawnictwo/catalog/view/PN.4188/115/85> 11. 12. 2024.
- Владимир. В. Лопатин, Игорь С. Улуханов *Словарь словообразовательных аффиксов современного русского языка*. Москва, 2016.
[Vladimir. V. Lopatin, Igor' S. Uluhanov *Slovar' slovoobrazovatel'nyh affiksov sovremennogo russkogo jazyka*. Moskva, 2016]
- Markežič T., Stramljič Breznik I. *Feminativi v slovenskem jeziku*. Maribor: Univerzitetna založba = University Press, 2021.
- Ološtiak M., Ivanová M. (ur.) *Slovník slovotvorných prostředků v slovenčine*. Prešov: Prešovská univerzita v Prešove, Filozofická fakulta, 2021.
- Sokolová M., Moško G., Šimon F., Benko V. *Morfematický slovník slovenčiny*. Prešov: Náuka, 1999.
- Skarżyński M. (ur.) *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego*, Knj. 3–4. Kraków: Towarzystwo Wydawnicze "Historia Iagellonica", 2004.
- Skok P. *Etimologijski rječnik hrvatskoga ili srpskoga jezika*. Zagreb: JAZU, 1971.
<https://archive.org/details/EtimologijskiRjecnikHrvatskogalliSrpskogaJezika> 10. 12. 2024.
- Snoj M. *Slovenski etimološki slovar³ (SES)*, 2016.
<https://fran.si/> 10. 12. 2024.
- Stramljič Breznik I. *Besednodružinski slovar slovenskega jezika: poskusni zvezek za iztočnice na b (BSSJ)*. Maribor: Slavistično društvo, 2004.
- Stramljič Breznik I. «Značilnosti izmedmetih priponskih nizov v Besednodružinskem slovarju slovenskega jezika za iztočnice na b.» *Slavia Centralis*. 17/1 (2024a) : str. 1–17.
- Stramljič Breznik I., Ledinek N. «Kvantitativni podatki o besedotvornih modelih in priponskih nizih izmedmetnih tvorjenk v Besednodružinskem slovarju slovenskega jezika za iztočnice na b.» *Slovenski jezik – Slovene linguistic studies*. 16 (2024b): str. 59–85.
- Šiška Z. *Bázový morfematický slovník češtiny*. Univerzita Palackého v Olomouci, Pedagogická fakulta, 2005.
- Toporišič J. ⁴2000: *Slovenska slovnica*. Maribor: Obzorja.
- Vidovič Muha A. *Slovensko skladenjsko besedotvorje ob primerih zloženek*. Ljubljana: Znanstveni inštitut Filozofske fakultete, 1988.
- Vidovič Muha A. *Slovensko skladenjsko besedotvorje*. Ljubljana: Znanstvena založba Filozofske fakultete, ²2011.

Besedišče slovenskega jezika

<http://bos.zrc-sazu.si/besedisc.html> 10. 11. 2024.

Fran. Slovarji inštituta za slovenski jezik Frana Ramovša ZRC SAZU.

<https://fran.si/> 10. 12. 2024.

Gigafida 2.0, referenčni korpus pisne standardne slovenščine.

<https://viri.cjvt.si/gigafida/> 10. 12. 2024.

Sloleks 2.0. Slovenski oblikoslovni leksikon.

<https://viri.cjvt.si/sloleks/slv/> 10. 12. 2024.

Slovenski pravopis 2001

<https://fran.si/> 10. 12. 2024.

SSKJ1: Slovar slovenskega knjižnega jezika. ¹1970–1991.

<https://fran.si/> 10. 12. 2024.

SSKJ2: Slovar slovenskega knjižnega jezika. ²2014.

<https://fran.si/> 10. 12. 2024.