

Slovenski jezik Slovene Linguistic Studies

16

2024

Znanstvenoraziskovalni center
Slovenske akademije znanosti in umetnosti
Inštitut za slovenski jezik Frana Ramovša
Ljubljana, Slovenija

in • and

Brigham Young University
College of Humanities
Provo, Utah, USA

Mladen Uhlik (ORCID: 0000-0003-4635-9273)

Univerza v Ljubljani, Filozofska fakulteta, Slovenija
ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
mladen.uhlik@zrc-sazu.si

Andreja Žele (ORCID: 0000-0002-6477-2590)

Univerza v Ljubljani, Filozofska fakulteta, Slovenija
ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
andreja.zele@zrc-sazu.si
DOI: <https://doi.org/10.3986/16.1.01>

ZGRADBE Z GLAVNIMI ŠTEVNIKI OD PET NAPREJ: SOPOSTAVITEV SLOVENŠČINE Z DRUGIMI JUŽNOSLOVANSKIMI JEZIKI IN RUŠČINO

V obravnavi poudarjamo posebnosti slovenskih zgradb z glavnimi števnikmi, večjimi od 5, in opozarjamo na podobnosti in razlike v primerjavi z ustreznici v štokavščini, makedonščini in ruščini. Sopostavitev različnih zgradb v slovanskih jezikih predstavlja različna skladijska razmerja znotraj količinskih besednih zvez. V predikacijskem razmerju pa je obravnavana količinska zveza v vlogi osebka.

KLJUČNE BESEDE: števnik, veliki števnik, količinske besedne zveze, ujemanje, vezava, slovanski jeziki, slovenščina, ruščina, štokavščina, makedonščina

The article focuses on the distinctive features of Slovenian constructions involving cardinal numbers from 5 and up. Additionally, we draw attention to both similarities and differences when comparing them with their counterparts in Štokavian (BCMS), Macedonian, and Russian. The comparison of diverse constructions in Slavic languages illustrates a variety of syntactic relationships within quantitative phrases. We examine the constructions in the context of a predicative agreement, where the quantifier phrase assumes the role of the subject.

KEYWORDS: numerals, high numerals, quantifier phrases, agreement, government, Slavic languages, Slovene, Štokavian, BCMS, Russian, Macedonian

UVOD

V prispevku obravnavamo rabo količinskih zgradb v slovenščini, štokavščini¹ in ruščini. V nekaterih primerih je dodana makedonščina kot slovanski jezik, ki ima med drugim tudi poenostavljen sklonski sistem. Obravnava se osredinja na zgradbe z glavnimi števnikami od 5 naprej oz. na zgradbe s t. i. velikimi števnikami.² V prid celotne predstavitve je potrebno najprej opozoriti na pomembne razlike med zgradbami s t. i. malimi števnikami³ in zgradbami z velikimi števnikami. Predstavljene so značilnosti enomestnih in dvomestnih števnikov, ki se v slovenščini, kot izhodiščnem jeziku obravnave, zapisujejo enobesedno.⁴

Po G. G. Corbettu (2009) so slovanski jeziki specifični ravno zaradi različnih skladijskih razmerij znotraj količinskih zgradb. Količinske zgradbe namreč vključujejo tako primere kanoničnega ujemanja kakor tudi primere odstopanja od ujemanja. Kanonično ujemanje je možno ponazoriti s količinskimi besedami z glavnimi števnikami od 1 do 4. Kot je razvidno iz (1), (2) in (3) se števniki v vlogi določila oz. nadzorovane besede ravna po kategorialnih lastnostih nadzornega jedra, ki je samostalniška beseda. V spodnjih zgledih se v vseh treh jezikih ob števnikiški obliki v imenovalniku uporabljajo imenovalniške oblike samostalnikov. Slovenščina že na ravni samostalniških oblik razlikuje med ednino, dvojino in množino (1). V štokavščini in ruščini dvojine ni in zato so ob števnikih od 2 do 4 rabljene posebne pavkalne oblike (2), (3), o tem glej tudi v prispevku Uhlik, Žele (2023).⁵

- (1) slvn. en_{NOM SG M} učbenik_{NOM SG M}, dva_{NOM M} učbenika_{NOM DU M}, trije_{NOM M}/štirje_{NOM M}
 učbeniki_{NOM PL M}
 ena_{NOM SG F} knjiga_{NOM SG F}, dve_{NOM F} knjigi_{NOM DU F}, tri_{NOM F}/štiri_{NOM F} knjige_{NOM PL F}
 eno_{NOM SG N} pismo_{NOM SG N}, dve_{NOM N} pismi_{NOM DU N}, tri_{NOM N}/štiri_{NOM N} pisma_{NOM PL N}

¹ Štokavščino razumemo kot nadsistem, ki zaradi istih jezikovnih pojavov združuje osrednji južnoslovanski prostor in zajema bosanski, črnogorski, hrvaški in srbski jezikovni standard. Našteti standardi se ne razlikujejo glede skladijskih razmerij v obravnavanih količinskih zgradbah, zato jih v besedilu ne obravnavamo posamezno.

² Termina *malí števnikí in veliki števnikí* prevzemamo po angleških izrazih »low numerals« in »high numerals«. V sodobni slovnici srbskega jezika (Piper, Klajn, Dragičević 2022: 149–150) se glavni števnik, večji od 1, delijo na pavkalne (2, 3 in 4) in nepavkalne (od 5 naprej).

³ O posebnostih količinskih zgradb s t. i. malimi števnikami gl. Uhlik, Žele 2023.

⁴ Iz obravnave so izvzeti večmestni števnikami, ki se končujejo z 1, 2, 3 in 4 (o teh gl. Uhlik, Žele 2023).

⁵ Pri glosiranju so uporabljeni mednarodni simboli in okrajšave, navedeni v poglavju na koncu prispevka OKRAJŠAVE.

- (2) štok. jedan_{NOM SG M} udžbenik_{NOM SG M}, dva_{NOM M/tri_{NOM}}/četiri_{NOM} udžbenika_{NOM PAU M}
 jedna_{NOM SG F} knjiga_{NOM SG F}, dvije_{NOM F/tri_{NOM}}/četiri_{NOM} knjige_{NOM PAU F}
 jedno pismo, dva_{NOM N/tri_{NOM}}/četiri_{NOM} pisma_{NOM PAU N}
- (3) rušč. один_{NOM SG M} učebnik_{NOM SG M}, два_{NOM M/tri_{NOM}}/четыре_{NOM} учебника_{NOM PAU M}
 одна_{NOM SG F} книга_{NOM SG F}, две_{NOM M/tri_{NOM}}/четыре_{NOM} книги_{NOM PAU F}
 одно_{NOM SG N} письмо_{NOM SG N}, два_{NOM M/tri_{NOM}}/четыре_{NOM} письма_{NOM PAU N}

V treh obravnavanih jezikih, tj. v slovenščini, štokavščini in ruščini, ugotavljamo odstopanja od ujemanja pri količinskih zvezah, ki vsebujejo števniki od vključno 5 naprej. Posebnosti omenjenih zgradb s števnikami od 5 dalje, ki jih prevzeto poimenujemo veliki števnikami (»high numerals«), bodo v nadaljevanju natančneje predstavljene.

1 SKLADENJSKA RAZMERJA ZNOTRAJ BESEDNE ZVEZE Z VELIKIMI ŠTEVNIKI

V količinskih zvezah z velikimi števnikami (od 5 naprej), pri katerih je glavni števniki v imenovalniku in tožilniku, se samostalniška beseda ne ujema s števnikom in je v slovenščini, štokavščini in ruščini v vseh spolih v roditeljski množini. Ker je imenovalniški ali tožilniški števniki jedro zveze, ki na izrazni ravni usmerja sklonsko obliko samostalnika, govorimo o vezavi (4):

- (4) slvn. pet_{NOM} učbenikov_{GEN PL M}, pet_{NOM} knjig_{GEN PL F}, pet_{NOM} pisem_{GEN PL N}
 štok. pet_{NOM} udžbenikâ_{GEN PL M}, pet_{NOM} knjigâ_{GEN PL F}, pet_{NOM} pisama_{GEN PL N}
 rušč. пять_{NOM} учебников_{GEN PL M}, пять_{NOM} книг_{GEN PL F}, пять_{NOM} писем_{GEN PL N}

V slovenščini in ruščini, kjer se števnikami tipa 5 sklanjajo,⁶ je to vezavno razmerje značilno samo za zgradbe, kjer je števniki v imenovalniku ali tožilniku. V vezavnem razmerju se števnikami, kot je 5, obnašajo kot merni samostalniki. Ob njih se merjena vsebina prav tako izraža z roditeljsko obliko ednine ali množine, npr. *vreča moke*_{GEN SG}, *vreča krompirja*_{GEN SG}, *vreča jabolka*_{GEN PL}.

⁶ Števnikami od 5 do 10 naj bi bili v baltoslovanščini še nesklonljivi, sklonljivi so postali v praslovanščini (Vaillant 1958: 632). V sodobni ruščini se števnikami, kot je 5, sklanjajo kot *i*-jevski samostalniki v ednini; v sodobni slovenščini pa v stranskih sklonih prevzemajo množinske končnice pridevnika. Kot zanimivost lahko omenimo, da so se v slovenščini 16. stoletja glavni števnikami od 5 občasno uporabljali kot nesklonljivi pridevniki (Ahačič 2017: 60).

Če pa je količinska zveza s števnikom, kot je 5, v stranskem netožilniškem sklonu, se med števnikom in samostalnikom pojavlja ujemalno razmerje (5). V tem primeru sta tako števnik kot samostalnik v istem sklonu.⁷ Samostalnik, ki označuje merjeno količino, je jedro količinske zveze. V zgledu (5) se celotna količinska zveza, znotraj katere velja ujemalno razmerje, ravna po konkretnem predlogu.

- (5) slvn. s petimi_{INS} učbeniki_{INS PL M}, s petimi_{INS} knjigami_{INS PL F}, s petimi_{INS} pismi_{INS PL N}
 rušč. с пятью_{INS} учебниками_{INS PL M}, с пятью_{INS} книгами_{INS PL F}, с пятью_{INS} письмами_{INS PL N}

slvn. v petnajstih_{LOC} obrokih_{LOC PL M}

rušč. в пятнадцати_{LOC} платежах_{LOC PL M}

Razlika med vezavo (4) in ujemanjem (5) nakazuje spremenljivo skladijsko obnašanje števnik, to je, da ima v imenovalniku in tožilniku vlogo skladijskega jedra (*pet knjig*), v stranskih sklonih pa vlogo nadzorovanega določila (*o petih knjigah*). Štokavščina se pri tem razlikuje od slovenščine in ruščine. V štokavščini so števniki, kot je 5, v nasprotju s slovenščino in ruščino, nesklonljivi. V primerih količinskih zvez z nesklonljivimi števnik, tipa 5, pa lahko pogojno govorimo o sklonljivosti celotne količinske zveze kot enovite skladijskopomenske enote. Zaradi lažje analize v spodnjih primerih (6)–(11) nesklonljivim števnikom pripisujemo isti sklon, kot bi ga imela sklonljiva zveza v istem položaju.

V štokavskih količinskih zvezah nesklonljivi števnik vzpostavlja vezavno razmerje, pri katerem je števnik skladijsko jedro zveze z rodilniško obliko samostalnika. To razmerje je tako v imenovalniški zvezi (6) kot pri zvezah v stranskih sklonih, npr. v tožilniku (7) in rodilniku (8).

- (6) štok. Pet_{NOM} udžbenikâ_{GEN PL M}/knjigâ_{GEN PL F}/pisama_{GEN PL N} nalazi se na stolu.

- (7) štok. Vidio je pet_{ACC} udžbenikâ_{GEN PL M}/knjigâ_{GEN PL F}/pisama_{GEN PL N}.

- (8) štok. Bojao se pet_{GEN} policajaca_{GEN PL M}/ženâ_{GEN PL F}.

Za dajalniške, orodniške in mestniške zveze pa velja, da se količinske zveze z velikimi števnik uporablajo predvsem v predložni rabi (9), (10), (11).

⁷ Po Ramovšu (1947–1949: 112) se števniki kot 5 v stranskih sklonih pregibajo kot pridevniki, v imenovalniku in tožilniku pa so kot samostalniki.

Predlogu sledi nesklonljiva oblika velikega števnik v stranskem sklonu, ki veže rodilniško obliko samostalnika množine (9), (10), (11). To vezavno razmerje lahko razložimo s tem, da so se veliki števnik razvojnno obnašali kot merni samostalniki, ob katerih se števna vsebina v vseh sklonih izraža z rodilniško obliko, npr. *s skupino prijatelj^{ev}_{GEN PL}*, *pri skupini prijatelj^v_{GEN PL}*.

(9) štok. prema udžbenik^â_{GEN PL M}/knjig^â_{GEN PL F}/pisama_{GEN PL N}

(10) štok. o pet_{LOC} udžbenik^â_{GEN PL M}/knjig^â_{GEN PL F}/pisama_{GEN PL N}

(11) štok. s pet_{INS} udžbenik^â_{GEN PL M}/knjig^â_{GEN PL F}/pisama_{GEN PL N}

Posebnost štokavščine je, da se brezpredložne količinske zveze ne uporabljajo v vezavno predvidljivem dajalniku ali orodniku (12)⁸ (o tem podrobneje gl. tudi N. Leko (2009:51)).

(12) štok. *Zavidim pet_{DAT} žen^â_{GEN PL F}. (neustrezna raba ob brezpredložnem dajalniku)

štok. *Prilazim pet_{DAT} žen^â_{GEN PL F}. (neustrezna raba ob brezpredložnem dajalniku)

štok. *Crtam pet_{INS} bojic^â_{GEN PL F}. (neustrezna raba ob brezpredložnem orodniku) vs.

Crtam s pet_{INS} bojic^â_{GEN PL F}. (ustrezna raba ob predložnem orodniku)

Kako razložiti to štokavsko posebnost? Števnik, kot je 5, se po sklonih v štokavščini ne pregibajo; v vseh sklonih jim sledi samostalnik v rodilniku. Šele predložna raba ob količinskih zvezah v dajalniku in orodniku razdvoumlja sporočilo. Naj ponazorimo s primerom: sredstvo dejanja se v štokavščini izraža z brezpredložnim orodnikom, npr. *Crtam bojicom_{INS SG F}* 'Rišem z barvico'. Če pa imamo v vlogi sredstva zvezo s 5, dobimo *Crtam pet_{INS/ACC} bojic^â_{GEN PL F}*. Toda prednostno razumevanje tega stavka je 'Rišem pet barvic', in šele potem pomislimo na drugo možnost 'Rišem s petimi barvicami'. Zato se zaradi razdvoumljanja v teh primerih uporablja predložna zveza, npr. *Crtam s pet_{INS} bojica_{GEN PL F}* 'Rišem s petimi barvicami'. Če pa v štokavščini števniki 5 nadomestimo s števniki samostalnikom *petorica*, se količinska zveza lahko uporablja tudi v brezpredložnem dajalniku – saj sklonskorazločevalna oblika števnikega samostalnika odpravlja dvoumnost (13).

⁸ Te zadrege pa ni v mestniku, ker je, tako kot v slovenščini in ruščini, vedno predložni, npr. *o pet_{LOC} udžbenik^â_{GEN PL M}* vs. **pet_{LOC} udžbenik^â_{GEN PL M}*.

(13) štok. Zavidim petorici_{DAT SG F} prijateljâ_{GEN PL M}. Ona prilazi petorici_{DAT SG F} prijateljâ_{GEN PL M}.

V makedonščini pri večini primerov (14), zlasti pri samostalnikih ženskega in srednjega spola, na samostalniško obliko ne vpliva, ali je pred samostalniškim delom zveze mali števnik (od 2 do 4) ali pa veliki števnik (od 5 naprej). V makedonskih količinskih zvezah z glavnimi števnik, večjimi od 1, se uporabljajo predvsem množinske oblike samostalnikov. Števnik, večji od 1, se torej ujema z množinsko obliko in zato lahko govorimo o **ujemanju**.

(14) mak. два/три/четири/пет учебници_{PL M}
 две/три/четири/пет книги_{PL F}
 две/три/четири/пет писма_{PL N}

Poseben tip množine je t. i. **števna množina** s končnico na *-a*.⁹ Ta tip množine je značilen za nekatere samostalnike v makedonščini, ki označujejo osebe moškega spola (15) ali pa so enozložni (16).¹⁰ Vendar je treba pripomniti, da je raba posebne števne množine v makedonščini največkrat fakultativna. Števna oblika na *-a* je doslednejša ob števnikih 2 in 3 pri enozložnih samostalnikih moškega spola (16).

(15) mak. два/три/четири/пет студента_{PL COUNT M}

(16) mak. два/три/четири/пет дена_{PL COUNT M}

Prikazano stanje s količinskimi zvezami, ki v slovenščini, štokavščini, ruščini in makedonščini vključujejo velike števnik od 5 naprej, lahko ponazorimo z razpredelnico, ki prikazuje razmerja znotraj količinske zveze: vezavno in ujemalno razmerje. V slovenščini in ruščini, kjer so števnik kot 5 sklonljivi, sta možna tako vezava s števnikom v imenovalniku in tožilniku kot ujemanje, ko je količinska zveza s števnikom v drugih stranskih sklonih. Za makedonščino kot jezik z reducirano sklonsko rabo je značilno ujemanje množinskega števnik in množinskega samostalnika. V štokavščini, kjer so števnik kot 5 nesklonljivi, pa se ob tovrstnih števnikih uporablja vezavno razmerje.

⁹ Samostalniške oblike na *-a* se razvojno razlagajo kot ostanki dvojine, ki so se posplošili tudi na količinske zveze z drugimi večjimi števnik.

¹⁰ Za komentarje makedonskega gradiva se zahvaljujeva Liljani Mitkovski in Eleni Bužarovski.

jezik	skladenjski odnos med števnikom in samostalnikom
slovenščina	vezava/ujemanje
ruščina	vezava/ujemanje
štokavščina	vezava
makedonščina	ujemanje

RAZPREDELNICA 1: Skladenjska razmerja v količinskih zvezah med glavnim števnikom in samostalnikom

1.1 KOLIČINSKE ZGRADBE Z 'VEČ'

Ob analizi skladenjskih razmerij v količinskih zgradbah z velikimi števnikami izpostavljamo tudi zveze s količinskim izrazom *več* ob števniki količini, saj tovrstne zgradbe v slovenščini in štokavščini vzpostavljajo podobna razmerja kot veliki števnik. Obenem se v zgradbah z 'več' pokažejo razlike med slovenščino in štokavščino.¹¹ Potrebno je opozoriti, da južnoslovanski leksemi slvn. *več*, štok. *više*, mak. *повеќе*, ki v rabi, brez primerjave, s števnim samostalnikom označujejo količino, večjo od 'nekaj' in manjšo od 'veliko', nimajo enakopomenskih ustreznice v ruščini. To postane očitno iz ruskih ustreznice v (17), kjer so omenjeni južnoslovanski količinski izrazi prevedeni z *несколько*. Ruski *несколько* pa izraža manjšo količino kot spodaj naštetih južnoslovanski količinski izrazi: slovenski in štokavski izraz lahko označujeta količino, večjo od 10, kar je za ruski *несколько* nenavadno.¹²

(17) slvn. Ana pozna **več**_{QUANT} **odvetnikov**_{GEN PL}.

štok. Ana zna **više**_{QUANT} **advokatâ**_{GEN PL}.

mak. Ана знае **повеќе**_{QUANT} **адвокати**_{PL}.

≈ rušč. Анна знает **нескольких**_{QUANT} **адвокатов**_{PL}.

V nadaljevanju sopostavljamo skladenjska razmerja v slovenskih in štokavskih zvezah z 'več'. Če je količinska zveza v imenovalniku ali tožilniku, se enako kot pri velikih števnikih v slovenščini in štokavščini vzpostavlja vezavno razmerje, pri makedonščini pa izrazu *повеќе* sledi množinska oblika samostalnika (17).

¹¹ Enaka razmerja v količinskih zgradbah vzpostavljajo tudi nedoločni števnik, kot npr. *malo*, *veliko*, *mного*.

¹² Ruski količinski izraz *несколько* je v pridevniški vlogi ob samostalniku pregiben in končniško ujemanen s samostalniškim jedrom, npr. *с несколькими*_{QUANT INS} *знакомыми*_{INS PL M}. Za komentarje ruskega gradiva se zahvaljujemo Mihailu Oslonu in Mihailu Sajenku.

Do razlik med slovenščino in štokavščino prihaja, ko so količinske zveze z 'več' v stranskih sklonih: slovenski *več* se obnaša tako kot nesklonljivi prilastek znotraj količinske zveze in ne vpliva na obliko samostalniške zveze. Ta samostalniška zveza pa se namreč v slovenščini vezavno prilagodi povedku in potencialno dodanemu predlogu (18):

(18) slvn. Ana se je pogovarjala z **več**_{QUANT} **znanci**_{INS PL M.}

štok. Ana je razgovarala s **više**_{QUANT} **poznanikâ**_{GEN PL M.}

mak. Ана разговарала со **повеќе**_{QUANT} **познаници**_{PL M.}

≈ rušč. Анна разговаривала с **несколькими**_{QUANT INS} **знакомыми**_{INS PL M.}

V štokavščini je s skladijskega stališča količinska zveza z izrazom *više* primerljiva s količinsko besedo z velikimi števnik, npr. *s više poznanikâ*_{GEN PL M} ≈ *s pet poznanikâ*_{GEN PL M}. V količinskih zvezah z *više* se v nasprotju s slovenščino v vseh sklonih pojavlja vezavno razmerje, pri katerem obliko rodilniškega samostalnika določa količinski izraz *više* znotraj količinske zveze, in ne povedek ali predlog, ki je zunaj količinske zveze.

Primerjava samostalniških oblik ob količinskem izrazu v slovenščini in štokavščini kaže na to,¹³ da je v slovenščini znotraj količinske zveze izrazita stalna težnja k ujemanju. To težnjo po ujemanju dodatno potrjuje govorjena različica količinske zveze s pregibnim *več* (19). V govorjenem jeziku je med količinskim izrazom in samostalniško besedo končniško ujemanje. V nekodificirani rabi se sklonljivi *več* prilagaja samostalniškemu delu količinske zveze.

(19) slvn. Pisal je **večim**_{DAT} **prijateljem**_{DAT PL M.} (pog.)

slvn. Potoval je z **večimi**_{INS} **prijatelji**_{INS PL M.} (pog.)

1.2 KOLIČINSKE ZGRADBE S PREDŠTEVNIŠKIMI PRILASTKI

Če hočemo opozoriti na razmerja znotraj količinskih besednih zvez, moramo izpostaviti tudi zveze, v katerih so količinski zvezi dodana predštevniška prilastkova določila, npr. *Tistih/ Vseh/Celih pet paketov je dostavil*.¹⁴ Kljub temu, da se prispevek osredotoča na količinske zgradbe s t. i. glavnimi velikimi

¹³ Primerjava skladijskih razmerij, ki se izraža s pomočjo sklonskih oblik, je smiselna za slovenščino in štokavščino. V makedonščini ni ohranjenih imenskih sklonskih oblik.

¹⁴ Suprun (1969: 156–157) ugotavlja, da takšna raba prilastkov naj ne bi bila značilna za praslovanske števnik od 1 do 4, ki naj bi bili takrat v vlogi pridevniških prilastkov. Iz tega lahko sledi, da je raba predštevniških prilastkov bila značilna šele od 5 naprej.

števnik, je v kontekstu rabe predštevniških prilastkov potrebno zaradi skladenskih posebnosti upoštevati tudi zgradbe z malimi števnikami (od 1 do 4). V slovenskih količinskih zvezah do vključno 4 namreč predštevniški prilastek kot modifikator s količinsko zvezo vzpostavlja ujemalno razmerje (20):

(20) slvn. tisti_{NOM SG M} eden_{NOM SG M} učbenik_{NOM SG M}

tista_{NOM SG F} ena_{NOM SG F} knjiga_{NOM SG F}

tisto_{NOM SG N} eno_{NOM SG N} pismo_{NOM SG N}

slvn. tista_{NOM DU M} dva_{NOM M} učbenika_{NOM DU M}

tisti_{NOM DU F} dve_{NOM F} knjigi_{NOM DU F}

tisti_{NOM DU N} dve_{NOM N} pismi_{NOM DU N}

slvn. tisti_{NOM PL M} trije_{NOM M}/štirje_{NOM M} učbeniki_{NOM PL M}

tiste_{NOM PL F} tri_{NOM F}/štiri_{NOM F} knjige_{NOM PL F}

tista_{NOM PL N} tri_{NOM N}/štiri_{NOM N} pisma_{NOM PL N}

V ustreznih štokavskih števnških zgradbah z malimi števnikami od 2 do 4 se predštevniški prilastek ujema s pavkalnimi oblikami samostalnikov (o tem gl. tudi Browne 2022). Kot je razvidno iz (21), se pavkalna oblika predštevniškega prilastka izrazno razlikuje od množinske zgolj v moškem spolu (22).

(21) štok. ona_{NOM PAU M} dva_{NOM M}/tri_{NOM}/četiri_{NOM} udžbenika_{NOM PAU M}

one_{NOM PAU F} dvije_{NOM F}/tri_{NOM}/četiri_{NOM} knjige_{NOM PAU F}

ona_{NOM PAU N} dva_{NOM N}/tri_{NOM}/četiri_{NOM} pisma_{NOM PAU} (oblike ob malih števnikih)

(22) štok. oni_{NOM PL M} udžbenici_{NOM PL M}

one_{NOM PL F} knjige_{NOM PL F}

ona_{NOM PL N} pisma_{NOM PL N} (množinske oblike)

V ruščini se predštevniški prilastek pred malimi števnikami navadno pri vseh spolih uporablja v imenovalniški obliki množine (23).¹⁵ Tako se predštevniški

¹⁵ Če je pred zvezo omenjenih števnškov in samostalnikov nedoločni količinski pridevnik *целый*, ima zaznamovan pomen 'cel, ki poudarja količino' in je njegova oblika enaka kot oblika pridevnškega prilastka, ki je med števnškom in samostalnikom: v vseh spolih je v roditelju množine *целых два (важных) учебника* 'cela dva (pomembna) učbenika', *целых два важных письма* 'celi dve pomembni pismi', *целых две важных*

prilastek v ruščini razlikuje od pridevniškega prilastka ob števnem samostalniku. V moškem in srednjem spolu ima pridevniški prilastek obliko, homonimno z rodilnikom množine, v ženskem spolu pa je najpogosteje sklonsko homonimen z imenovalnikom množine.

- (23) rušč. **эти**_{NOM PL} два/три/четыре **важных** учебника_{NOM PAU M}
 rušč. **эти**_{NOM PL} две/три/четыре **важные/важных** книги_{NOM PAU F}
 rušč. **эти**_{NOM PL} два/три/четыре **важных** письма_{NOM PAU N}

V slovenščini in štokavščini se v količinskih zvezah s števnikami od vključno 5 naprej, tako kot pri samostalniških besedah, skladijsko razmerje spremeni. V npr. *tistih pet hiš* je predštevniški prilastek *tisti* podrejen števniku, hkrati pa se ujema s samostalniškim določilom, ki je v rodilniški množinski obliki. Lahko bi rekli, da števniki vezavno usmerja obe določili, samostalniško in predštevniško prilastkovo. V zvezah *tistih pet hiš* in *pet tistih hiš* sta rabljeni isti skladijski razmerji (24).

- (24) slvn. *tistih*_{GEN PL} *pet*_{NOM} učbenikov_{GEN PL M}
*tistih*_{GEN PL} *pet*_{NOM} knjig_{GEN PL F}
*tistih*_{GEN PL} *pet*_{NOM} pisem_{GEN PL N}

Isto velja za zglede v štokavščini s števnikom 5 v imenovalniku, kjer se predštevniški prilastek *oni* vedno, ne glede na spol, pojavlja v rodilniški množinski obliki (25).

- (25) štok. *onih*_{GEN PL} *pet*_{NOM} udžbenikâ_{GEN PL M}
*onih*_{GEN PL} *pet*_{NOM} knjigâ_{GEN PL F}
*onih*_{GEN PL} *pet*_{NOM} pisamâ_{GEN PL N}

Predštevniški prilastek v štokavščini ima, tako kot samostalnik ob števniku, v vseh sklonih količinske zveze s 5 isto rodilniško obliko, npr. *onih pet udžbenika*, *od onih pet udžbenikâ*, *prema onih pet udžbenikâ*, *onih pet udžbenikâ*, *o onih pet udžbenikâ*, *s onih pet udžbenikâ*. Medtem pa se *tistih* v slovenskih

книги 'celi dve pomembni knjigi'. Če pa ima pridevnik *целый* pomen 'nepoškodovan', je njegov položaj v količinski zvezi za števnikom.

količinskih zvezah s 5 ravna po samostalniškem delu količinske zveze, tj. razen v imenovalniku in tožilniku je s samostalnikom v ujemalni obliki, npr. *tistih pet učbenikov, od tistih petih učbenikov, proti tistim petim učbenikom, tistih pet učbenikov, o tistih petih učbenikih, s tistimi petimi učbeniki*.

V makedonščini se ob vseh števnikih, večjih od 1, pojavlja identična množinska oblika predštevniškega prilastka *оние* (26), (27):

(26) mak. *оние*_{PL} *два* *учбеника*_{PL COUNT M/учебници}_{PL M}

*оние*_{PL} *две* *книги*_{PL F}

*оние*_{PL} *две* *писма*_{PL N}

(27) mak. *оние*_{PL} *пет* *учебници*_{PL M}

*оние*_{PL} *пет* *книги*_{PL F}

*оние*_{PL} *пет* *писма*_{PL N}

Predštevniški prilastek se v ruščini ob velikih števnikih (od 5 naprej), če je količinska zveza v imenovalniku, tako kot pri malih števnikih, uporablja v imenovalniški pridevniški obliki množine *эти*. To pomeni, da je v nasprotju s slovenščino in štokavščino njegova vloga skladiščno neodvisna od števnikova oziroma količinske zveze, ki mu sledi (28).

(28) rušč. *эти*_{NOM PL} *пять*_{NOM} *учебников*_{GEN PL M}

*эти*_{NOM PL} *пять*_{NOM} *книг*_{GEN PL F}

*эти*_{NOM PL} *пять*_{NOM} *писем*_{GEN PL N}

Imenovalniška množinska oblika *эти* 'tisti' je značilna zgolj za predštevniški prilastek, in ne za pridevniški prilastek, ki je med števnikom in samostalnikom v količinski zvezi. Ta prilastek se namreč tako kot samostalnik vezavno podreja števniku in ima stalno rodilniško množinsko obliko (gl. *этих* v (29)).

(29) rušč. *пять*_{NOM} *этих*_{GEN PL} *учебников*_{GEN PL M}

*пять*_{NOM} *этих*_{GEN PL} *книг*_{GEN PL F}

*пять*_{NOM} *этих*_{GEN PL} *писем*_{GEN PL N}

2 PREDIKACIJSKO RAZMERJE Z VELIKIMI ŠTEVNIKI V OSEBKVI VLOGI

Količinska zveza z glavnimi velikimi števnikmi v stavčni vlogi osebkva odpira vprašanja, kako količinska zveza vpliva na obliko povedka in posledično, katera razmerja se med osebkom in povedkom lahko vzpostavljajo. Pri tem je treba opozoriti tudi, v kolikšni meri se količinske zveze z glavnimi velikimi števnikmi razlikujejo od količinskih zvez z malimi števnikmi. Opozorili bomo na posebnosti in razlike, ki se kažejo zlasti med slovenščino in štokavščino, sopostavitveno bosta dodani še ruščina in makedonščina.

2.1 PREDIKACIJSKO RAZMERJE Z VELIKIMI ŠTEVNIKI V SLOVENŠČINI IN ŠTOKAVŠČINI:

PREVLADA EDNINSKEGA POVEDKA

Predikacijsko ujemanje s količinskimi zvezami, ki v slovenščini in štokavščini vključujejo male števnikme, se izraža z **needninskimi oblikami povedka**. Iz slovenskih zgledov (30), (31) je razvidno, da sam števniki v tovrstnih količinskih zvezah ne vpliva na obliko povedka. Obliko povedka v teh primerih določa samostalniški del količinske zveze, ki je v osebkvi vlogi. V (30) obliko povedka določa samostalniški del zveze *fanta* v dvojini, v zgledu (31) pa samostalniški del *fantje* v množini. Med količinsko zvezo v osebkvi in povedkom se vzpostavlja pomensko-izrazno ujemanje, ki ga lahko imenujemo **navadno/kanonično ujemanje**.

(30) slvn. (Dva_{NOM M}) Fanta_{NOM DU M} odhajata_{PRS DU} na zabavo. (Dva_{NOM M}) Fanta_{NOM DU M} sta_{AUX PRS DU} odšla_{LPTCP DU M} na zabavo.

(31) slvn. Trije_{NOM M}/štirje_{NOM M} fantje_{NOM PL M} odhajajo_{PRS PL} na zabavo. Trije/štirje fantje_{PL} so_{AUX PRS DU} odšli_{LPTCP PL M} na zabavo.

V štokavskih ustreznica se uporabljajo needninske oz. pavkalne oblike (32), ki praviloma sovpadajo z množinskimi oblikami (34). Posebnost so pretekliške zložene glagolske oblike, pri katerih ima lahko moška oblika značilno pavkalno končnico *-a* (33).¹⁶ Naj dodamo, da v štokavščini ni razlike med oblikami povedkov, ki se navezujejo na količinske zveze z 2 in na količinske zveze s 3 in 4 (32), (33).

¹⁶ Podrobneje o tem gl. Uhlik, Žele 2023.

- (32) štok. Dva_{NOM M}/tri_{NOM}/čtetiri_{NOM} momka_{NOM PAU M} odlaze_{PRS PAU} na zabavu.
- (33) štok. Dva_{NOM M}/tri_{NOM}/čtetiri_{NOM} momka_{NOM PAU M} sua_{AUX PRS PL} otišla_{LPTCP PAU M}/otišli_{LPTCP PL M} na zabavu.
- (34) štok. Momci_{NOM M PL} odlaze_{PRS PL} na zabavu. Momci_{NOM M PL} su_{AUX PRS PL} otišli_{LPTCP PL M} na zabavu.

Pri količinskih zvezah z velikimi števnikami pa se razmerje med osebkom in povedkom spremeni oz. ni več navadnega kanoničnega ujemanja. Samostalniški del količinske zveze ob velikih števnikih, npr. slvn. *pet prijatelj*_{GEN PL M}, *pet prijateljâ*_{GEN PL M}, ni v imenovalniku. V teh zgradbah oblike povedka ne more narekovati neimenovalniški samostalnik, ampak jo lahko določa le celotna količinska zveza. Za predikacije s količinskimi zvezami z velikimi števnikami je zato značilno posebno t. i. **izrazno (ad formam), slovnično ali privzeto ujemanje**, pri katerem količinska zveza deluje kot enovita celota oz. kot posebna zgradba, ki povedku določa edninsko srednjepolsko obliko za vse spole. Način privzetega ujemanja z edninskimi povedki v srednjem spolu je značilen za slovenščino (35) in prevladujoč tudi v štokavščini (36).

- (35) slvn. Pet_{NOM} fantov_{GEN PL M} je_{AUX PRS 3SG} odšlo_{LPTCP SG N} na zabavo.
 Pet_{NOM} žen_{GEN PL F} je_{AUX PRS 3SG} odšlo_{LPTCP SG N} na zabavo.
 Pet_{NOM} pisem_{GEN PL N} je_{AUX PRS 3SG} odšlo_{LPTCP SG N} v tujino.
- (36) štok. Pet_{NOM} momaka_{GEN PL M} je_{AUX PRS 3SG} otišlo_{LPTCP SG N} na zabavo.
 Pet_{NOM} ženâ_{GEN PL F} je_{AUX PRS 3SG} otišlo_{LPTCP SG N} na zabavo.
 Pet_{NOM} pisama_{GEN PL F} je_{AUX PRS 3SG} otišlo_{LPTCP SG N} u inozemstvo.

Veliki števnikami tipa 5 so bili razvojno gledano samostalniki,¹⁷ ki so sčasoma izgubili svoje samostalniške kategorialne lastnosti (v štokavščini so tovrstni števnikami izgubili pregibnost). Veliki števnikami imajo danes enako skladijsko vlogo kot količinski izrazi tipa *veliko* in *malo* in količinske zveze z obojimi zahtevajo povedek v srednjem spolu ednine (37).

- (37) slvn. Veliko_{QUANT}/Malo_{QUANT} fantov_{GEN PL M} je_{AUX PRS 3SG} odšlo_{LPTCP SG N}.
 štok. Puno_{QUANT}/Malo_{QUANT} momaka_{GEN PL M} je_{AUX PRS 3SG} otišlo_{LPTCP SG N}.

¹⁷ Suprun (1969: 171) npr. navaja, da so bili števnikami, kot je 5, *i*-jevski samostalniki ženskega spola, ki so lahko tudi določali žensko obliko zloženega povedka. Tovrstnega skladijskega razmerja danes v slovanskih jezikih ni.

Naj poudarimo, da privzeti tip ujemanja s srednjespolskimi končnicami ne presega mej stavka: če matičnemu stavku z glavnim velikim števnikom v osebkovi vlogi dodamo prislovni odvisni stavek, je oblika povedka v odvisnem stavku z neizraženo količinsko zvezo v množinski obliki. V zgledu (38) se namreč množinska oblika povedka v vzročnem odvisniku pomensko ujema z neizraženim množinskim osebkom (slvn. *oni fantje*, štok. *oni dječaci*), ki sovпада z osebkom matičnega stavka.

- (38) slvn. Pet_{NOM} fantov_{GEN PL M} je_{AUX PRS 3SG} ostalo_{LTCP SG N} brez večerje, ker so_{AUX PRS 3PL} zamudili_{LTCP PL M}.
 štok. Pet_{NOM} dječakâ_{GEN PL M} ostalo_{LTCP SG N} je_{AUX PRS 3SG} bez večere jer su_{AUX PRS 3PL} zakasnili_{LTCP PL M}.

Sicer pa se v štokavščini ob prevladujočem privzetem ujemanju v matičnem stavku s količinskimi zvezami z velikimi števnikmi pojavlja tudi tip ujemanja, ko osebkova količinska zveza zahteva množinsko obliko povedka (39), (40). V tem primeru, ko se povedek ravna po množinskem pomenu samostalniške zveze v osebkovi vlogi, lahko govorimo o **pomenskem ujemanju (ad sensum)**. Pojav pomensko-izraznega usklajevanja povedka z množinskim osebkom se v sodobnem jezikoslovju imenuje pomensko ujemanje (»semantic agreement«).

- (39) štok. Prvih_{GEN PL} pet_{NOM} nositeljâ_{GEN PL M} turnira su_{AUX PRS 3PL} najbolji_{NOM PL M} na svijetu.
 (Hrvatski jezični korpus Riznica)
 (40) štok. Svih_{GEN PL} pet_{NOM} igračâ_{GEN PL M} imaju_{PRS 3PL} mogućnost kreacije. (Hrvatski jezični korpus Riznica)

V zgledih (39) in (40) se množinske oblike povedka pojavljajo ob samostalniških zvezah, za katere je značilna določena vsebinska konkretnost, ki jo potrjuje predštevniški prilastek, npr. *prvih*_{GEN PL} ali *svih*_{GEN PL}. Ta način določanja in izvzemanja posamezne enote iz ostale množice je lahko v štokavščini eden izmed dejavnikov, ki vpliva na rabo množinske, in ne edninske srednjespolske, oblike povedka (prim. Milosavljević 2018).

2. 2 PREDIKACIJSKO RAZMERJE V RUŠČINI: VARIABILNOST

Ruske količinske zveze z malimi in velikimi števnikmi, v nasprotju s slovenščino in štokavščino, dopuščajo oba tipa ujemanja: ob količinskih zvezah s števnikmi,

večjimi od 1, v osebki vlogi je povedek ali v množinski ali v edninski tretjeosebni obliki. Obstaja zgolj težnja, da količinska zveza z malimi števnikami določa množinski povedek, količinska zveza z velikimi števnikami pa edninski srednjepolski povedek. Navadno na izbiro povedkove oblike lahko vplivajo različni pomenski, skladijski in pragmatični dejavniki (Neset, Janda 2023, Циммерлинг 2019, Uhlik, Žele 2023).

Najdoslednejše pravilo je, da količinska zveza z imenovalniškim predštevniškim prilastkom zahteva povedkovo obliko v množini (41), (42).

(41) rušč. Эти_{NOM PL} пять_{NOM} женщин_{GEN PL F} всегда задают_{PRS 3PL} неприличные вопросы.
(tipski primer)

slvn. Teh_{GEN PL} pet_{NOM} žensk_{GEN PL F} vedno zastavlja_{PRS 3SG} nespodobna vprašanja.

(42) rušč. И тут появились_{РСТ PL} наши_{NOM PL} пять_{NOM} подруг_{GEN PL F}. (tipski primer)

slvn. Tu se_{REF} je_{AUX PRS 3SG} pojavilo_{LPTCP SG N} naših_{GEN PL} pet_{NOM} prijateljic_{GEN PL F}.

Na prednostno rabo množinske oblike vplivajo še manj odločilni dejavniki, kot so besedni red SV (stava osebka pred povedkom), živost in višja vršilskost osebka (43).

(43) rušč. При появлении подполковника пять_{NOM} человек_{GEN PL M} поднялись_{РСТ PL C} с мест.
slvn. Ob pojavitvi podpolkovnika je_{AUX PRS 3SG} pet_{NOM} ljudi_{GEN PL M} vstalo_{LPTCP SG N}.

V ruščini na izbiro edninskega povedka v srednjem spolu vplivajo zlasti besedni red VS (stava povedka pred osebkom), nizka vršilskost in neživost osebka. Zgled (44) kaže, kako je srednjepolska oblika povedka povezana s stavčnim položajem osebka; v prvem stavku, kjer je povedek pred osebkom (VS), je v rabi srednjepolska edninska oblika, medtem ko je v drugem stavku z navadno stavo osebek – povedek (SV) prevladujoča raba množinske oblike povedka.

(44) rušč. Тогда от токсического гепатита в Еланске умерло_{РСТ SG N} пять_{NOM} человек_{GEN PL M}, а несколько_{QUANT} десятков_{GEN PL} остались_{РСТ PL} инвалидами. (НКРЯ)
slvn. Takrat je_{AUX PRS 3SG} zaradi hepatitisa v Jelansku umrlo_{LPTCP SG N} pet_{NOM} ljudi_{GEN PL M}, nekaj_{QUANT} deset_{NOM} pa jih je_{AUX PRS 3SG} ostalo_{LPTCP SG N} invalidnih/invalidov.¹⁸

¹⁸ Posebnost slovenščine je, da količinskemu izrazu *nekaj* sledi števniki *deset* v imenovalniku, npr. *nekaj*_{QUANT} *deset*_{NOM} *ljudi*. Ta skladijska raba se torej razlikuje od zveze *nekaj* s samostalnikom, ki bi v nadaljevanju zahteval množinsko obliko rodilnika, npr. *nekaj*_{QUANT} *deset*_{GEN PL} *vojakov*.

2. 3 PREDIKACIJSKO RAZMERJE V MAKEDONŠČINI: MNOŽINSKI POVEDEK

V makedonščini ni nihanja v rabi med edninsko in množinsko obliko povedka: ob količinskih zvezah z glavnimi števnikami, večjimi od 1, sledi vedno množinska oblika povedka; v zgledu (45) je povedek v množinski obliki aorista. Naj spomnimo, da v makedonščini v količinskih zvezah samostalniška zveza ni izražena s stranskim sklonom (gl. POGlavJE 1). Torej v nasprotju s slovenščino množinski samostalnik v izhodiščni obliki narekuje zgolj množinsko obliko povedka (45). Zato v tem primeru govorimo o kanoničnem ujemanju, pri katerem pomenskost osebka tudi končniško izraža.

(45) mak. Деcет расипници_{PL M} уништија_{AOR PL} илјадници фирми.

slvn. Deset_{NOM} pokvarjenec_{GEN PL M} je_{AUX PRS 3SG} uničilo_{LPTCP SG N} na tisoče podjetij.

Prikazano stanje v slovenščini, štokavščini, ruščini in makedonščini lahko ponazorimo z razpredelnico, v kateri je poudarjena uporabljena povedkova oblika ob količinskih zvezah z velikimi števnikami. Očitno je, da skrajni mesti v razpredelnici zavzemata slovenščina, ki dovoljuje izključno edninski povedek, in makedonščina z izključno množinsko rabo povedka.

jezik	oblika povedka
slovenščina	V _{SG N} (privzeto ujemanje)
štokavščina	V _{SG N} / izjemoma V _{PL} (prevladujoče privzeto ujemanje)
ruščina	V _{PL} & V _{SG N} (soobstoj privzetega in pomenskega ujemanja)
makedonščina	V _{PL} (kanonično ujemanje)

RAZPREDELNICA 2: Skladenjska razmerja med količinsko zvezo v vlogi osebka in povedkom

2. 4 ZGRADBE Z VELIKIMI ŠTEVNIKI OB POVEDKU Z *BITI* + PRIDEVNIŠKA BESEDA

V slovenščini in štokavščini se v predikativnih zgradbah *biti* + *ADJ* v povezavi s količinsko zvezo z velikimi števnikami v osebki vlogi uporablja srednjepolska oblika pomožnega glagola.¹⁹

¹⁹ Kot izjemo v štokavščini lahko omenimo zgradbe, kjer je ob glavnem števniku predštevniški modifikator, ki lahko vpliva na množinsko obliko povedka. V slovnici Piper, Klajn, Dragičević (2022) je naveden primer količinske zveze brez samostalnika: *Tih poslednjih pet bile su najuspješnije* (2022: 349).

V štokavščini je tudi pridevniški/deležniški del predikativne zgradbe izražen s srednjim spolom ednine:

(46) štok. Primjetili smo da je_{AUX PRS 3SG} otvoreno_{PTCP SG N} pet_{NOM} prozorâ_{GEN PL M}. (rezultatívno stanje).

slvn. Opazili smo, da je_{AUX PRS 3SG} bilo_{LPTCP SG N} odprtih_{PTCP GEN PL} pet_{NOM} oken_{GEN PL N}.

(47) štok. Prošle godine je_{AUX PRS 3SG} izgrađeno_{PTCP SG N} pet_{NOM} kućâ_{GEN PL F}. (trpnik)

slvn. Lani je_{AUX PRS 3SG} bilo_{LPTCP SG N} izgrajenih_{PTCP GEN PL}/izgrajeno_{PTCP NOM SG N} pet_{NOM} hiš_{GEN PL F}.

Pri predikativnih zgradbah s pomožnimi ali oslabljenimi glagoli in pridevniškimi besedami se slovenščina od drugih izbranih slovanskih jezikov razlikuje. V zgradbi, ki izraža stanjskost, količinska zveza z velikimi števnikami v povedku določa rodilniško množinsko obliko pridevniške besede, npr. *Pet oken je bilo odprtih* (46). V tej zgradbi se rodilniška oblika *odprtih* ujema z rodilnikom samostalniškega dela količinske zveze *oken* torej se obnaša enako, kot da bi bila prilastek znotraj količinske zveze, npr. *pet odprtih oken*.²⁰ V trpniški zgradbi (47) pa slovenščina dopušča dve možnosti izražanja pridevniške besede znotraj povedka: pridevniška beseda je lahko v rodilniku množine, kot pri stanjski zgradbi, ali pa v imenovalniku srednjega spola, kot v štokavščini.

V ruščini med možnimi ustreznimi zgradbami obstaja tekmovalnost oblik: povedek in pridevniška beseda sta lahko v množinski ali pa v edninski srednjespolski obliki. Na to izbirno možnost, tako kot pri drugih tipih povedkov (gl. 2. 2), vpliva več dejavnikov, med njimi tudi položaj osebka glede na povedek: osebek za povedkom vpliva na večjo možnost rabe edninskega povedka (48), osebek pred povedkom pa določa množinsko obliko povedka (49).

(48) rušč. По официальным данным было_{PST SG N} убито_{PTCP SG N} пять_{NOM} человек_{GEN PL M}.

slvn. Po uradnih podatkih je_{AUX PRS 3SG} bilo_{LPTCP SG N} ubitih_{PTCP GEN PL} pet_{NOM} ljudi_{GEN PL M}.

(49) rušč. Девять_{NOM} человек_{GEN PL M} были_{PST PL} убиты_{PTCP PL} на месте, один скончался в больнице.

slvn. Deset_{NOM} ljudi_{GEN PL M} je_{AUX PRS 3SG} bilo_{LPTCP SG N} ubitih_{PTCP GEN PL}/ubito_{PTCP NOM SG N} takoj, eden je umrl v bolnišnici.

²⁰ Podoben primer, kjer se oblika pridevnika znotraj povedka ujema z rodilniško obliko samostalniškega dela količinske zveze, se rabi v slovenščini, če je v količinsko zvezo, ki je v osebku vlogi, vključen nedoločni števniki izraz, npr. Nekaj **džem-a**_{GEN SG M} je (bil-o) **pokvarjen-ega**_{GEN SG M}.

V makedonščini se tako glagolski kot pridevniški del povedka uporablja v množini (50), torej se *biti-stavki* glede oblike povedka ne razlikujejo od stavkov s polnompomenskimi glagolom.

(50) mak. Девет куќи_{PL F} беа_{IMPERF.3PL} потпојно уништени_{PTCP PL}.

slvn. Deset_{NOM} hiš_{GEN PL F} je_{AUX PRS 3SG} bilo_{LPTCP SG N} popolnoma uničenih_{PTCP GEN PL}.

V tu obravnavanih jezikih zgradbe z nepolnompomenskimi oz. pomožniškimi glagoli v povedku, ki jim sledijo pridevniške besede, do neke mere ponovijo opisano rabo s polnompomenskimi glagoli (gl. RAZPREDELNICO 2). Razlike med jeziki se kažejo pri oblikah pridevniške besede v povedku: slovenščina je med izbranimi jeziki edini jezik, ki lahko pridevniško besedo izraža z množinskim rodilnikom (za stanje in trpnik) ali v imenovalniku ednine v srednjem spolu (za trpnik). Na tak način se v slovenščini trpniške zgradbe s pomožnikom lahko razlikujejo od stanjskih zgradb.

jezik	oblika povedka
slovenščina	AUX _{SG N} + ADJ _{GEN PL/NOM SG N}
štokavščina	AUX _{SG N} + ADJ _{NOM SG N}
ruščina	AUX _{PL} + ADJ _{NOM PL} ; AUX _{SG} + ADJ _{NOM SG N}
makedonščina	AUX _{PL} + ADJ _{PL}

RAZPREDELNICA 3: Razmerja v predikaciji *biti* + pridevniška beseda

3 SKLEP

Slovenski jeziki so specifični zaradi skladenjskih značilnosti količinskih zgradb, ki vključujejo glavne števnik in samostalnik kot šteto vsebino. V prispevku obravnavamo dva tipa skladenjskih razmerij, in sicer razmerja med glavnim števnikom in samostalnikom znotraj količinskih zgradb in razmerja med količinskimi zgradbami v osebkovi vlogi in povedkom znotraj predikacije.

Znotraj količinskih zgradb se med glavnim števnikom in samostalnikom vzpostavljajo primeri kanoničnega ujemanja, hkrati pa prihaja tudi do odstopanj od ujemanja, in s tem do prehoda v vezavno razmerje. V slovenskih, štokavskih in ruskih besednih zvezah s števniki od 1 do 4 je značilno ujemanje števnik in samostalniškega dela količinske zveze. V slovenščini in ruščini se za velike števnik (od 5 naprej) v imenovalniku ali tožilniku uporablja vezava. V teh

primerih veliki števniki zahteva množinsko obliko samostalnika v rodilniku. Če pa je veliki števniki v slovenščini ali ruščini v netožilniških stranskih sklonih, se med njim in samostalnikom vzpostavlja ujemanje. V nasprotju s slovenščino in ruščino se v štokavščini, kjer so veliki števnikami nesklonljivi, uporablja zgolj vezavno razmerje: nesklonljivi števniki namreč vedno zahteva množinsko obliko v rodilniku. Posebnost štokavščine je, da se količinske zveze s števnikami, kot je 5, ne uporabljajo v brezpredložnem dajalniku ali brezpredložnem orodniku. V makedonščini s poenostavljenim sklonskim sistemom se ob števnikih, večjih od 1, pretežno uporabljajo množinske oblike samostalnikov. Posebne množinske oblike so t. i. števnice oblike samostalnikov moškega spola na *-a*, ki so najpogostejše pri enozložnih samostalnikih.

Podobno kot količinske zgradbe z velikimi števnikami se v slovenščini in štokavščini obnašajo tudi zveze s količinskimi izrazi z *več* (slvn.) in *više* (štok.). V teh zvezah se v slovenščini kaže težnja k ujemanju razmerju, ker *več* nima statusa jedra in se posebej v pogovornem jeziku ravna po sklonu samostalnika, npr. z *večimi prijatelji*. Nasprotno pa je štok. *više* skladenjsko jedro količinske zveze in zato samostalniku vedno določa množinsko rodilniško obliko, npr. s *više prijatelja*.

Znotraj predikacijskega razmerja se v obravnavanih jezikih med količinsko zvezo v osebkovi vlogi in povedkom vzpostavljata dva tipa ujemanja: izrazno oz. privzeto ujemanje z edninskim povedkom, in pomensko ujemanje z množinskim povedkom.

V slovenščini je ob količinskih zvezah s 5 v osebkovi vlogi stalnica t. i. izrazno oz. privzeto ujemanje. To ob omenjenih količinskih zvezah zahteva edninsko obliko povedka v srednjem spolu, npr. *Pet fantov je odšlo na zabavo*. V štokavščini prevladuje raba privzetega ujemanja, npr. *Pet momakâ je otišlo na zabavu*, ni pa izključena raba pomenskega ujemanja, npr. *Svih pet momaka su otišli na zabavu*, ko količinska zveza označuje konkretno in bolj določeno količino. V ruščini pa se glede na različne dejavnike, kot sta živost ali stava osebkovi glede na povedek, uporabljata tako pomensko kot privzeto ujemanje. V makedonskih količinskih zvezah se rabi izhodiščna množinska oblika samostalnika, ki vedno določa obliko povedka, npr. *Пет момци отидоа на забава*; to razmerje med osebkom in povedkom je primer kanoničnega ujemanja.

Ko je v predikacijsko razmerje vključen pomožni ali oslabljeni glagol s pridevniško besedo, se med jeziki v povedku, in sicer v obliki pridevniške besede, kažejo razlike. Slovenščina se pri tem od drugih izbranih slovanskih jezikov razlikuje in dovoljuje rabo dveh možnosti. V stanjskih zgradbah, npr. *Pet oken*

je bilo odprtih, je pridevniška oblika izražena izključno z rodilnikom množine in se tako ujema s samostalniškim delom količinske zveze, in ne s srednjepolskim edninskim pomožnikom. Nasprotno pa sta v trpniških zgradbah, kot je *Lani je bilo izgrajenih/izgrajeno pet hiš*, možni dve obliki pridevniške besede: ali množinska rodilniška (*izgrajenih*) ali edninska imenovalniška (*izgrajeno*). Štokavščina tovrstnega razlikovanja med stanjem in trpnikom ne pozna: pridevniška beseda je v obravnavanih zgradbah vedno ujemalna s pomožnikom, in je v srednjepolski edninski obliki, npr. *izgrađeno v Pet kuća je bilo izgrađeno prošle godine*. V ruščini se v predikacijskem razmerju z obravnavanimi količinskimi zvezami v osebku v razmerju s povedkom lahko uporabljata dva tipa ujemanja: glede na položaj osebka in druge dejavnike so sestavine povedka, tj. pomožnik in pridevniška beseda, ali v edninski srednjepolski obliki (privzeto ujemanje), npr. *Вчера было убито пять человек*, ali v množinski obliki (pomensko ujemanje), npr. *Пять человек были убиты на месте*. V makedonščini se tako glagolski kot pridevniški del povedka uporabljata izključno v množini, npr. *Вчера беа убиени пет луѓе*.

OKRAJŠAVE

M = moški spol; F = ženski spol; N = srednji spol; SG = ednina; DU = dvojina; PAU = pavkalna oblika; PL = množina; PL COUNT = števna množina (makedonščina); NOM = imenovalnik; GEN = rodilnik; DAT = dajalnik; ACC = tožilnik; LOC = mestnik; INS = orodnik; AOR = aorist; IMPER = imperfekt; PRS = sedanjik; PST = preteklik; AUX = pomožnik; LPCP = opisni deležnik na -i; PTCP = pridevniški deležnik; REF = morfem povratnosti; QUANT = količinski izraz.

Okrajšave za jezike in govore: slvn. = slovenščina; štok. = štokavski standardi; mak. = makedonščina; rušč. = ruščina.

REFERENCE

- Ahačič, Kozma. 2017. *Slovnica na kvadrat. Slovenska slovnica za srednjo šolo*. Ljubljana: Založba Rokus Klett, d. o. o.
- Browne, Wayles E. 2022. Premodifiers On Bosnian-Croatian-Serbian “Numeral + Noun” Phrases. In: L. Iomdin, J. Miličević, A. Polguère (eds.). *Lifetime Linguistic Inspirations: To Igor Mel'čuk from Colleagues and Friends for his 90th Birthday*. *Wiener Slawistischer Almanach, Linguistische Reihe* 101. Berlin: Peter Lang. 129–136.
- Corbett, Greville G. 2009. Agreement in Slavic. In: T. Berger, K. Gutschmidt, S. Kempgen, P. Kosta (eds.). *Die slavischen Sprachen (Ein internationales Handbuch zu ihrer Struktur, ihrer Geschichte und ihrer Erforschung)*. Berlin, New York: Walter de Gruyter. 342–354.

- Gigafida 2.0: Korpus pisne standardne slovenščine*. https://www.clarin.si/noske/run.cgi/first_form?corpname=gfida20_dedup;align=
- Hrvatski jezični korpus Riznica Instituta za hrvatski jezik i jezikoslovje*. https://www.clarin.si/noske/run.cgi/first_form?corpname=riznica;align=
- Leko, Nedžad. 2009. The Syntax of Numerals in Bosnian. *Studies in Slavic Linguistics* 33. München: LINCOM GmbH.
- Milosavljević, Stefan. 2018. Kongruencija glagolskog predikata sa subjektskim sintagmama konstituisanim osnovnim brojem. *Zbornik Matice srpske za filologiju i lingvistiku*. 43–56.
- Neset, Tore, Janda, Laura A. 2023. A network of allostructions: quantified subject constructions in Russian. *Cognitive Linguistics*. 67–97. <https://doi.org/10.1515/cog-2021-0117>
- Piper, Predrag, Klajn, Ivan, Dragičević, Rajna. 2022. *Normativna gramatika srpskoga jezika*. Četvrto, izmenjeno i dopunjeno izdanje. Novi Sad: Matica srpska.
- Ramovš, Fran. 1947–1949. *Morfologija slovenskega jezika*. Ljubljana: DZS.
- Uhlik, Mladen, Žele, Andreja. 2023. Ujemanje v zgradbah z glavnimi števnikami: slovenščina v sopostavitvi z drugimi zahodnimi južnoslovanskimi jeziki. *Slavistična revija* 71/4. 337–354. <https://doi.org/10.57589/srl.v71i4.4130>
- Vaillant, André. 1958. *Grammaire comparée des langues slaves. Tome 2. Morphologie. Deuxième partie: Flexion pronominale*. Lyon: IAC.
- Циммерлинг, Антон. 2019. Корпусная грамматика количественных групп в русском языке. *Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2019»* 18. 781–799. [Cimmerling, Anton. 2019. Korpusnaja grammatika količestvennyh grupp v russkom jazyke. *Komp'juternaja lingvistika i intelektual'nye tehnologii: po materialam meždunarodnoj konferencii «Dialog 2019»* 18. 781–799.]
- Национальный корпус русского языка. НКРЯ*. [Nacional'nyj korpus russkogo jazyka. NKRJa.] <https://ruscorpora.ru>
- Супрун, Адам Е. 1969. *Славянские числительные. Становление числительных как особых частей речи*. Минск: Издательство БГУ им. В. И. Ленина. [Suprun, Adam E. 1969. *Slavjanske čislitel'nye. Stanovlenie čislitel'nyh kak osobyh častej reči*. Minsk: Izdatel'stvo BGU im. V. I. Lenina.]

Prispelo januarja 2024, sprejeto junija 2024.

Received January 2024, accepted June 2024.

ZAHVALE

Članek je nastal v okviru programa P6-0038 Slovenski jezik v sinhronem in diahronem razvoju, ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije. Avtorja se posebej zahvaljujeta Mislavu Beniću, Marcu L. Greenbergu, E. Waylesu Brownu, Matetu Kapoviću,

Liljani Mitkovski, Vladimirju A. Plungjanu in Danku Šipki za vse komentarje. Vso odgovornost za morebitne napake pri analizi in interpretaciji prevzemata avtorja.

POVZETEK

ZGRADBE Z GLAVNIMI ŠTEVNIKI OD PET NAPREJ: SOPOSTAVITEV SLOVENŠČINE Z DRUGIMI JUŽNOSLOVANSKIMI JEZIKI IN RUŠČINO

Slovenski jeziki so specifični zaradi skladenjskih značilnosti količinskih zgradb, ki vključujejo glavne števnike in samostalnike kot šteto vsebino. V prispevku obravnavamo dva tipa skladenjskih razmerij, in sicer razmerja med glavnim števnikom in samostalnikom znotraj količinskih zgradb in razmerja med količinskimi zgradbami v osebkuvi vlogi in povedkom znotraj predikacije.

Znotraj količinskih zgradb se med glavnim števnikom in samostalnikom vzpostavljajo primeri kanoničnega ujemanja, hkrati pa prihaja tudi do odstopanj od ujemanja, in s tem do prehoda v vezavno razmerje.

V slovenskih, štokavskih in ruskih besednih zvezah s števniki od 1 do 4 je značilno ujemanje števnik in samostalniškega dela količinske zveze. V slovenščini in ruščini se za velike števnike (od 5 naprej) v imenovalniku ali tožilniku uporablja vezava. V teh primerih veliki števniki zahteva množinsko obliko samostalnika v rodilniku. Če pa je veliki števniki v slovenščini ali ruščini v netožilniških stranskih sklonih, se med njim in samostalnikom vzpostavlja ujemanje. V nasprotju s slovenščino in ruščino se v štokavščini, kjer so veliki števniki nesklonljivi, uporablja zgolj vezavno razmerje: nesklonljivi števniki namreč vedno zahteva množinsko obliko v rodilniku. Posebnost štokavščine je, da se količinske zveze s števniki, kot je 5, ne uporabljajo v brezpredložnem dajalniku ali brezpredložnem orodniku.

V makedonščini s poenostavljenim sklonskim sistemom se ob števniki, večjih od 1, pretežno uporabljajo množinske oblike samostalnikov. Posebne množinske oblike so t. i. števnike oblike samostalnikov moškega spola na *-a*, ki so najpogostejše pri enozložnih samostalnikih.

Podobno kot količinske zgradbe z velikimi števniki se v slovenščini in štokavščini obnašajo tudi zveze s količinskimi izrazi z slvn. *več* in štok. *više* 'več'. V teh zvezah se v slovenščini kaže težnja k ujemalnemu razmerju, ker *več* nima statusa jedra in se posebej v pogovornem jeziku ravna po sklonu samostalnika, npr. *z večimi prijatelji*. Nasprotno pa je štok. *više* skladenjsko jedro količinske zveze in zato samostalniku vedno določa množinsko rodilniško obliko, npr. štok. *s više prijatelja* 'z več prijatelji'.

Znotraj predikacijskega razmerja se v obravnavanih jezikih med količinsko zvezo v osebkovi vlogi in povedkom vzpostavljata dva tipa ujemanja: izrazno oz. privzeto ujemanje z edninskim povedkom, in pomensko ujemanje z množinskim povedkom.

V slovenščini je ob količinskih zvezah s 5 v osebkovi vlogi stalnica t. i. izrazno oz. privzeto ujemanje. To ob omenjenih količinskih zvezah zahteva edninsko obliko povedka v srednjem spolu, npr. *Pet fantov je odšlo na zabavo*. V štokavščini prevladuje raba privzetega ujemanja, npr. štok. *Pet momakâ je otišlo na zabavu* 'Pet fantov je odšlo na zabavo', ni pa izključena raba pomenskega ujemanja, npr. štok. *Svih pet momaka su otišli na zabavu* 'Vseh pet fantov je odšlo na zabavo', ko količinska zveza označuje konkretno in bolj določeno količino. V ruščini pa se glede na različne dejavnike, kot sta živost ali stava osebkâ glede na povedek, uporabljata tako pomensko kot privzeto ujemanje. V makedonskih količinskih zvezah se rabi izhodiščna množinska oblika samostalnika, ki vedno določa obliko povedka, npr. mak. *Пет девојку отудоа на забава* 'Pet deklet je odšlo na zabavo'; to razmerje med osebkom in povedkom je primer kanoničnega ujemanja.

CONSTRUCTIONS WITH HIGH NUMERALS: A COMPARISON OF SLOVENE WITH OTHER SOUTH SLAVIC LANGUAGES AND RUSSIAN

The Slavic languages hold a specific typological position due to the syntactic characteristics of quantifying phrases (QP), which encompass cardinal numbers and nouns as counted entities. In this article we analyze these characteristics involving both syntactic relationships within QP, containing a cardinal number and a nominal part, as well as predicate agreement with a quantified subject.

In Slovene, BCMS, and Russian, canonical agreement is established within QP containing numerals from 1 to 4. In this agreement, a noun functions as the head, and numerals serve as attributes.

When dealing with high numerals, such as 5 and up, in Slovenian and Russian, when the QP is in the nominative or accusative case, we encounter binding in which numerals act as a syntactic head that elicits a genitive complement. However, in the non-accusative oblique case, agreement is established between the high numeral and the noun.

In contrast to Slovenian and Russian, BCMS, where high numerals are indeclinable, relies solely on binding relations irrespective of case: an indeclinable high numeral always takes the plural noun phrase in the genitive. A distinctive feature of BCMS is that QPs with high numerals are not employed in non-prepositional dative or non-prepositional instrumental cases.

In Macedonian, which features a simplified case system, the plural forms of nouns with numerals greater than 1 are primarily utilized within the QP. Additionally, there are specific plural forms known as count forms, distinguished by the ending *-a*, and these are most prevalent among monosyllabic masculine nouns.

We analyze certain features of quantifying phrases, including quantifiers such as Slovene *več* and his BCMS equivalent *više* ‘a number of’. These constructions exhibit some common features shared with those of QP with high numerals. In constructions with this quantifier, Slovene tends towards agreement, where the quantifier *več* does not hold the status of a head and functions as an indeclinable attribute, e. g. *z več prijatelji* ‘with several friends’. This tendency towards agreement is most evident in colloquial Slovene, where the quantifier *več* takes on adjectival endings, e. g. *z večimi prijatelji* ‘with several friends’.

In predicative agreement with a quantified subject, including high numerals and a plural NP in the genitive, in most of the analyzed languages, we distinguish two types of agreement patterns. The quantified subject that incorporates numerals as 5 can elicit the predicate rendered in the neuter singular form (the default agreement pattern), or it can take the plural form of the predicate (the semantic agreement).

In Slovene, the so-called default agreement with a singular predicate is consistent, e.g., *Pet fantov je odšlo na zabavo* ‘Five boys went to the party’.

In BCMS, within the same type of construction, the use of default agreement predominates, e.g., *Pet momaka je otišlo na zabavu* ‘Five boys went to the party’. However, BCMS does not exclude the semantic agreement with the plural predicate, e.g., *Svih pet momaka su otišli na zabavu* ‘All five guys went to the party’, particularly when the QP denotes a concrete animate subject, such as ‘all five guys’.

In Russian, both types of agreement are used, such as the semantic agreement with the plural predicate *Пять девушек ушли на вечеринку* ‘Five girls went to a party’ or the default agreement implying the singular predicate *В аварии пострадало пять человек* ‘Five people were injured in the accident’. The choice of type depends on various pragmatic and semantic factors, including the animacy of the subject, the position of the subject in relation to the predicate, and the type of the predicate.

The variation of agreement patterns is not characteristic of Macedonian, where the plural form of the noun phrase always elicits the plural form of the predicate, e.g., *Пет девојки отидоа на забава* ‘Five girls went to a party’.

Kristina Gregorčič (ORCID: 0000-0002-4255-5464)

Univerza v Ljubljani, Filozofska fakulteta, Slovenija

kristina.gregorcic@ff.uni-lj.si

DOI: <https://doi.org/10.3986/16.1.02>

THE USE, MEANING AND FUNCTIONS OF SLOVENE BARE PRONOUNS

The paper presents selected aspects of the use and meaning of Slovene bare pronouns. An extensive quantitative and qualitative analysis of concordances from the reference corpus *Gigafida 2.0* allows for a new, more accurate classification of these pronouns in terms of their indefinite pronoun functions (Haspelmath 1997), as well as their categorization as epistemic indefinite pronouns and Negative Polarity Items. Furthermore, the corpus data show that the placement of bare pronouns in a clause is predictable, typically depending on the complexity of the pronominal phrase and the speaker's communicative intentions.

KEYWORDS: Slovene bare pronouns, epistemic indefinite pronouns, Negative Polarity Items, implicational map for indefinite pronoun functions, corpus analysis

Prispevek predstavlja izbrane vidike rabe in pomena poljubnostnih zaimkov. Obsežnejša količinska in kakovostna analiza konkordanc iz referenčnega korpusa *Gigafida 2.0* omogoča novo, točnejšo opredelitev poljubnostnih zaimkov z vidika funkcij nedoločnih zaimkov (Haspelmath 1997) ter njihovo uvrstitev med epistemične nedoločne zaimke in k negativni polarnosti usmerjene izraze. Poleg tega korpusni podatki kažejo na predvidljivost stave poljubnostnih zaimkov v stavku; ta je običajno odvisna od kompleksnosti zaimenske zveze in govorčevih sporočevalnih namenov.

KLJUČNE BESEDE: poljubnostni zaimki, epistemični nedoločni zaimki, k negativni polarnosti usmerjeni izrazi, implikacijski zemljevid funkcij nedoločnih zaimkov, korpusna analiza

1 INTRODUCTION¹

In his monograph based on an analysis of indefinite pronouns in 140 natural languages, Haspelmath (1997) identifies nine functions that indefinite pronouns can perform in any language. An attempt to identify the functions of Slovene

¹ This paper is based on a broader and more comprehensive study that is detailed in *Semantični in pragmatični vidiki k negativni polarnosti usmerjenih nedoločnih zaimkov* (Gregorčič 2023).

indefinite pronouns à la Haspelmath (1997) has already been made by Willis (2013: 394). However, this attempt does not fully capture the various dimensions of use and meaning displayed by Slovene indefinite pronouns. Furthermore, Slovene grammars ascribe the status of indefinite pronouns either only to some pronoun series with characteristics of indefinite pronouns (Toporišič 2000), or to pronoun series that do not belong to the typological category of indefinite pronouns (Janežič 1863; Breznik 1916 and 1934; Bajec et al. 1973). It is therefore necessary to re-examine Slovene indefinite pronouns to provide an updated, more accurate classification.

Among the Slovene indefinite pronouns that have already been the subject of more detailed discussions are *ni*-pronouns (e.g., *nihče* ‘no one’, *nič* ‘nothing’), which are Negative Concord Items (Ilc 2019),² and *koli*-pronouns (e.g., *kdorkoli* ‘anyone’, *karkoli* ‘anything’), which belong to the group of Negative Polarity Items triggering the Free Choice implicature in non-negative modal contexts (Gregorčič 2021).³ There has been considerably less focus on *ne*-pronouns (e.g., *nekdo* ‘someone’, *nekaj* ‘something’) and bare pronouns (e.g., *kdo* ‘someone/anyone’, *kaj* ‘something/anything’). The purpose of our paper is to define the use, meaning and functions of the latter. SECTION 2 provides an overview of the treatment of bare pronouns in selected Slovene grammars. SECTION 3 outlines Haspelmath’s typology of indefinite pronoun functions (1997). SECTION 4 presents the results of our quantitative and qualitative study of randomly selected concordances with bare pronouns from the reference corpus of written standard Slovene *Gigafida 2.0*, focusing on the functions of bare pronouns in terms of Haspelmath’s typology (1997) and on the position of bare pronouns in a clause. SECTION 5 summarizes the main findings and concludes the paper.

² Negative Concord Items (NCIs) are expressions that require the presence of clausemate propositional negation for grammaticality. Although these expressions can be morphologically negative (Slovene NCIs, for instance, contain the negative morpheme *ni*-), negated clauses with one or more NCIs typically yield single negation interpretations, as opposed to languages without Negative Concord, such as standard English, where each negative pronoun contributes its own negative interpretation. A detailed discussion of Slovene NCIs is provided in Ilc (2019).

³ Negative Polarity Items (NPIs) are expressions that typically denote minimal values and are licensed in semantically negative environments, such as the scope of negation and other semantically negative expressions (e.g., *without* and *doubt*). The NPI-licensing environments were first described as a homogeneous group in Ladusaw (1980). A more detailed discussion of Slovene NPIs can be found in Gregorčič (2018, 2021 and 2023).

2 BARE PRONOUNS IN SLOVENE GRAMMARS⁴

Bare pronouns (for an overview, see TABLE 1), formed from interrogative pronouns by conversion, have not been the subject of detailed discussion in technical and scientific linguistic literature. Brief overviews of their use and meaning can be found in the grammars of Janežič (1863), Breznik (1916 and 1934), Bajec et al. (1973) and Toporišič (2000).

Janežič (1863) classifies bare pronouns within the category of indefinite pronouns, which include *ni*-pronouns (*nihče* ‘no one’, *nič* ‘nothing’), quantifying pronouns (*marsikdo* ‘many a person’, *marsikaj* ‘many a thing’), universal pronouns (*vsakdo* ‘everyone’, *vse* ‘everything’) and *ne*-pronouns (*nekdo* ‘someone’, *nekaj* ‘something’). He presents bare pronouns as descriptions of “possible, uncertain or merely imaginary” referents (Janežič 1863: 225), illustrating their use in descriptions of generic, iterative or stative eventualities (1a), in subordinate clauses of contingency (1b), and in negated modal existential *wh*-constructions with the verb *imeti* ‘have’ (1c).⁵ In the examples below, the relevant bare pronouns are bolded for the sake of clarity.

(1)

a. *Včasih se vržejo otroci po **kom** iz bližnje rodbine.*
 sometimes REFL throw.3PL.PRS children after NPL.person.LOC from near family
 ‘Sometimes children take after someone from their close family.’

(Janežič 1863: 225)

b. *Kdor očetu ali materi **kaj** vzame pa pravi:*
 who.REL father.DAT or mother.DAT NPL.thing.ACC take.3SG.PRS and say.3SG.PRS
ni greh – tovarš tolovajev je tak.
 NEG.be.3SG.PRS sin comrade bandits.GEN be.3SG.PRS like.this

‘Whoever takes anything from their father or mother, saying, “It is not a sin” – they are a comrade of bandits.’

(Janežič 1863: 225)

⁴ We gloss bare pronouns as follows: *NPL.ontological category*. We have adopted this approach to avoid confusion between Slovene bare pronouns and English indefinite pronouns of the *any*- and *some*-series. Even though both *any*-pronouns and bare pronouns belong to the category of NPIs, they yield different interpretations in non-negative modal contexts (see Gregorčič 2023). *Some*-pronouns, on the other hand, can refer to specific referents, which is not typical of bare pronouns. Consequently, Slovene bare pronouns do not consistently translate into English as *any*- or *some*-pronouns, as evidenced by our translations of the Slovene example sentences.

⁵ For more information about modal existential *wh*-constructions, see Grosu (2004).

c. *nima kaj prigrizniti*

NEG.have.3SG.PRS NPI.thing.ACC snack.INF

‘(s)he has nothing to snack on’

(Janežič 1863: 165)

Breznik (1916 and 1934) adopts Janežič’s classification (1863) of bare pronouns. In the first edition of his grammar, he defines bare pronouns as words with a “more general, indefinite meaning” compared to *ne*-pronouns (Breznik 1916: 108). In the second edition, he replaces his initial semantic description of bare pronouns with a more prescriptive usage description (Breznik 1934: 104). He emphasizes the preference for substituting the indefinite-article-like numeral *e(de)n* ‘one’ with bare pronouns in dependent and independent questions (2a), in structures expressing commands and wishes (2b), and in descriptions of future events (2c).

(2)

a. *Ali bi hotel priti kak večer (ne: en večer!)*

Q COND want.PTCP come.INF NPI.DET.ACC evening.ACC NEG one evening

k meni na prejo?

to me.DAT on spinning

‘Would you like to come to my place some evening (*not: one evening!*) to do a bit of spinning?’

(Breznik 1934: 104)

b. *No, Luka, pa reci kako (kakšno) pametno*

well Luka PTCL say.2SG.IMP NPI.DET.ACC NPI.DET.ACC smart.ACC

(*ne: eno pametno*).

NEG one smart.ACC

‘Well, Luka, say something smart then (*not: one smart thing*).’

(Breznik 1934: 104)

c. *Vam se bo enkrat še dobro godilo (prav: kedaj).*

YOU.DAT REFL AUX.3SG.FUT one.day PTCL well live.PTCP correct NPL.time

‘You will do well one day (*correct: someday*).’

(Breznik 1934: 104)

Bajec et al. (1973: 185) continue the tradition of classifying bare pronouns as indefinite pronouns. The authors define indefinite pronouns as descriptions of undefined or prototypical entities, with *ne*-pronouns referring to specific unnamed entities and bare pronouns referring to non-specific entities (Bajec et al. 1973: 189). The authors thus recommend the use of bare pronouns in generic (3a) and modal descriptions (3b), as well as negated clauses (3c). They also illustrate the use of bare pronouns in dependent and independent questions (3d), and in conditional clauses (3e).

(3)

a. *Po njej se pretaka življenje,*
 through her.LOC REFL flow.3SG.PRS life
ki nekaj velja (nam. ki kaj velja).
 REL something count.3SG.PRS instead REL NPI.quantity count.3SG.PRS
 ‘Through it flows life that is worth something.’

(Bajec et al. 1973: 189)

b. *Vprašati morate nekoga (nam. koga),*
 ask.INF must.2PL.PRS someone.ACC instead NPI.person.ACC
ki je dalj časa tu.
 REL be.3SG.PRS longer time here

‘You need to ask someone who has been here longer.’

(Bajec et al. 1973: 189)

c. *Nečesa takšnega ne zmore ves svet*
 something.GEN like.this NEG can.3SG.PRS whole world
(nam. kaj takega ali česa takega).
 instead NPI.thing.ACC like.this or NPI.thing.GEN like.this

‘The whole world cannot do anything like this.’

(Bajec et al. 1973: 189)

d. *Starost je priča, kako je kdo preživel mladost.*
 age be.3SG.PRS witness how AUX.3SG.PRS NPI.person.NOM spend.PTCP youth

‘Age bears witness to how someone has spent their youth.’

(Bajec et al. 1973: 185)

e. *Kaj bo, če pride kak vihar?*
 what be.3SG.FUT if come.3SG.PRS NPL.DET.NOM storm.NOM

‘What if there is a storm?’

(Bajec et al. 1973: 185)

Toporišič (2000) uses the term indefinite pronoun only to refer to *ne*-pronouns. For bare pronouns, he introduces the term randomness pronouns, as they are used when there are no restrictions on the choice of potential referents. As with all pronoun series, the author categorizes individual bare pronouns according to their word-class properties and the ontological categories to which they refer. An overview of bare pronouns, based on Toporišič (2000: 311, 335, 406), is given in TABLE 1.

NOMINAL		ADJECTIVAL			
PERSON	THING	PROPERTY	CLASS	POSSESSION	QUANTITY
<i>kdó</i>	<i>kàj</i>	<i>kàk(šen)</i>	<i>katéri</i>	<i>čigáv</i>	<i>kóliko (kàj)</i>
ADVERBIAL					
PLACE			TIME	PROPERTY	
LOCATION	DESTINATION	DISTRIBUTION		MANNER	QUANTITY
<i>kjé</i>	<i>kàm</i>	<i>kód</i>	<i>kdàj</i>	<i>kakó</i>	<i>kóliko</i>

TABLE 1: Bare pronouns according to their word-class properties and the ontological categories to which they refer

Toporišič (2000: 311–312) notes that – in contrast to *ne*-pronouns, which refer to specific entities that are known or unknown, but never named – bare pronouns refer to non-specific entities that can be freely chosen from the pronoun’s referential domain. In one of the examples, the author shows that both *ne*- and bare pronouns can refer to small quantities (4), but he explicitly characterises their mutual substitution as not recommended (2000: 344). Instead, he notes the interchangeability of bare pronouns with *koli*-pronouns,⁶ with the phrase *ta in oni* ‘this and that one’, and with the phrase consisting of a bare pronoun preceded by the emphatic particle *sploh* ‘even’ (2000: 312).

⁶ The interchangeability and semantic affinities between bare and *koli*-pronouns are also pointed out in Bajec (1972).

(4)

Kakšne orehe boš našla v skrinji ('nekaj orehov').

NPL.DET.ACC walnuts.ACC AUX.2SG.FUT find.PTCP in chest some walnuts

'You will probably find some walnuts in the chest.'

(Toporišič 2000: 344)

According to Toporišič (2000: 312), the typical environments for bare pronouns are conditional, desiderative, interrogative and exclamatory clauses. The author illustrates their use in descriptions of general truths, customs, predictions and wishes (5a–b), in conditional and semantically similar temporal clauses (5c), and in dependent and independent questions (5d).

(5)

a. **Močilar** mi je včasih **kaj** razkladal

Močilar me.DAT AUX.3SG.PRS sometimes NPL.thing.ACC tell.PTCP

o nekdanjih časih.

about past times

'Močilar used to tell me some stories about the old days.'

(Toporišič 2000: 311)

b. **Ko** bi mi **kdo** hotel pomagati!

if.only COND me.DAT NPL.person.NOM want.PTCP help.INF

'If only someone would help me!'

(Toporišič 2000: 312)

c. **Če** je **kdo** pred vrati, ga povabi noter.

if be.3SG.PRS NPL.person.NOM in.front.of door him.ACC invite.2SG.IMP inside

'If there is anyone at the door, invite them in.'

(Toporišič 2000: 312)

d. **Ali** je **kdo** pred vrati?

Q be.3SG.PRS NPL.person.NOM in.front.of door

'Is there anyone at the door?'

(Toporišič 2000: 312)

3 HASPELMATH'S TYPOLOGY OF INDEFINITE PRONOUNS (1997)

As illustrated in SECTION 2, Toporišič (2000) restricts the category of indefinite pronouns to *ne*-pronouns. In contrast, Janežič (1863), Breznik (1916 and 1934) and Bajec et al. (1973) propose an expanded category of indefinite pronouns that encompasses additional pronoun series, including bare, universal, quantifying and *ni*-pronouns. Vidovič Muha (2013: 293–298) presents a classification of pronouns based on their definiteness, which is determined by textual actualization. As textually indefinite pronouns, she characterizes *ne*-pronouns, bare pronouns, quantifying pronouns, as well as pronouns expressing otherness (*drugi* ‘other’) and paucity (*redkokdo* ‘few people’, *redkokaj* ‘few things’), as they all convey the speaker’s inability to pinpoint the identity and/or number of potential referents.

The use of the term indefinite pronoun presented so far does not align with Haspelmath’s typological definition (1997), which is based on an in-depth analysis of a more narrowly studied sample of 40 languages and an additional, more roughly studied sample of 100 languages.⁷ Haspelmath’s study shows that, crosslinguistically, indefinite pronouns share morphological, semantic and syntactic properties. Morphologically, they are derived from the numeral *one*, interrogative pronouns or generic expressions; their indefiniteness is indicated by special affixes, particle clusters, reduplication and/or root transformations (Haspelmath 1997: 22–29). Semantically, indefinite pronouns refer to one of the basic ontological categories: person, thing, place, direction, time, manner, cause, quantity, determiner or property (Haspelmath 1997: 21–22, 29–31). Syntactically, they are characterized by multifunctionality, which means that they are compatible with a number of different syntactic environments in

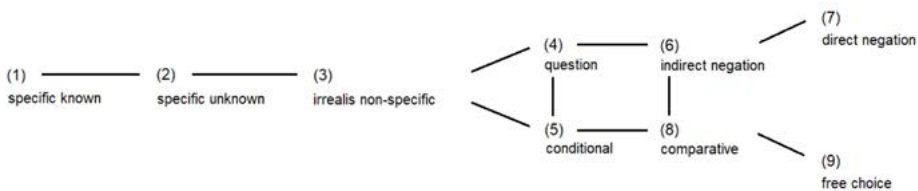


FIGURE 1: Implicational map for indefinite pronoun functions (Haspelmath 1997: 64)

⁷ The sampling method and the languages studied are detailed in Chapter 2 (Haspelmath 1997: 15–20). The monograph does not provide an analysis of Slovene indefinite pronouns.

which they yield fully predictable interpretations (Haspelmath 1997: 58–77). A single indefinite pronoun series is therefore expected to perform one or more of the nine functions presented on Haspelmath's implicational map for indefinite pronoun functions (1997: 64), illustrated in FIGURE 1. Since semantically related functions are adjacent on the map, Haspelmath predicts that a particular pronoun series can only perform the combinations of functions that form a continuous field. On the far left of the implicational map are the functions of pronouns whose main characteristic is reference to a single, i.e., specific existing referent. This may be a known referent that the speaker does not wish to identify (6) or an unknown referent that they cannot identify (7).

(6)

Somebody called while you were away: guess who!

(Haspelmath 1997: 2, (5))

(7)

*I heard **something**, but I couldn't tell what kind of sound it was.*

(Haspelmath 1997: 2, (6))

The semantically related *irrealis non-specific* function is located to the right of the *specific (un)known* functions. It is typical of pronouns used in descriptions of hypothetical, potentially viable, but not actual situations, which name a referent that cannot exist in the real world (8).

(8)

*Please try **somewhere** else.*

(Haspelmath 1997: 2, (7))

The *irrealis non-specific* function is followed on the right by the *question, conditional, indirect negation, direct negation* and *comparative* functions. These presuppose the use of indefinite pronouns in irrealis (interrogative, conditional and negated clauses) and realis environments (complements of comparatives) that allow for the quantitative use of superlatives enabling inferences from minimum to maximum values (see Fauconnier 1975). This kind of use is illustrated in (9): if a person does not have even the slightest difficulty, this necessarily implies that they do not have any major difficulty either.

(9)

Imeli ne boste niti najmanjših težav. = Imeli ne boste ne najmanjših ne majhnih ne velikih ne največjih možnih težav.

‘You will not have the slightest problem. = You will not have the slightest problem, nor a small problem, nor a big problem, nor the biggest problem possible.’

In addition to enabling the quantitative use of superlatives, the environments typical of the *question*, *conditional*, *indirect negation*, *direct negation* and *comparative* functions are compatible with Negative Polarity Items (Haspelmath 1997: 111–122).

Indefinite pronouns perform the *question* function if they are used in polar or content questions, whether rhetorical or not (10). The *indirect negation* function is characterized by the use of indefinite pronouns in the scope of semantically negative expressions (e.g., *without*, *doubt*, *be afraid of*) or in the scope of non-clausemate negation (11). In contrast, the *direct negation* function presupposes the use of indefinite pronouns in the scope of clausemate propositional negation (12).⁸

(10)

Did anybody tell you anything about it?

(Haspelmath 1997: 2, (8))

(11)

I don't think that anybody knows the answer.

(Haspelmath 1997: 2, (12))

(12)

Nobody knows the answer.

(Haspelmath 1997: 2, (11))

⁸ A proposition (or a part of it) is in the scope of an expression if the latter affects its interpretation. In example (i), the proposition *Anja ima čas* ‘Anja has time’ is in the scope of negation, which denies its truth. In example (ii), the proposition is in the scope of the conjunction *če* ‘if’, which places it in the category of hypothetical descriptions.

(i) *Anja nima časa. = Ne drži, da ima Anja čas.*

‘Anja does not have time. = It is not true that Anja has time.’

(ii) *Če ima čas, bo prišla na obisk. = Na obisk bo prišla pod pogojem, da ima čas.*

‘If she has time, she will come to visit. = She will come to visit on condition that she has time.’

Haspelmath (1997) illustrates the *conditional* function with examples of indefinite pronouns in dependent conditional clauses introduced by *if* and its equivalents in other languages (13). However, conditional relationship can also be conveyed with the restrictor of a universal quantifier (see Liu 2010), with unconditionals, which name sets of conditions (see Cazinkić 2002; Rawlins 2013), and with restrictive temporal clauses (see Farkas and Sugioka 1983). These are exemplified in (14a–c).

(13)

*If you see **anything**, tell me immediately.*

(Haspelmath 1997: 2, (9))

(14)

a. *Vsak učenec, ki bo pravilno rešil nalogo, bo nagrajen. ≈ Če bo učenec pravilno rešil nalogo, bo nagrajen.*

‘Every student who solves the task correctly will be rewarded. ≈ If a student solves the task correctly, they will be rewarded.’

b. *Kdorkoli bo pravilno rešil nalogo, bo nagrajen. ≈ Če jo bo pravilno rešil učenec a, če jo bo pravilno rešil učenec b, če jo bo pravilno rešil učenec c ..., bo nagrajen.*

‘Whoever solves the task correctly will be rewarded. ≈ If student a solves it correctly, if student b solves it correctly, if student c solves it correctly ..., they will be rewarded.’

c. *Ko učenec pravilno reši nalogo, je nagrajen. ≈ Če učenec pravilno reši nalogo, je nagrajen.*

‘When a student solves the task correctly, they are rewarded. ≈ If a student solves the task correctly, they are rewarded.’

The *comparative* function is characterized by the use of an indefinite pronoun in the nominal or clausal complement of a comparative. The indefinite pronoun establishes the standard of comparison which includes all alternatives of the compared item. In (15), the indefinite pronoun *anywhere* refers to all German cities except Freiburg, which is the compared item.

(15)

*In Freiburg the weather is nicer than **anywhere** in Germany.*

(Haspelmath 1997: 2, (10))

In the bottom right-hand corner of the implicational map is the *free choice* function, which is typical of indefinite pronouns that trigger the Free Choice implicature. They are used to convey a completely unrestricted choice between referential alternatives (see also Chierchia 2013), as illustrated in (16).

(16)

***Anybody** can solve this simple problem.*

(Haspelmath 1997: 3, (13))

Like the contexts of the *question*, *conditional*, *indirect negation*, *direct negation* and *comparative* functions, the contexts of the *free choice* function enable the quantitative use of superlatives. However, unlike the former, the latter allow inferences from maximum to minimum values, as in (17). If a detergent removes the toughest stain, it is expected to remove the less persistent stains as well.

(17)

To čistilno sredstvo lahko odstrani najtrdovratnejše madeže. = To sredstvo lahko odstrani najtrdovratnejše, trdovratne in netrdovratne madeže.

‘This detergent can remove the toughest stains. = This detergent can remove the toughest, tough, less persistent and the least persistent stains.’

On the basis of the properties of indefinite pronouns just presented, Haspelmath (1997: 11–12) excludes from this group of pronouns quantifiers of intermediate value such as *few*, generic pronouns such as *man*, universal quantifiers such as *all*, and pronouns of identity and otherness such as *same* and *other*. He argues that the expression of quantity, which is the central semantic property of quantifiers, is not a key feature of indefinite pronouns. Furthermore, universal quantifiers, which refer to all elements in a set, as well as identity and otherness expressions, which specify the identity of a referent, are definite and therefore fundamentally different from indefinite pronouns. Haspelmath’s definition of indefinite pronouns thus differs significantly from

the definitions in traditional grammars of Western languages, including the Slovene grammars listed in SECTION 2.

Slovene indefinite pronouns have already been classified according to Haspelmath's typology by Willis (2013).⁹ His implicational map for indefinite pronoun functions is presented in FIGURE 2.

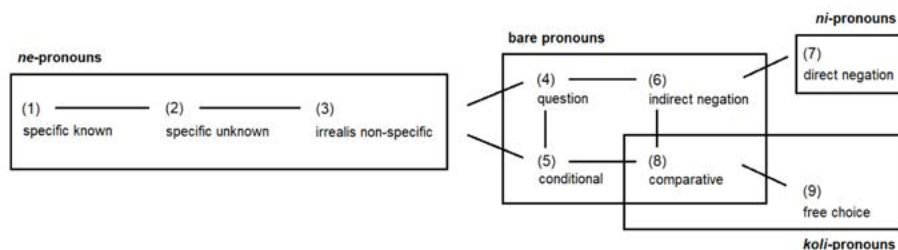


FIGURE 2: Implicational map for Slovene indefinite pronoun functions by Willis (2013: 394)

Willis (2013) does not provide a detailed analysis of Slovene pronouns, as his research aims to present the use of indefinite pronouns in all South Slavic languages. Consequently, the classification of Slovene indefinite pronouns in his study is not entirely accurate, as evidenced by the examples from Slovene grammars presented in SECTION 2. The functions of Slovene bare pronouns therefore need to be redefined.

4 SURVEY: BARE PRONOUNS IN THE REFERENCE CORPUS *GIGAFIDA 2.0*

4.1 RESEARCH GOALS AND METHODOLOGY

The referential quality of pronouns is contingent on their co-text and wider context (Vidovič Muha 2013). It is therefore important to define them using a sufficiently large representative sample. In the existing linguistic literature,

⁹ In light of the characteristics of indefinite pronouns described above, the colloquial pronoun *en* 'one' can also be considered as an indefinite pronoun. However, it is important to note that due to desemanticization, partial morphological and phonological reduction, and loss of syntactic flexibility, *en* is gradually acquiring characteristics of indefinite articles (see Bažec 2012). Similar processes can also be observed in the bare pronoun *kakšen* 'some/any', which shows signs of phonological reduction in the (colloquial) example below:

(i) *Včash probam kšn nov trend, včash mi je ful všeč, včash tut prbližn ne.*

'Sometimes, I try out a new trend; sometimes I like it a lot, sometimes not at all.'

Source: www.instagram.com/p/CT2g_zC1b08/?img_index=1 (published on 15 September 2021).

bare pronouns have only been the subject of rough and partly misleading descriptions (see SECTIONS 2 and 3). The aim of our research was to describe their use, meaning and functions on the basis of a more substantial set of examples.

Our research was based on a quantitative and qualitative analysis of samples obtained from the corpus of written standard Slovene *Gigafida 2.0* using the *SketchEngine* tool.¹⁰ We examined randomly generated samples of 700–800 concordances with the bare pronouns listed in TABLE 1. The sampling process proved to be rather challenging, as bare pronouns are homographs of interrogative and some relative pronouns, and they are also tagged as either interrogative or relative pronouns in *Gigafida 2.0*. Consequently, the sampling was carried out in two to three steps.¹¹

The first step consisted of an advanced search by lemmas (*Advanced > Query type: Lemma*), followed by the extraction of random samples of 1,000 concordances (*Get a random sample*). The random samples were manually filtered to exclude any examples that were irrelevant to the study, i.e., those that contained interrogative or relative pronouns, but no bare pronouns. After eliminating irrelevant and duplicate concordances, different sample sizes were obtained, each containing fewer than 700 concordances. As a result, a second sampling step was conducted, in which most of the unwanted concordances containing interrogative and relative pronouns were filtered out using the advanced filtering function (*Filter > Advanced*). The sizes of the random samples extracted in this step were set in such a way that the final number of concordances gathered in both steps combined was between 700 and 800 per individual bare pronoun. This process yielded the final samples for the pronouns *kdo*, *kaj*, *kakšen*, *čigav*, *kdaj*, *kje* and *kam*.

The second sampling step for the pronouns *kod*, *kako* and *koliko* yielded only 226, 47 and 23 relevant examples of use, respectively, from the total of 1,000 randomly selected concordances. The third sampling step was thus needed, in which the queries were further adjusted to ensure a higher percentage of relevant concordances in the random samples. From the random samples of 1,000

¹⁰ Although *Gigafida 2.0* is considered a corpus of standard language, it should be noted that its composition is not balanced in terms of text genres: 64.3% of the corpus consists of articles from printed journals and newspapers, while 28.0% of the texts are from online sources, the most prominent of which are again media publications (Krek et al. 2019). It would be beneficial for future research to analyse the use of bare pronouns in non-journalistic and spoken texts as well.

¹¹ The methodology with all search queries is described in detail in Gregorčič (2023).

concordances, we retrieved 241 examples with the pronoun *kod*, 163 examples with the pronoun *kako* and 44 examples with the pronoun *koliko*. Given the low frequency of occurrence, we deemed it inappropriate to pursue further sampling, as speakers seem to use these pronouns much less frequently than other bare pronouns.

TABLE 2 presents the number of concordances included in each sample.

PRONOUN	<i>kdo</i>	<i>kaj</i> ¹²	<i>kakšen</i>	<i>kateri</i>	<i>čigav</i>	<i>kdaj</i>	<i>kje</i>	<i>kam</i>	<i>kod</i>	<i>kako</i>	<i>koliko</i>	TOTAL
NUMBER OF CONCORDANCES	723	747	703	734	270	718	769	758	421	220	58	6,121

TABLE 2: Sample sizes

We first classified the concordances according to Haspelmath's typology of indefinite pronouns (1997). As we observed a relatively predictable placement of bare pronouns, we subsequently classified all the concordances according to the position of the bare pronoun in a clause. We were particularly interested in whether the pronoun occurs in the clause-final position, which in the case of a stylistically unmarked sentence typically contains new information or the focus, or whether it occurs in the non-final position in a clause, which is usually reserved for the topic and the transition (Toporišič 2000: 668–678).

4.2 RESEARCH RESULTS AND DISCUSSION

In SECTION 4.2.1, we present the use, meaning and functions of bare pronouns as defined by Haspelmath (1997). In SECTION 4.2.2, we focus on the placement of bare pronouns in a clause. All examples in this section are taken from *Gigafida 2.0*, unless otherwise stated. They have not been modified in any way, except for the shortening of longer passages that are not relevant to our present research.

4.2.1 THE USE, MEANING AND FUNCTIONS OF BARE PRONOUNS ACCORDING TO HASPELMATH (1997)

The first column of TABLE 3 presents the composition of all the samples of bare pronouns combined, while the subsequent columns show the composition of

¹² Our analysis focuses on the nominal pronoun *kaj*, not the homographic adverbial pronoun.

each individual bare pronoun sample in terms of Haspelmath's classification of indefinite pronoun functions (1997).

PRONOUN FUNCTION	TOTAL	kdo	kaj	kakšen	kateri	čigav
FREE CHOICE	/	/	/	/	/	/
COMPARATIVE	/	/	/	/	/	/
DIRECT NEGATION	14.1%	5.8%	23.6%	13.9%	6.3%	2.2%
INDIRECT NEGATION	8.3%	11.3%	11.9%	4.7%	8.2%	2.2%
CONDITIONAL	17.7%	27.1%	14.3%	15.1%	24.7%	3.7%
QUESTION	17.6%	17.4%	12.9%	11.9%	15.9%	75.2%
IRREALIS NON-SPECIFIC	42.3%	38.3%	37.3%	54.3%	45.0%	16.7%
SPECIFIC UNKNOWN	/	/	/	/	/	/
SPECIFIC KNOWN	/	/	/	/	/	/

	kdaj	kako	kje	kam	kod
FREE CHOICE	/	/	/	/	/
COMPARATIVE	/	/	/	/	/
DIRECT NEGATION	1.7%	14.5%	13.8%	30.7%	26.6%
INDIRECT NEGATION	17.1%	6.4%	4.9%	6.9%	3.1%
CONDITIONAL	16.2%	25.0%	20.8%	12.3%	13.5%
QUESTION	22.6%	10.9%	13.3%	6.7%	12.6%
IRREALIS NON-SPECIFIC	42.5%	43.2%	47.2%	43.4%	44.2%
SPECIFIC UNKNOWN	/	/	/	/	/
SPECIFIC KNOWN	/	/	/	/	/

TABLE 3: Structure of the samples according to Haspelmath's typology (1997)¹³

TABLE 3 indicates that bare pronouns do not name specific (un)known referents. In the majority of concordances in our samples, they perform the *irrealis non-specific* function,¹⁴ as they occur in a wide variety of modal environments and generic descriptions. The corpus data suggest that speakers tend

¹³ The data pertaining to the pronoun *koliko* are not reported in TABLE 3 due to the insufficient number of concordances in *Gigafida 2.0*. The scarcity of data suggests that this pronoun is only rarely used in modern Slovene.

¹⁴ The only exception is the sample of the pronoun *čigav*, whose usage patterns deviate from those exhibited by other bare pronouns in a number of aspects. In modern Slovene, *čigav* is used relatively infrequently, which is reflected in the small number of concordances in *Gigafida 2.0*. The deviation from the expected usage patterns is thus most likely a reflection of the pronoun's fossilization in a limited set of environments.

to associate bare pronouns with descriptions of possible events (18a), as well as with descriptions of states, habits and recurrent events (18b).¹⁵

(18)

a. *Morda lahko komu pomagata prav vi!*
 maybe easily NPL.PERSON.DAT help.2PL.PRS exactly you
 ‘Maybe it is you who can help someone!’

b. *Vsak dan se zgodi kaj zanimivega.*
 every day REFL happen.3SG.PRS NPL.thing.NOM interesting
 ‘Something interesting happens every day.’

In all concordances that illustrate the *irrealis non-specific* function, bare pronouns occur in the scope of a modal or generic operator (19). In this respect, bare pronouns are fundamentally distinct from *koli*-pronouns, which generally avoid the scope of modal operators (20) (Gregorčič 2023: 104).

(19)

Če ste vajeni spletnega nakupovanja, lahko kupite katero od zmagovalnih torbic.
 if be.2PL.PRS used.to online shopping easily buy.2PL.PRS
 NPL.class.ACC from winning handbags

‘If you are used to online shopping, you can buy one of the winning handbags.’

= It is possible for you to buy handbag a or handbag b or handbag c ...

(20)

Domišljija lahko v katerikoli situaciji zapolni praznino.
 imagination easily in NPL.class.LOC situation.LOC fill.3SG.PRS void

‘Imagination can fill the void in any situation.’

= For situation a and situation b and situation c ..., it is the case that imagination can fill the void in that situation.

¹⁵ A review of the concordances illustrating bare pronouns in the *irrealis non-specific* function reveals that 36.9% feature bare pronouns in the scope of expressions of probability and possibility (e.g., *verjetno* ‘probably’, *morda* ‘maybe’, *predvidevati* ‘suppose’), while 23.3% illustrate their use in descriptions of general truths and habits/recurrent events.

Modal environments highlight another semantic difference between bare and *koli*-pronouns. While the latter trigger the Free Choice implicature, suggesting unlimited freedom of choice between potential referents (see Gregorčič 2021 and 2023), the former are unable to do so, as is evident from the paraphrases of (19) and (20). Bare pronouns typically foreground the speaker’s ignorance or indifference, which is the defining characteristic of epistemic indefinite pronouns (see Alonso-Ovalle and Menéndez-Benito 2015). This semantic feature of bare pronouns aligns with the *irrealis non-specific* function, which has been overlooked by Willis (2013: 394).

The corpus data also suggest that there is no reason to attribute the *comparative* function to bare pronouns, contrary to what has been proposed by Willis (2013: 394). In this function, pronouns are used in the complement of the comparative and are expected to establish the standard of comparison that includes all alternatives of the compared item (see (15)). Although bare pronouns can be found in complements of comparatives, they do not yield the predicted interpretation, as indicated in (21). The unbracketed part of the example is from *Gigafida 2.0*. In brackets is our own continuation of the corpus example that is semantically and pragmatically acceptable, and that precludes the interpretation that the phrase *katero drugo tekmovalje* refers to all other competitions.

(21)

<i>Na</i>	<i>tej</i>	<i>olimpijadi</i>	<i>je</i>	<i>bito</i>	<i>lažje</i>	<i>priti</i>
on	this	Olympiad	AUX.3SG.PRS	be.PTCP	easier	come.INF
<i>do</i>	<i>nagrade</i>	<i>kot</i>	<i>na</i>	<i>katerem</i>	<i>drugem</i>	<i>tekmovalju.</i>
to	prize	than	on	NPL.CLASS.LOC	other	competition.LOC
<i>(Ne</i>	<i>pa</i>	<i>lažje</i>	<i>kot na</i>	<i>vsakem</i>	<i>od</i>	<i>njih.)</i>
NEG	but	easier	than on	every	from	them

‘It was easier to win a prize at this Olympiad than at another competition.
(But not easier than at every other competition.)’

According to Willis (2013: 394), bare pronouns perform the function of *indirect negation*, but not *direct negation*. This is not entirely consistent with the corpus data. In fact, *Gigafida 2.0* contains a non-negligible number of concordances with bare pronouns in negated clauses. Bare pronouns interpreted in the scope of clausemate propositional negation typically form complex

phrases in which they are postmodified by adjectives, adverbs, pronouns or relative clauses. In (22), the phrase headed by the bare pronoun *česa* (i.e., *česa podobnega*) is in the scope of negation, as evidenced by the English translation: note that the English *anything* is a Negative Polarity Item (see FOOTNOTE 3), so it must occur in the scope of the negative pronoun *no one*.

(22)

Mislím, da česa podobnega ni poskusil še nihče.
 think.1SG.PRS that NPI.thing.GEN similar NEG.AUX.3SG.PRS try.PTCP yet no one
 ‘I think that no one has tried anything similar yet.’

The presence of a postmodifier plays an important role in determining the scopal interaction between the negator and the bare pronoun. When the bare pronoun lacks a postmodifier, it is interpreted outside the scope of clausemate propositional negation, as in (23). In the English translation of the sentence, the negator appears in a clause that is syntactically subordinate to the clause with the indefinite pronoun.

(23)

Če česa ni znala, je segla po strokovnih knjigah.
 if NPI.thing.GEN NEG.AUX.3SG.PRS know.PTCP AUX.3SG.PRS reach.PTCP along specialist books

‘If there was something that she did not know, she turned to specialist books.’

The relevance of the postmodifier in determining the scopal interaction between clausemate propositional negation and bare pronouns can also be verified by a minimal modification of corpus example (22). If the postmodifier *podobnega* is removed from the complex phrase headed by the bare pronoun, as in (22'), the bare pronoun can only be interpreted outside the scope of negation.

(22')

Mislím, da česa ni poskusil še nihče.
 think.1SG.PRS that NPI.thing.GEN NEG.AUX.3SG.PRS try.PTCP yet no one
 ‘I think there must be something that no one has tried yet.’

This phenomenon is similar to the inverse scope of negation illustrated in (24) and (25). Both examples feature the propositional negator *not* and have the subject realized by an indefinite nominal phrase preceding the negator (*anyone* in (24) and *a doctor who knew anything about acupuncture* in (25)). Despite the syntactic parallels between the two examples, the subject is in the scope of negation in (25), but not in (24).

(24)

**Anyone did not talk to me.*

(de Swart 1998: 178, (5b))

(25)

A doctor who knew anything about acupuncture was not available.

(Linebarger 1980: 227, (21a), cited in de Swart 1998: 180, (13c))

The scopal properties of the propositional negation in (24) and (25) are indicated by the (un)grammaticality of these examples. The indefinite pronouns *anyone* in (24) and *anything* in (25) are Negative Polarity Items, whose acceptability depends on whether they occur in the scope of semantically negative expressions (see FOOTNOTE 3). Given that the only negative expression in (24) and (25) is the negator *not*, we can conclude that *anyone* in (24) is not in the scope of the negator, since its use is ungrammatical. In contrast, *anything* in (25) is in the scope of the negator; if it were not, it would be just as ungrammatical as *anyone* in (24).

De Swart (1998) shows that the inverse scope of negation arises from the interplay of Grice's maxims of quantity and relevance.¹⁶ Extending the scope of negation reduces the informative value of an utterance.¹⁷ Such a reduction can be justified only if it is compensated for. This can be achieved by increasing the complexity of the message. The utterance thus becomes longer, which is disadvantageous from the point of view of the maxim of quantity, but this negative effect is counterbalanced by an enhanced informative value of the

¹⁶ According to the maxim of quantity, speakers must limit themselves to communicating only what is necessary; according to the maxim of relevance, they must convey what is most relevant to the topic of the communicative exchange (see Grice 1975).

¹⁷ There are many more true negative statements than true affirmative statements about the world, because the latter are more specific and thus easier to falsify (see Horn 1989).

message. In accordance with the maxim of relevance, the addressees assume that the selection of a longer utterance is crucial for the accurate interpretation of the message. Consequently, they interpret the more complex structures of the utterance contrastively. This generates a positive implicature, which increases the informative value of the utterance. The positive implicature of (25) is illustrated in (25').

(25')

Some doctor was available.

De Swart (1998) suggests that the entire subject in (25) is in the semantic scope of negation, but that the positive implicature – a pragmatic element of the utterance – leads the addressees to interpret the more complex part of the sentence non-negatively. The interpretation of (25), shaped by the interplay of semantic and pragmatic factors, is presented in (25'').

(25'')

A doctor was available, but not one who would know anything about acupuncture.

In light of de Swart's findings (1998), the interaction between negation and the bare pronoun in (22), which is repeated below, can be explained from a pragmatic point of view. Although the entire nominal phrase *česa podobnega* occurs in the semantic scope of negation, the bare pronoun *česa* is simultaneously interpreted in a non-negative pragmatic environment created by the positive implicature that someone must have tried something before. This enables the interpretation that while the predicate is not true of the set described by the complex phrase *česa podobnega*, this does not mean that it is also not true of its superset described by the bare pronoun *česa* alone (22'').

(22)

Mislim, da česa podobnega ni poskusil še nihče.

think.1SG.PRS that NPI.thing.GEN similar NEG.AUX.3SG.PRS try.PTCP yet no one

‘I think that no one has tried anything similar yet.’

(22'')

Someone must have tried something before, but no one has tried anything similar.

The existence of this interpretation is supported by the examples from *Gigafida 2.0* in which the positive implicature is explicitly stated, as in (26).

(26)

<i>Videl</i>	<i>sem</i>	<i>nekaj</i>	<i>izjemnih</i>	<i>tekem,</i>
see.PTCP	AUX.1SG.PRS	some	extraordinary	matches
<i>česa</i>	<i>takega</i>	<i>pa</i>	<i>še</i>	<i>nikoli.</i>
NPI.thing.GEN	like.this	but	yet	never

‘I have seen some extraordinary matches, but never anything like this.’

It is possible for unmodified bare pronouns to occur in the scope of clause-mate propositional negation as well, but in a completely different context. In our samples, more than half of the examples illustrating the *direct negation* function show the use of bare pronouns in modal existential *wh*-constructions introduced by the modal verbs *imeti* ‘have’ and *biti* ‘be’ (27).

(27)

<i>Hudo je,</i>	<i>da se</i>	<i>v krizi</i>	<i>nimaš</i>	<i>na koga</i>	<i>obrniti.</i>				
bad	be.3SG.PRS	that	REFL	in	crisis	NEG.have.2SG.PRS	on	NPI.person.ACC	turn.INF

‘It is terrible not having anyone to turn to in a crisis.’

However, the presence of the negator does not seem obligatory for the acceptability of bare pronouns in such contexts. Several concordances in our samples illustrate the use of bare pronouns in non-negated modal existential *wh*-constructions (28), whereby the bare pronoun assumes a hyperbolic meaning.¹⁸

¹⁸ We have verified this observation by searching the entire corpus. The results show that bare pronouns appear in 15,357 concordances with negated modal existential *wh*-constructions (query (i)), and in 2,471 concordances with non-negated constructions of this type (query (ii)). In the latter case, the modal existential *wh*-construction typically does not occur in the scope of any other semantically negative expression.

(i). [word="ne"][lemma="biti"&tag="G.....n.*"]|[lemma="imeti"]|[lemma="kaj|kdo|kakšen|kateri|čigav|kdaj|kje|kod|kam|kako|koliko"]|[tag="G..n.*"] | [lemma="biti"&tag="G.....d.*"]|[lemma="imeti"]|[lemma="kaj|kdo|kakšen|kateri|čigav|kdaj|kje|kod|kam|kako|koliko"]|[tag="G..n.*"] | [lemma="imeti"&tag="G.....d.*"]|[lemma="kaj|kdo|kakšen|kateri|čigav|kdaj|kje|kod|kam|kako|koliko"]|[tag="G..n.*"]

(ii). [lemma="imeti"&tag="G...s...n.*"]|[lemma="kaj|kdo|kakšen|kateri|čigav|kdaj|kje|kod|kam|kako|koliko"]|[tag="G..n.*"] | [word!="ne"][lemma="biti"&tag="G.....n.*"]|[lemma="imeti"]|[lemma="kaj|kdo|kakšen|kateri|čigav|kdaj|kje|kod|kam|kako|koliko"]|[tag="G..n.*"]

(28)

<i>Dandanes</i>	<i>se</i>	<i>ima</i>	<i>kam</i>	<i>ozreti</i> ;
nowadays	REFL	have.3SG.PRS	NPI.direction	look.INF
<i>tovarni</i>	<i>je</i>	<i>veliko</i>		<i>dal.</i>
factory.DAT	AUX.3SG.PRS	much		give.PTCP

‘Today, he has somewhere to look; he has given a lot to the factory.’

The question thus arises as to whether it is reasonable to assign the *direct negation* function to bare pronouns at all. We have shown that their use in the scope of clausemate propositional negation is enabled by the presence of a positive implicature; in its absence, the pronoun escapes the scope of negation. The exception to this is the modal existential *wh*-construction, in which bare pronouns always occur in the scope of the clausemate propositional negator, but do not depend on its presence.

Haspelmath (1997) does not examine the examples of use in individual languages in much detail, so it is not entirely clear what status he would assign to Slovene bare pronouns. He claims that pragmatic parameters, which are rooted in the expectations of discourse participants, have no bearing on whether a pronoun assumes a particular function on his implicational map (Haspelmath 1997: 82–86); the only relevant condition is that the pronoun occurs in the appropriate context and yields the expected interpretation. In the case of bare pronouns in the *direct negation* function, this means that they are expected to be interpreted in the scope of clausemate propositional negation with a non-specific meaning. In this respect, it seems reasonable to assign them the *direct negation* function. However, the examples provided by Haspelmath to illustrate the pragmatic effects triggered by indefinite pronouns are not entirely identical to those governing the use of bare pronouns in negated clauses. In (22) and (26) above, the presence of the positive implicature is necessary for the bare pronoun to be interpreted in the scope of clausemate negation at all. In contrast, the use of the English pronoun *some* in interrogative sentences such as (29), which Haspelmath (1997) cites to illustrate the pragmatic parameters influencing the interpretation of indefinite pronouns, is always possible and grammatical. The pragmatic effect of signalling the expectation of a positive answer is triggered by the pronoun itself, not by its context, as is the case in Slovene.

(29)

*Do you think those men want to do **some** work?*

(Lakoff 1969: 609–611, cited in Haspelmath 1997: 82, (151a))

Based on the results of our corpus analysis, we propose a revised implicational map for the functions of Slovene bare pronouns in FIGURE 3. The revised map differs from Willis' (2013) in three respects: bare pronouns (i) perform the *irrealis non-specific* function, (ii) do not perform the *comparative* function, and (iii) perform the *direct negation* function if certain pragmatic conditions are met. The black line in FIGURE 3 indicates the functions that can be attributed to bare pronouns based on corpus data, while the grey line indicates the functions attributed to these pronouns by Willis (2013: 394).

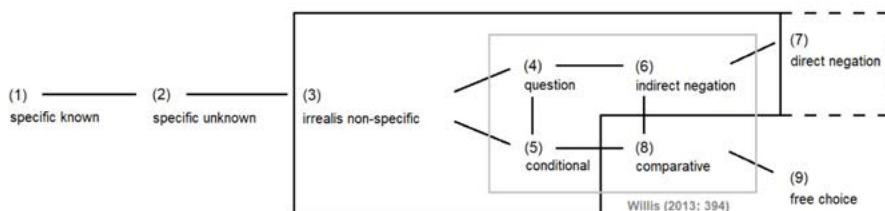


FIGURE 3: New implicational map for the functions of Slovene bare pronouns

4.2.2 THE POSITION OF BARE PRONOUNS IN A CLAUSE

TABLE 4 presents the proportion of nominal and adverbial bare pronouns in our samples, classified according to whether the phrase which they head occurs in the final or non-final position in a clause.

		Number of examples	Percentage
FINAL POSITION	WITH A POSTMODIFIER	1,440	28.2%
	WITHOUT A POSTMODIFIER	519	10.2%
	TOTAL	1,959	38.4%
NON-FINAL POSITION	WITH A POSTMODIFIER	534	10.4%
	WITHOUT A POSTMODIFIER	2,619	51.2%
	TOTAL	3,153	61.6%

TABLE 4: Nominal and adverbial bare pronouns according to their position in a clause

Speakers most often place phrases headed by bare pronouns in a non-final position in a clause. Their decision seems to be at least partly influenced by the presence of a postmodifier of the bare pronoun. When the postmodifier is present, speakers are more likely to place the phrase with the bare pronoun in the clause-final position. Bare pronouns without a postmodifier appear in the clause-final position only in a limited range of contexts: they are often reinforced by additive particles (e.g., *še* ‘also’, *tudi* ‘also’) (30a), or they occur in short clauses in which the clause-final position is the only available option for them. The most common types of construction in which this occurs are interrogative (30b), imperative (30c) and elliptical clauses, particularly elliptical negated modal existential *wh*-constructions (30d) and elliptical conditional clauses (30e).

(30)

- a. *Upam, da bo prišel še kdaj.*
 hope.1SG.PRS that AUX.3SG.FUT come.PTCP also NPI.time
 ‘I hope he will come back another time.’
- b. *Nikoli ne veš, kdo bo kje.*
 never NEG know.2SG.PRS who be.3SG.FUT NPI.location
 ‘You never know who will be somewhere.’
- c. *Storite kaj.*
 do.2PL.IMP NPI.thing.ACC
 ‘Do something.’
- d. *So ljudje, ki nimajo kam.*
 be.3PL.PRS people REL NEG.have.3PL.PRS NPI.direction
 ‘There are people who do not have anywhere to go.’
- e. *Če kdo, je Jože vedno vedel odgovor.*
 if NPI.person.NOM AUX.3SG.PRS Jože always know.PTCP answer
 ‘If anyone, Jože always knew the answer.’

The examination of individual concordances reveals that speakers select the position of a bare pronoun according to the information to which they want to draw the addressee’s attention. In the clause-final (i.e., focus) position, the

phrase headed by a bare pronoun is emphasized and understood as contributing new information; in the clause-initial (i.e., topic) position, it most frequently refers to the preceding clause or to referents whose identity is not essential to the message (see Toporišič 2000: 668–678). A phrase headed by a bare pronoun is part of the focus if the pronoun is postmodified, which narrows the set of potential referents of the pronominal phrase and thus contributes to the greater specificity of the message (31a), or if the speaker focuses on emphasizing the paucity of potential referents expressed by the pronoun (31b).¹⁹

(31)

- a. *Vsak dan se zgodi kaj zanimivega.*
 every day REFL happen.3SG.PRS NPI.thing.NOM interesting
 ‘Something interesting happens every day.’
- b. *Kar naprej se mi zdi, da je moja omara prazna in da nujno rabim še kaj.*
 just forward REFL me.DAT seem.3SG.PRS that be.3SG.PRS my wardrobe
 empty and that urgently need.1SG.PRS also NPI.thing.ACC
 ‘I keep feeling that my wardrobe is empty and that I urgently need something else.’

In general, speakers use bare pronouns more often in the topic position than in the focus position. When the bare pronoun is part of the topic, the focus is either on the participant of the event that is not referred to with the bare pronoun (32a), or on the event as a whole (32b). In (32b), the focus is on losing and forgetting in general; the exact identification of the affected entity (*kakšna reč*) or the setting (*kje*) is not central to the message.

(32)

- a. *Če bo kdo potreboval mojo pomoč, ve, kje me lahko najde.*
 if AUX.3SG.FUT NPI.person.NOM need.PTCP my help
 know.3SG.PRS where me.ACC easily find.3SG.PRS
 ‘If anyone needs my help, they know where they can find me.’

¹⁹ The expression of quantity is most often indicated by the presence of the additive particle *še* ‘also’ and the use of bare pronouns in emotionally marked elliptical structures (see (30a) and (30d–e) above).

- b. *Včasih kakšno reč izgubimo ali kje pozabimo,*
 sometimes NPL.DET.ACC thing.ACC lose.1PL.PRS or NPL.location forget.1PL.PRS
pa tega še opazimo ne.
 but this PTCL notice.1PL.PRS NEG
 ‘Sometimes we lose something or forget it somewhere, and we do not even notice it.’

Given that non-nominalized adjectival bare pronouns are incapable of forming independent phrases with sentence functions, their position was observed separately. The results are presented in TABLE 5.

	Number of examples	Percentage
FINAL POSITION	789	78.2%
NON-FINAL POSITION	220	21.8%

TABLE 5: Adjectival bare pronouns according to their position in a clause

Speakers usually use nominal phrases with adjectival bare pronouns in the clause-final position. However, the informative focus of such a nominal phrase is typically not on the pronoun, but rather on the nominal head and its non-pronominal modifiers. In (33), the bare pronoun *kakšna* could therefore be omitted without significantly affecting the message (see also FOOTNOTE 9 about the article-like behaviour of adjectival bare pronouns).

(33)

- Nimamo dokaza, da je na Marsu kdaj obstajala*
 NEG.have.1PL.PRS evidence that AUX.3SG.PRS on Mars NPL.time exist.PTCP
kakšna višja oblika življenja.
 NPL.DET.NOM higher.NOM form.NOM life.GEN

‘We have no evidence that any superior life form has ever existed on Mars.’

Speakers can use adjectival bare pronouns to express small quantities, particularly in modal and generic contexts. In such instances, the speaker’s intention is often discernible through the presence of the additive particle *še* ‘also’ or other contextual cues, as in (34).

(34)

Rokavi reke so ena sama velika prometnica, po kateri se
 arms river be.3PL.PRS one sole great thoroughfare along which REFL
vsak dan premika na tisoče čolničev, stotine malo večjih
 every day move.3SG.PRS on thousands boats hundreds little larger
barkač, nekaj deset velikih ladij in kakšna čezoceanka.
 tubs some ten large ships and NPL.DET.NOM ocean.liner.NOM

‘The river’s arms are great thoroughfares, carrying thousands of small boats, hundreds of slightly larger tubs, a few dozen large ships and the odd ocean liner every day.’

5 CONCLUSION

We have presented the use, meaning and functions of bare pronouns, identified through an extensive analysis of examples from *Gigafida 2.0*. In contrast to Toporišič’s classification (2000), we propose to classify bare pronouns as indefinite pronouns, which is in line with Haspelmath’s typological characterization (1997). The corpus data indicate that bare pronouns perform more functions on the implicational map for indefinite pronoun functions than proposed by Willis (2013). In modal environments, they perform the *irrealis non-specific* function and reveal the speaker’s ignorance or indifference, which makes them epistemic indefinite pronouns. In addition, they perform the *conditional*, *question*, *indirect negation* and *direct negation* functions, which are typical of Negative Polarity Items.

Nominal and adverbial bare pronouns typically occur in a non-final position in a clause, with the purpose of referring to event participants and circumstances whose precise identification is not essential to the informative value of the utterance. In the clause-final position, they are typically postmodified, which enhances the informativity of the phrases they form. Alternatively, they may be unmodified, in which case they emphasize the paucity of referents.

REFERENCES

- Alonso-Ovalle, Luis, Menéndez-Benito, Paula. 2015. *Epistemic indefinites: exploring modality beyond the verbal domain*. Oxford: Oxford University Press.
- Bajec, Anton. 1972. Relativa in indefinita v slovenščini. In: *VIII. seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta UL. 24–34.

- Bajec, Anton, Kolarič, Rudolf, Rupel, Mirko, Šolar, Jakob. 1973. *Slovenska slovnica*. Ljubljana: DZS.
- Bažec, Helena. 2012. Gramatikalizacija nedoločnega člena v slovenščini. *Annales* 22/2: 461–470.
- Breznik, Anton. 1916. *Slovenska slovnica za srednje šole*. Celovec: Družba sv. Mohorja.
- Breznik, Anton. 1934. *Slovenska slovnica za srednje šole*. Celje: Družba sv. Mohorja.
- Cazinkič, Robert. 2002. *Oziralni odvisniki v sodobnem slovenskem knjižnem jeziku*. MA thesis. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Chierchia, Gennaro. 2013. *Logic in grammar: polarity, free choice, and intervention*. Oxford: Oxford University Press.
- Farkas, Donka F., Sugioka, Yoko. 1983. Restrictive if/when clauses. *Linguistics and Philosophy* 6: 225–258.
- Fauconnier, Gilles. 1975. Pragmatic scales and logical structure. *Linguistic Inquiry* 6/3: 353–375.
- Gigafida 2.0: Korpus pisne standardne slovenščine*. https://www.clarin.si/noske/run.cgi/first_form?corpname=gfida20_dedup;align=
- Gregorčič, Kristina. 2018. *Affective polarity items in Slovene*. MA thesis. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Gregorčič, Kristina. 2021. Pomenske lastnosti in značilnosti rabe slovenskih oziralnopoljubnostnih zaimkov. *Slovenski jezik / Slovene Linguistic Studies* 13: 61–84. <https://doi.org/10.3986/sjls.13.1.04>
- Gregorčič, Kristina. 2023. *Semantični in pragmatični vidiki k negativni polarnosti usmerjenih nedoločnih zaimkov*. Dissertation. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Grice, Paul. 1975. Logic and conversation. In: P. Cole, J. L. Morgan (eds.). *Syntax and semantics 3: speech acts*. New York: Academic Press. 41–58.
- Grosu, Alexander. 2004. The syntax-semantics of modal existential wh-constructions. In: O. M. Tomić (ed.). *Balkan syntax and semantics*. Amsterdam: John Benjamins. 405–438.
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Horn, Larry R. 1989. *A natural history of negation*. Chicago: University of Chicago Press.
- Ilc, Gašper. 2019. *Aspects of negation*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Janežič, Anton. 1863. *Slovenska slovnica za domačo in šolsko rabo*. Celovec: [Janez Leon].
- Krek, Simon, Arhar Holdt, Špela, Čibej, Jaka, Repar, Andraž, Ljubešič, Nikola. 2019. *Specifikacije izdelave korpusa Gigafida 2.0*. https://www.cjvt.si/gigafida/wp-content/uploads/sites/10/2019/06/Gigafida2.0_specifikacije.pdf (30 September 2023).
- Ladusaw, William A. 1980. *Polarity sensitivity as inherent scope relations. Outstanding dissertations in linguistics*. New York: Garland Publications.
- Liu, Mingya. 2010. Universal quantification and NPI licensing. In: M. Prinzhorn, V. Schmitt, S. Zobel (eds.). *Proceedings of Sinn und Bedeutung* 14. 273–288. <https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/472> (10 June 2023).
- Rawlins, Kyle. 2013. Conditionals. *Natural Language Semantics* 21/2: 111–178.
- de Swart, Henriëtte. 1998. Licensing of negative polarity items under inverse scope. *Lingua* 105/3–4: 175–200. [https://doi.org/10.1016/S0024-3841\(98\)00021-7](https://doi.org/10.1016/S0024-3841(98)00021-7)

- Toporišič, Jože. 2000. *Slovenska slovnica*. Maribor: Obzorja.
- Vidovič Muha, Ada. 2013. *Slovensko leksikalno pomenoslovje*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Willis, David. 2013. Negation in the history of the Slavonic Languages. In: D. Willis, C. Lucas, A. Breitbarth (eds.). *The history of negation in the languages of Europe and the Mediterranean. Volume I: Case studies*. Oxford: Oxford University Press. 341–398.

Received August 2024, accepted September 2024.

Prispelo avgusta 2024, sprejeto septembra 2024.

SUMMARY

THE USE, MEANING AND FUNCTIONS OF SLOVENE BARE PRONOUNS

The paper presents selected aspects of the use and meaning of Slovene bare pronouns, identified through an extensive analysis of examples from the reference corpus *Gigafida 2.0*. The corpus data confirm our hypothesis regarding the inaccuracy of the existing classification of bare pronouns according to their indefinite pronoun functions (Willis 2013: 394). In most examples from our samples, bare pronouns perform the *irrealis non-specific* function, revealing the speaker's ignorance, which categorizes them as epistemic indefinite pronouns. They also perform the *conditional*, *question*, *indirect negation* and *direct negation* functions, typical of Negative Polarity Items. Their use in the scope of clausemate propositional negation seems to be restricted to modal existential *wh*-constructions, and to instances of the inverse scope of negation, stemming from the complexity of pronominal phrases.

The results of our analysis also show that the placement of nominal and adverbial bare pronouns in a clause is predictable. It primarily depends on the complexity of the pronominal phrase and the speaker's communicative intentions. Bare pronouns in clause-initial or medial positions indicate referents whose identity is not crucial to the message, whereas bare pronouns in the final position either emphasize the small number of referents or are part of more complex phrases in which the bare pronoun is followed by a postmodifier.

RABA, POMEN IN FUNKCIJE POLJUBNOSTNIH ZAIMKOV

Prispevek predstavlja izbrane vidike rabe in pomena poljubnostnih zaimkov, ki izhajajo iz obsežnejše analize zgledov iz korpusa *Gigafida 2.0*. Korpusno gradivo potrjuje izhodiščno hipotezo, da obstoječa opredelitev funkcij poljubnostnih zaimkov (Willis 2013: 394) ni točna. Poljubnostni zaimki namreč v največ zgledih v preučevanih

vzorci opravljajo funkcijo 'nespecifični nosnik v nestvarnih opisih'. V tej funkciji razkrivajo govorcevo nevednost, kar jih uvršča med epistemične nedoločne zaimke. Hkrati opravljajo funkcije 'pogoj', 'vprašanje', 'posredno zanikanje' in 'neposredno zanikanje', značilne za k negativni polarnosti usmerjene izraze. Korpusni podatki sicer kažejo, da se poljubnostni zaimki v neposrednem dosegu nikalnice ne pojavljajo prosto, temveč le v primeru rabe v modalnih strukturah z glagoloma *imeti* in *biti* ter ob pojavu obratnega dosega zanikanja, ki je posledica kompleksnosti zaimenske zveze.

Rezultati analize kažejo tudi predvidljivost stave samostalniških in prislovnih poljubnostnih zaimkov v stavku; ta je običajno odvisna od kompleksnosti zaimenske zveze in govorcevih sporočevalnih namenov. Na nekončnem mestu rabljeni poljubnostni zaimki označujejo nosnike, katerih identiteta za sporočilo ni bistvena. Na končnem mestu rabljeni poljubnostni zaimki izpostavljajo maloštevilnost nosnikov ali pa so del kompleksnejše besedne zveze, v kateri zaimku sledi desni prilastek.

Irena Stramljič Breznik (ORCID: 0000-0002-7267-6507)

Univerza v Mariboru, Filozofska fakulteta, Slovenija

irena.stramljic@um.si

Nina Ledinek (ORCID: 0000-0003-1068-3856)

ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija

nina.ledinek@zrc-sazu.si

DOI: <https://doi.org/10.3986/16.1.03>

KVANTITATIVNI PODATKI O BESEDOTVORNIH MODELIH IN PRIPONSKIH NIZIH IZMEDMETNIH TVORJENK V *BESEDNODRUŽINSKEM SLOVARJU SLOVENSKEGA JEZIKA ZA IZTOČNICE NA B*

V prispevku predstavimo starejšo in novejšo pomensko tipologizacijo medmetov ter njihovo besedotvorno zmožnost za tvorbo lastne ali drugih besednih vrst. Poskusni besednodružinski slovar, ki znotraj iztočnice združuje vse njene možne različnostopenjske nize, umestimo v kontekst morfotaktičnih raziskav v slovenistiki. Nato na podlagi devetnajstih v besednodružinski slovar vključenih medmetnih iztočnic na b predstavimo številske značilnosti njihovih priponskih nizov glede na stopnjo tvorbe, dolžino, priponsko kombinatoriko, število tvorjenk v slovarju in število njihovih potrditev v korpusu metaFida.

KLJUČNE BESEDE: medmet, besedotvorje, tvorjenke, izmedmetni stopenjski priponski nizi, morfotaktika

This article discusses the older and newer semantic typologies of interjections and their word formation properties – in relation to both the formation of other interjections and other parts of speech. We show how the Word Family Dictionary of the Slovene Language, which contains within the individual entries all possible word formation sequences at different levels, can be used for morphotactic research in Slovene. This article provides a statistical analysis of the word formation sequences derived from all nineteen interjection entries in the Word Family Dictionary of the Slovene Language. The analysis focuses on different stages in word formation, the length of suffix chains, the ability of suffixes to combine in suffix chains and the number of different multilevel derivations as well as their frequency in the metaFida corpus.

KEYWORDS: interjection, word formation, derivatives, word formation sequences of the interjection, morphotactics

1 UVOD

Kot ugotovljeno (Snoj 2022: 237), je bilo medmetom v sodobnem slovenskem jezikoslovju v zadnjih desetletjih posvečene precej pozornosti v izdajah slovenske slovnice od 1976 do 2004 (Toporišič ⁴2000: 450–462), leksikalnega pomenoslovja od 2000, 2013 do e-izdaje 2021 (Vidovič Muha 2021), v monografiji, posvečeni medmetom (Stramljič Breznik 2014a), in člankih (Stramljič Breznik 2012, 2013a, 2013b, 2013c, 2014b; Krvina, Žele 2017).

Raziskave sledijo spoznanju, da izhajajo kategorialne lastnosti leksemov iz njihove skladenjske vloge. Medmeti ali interjekcije sodijo v tisto skupino besednih vrst, ki imajo nestavčnočlensko vlogo, saj so modificiran izraz stavčne povedi, njihov denotat je govorno dejanje (Vidovič Muha 2000: 30–32). Zato jih lahko uporabimo kot samostojne pastavke, ki so v določenem kontekstu, skupnem tvorcu in naslovniku, popolnoma smiselni (Toporišič 2000: 450).

2 O DELITVI MEDMETOV

V slovenistiki (Toporišič 2000: 450–462) je utrjena delitev medmetov na: (a) razpoloženske, ki izražajo človekovo razpoloženje; (b) posnemovalne, ki posnemajo šume v zvezi s človekom, živalsko oglašanje, zvoke naprav in slušni vtis kretnje ali giba; in (c) velelne, s katerimi velevamo ljudem in živalim; z vabnimi klici vabimo živali; z zvalnicami in pozdravi pa vzpostavljamo medčloveški družbeni stik. Nekoliko drugačna, vendar ne bistveno, je delitev (Vidovič Muha 2000: 87), ki medmete prav tako razvršča v tri pomenske skupine: (a) razpoloženske, Toporišičevi velelni so preimenovani v (b) namerne, ker z določenimi variantami glasov ali njihovih povezav izražajo določeno namero v zvezi s človekom (zavrnitev, potrditev, dvom ...), v zvezi z živalmi (vabni klici, ukazi za določeno vedenje ...), in (c) posnemovalne (onomatopejske).

Po svoji naravi so medmeti lahko popolnoma samostojno besedilo, ki ne potrebuje niti diskurza niti jezikovnega konteksta, potrebuje le enega govorca, ki ubesedi svoj odnos do zunajjezikovne pojavnosti. Medmetno sporočilo praviloma ne pričakuje naslovnikovega odziva in pogosto ne potrebuje niti naslovnika oziroma je naslovnik govorec sam. Predstavljena delitev medmetov velja v mnogih primerih, vendar ni absolutna, saj medmeti – razen prvinskih iz prve skupine – niso zelo stabilne besede, saj se npr. od 103 enobesednih

medmetov, ki jih v svojih jezikoslovnih delih navaja Marko Pohlin v drugi polovici 18. stoletja, v SSKJ navaja le še 62 (Snoj 2022: 241).¹

Redakcija medmetnih iztočnic v *eSSKJ: Slovarju slovenskega knjižnega jezika, tretja izdaja*, je zato vodila do predloga njihove novejšje tipologizacije (Snoj 2022: 241–244), ki ji sledijo tudi slovarske razlage s pomenskouvrščevalnimi sestavinami *posnema, ponazarja, izraža* ali *se uporablja*:

- a) posnemovalni medmeti posnemajo glas/zvok; njihova podvrsta so (a1) ponazoritveni, s katerimi glasovno ponazorimo dejanje ali okoliščino;
- b) razpoloženski izražajo razpoloženje s podvrstami (b1) vzkličnih, s katerimi izražamo odnos do zaznavanja s čutili; (b2) kletvic, ki izražajo nejevoljo na bolj ali manj družbeno sprejemljiv način; in (b3) mašil, s katerimi zapolnjujemo premore v govoru ali nakazujemo govorčevo obotavljanje ali nadomeščamo zloge besedila, zlasti v refrenu pesmi;
- c) pozivni medmeti se od zgornjih skupin razlikujejo po navadno nižji stopnji ekspresivnosti ali njeni odsotnosti, pa tudi po večji usmerjenosti na naslovnika, soudeleženca v govornem dejanju, saj pozivajo človeka, boga ali živali k določenemu dejanju ali jih od njega odvrčajo, njihova podskupina so (c1) pozdravni medmeti in (c2) zvalnice, s katerimi želi govorec pritegniti sogovornikovo pozornost.

Medmeti so lahko samostojni stavki ali pastavki, in to jih družijo s členki, ki se od medmetov razlikujejo po tem, da se uporabljajo le znotraj jezikovnega konteksta, v dialogu, da se torej sklicujejo na vsebino aktualnega diskurza, poleg tega so členkovni (pa)stavki praviloma eliptični. Temeljna razlika med besednima vrstama po mnenju A. Žele je ta, da medmet ubeseduje subjektivno dožemanje dogodka, členek pa je opis subjektivnega dožemanja okoliščine (Snoj 2022: 241).

Tako kot ni dokončno utrjena tipologija medmetov in členkov, tudi uvrščanje izrazov v eno ali drugo besedno vrsto ni enostavno, še posebej v primerih, ko imajo taki izrazi del kategorialnih lastnosti medmeta in del kategorialnih lastnosti členska. Prav zato Snoj in Žele (2022: 558–572) podrobno predstavita osemindvajset meril za razpoznavanje razlikovalnih lastnosti medmetov in členkov na ravni izraznih (segmentalne, nadsegmentalne in tvorbenne lastnosti), besedilnoskladenjskih (medmeti kot samostojna sporočila za razliko od členkov,

¹ V določeni meri so razlike tudi posledica različnih konceptualnih predpostavk slovarjev glede vključitve iztočnic v slovar.

ki so lahko le znotrajbesedilni) in pomenskih značilnosti (zmožnost večpomenskosti, abstrakcije in vzpostavljanje sinonimnega razmerja). Predstavljena merila se uporabljajo pri uvrstitvi besed, pri katerih jim besednovrstnosti medmeta ali členka ni mogoče pripisati zlahka. Številska prevlada medmetnih oz. členkovnih lastnosti pa odloča o dokončnem pripisu eni ali drugi besedni vrsti, zlasti za potrebe besednovrstne kategorizacije leksemov v eSSKJ.

3 MEDMETI IN NJIHOV TVORBENI POTENCIAL

Medmeti lahko tvorijo lastno besedno vrsto, in to s postopki sklapljanja, ki so rezultat podvojevanja (*čiv-čiv*)² ali potrojevanja (*ha-ha-ha*) istih ali različnih medmetov (*hokus-pokus*), redkeje s sestavljanjem s predponskim obrazilom *pre-* (*pre-joj*) in z modifikacijsko izpeljavo, ko je medmet intenziviran s sestavinami *-ej*, *-la*, *-sa*, *-ta* ipd. (*juh – juh-ej – juhej-sa*; *hoj – hoj-la*; *jeh – jeha-ta*).

Vendar so medmeti mnogo pogostejši v vlogi reprodukcijских prvin za druge besedne vrste. Medmet kot motivirajoča beseda uporablja dva tvorbeni načina, tj. konverzijo³ in izpeljavo, s katero najpogosteje motivira druge besedne vrste in tvori obsežnejše besedne družine, še posebej zaradi sistemske tvorbe izmedmetnih glagolov, ki po predvidljivih vzorcih svoje besedne vrste tvorijo nadaljnje višjestopenjske tvorjenke (Stramljič Breznik 2014a: 47–64).

V slovenščini je izpeljava glagolov iz medmetov produktivni tvorbeni vzorec (Toporišič 2000: 213); tako nastali glagoli izražajo naslednje pomene: a) slušni vtis, ki ga povzroča predmet, z obrazili: *-ati* (*bevsكاتi*), *-niti* (*čofniti*); b) oglašanje živega bitja: *-tati* (*hehetati*), *-ljati* (*fefljati*), *-utati* (*cvrkatati*), *-ati* (*gagati*), *-oleti* (*žvrgoleti*), *-liti* (*gruliti*); c) glas pri premikanju: *-ati* (*frčati*), *-ljati* (*frfljati*), *-leti* (*frleti*), *-oleti* (*frfoleti*); č) glas ob kinetičnem dejanju: *-ati* (*bumfati*), *-kati* (*hukati*); d) glas ob govorjenju značilnega medmeta: *-ati* (*ahati*), *-niti* (*ahniti*) in e) glas, ki nakazuje premikanje več bitij hkrati: *-oleti* (*mrgoleti*). Glede na sinhrono motivacijo so lahko problematični nekateri primeri glagolov, pri katerih ne moremo vzpostaviti obstoječega medmeta kot motivirajoče besede, npr. **fef*, **žvrg*, zato lahko zanje samo diahrono dokažemo, da so onomatopejskega izvora.

² Z vezajem so označene morfemske meje. Sklopljeni medmeti se namreč lahko zapisujejo skupaj (*čivčiv*, *hokuspokus*), lahko tudi kot besedna zveza (*hokus pokus*), prim. Fran.si.

³ V tem prispevku ji ne bomo posvečali pozornosti. Več v: Stramljič Breznik 2014a; Snoj 2022; Snoj, Žele 2022.

Za tvorbo glagolov sta najproduktivnejši skupini posnemovalnih (*beketati*, *bevskati*, *brbrati*, *brenkati*, *brundati*, *čivkati*, *drdrati*, *hrzati*, *javkati*, *kikirikati*, *kokodajsati*, *kukati*, *kvakati*, *meketati*, *škrtati* ...) in razpoloženskih medmetov (*ahati*, *bogmati se*, *dušati se*, *hudičati*, *madonati*, *ohati*, *primojdušati se*, *sakramentirati* ...) ⁴. Tovrstno tvorbeno aktivnost, ki pa je manjša, kažejo tudi velelni medmeti (*ajati* otr. 'spati', *eksati* 'izpiti do dna', *hopati* 'poskakovati', *hopsati* 'poskakovati', *pardonirati* 'oproščati', (*s/š*)*topati* 'ustavljati avtomobile'). Še bolj tvorbeno omejeni pa so pozdravni medmeti, pri katerih smo ugotovili le eno tovrstno tvorjenko (*hajlati* 'pozdravljati z nacističnim pozdravom heil') (Stramljič Breznik 2014a: 47–64).

4 O ZAČETKIH MORFOTAKTIČNIH RAZISKAV V SLOVENISTIKI IN VIRIH ZANJE

V sodobnem slovenskem jezikoslovju se do preloma stoletja besedotvorne raziskave ne ukvarjajo z obrazilno kombinatoriko ali pa jo omenjajo le obrobno (Vidovič Muha 1988; Toporišič 2000), čeprav so bile za večino slovanskih jezikov (zlasti češčino, slovaščino, poljščino in ruščino) že izdelane morfemske analize (prim. Stramljič Breznik 2004: 8), na osnovi katerih so bili v nadaljevanju izdani izčrpni oblikoslovno-besedotvorni ali samo besedotvorni morfološki slovarji, npr. slovaški *Morfematský slovník slovenčiny* M. Sokolove idr. (1999), poljski *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego* H. Jadacke idr. (2001), M. Skarżyńskiego (2004), češki *Bazový morfematský slovník češtiny* Z. Šiška (2005), poljski *Słownik gramatycznego języka polskiego* Z. Salonija idr. (2015), ruski *Словарь словообразовательных аффиксов современного русского языка* V. V. Lopatina in I. S. Uluhanova (2016), *Slovník slovotvorných prostředků v slovenčine* M. Ološtiaka in M. Ivanove (2021).

Šele v zadnjih dveh desetletjih se je v slovenistiki pozornost tej problematiki sistematičneje posvetila z besednodružinskim oz. stopenjskim pristopom. Metodologija stopenjskega besedotvorja, ki predstavlja eno od možnih izhodišč za morfološke raziskave, je bila v slovenskem jezikoslovnem prostoru preizkušena na gradivu besednih družin iztočnic na črko b (Stramljič Breznik 2004) ter na primeru glagolov čutnega zaznavanja (Kern 2017, 2020).

⁴ Tukaj v primerih npr. *hudičati*, *madonati* razumemo, da so prvotni samostalniki po konverziji prešli med medmete (Toporišič 2000: 457; Stramljič Breznik 2014a: 49).

Iz besednodružinskega slovarja so bili analizirani tudi feminativi v izsamostalniških nizih (Kern 2023).⁵

Besednodružinski slovar slovenskega jezika – Poskusni zvezek za iztočnice na b (BSSJ) je zasnovan kot tip enojezičnega specialnega slovarja, ki prinaša podatke o tvorbeni povezanosti besed. To pomeni, da so ob nemotivirani iztočnici v ustreznih tvorbenih povezavah prikazane vse tvorjenke s skupnim korenem, ki so dosledno členjene na podstavni in obrazilni del. Knjižni slovar, izdelan na podlagi elektronskega, obsega iztočnice na b s 666 besednimi družinami, v katerih je razvrščenih 11.136 enot. Slovar je bil ob nastanku pri vključevanju gradiva zavestno in hote izčrpen z namenom, da bi se zajela vsa tvorbeno problematika in nakazale njene potencialne rešitve, ki bi lajšale nadaljnjo sestavo. Temeljni gradivni viri so bili: *Slovar slovenskega knjižnega jezika* (SSKJ 1), *Besedišče slovenskega jezika* (BSJ), *Slovenski pravopis 2001* (SP) ter po referenčnem korpusu Fida⁶ paberkovalno zbrane nekatere besede črke b, manjkajoče v naštetih virih (Stramljič Breznik 2004). Tako zasnovan slovar predstavlja dobro izhodišče za ugotavljanje temeljne kombinatorike besedotvornih obrazil in njihovega preverjanja v drugih leksikalnih (slovarskih) virih in korpusih.

Med 666 iztočnicami je v slovar vključenih 19 medmetov, kar predstavlja 2,85 % vseh iztočnic. Medmetnih tvorjenk je le 38, in to predstavlja le 0,37-odstotni delež, od tega jih je 34 na prvi in 4 na drugi stopnji (Stramljič Breznik 2005: 505–520).

Celovite morfosintaktične raziskave v zadnjem času potekajo tudi pri projektu KOBOS.⁷ Trenutni projektni izsledki,⁸ zamejeni na uporabo podatkov iz besednodružinskega slovarja (Stramljič Breznik 2004), v katerem je vključenih 19 medmetnih iztočnic, prinašajo v nadaljevanju predstavljene kvantitativne podatke o besedotvornih nizih in kombinatoriki desnih, tj. priponskih, obrazil izmedmetnih tvorjenk, katerih pojavitve preverjamo v združenem korpusu metaFida 1.0 (v nadaljevanju: mFida), ki zajema 6 milijard pojavnic in združuje najpomembnejše javno dostopne slovenske korpuse ter omogoča enovito ter pregledno iskanje po njih.

⁵ V okviru projekta KOBOS je nastalo že več analiz, ki jih podrobneje predstavljamo v Stramljič Breznik 2024.

⁶ V času nastajanja slovarja sta bila edina razpoložljiva korpusa Fida (Filozofska fakulteta Univerze v Ljubljani, Inštitut Jožef Stefan, Založba DZS, d. d., Amebis, d. o. o.) in Nova beseda (Inštitut za slovenski jezik Frana Ramovša ZRC SAZU).

⁷ Partnerji v projektu KOBOS (2021–2024) so: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, Inštitut Jožef Stefan in Filozofska fakulteta, Univerza v Mariboru.

⁸ Projekt vključuje tudi jezikovnotehnološki del, v katerem bo oblikovan model za strojno besedotvorno analizo slovenskih besed. Prvi rezultati raziskav so na voljo v Pollak idr. 2023 in Erjavec idr. 2023.

5 KVANTITATIVNE ZNAČILNOSTI IZMEDMETIH PRIPONSKIH NIZOV⁹

V primeru izmedmetnih stopenjskih tvorjenk je maksimalno število priponskih obrazil v nizu 4.¹⁰ Obstajata le dva različna priponska niza s prvostopenjsko glagolsko tvorjenko, vsak od njiju je potrjen z eno tvorjenko v BSSJ. Štiripriponski nizi obrazil pri izmedmetnih tvorjenkah so torej redki. Kot že ugotovljeno, obstajata le dva, ki sta brez potrjenih tvorjenk v združenem korpusu slovenskega jezika in predstavljata 2,63-odstotni delež vseh izmedmetnih nizov. Pričakovano več pa je tri- (20, tj. 26,32 %) oz. dvopriponskih nizov (54, ki predstavljajo 71,05 %), saj se njihovo število glede na tripriponske več kot podvoji. Sorazmerni temu so tudi deleži nastalih tvorjenk in njihovih potrditev v korpusu. Dvopriponski nizi s 75 tvorjenkami v BSSJ ustvarijo 75,76-odstotni delež izmedmetnih tvorjenk, ki so v 20,20-odstotnem deležu (20 tvorjenk) tudi korpusno potrjene. Pri tripriponskih znašata deleža 22,22 % (22 tvorjenk) in 7 korpusno potrjenih (7,07 %), pri dveh štiripriponskih nizih z 2 tvorjenkama v BSSJ pa se delež tvorjenk drastično zmanjša (2,02 %) in tvorjenki tudi korpusnih potrditev nimata. Skupno je iz BSSJ korpusno potrjena dobra četrtnina (27,27 %) tvorjenk iz 76 dvo-, tri- ali štiripriponskih nizov.

Dolžina nizov	Št. različnih priponskih nizov	Delež v %	Št. tvorjenk	Delež v %	Št. v mFidi potrjenih tvorjenk	Delež v %
Dvopriponski	54	71,05	75	75,76	20	74,07
Tripriponski	20	26,32	22	22,22	7	25,93
Štiripriponski	2	2,63	2	2,02	0	0,00

PREGLEDNICA 1: Število različnih priponskih kombinacij v priponskih nizih

Zanimivo je pogledati, kakšni so najtipičnejši besedotvorni modeli za izmedmetne podstave. Potrjuje se ugotovitev, znana tudi iz besedotvornih opisov drugih slovanskih jezikov, da medmet kot besedotvorno izhodišče na prvi stopnji najpogosteje generira glagolske tvorjenke. V vseh šestinsedemdesetih dvo-,

⁹ Zgled za analizo kvantitativnih lastnosti izmedmetnih tvorjenk predstavlja monografija Iwone Burkacke z naslovom *Kombinatoryka sufiksalna w polskiej derywacji odrzeczownikowej* (2012).

¹⁰ Pri interpretaciji nizov, v katerih eno stopnjo predstavljajo sestavljenke (npr. *brenk*: *-ati + po-Se + -ava-ti + -anje*), upoštevamo, da je niz tripriponski, saj nas v analizi zanima dolžina nizov, ki jih tvorijo izključno priponska obrazila. Zapis *po-Se* pomeni sestavljenka s predpionskim obrazilom *po-*.

tri- in štiripriponskih nizih se to zgodi kar v sedemdesetih primerih (92,17 %). V dvopriponskih izmedmetnih nizih sta najpogostejši besednovrstni zaporedji tvorjenk glagol in nato samostalnik (I,V,S)¹¹ z 28 različnimi priponskimi nizi. S po 7 nizi so zastopani modeli, ki imajo na drugi stopnji vmesno predponsko obraziljeno glagolsko sestavljenko, zato je niz dejansko dvopriponski: I,V,V,Ad (npr. *-ati* + (*na-Se*) + *-n*; *nabunkan*); I,V,V,S (npr. *-ati* + (*od-Se*) + *-anje*; *odbrenkanje*); I,V,V,V največkrat tudi z drugostopenjsko vmesno predponsko obraziljeno glagolsko sestavljenko (npr. *-ati* + (*po-Se*) + (*o/a*)*va-ti* ali *-lj-ati*; *pobrenk(o/a)vati*, *pribrenkljati*), le enkrat se dvopriponski model zaključi z glagolsko sestavljenko (*-ati* + *-k-ati* + *po-Se*; *pobuckati*).

V tripriponskih nizih po številu različnih priponskih kombinacij (največkrat gre za po 4 različne kombinacije)¹² zelo izenačeno nastopajo besedotvorni modeli I,V,Ad,S (npr. *-ati* + *-av* + *-ec*; *brbravec*); I,V,S,Ad (npr. *-ati* + *-ač* + *-ev*; *brenkačev*) in I,V,S,S (npr. *-ati* + *-ar* + *-stvo*; *brenkarstvo*). Med osamele besedotvorne modele s po enim priponskim nizom sodijo I,V,S,V (*-ati* + *-ika* + *-ariti*; *bucikariti*); I,V,V,Ad (*-ati* + *-lj-ati* + *-av*; *brenkljav*) in I,V,V,S (*-ati* + *-lj-ati* + *-anje*; *brenkljanje*). Modela I,V,V,V,Ad (*-ati* + *po-Se* + *-ava-ti* + *-oč*; *pobrenkavajoč*) in I,V,V,V,S (*-ati* + *po-Se* + *-ava-ti* + *-anje*; *pobrenkavanje*) imata v nizu na 2. stopnji tudi glagolsko sestavljenko.

Štiripriponski nizi, kot že omenjeno, so redki. Zastopana sta le dva besedotvorna modela, vsak s po enim nizom: I,V,S,S,Ad (*-ati* + *-a* + *-ica* + *-ast*; *bunčičast*) in I,V,S,V,S (*-ati* + *-ika* + *-ariti* + *-enje*; *bucikarjenje*). Prav tako imata oba modela na prvi stopnji glagolsko tvorjenko.

Redki so modeli, ko je prvostopenjska tvorjenka medmetni sklop. Taka modela sta: I, I,V,S (npr. *Sk*l + *-ati* + *-anje*; *štrbunkanje*), I, I,V,V (*Sk*l + *-ati* + *-lj-ati*; *štrbunkljati*).

Medmeti, ki na prvi stopnji generirajo tvorjenko iste besedne vrste, tvorijo le v dveh primerih tudi četrstopenjsko tvorjenko z modelom I, I,V,V,S in priponskim nizom *Sk*l + *-ati* + *-lj-ati* + *-anje* (*štrbunkljanje*)¹³ brez korpusne potrditve ter I,I,S,Ad,S (*Sk*l + *-ø* + *-ski* + *-ina*; *bavbavščina*), ko medmetni

¹¹ Krajšave za besedne vrste izhajajo iz latinskih poimenovanj: Ad = pridevnik, Adv = prislov, S = samostalnik, V = glagol, I = medmet. Druge krajšave: Se = sestavljenka, Skl = sklop, Z = zloženka.

¹² Kot ponazoritev izpostavljamo le eno priponsko kombinacijo.

¹³ V tem primeru gre za t. i. sklopno neglasovno obrazilo. SES navaja, da gre za onomatopejo, ki posnema pri padcu v vodo nastale zvoke. Tudi pri samostalniku *bunka* se kot izhodišče omenja onomatopeja *bunk* v pomenu 'posnemanje glasu pri udarcu ali padcu', sklicujoč se na *štrbunk* iz Pleteršnikovega slovarja (ESSJ I: 54). Torej lahko sklepamo, da je medmet *štrbunk* dvodelen in rezultat sklapljanja dveh sestavin.

sklop konvertira v samostalnik, iz katerega nastane vrstni pridevnik, iz njega pa ponovno samostalnik.

Besedotvorni model	Dvopriponski nizi	Primeri v BSSJ	Št. tvorjenk	Št. potrjenih tvorjenk v mFidi
I,V,S	-ati + -ø	brenk, bevs ¹⁴	2	2
	-ati + -a	bunka, brbra, bumfa	3	2
	-ati + -ač	brenkač, bevskač, brbrač	3	2
	-ati + -aj	brenkaj	1	0
	-ati + -alec	brenkalec	1	1
	-ati + -alo	brenkalo	1	1
	-ati + -anje	brenkanje, bevskanje, bunkanje, brbranje, bumfanje, blunkanje, babljanje	7	4
	-ati + -ar	brenkar	1	0
	-ati + -arija	brenkarija	1	0
	-ati + -e	brbre	1	0
	-ati + -elj	bunkelj	1	1
	-ati + -ež	bunkež	1	1
	-ati + -ica	brenčica	1	0
	-ati + -ika	bucika	1	1
	-ati + -ljaj	brenkljaj	1	0
	-ati + -ot	brenkot	1	0
	-jati + -anje	blejanje	1	1
	-jati + -če	blejče	1	0
	-kati + -ø	bevk	1	0
	-kati + -anje	bavkanje, bekanje, bevkaj	3	0
	-kati + -la	bevkla	1	0
	-kati + -ot	bevkot	1	0
	-ketati + -ø	beket	1	1
	-ketati + -ača	beketača	1	0
	-ketati + -alec	beketalec	1	0

¹⁴ Izhajamo iz medmeta (*bevs*), njegove izglagolske tvorjenke (*bevs-ati*), iz katere je edino mogoče tvoriti samostalnik s pomenom dejanja ali rezultat dejanja (*bevs-0*).

	-ketati + -anje	beketanje	1	1
	-ketati + -aš	beketaš	1	0
	-niti + -enje	brenknjenje, bumnjenje, bunknjenje	3	0
I,V,S,S	-ati + -a + Z	polbunka	1	0
I,V,V,Ad	-ati + iz-Se + -n	izbevskan	1	0
	-ati + na-Se + -n	nabunkan	1	0
	-ati + od-Se + -n	odbrenkan	1	0
	-ati + o-Se + -n	obunkan, obrenkan	2	0
	-ati + po-Se + -n	pobunkan	1	0
	-ati + pre-Se + -n	prebunkan	1	0
	-ati + za-Se + -n	zabunkan, zabrenkan	2	0
I,V,V,S	-ati + na-Se + -anje	nabunkanje	1	0
	-ati + od-Se + -anje	odbrenkanje	1	0
	-ati + o-Se + -anje	obrenkanje, obunkanje	2	0
	-ati + pre-Se + -anje	prebunkanje	1	0
	-ati + pri-Se + -anje	pribevskanje	1	0
	-ati + raz-Se + -anje	razbrbranje	1	0
	-ati + za-Se + -anje	zabrenkanje, zabrbranje, zabunkanje, zabevskanje	4	0
I,V,V,V	-ati + -k-ati + po-Se	pobuckati	1	0
	-ati + o-Se + -ava-ti	obrenkavati	1	0
	-ati + po-Se + -ova-ti	pobrenkovati	1	0
	-ati + po-Se + -ava-ti	pobrenkavati	1	1
	-ati + pri-Se + -ava-ti	pribrenkavati	1	0
	-ati + pri-Se + -lj-ati	pribrenkljati	1	0
	-ati + za-Se + -lj-ati	zabrenkljati	1	0
I,I,V,S	Sk1 + -ati + -anje	štrbunkanje	1	0
	Sk1 + -niti + -ø	štrbunk ¹⁵	1	1
	Sk1 + -niti + -elj	štrbunkelej	1	0
I,I,V,V	Sk1 + -ati + -lj-ati	štrbunkljati	1	0

PREGLEDNICA 2: Besedotvorni modeli in frekvenca dvopriponskih nizov

¹⁵ Izhajamo iz sklopljenega medmeta (*štrbunk*), njegove izglagolske tvorjenke (*štrbunk-ati*), iz katere je edino mogoče tvoriti samostalnik s pomenom dejanja ali rezultat dejanja (*štrbunk-0*).

Besedotvorni model	Tripriponski nizi	Primeri iz BSSJ	Št. tvorjenk	Št. potrjenih tvorjenk v mFidi
I,V,Ad,S	-ati + -av + -ec	brbravec	1	0
	-ati + -av + -ost	brbravost	1	0
	-jati + -alen + -ica	blejalnica	1	0
	-kati + -av + -ec	bevkavec	1	0
I,V,S,Ad	-ati + -a + -ast	bunkast, bumfast	2	1
	-ati + -ač + -ev	brenkačev	1	0
	-ati + -ika + -ast	bucikast	1	1
	-ati + -ika + -in	bucikin	1	1
I,V,S,S	-ati + -a + -ača	bunkača	1	0
	-ati + -a + -ica	bunkica, bunčica	2	2
	-ati + -a + -ula	bunkula	1	1
	-ati + -ar + -stvo	brenkarstvo	1	0
I,V,S,V	-ati + -ika + -ariti	bucikariti	1	0
I,V,V,Ad	-ati + -lj-ati + -av	brenkljav	1	0
I,V,V,S	-ati + -lj-ati + -anje	brenkljanje	1	0
I,V,V,V,Ad	-ati + po-Se + -ava-ti + -oč	pobrenkavajoč	1	0
I,V,V,V,S	-ati + po-Se + -ava-ti + -anje	pobrenkavanje	1	1
	-ati + po-Se + -ova-ti + -anje	pobrenkovanje	1	0
I, I,V,V,S	Skl + -ati + -lj-ati + -anje	štrbunkljanje	1	0

PREGLEDNICA 3: Besedotvorni modeli in frekvenca tripriponskih nizov

Besedotvorni model	Štiripriponski nizi	Primeri iz BSSJ	Št. tvorjenk	Št. potrjenih tvorjenk v mFidi
I,V,S,S,Ad	-ati + -a + -ica + -ast	bunčičast	1	0
I,V,S,V,S	-ati + -ika + -ariti + -enje	bucikarjenje	1	0

PREGLEDNICA 4: Besedotvorni modeli in frekvenca štiripriponskih nizov

V nadaljevanju predstavljamo, kateri besedotvorni modeli so najpogostejši po številu priponskih kombinacij, po številu tvorjenk v BSSJ in po številu potrditev v korpusu mFida. Izmed šestnajstih v BSSJ prepoznanih besedotvornih modelov po številu priponskih kombinacij opazno prevladuje model I,V,S,

ki z 28 kombinacijami predstavlja 36,84-odstotni delež s 43 tvorjenkami, od katerih jih je manj kot polovica (18) potrjenih v korpusu.

Z 10,52-odstotnim deležem (po 8 priponskih kombinacij) nastopata modela I,V,V,Ad in I,V,V,S s po 10 oz. 12 tvorjenkami, vendar vmesno stopnjo lahko predstavljajo tudi predpanska obrazila sestavljenk, npr. pri prvem (*-ati + za-Se + -n*: *bunkati* → *zabunkati* → *zabunkan* in podobno *zabrenkan*) oz. npr. pri drugem modelu (*-ati + za-Se + -anje*: *brenkati* → *zabrenkati* → *zabrenkanje*, enako *zabrbranje*, *zabunkanje*, *zabevskanje*). Tvorjenke nimajo potrditev v korpusu, kar ni presenetljivo. Marvin Derganc (2019: 251–266) je namreč v SSKJ 1 in v korpusu Gigafida preverjala pojavnost glagolnikov na *-je*¹⁶ iz dovršnih glagolov s predpono *iz-/z-/s-/po-*. Ugotovila je, da se v slovarju glagolniki, tvorjeni iz dovršne osnove, pojavljajo zelo redko. V korpusu Gigafida je raba sicer pogostejša (od analiziranih primerov s predpono *iz-/z-/s-* jih ima v korpusu kar 50 % vsaj en zabeležen glagolnik iz dovršnika, s predpono *po-* pa 30 %), vendar gre večinoma za majhno frekvenco pojavljanja (manj kot 10 pojavnic), še posebej pri predponi *po-*. Iz analize omenjenega korpusa je razvidno, da govorci v sodobni slovenščini, sicer v omejenem obsegu, tvorijo glagolnike iz dovršnih glagolov s predponami z vsemi tremi osnovnimi pomeni, to so krajevni pomen, pomen vrste glagolskega dejanja ter pomen dovršnosti. Nekateri podpomeni vrste glagolskih dejanj (delimitativni in inkohativni pri predponi *po-*) ter dovršni pomen (pri predponi *po-*) so za tvorbo iz dovršne osnove manj dovtetni; v korpusu nastopi le nekaj osamljenih, zanemarljivih pojavitev, saj so dovršni glagolniki namreč mnogokrat rabljeni kot iztočnice za gesla v križankah (*potratenje*, *porabljenje*) ali se pojavljajo v neelektoriranih besedilih neznanih avtorjev, ko govorec želi poudariti dovršnost glagolskega dejanja kot besedotvornega izhodišča, npr. *pobranje*, *pojedenje* (Marvin Derganc 2019: 265–266).

Po številu v BSSJ potrjenih besedotvornih kombinacij (7) in tvorjenk (7) izstopa še model I,V,V,V (npr.: *-ati + za-Se + -lj-ati*; *zabrenkljati*) – tudi tukaj eno stopnjo tvorbe predstavlja sestavljenka –, vendar le z eno korpusno potrditvijo take tvorjenke. Preostali modeli I,V,S,S (5); I,V,Ad,S (4); I,V,S,Ad (4); I,I,V,S (3); I,V,V,V,S (2) imajo uresničenih od 5 do 2 priponski kombinaciji s številom tvorjenk med 6 in 2, vendar z najmanj eno korpusno potrditvijo, razen modela I,V,Ad,S, ki je nima. Po številu korpusno potrjenih tvorjenk (3)

¹⁶ To vključuje obrazila *-(a/e)n-je* ali *-t-je*.

sta izstopajoča modela I,V,S,S in I,V,S,Ad. Preostalih sedem besedotvornih modelov predstavlja praktično osamelce z eno priponsko kombinacijo, ki je v viru (BSSJ) potrjena z eno tvorjenko, a brez potrditve v korpusu mFida.

Besedotvorni model	Št. priponskih kombinacij	Št. tvorjenk	Št. tvorjenk s pojavitvami v mFidi
I,V,S	28	43	18
I,V,V,Ad	8	10	0
I,V,V,S	8	12	0
I,V,V,V	7	7	1
I,V,S,S	5	6	3
I,V,Ad,S	4	4	0
I,V,S,Ad	4	5	3
I,I,V,S	3	3	1
I,V,V,V,S	2	2	1
I,I,V,V	1	1	0
I,I,V,V,S	1	1	0
I,S,Ad,S	1	1	0
I,V,S,S,Ad	1	1	0
I,V,S,V	1	1	0
I,V,S,V,S	1	1	0
I,V,V,V,Ad	1	1	0

PREGLEDNICA 5: Besedotvorni modeli glede na število priponskih kombinacij in število tvorjenk

Pogledali bomo še, katera obrazila so na začetku priponskih nizov in katera so najbolj produktivna. Medmet na prvi stopnji najpogosteje generira glagol, ki v dvo-, tri- in štiripriponskih nizih največkrat nastopa z obrazilom *-ati*. Poleg *-ati* v dvopriponskih nizih nastopajo še obrazila *-kati*, *-ketati* in *-niti*.

Kot že opaženo, z obrazilom *-ati* izstopa dvopriponski niz *-ati* + *-anje* z 8 tvorjenkami, od katerih je 5 potrjenih v mFidi. S po dvema korpusnima potrditvama nastopata iz BSSJ še dvopriponska niza *-ati* + *-ač* (npr. *brenkač*), *-ati* + *-a* (npr. *bunka*) in *-ati* + *-ø* (npr. *bevsk*), s po eno pa *-ati* + *-alec* (*brenkalec*), *-ati* + *-alo* (*brenkalo*), *-ati* + *-elj* (*bunkelj*), *-ati* + *-ež* (*bunkež*) in *-ati* + *-ika* (*bucika*). Preostali nizi s tem začetnim glagolskim obrazilom so enkratnice v viru in zato pričakovano brez korpusnih potrditev.

Med obrazili *-kati*, *-ketati* in *-niti* s po tremi primeri v BSSJ nastopata le niza *-kati + -anje* (npr. *bavkanje*) in *-niti + -enje* (npr. *brenknjenje*), vendar brez potrditev v rabi. Vsi drugi nizi z izpostavljenimi tremi začetnimi glagolskimi obrazili imajo v viru le eno potrditev, izmed njih pa sta z eno pojavitvijo v korpusu potrjeni samo obrazilni kombinaciji *-ketati + -ø* (*beket*) in *-ketati + -anje* (*beketanje*).

Prvo priponsko obrazilo	Priponski niz	Realizacija niza	Primer	Št. tvorjenk	Št. potrjenih tvorjenk v mFidi
-ati	-ati + -ø	-ø	bevsk	2	2
	-ati + -a	-a	bunka	3	2
	-ati + -a + Z	-a	polbunka	1	0
	-ati + -ač	-ač	brenkač	3	2
	-ati + -aj	-aj	brenkaj	1	0
	-ati + -alec	-alec	brenkalec	1	1
	-ati + -alo	-alo	brenkalo	1	1
	-ati + -anje	-anje	brenkanje	8	5
	-ati + -ar	-ar	brenkar	1	0
	-ati + -arija	-arija	brenkarija	1	0
	-ati + -če	-če	blejče	1	0
	-ati + -e	-e	brbre	1	0
	-ati + -elj	-elj	bunkelj	1	1
	-ati + -ež	-ež	bunkež	1	1
	-ati + -ica	-ca	brenčica	1	0
	-ati + -ika	-ika	bucika	1	1
	-ati + iz-Se + -n	-an	izbevskan	1	0
	-ati + -k-ati + po-Se	-kati	pobuckati	1	0
	-ati + -ljaj	-ljaj	brenkljaj	1	0
	-ati + na-Se + -anje	-anje	nabunkanje	1	0
	-ati + na-Se + -n	-an	nabunkan	1	0
	-ati + od-Se + -anje	-anje	odbrenkanje	1	0
	-ati + od-Se + -n	-an	odbrenkan	1	0
	-ati + o-Se + -anje	-anje	obrenkanje	2	0
	-ati + o-Se + -ava-ti	-avati	obrenkavati	1	0
	-ati + o-Se + -n	-an	obunkan	2	0

	-ati + -ot	-ot	brenkot	1	0
	-ati + po-Se + -n	-an	pobunkan	1	0
	-ati + po-Se + -ova-ti	-ovati	pobrenkovati	1	0
	-ati + po-Se + -ava-ti	-avati	pobrenkavati	1	1
	-ati + pre-Se + -anje	-anje	prebunkanje	1	0
	-ati + pre-Se + -n	-an	prebunkan	1	0
	-ati + pri-Se + -anje	-anje	pribevskanje	1	0
	-ati + pri-Se + -ava-ti	-avati	pribrenkavati	1	0
	-ati + pri-Se + -lj-ati	-lj-ati	pribrenkljati	1	0
	-ati + raz-Se + -anje	-anje	razbrbranje	1	0
	-ati + za-Se + -anje	-anje	zabrenkanje	4	0
	-ati + za-Se + -lj-ati	-ljati	zabrenkljati	1	0
	-ati + za-Se + -n	-an	zabunkan	2	0
	Skl + -ati + -anje	-anje	štrbunkanje	1	0
	Skl + -ati + -lj-ati	-ljati	štrbunkljati	1	0
-kati	-kati + -ø	-ø	bevk	1	0
	-kati + -anje	-anje	bavkanje	3	0
	-kati + -la	-la	bevkla	1	0
	-kati + -ot	-kot	bevkot	1	0
-ketati	-ketati + -ø	-ket	beket	1	1
	-ketati + -ača	-ketača	beketača	1	0
	-ketati + -alec	-ketalec	beketalec	1	0
	-ketati + -anje	-ketanje	beketanje	1	1
	-ketati + -aš	-ketaš	beketaš	1	0
-niti	-niti + -enje	-njenje	brenknjenje	3	0
	Skl + -niti + -ø	-ø	štrbunk	1	1
	Skl + -niti + -elj	-elj	štrbunkelj	1	0

PREGLEDNICA 6: Dvopriponski nizi z glagolskim začetnim priponskim obrazilom

Med tripriponskimi nizi na prvi stopnji izmedmetne tvorbe nastopata le obrazili *-ati* in *-kati*. Pričakovano je prvo tvorno za največ priponskih nizov (17), ki lahko imajo v svojem sestavu tudi za nas nerelevantno predponsko obrazilo. Pri obrazilu *-ati* imata v BSSJ potrjeni 2 pojavitvi le *-ati + -a + -ast* in *-ati + -a + -ica* (npr. *bunkast* in *bunkica*), v drugih primerih gre za enkratne tvorbene nize. Zanimivo pa je, da je med njimi kljub temu več njih tudi korpusno

potrjenih: *-ati + -a + -ica* (2 pojavitvi) in *-ati + -a + -ast*, *-ati + -ika + -ast*, *-ati + -ika + -in* in *-ati + -a + -ula* (po 1 pojavitev). Obrazilo *-kati* je začetno le v nizu *-kati + -av + -ec* (*bevkavec*), ki pa nima potrditve v korpusu.

Prvo priponsko obrazilo	Priponski niz	Realizacija niza	Primer	Št. tvorjenk	Št. potrjenih tvorjenk v mFidi
-ati	-ati + -alen + -ica	-alnica	blejalnica ¹⁷	1	0
	-ati + -av + -ost	-avost	brbravost	1	0
	-ati + -av + -ec	-avec	brbravec	1	0
	-ati + -a + -ast	-ast	bunkast	2	1
	-ati + -ika + -ast	-ikast	bucikast	1	1
	-ati + -ika + -in	-ikin	bucikin	1	1
	-ati + -ač + -ev	-ačev	brenkačev	1	0
	-ati + -a + -ica	-ica	bunkica	2	2
	-ati + -a + -ula	-ula	bunkula	1	1
	-ati + -ar + -stvo	-arstvo	brenkarstvo	1	0
	-ati + -a + -ača	-ača	bunkača	1	0
	-ati + -ika + -ariti	-ikariti	bucikariti	1	0
	-ati + -lj-ati + -av	-ljav	brenkljav	1	0
	-ati + -lj-ati + -anje	-ljanje	brenkljanje	1	0
	-ati + po-Se + -ava-ti + -oč	-avajoč	pobrenkavajoč	1	0
	-ati + po-Se + -ava-ti + -anje	-avanje	pobrenkavanje	1	1
-ati + po-Se + -ova-ti + -anje	-ovanje	pobrenkovanje	1	0	
-kati	-kati + -av + -ec	-kavec	bevkavec	1	0

PREGLEDNICA 7: Tripriponski nizi z glagolskim začetnim priponskim obrazilom

Štiripriponski nizi v izmedmetni tvorbi so redkost, zato sta za prvostopenjske glagolske tvorjenke ugotovljena le dva niza z začetnim obrazilom *-ati* (*-ati + -a + -ica + -ast* in *-ati + -ika + -ariti + -enje*), ki v mFidi pričakovano ostajata brez potrditev.

¹⁷ Iz medmeta *ble* je tvorjen glagol *blej-ati*. Na morfemski meji zev med samoglasnikom podstave in obrazila zapira *j*, ki je v tem primeru upoštevan kot del podstave in ne obrazila.

Prvo priponsko obrazilo	Priponski niz	Realizacija niza	Primer	Št. tvorjenk	Št. potrjenih tvorjenk v mFidi
-ati	-ati + -a + -ica + -ast	-ičast	bunčičast	1	0
	-ati + -ika + -ariti + -enje	-ikarjenje	bucikarjenje	1	0

PREGLEDNICA 8: Štiripriponski nizi z glagolskim začetnim priponskim obrazilom

Pri prvostopenjski izmedmetni tvorbi glagolov se na omejenem gradivu medmetov na črko *b* pokaže, da je prevladujoče obrazilo glagolskih tvorjenk *-ati*, saj se *-iti*¹⁸ pojavi le v štirih primerih. Tako razmerje potrjuje tudi SSKJ 2, v katerem nastopa enkrat več glagolov z *-ati* (10015) kot z *-iti* (5483).

Povezava med besednovrstnostjo tvorjenk in pogostostjo priponskih nizov kaže pričakovana razmerja. Samostalniška končna obrazila nastopajo v največ priponskih nizih (53), s čimer predstavljajo skoraj 70-odstotni delež (69,74 %) zastopanosti. Temu primerno je v viru visoko tudi število tvorjenk (73), od tega je korpusno potrjenih 23, kar glede na druge besedne vrste predstavlja najvišji delež (85,19 %). V 14 priponskih nizih nastopajo pridevniška končna obrazila. V primerjavi z zastopanostjo samostalniških obrazil je teh nizov 4-krat manj in predstavljajo 18,42 %. Število tvorjenk v viru (BSSJ) je 17, od tega so korpusno potrjene le 3, kar znaša 11,11 % vseh potrjenih tvorjenk. Glagolska priponska obrazila nastopajo v 9 nizih (11,84 %) z 9 tvorjenkami, od katerih je korpusno potrjena le 1 (3,7 %).

Priponsko obrazilo	Št. različnih priponskih nizov	Delež v %	Št. tvorjenk	Delež v %	Št. potrjenih tvorjenk v mFidi	Delež v %
Samostalniško	53	69,74	73	73,74	23	85,19
Pridevniško	14	18,42	17	17,17	3	11,11
Glagolsko	9	11,84	9	9,09	1	3,70

PREGLEDNICA 9: Povezanost besednovrstne pripadnosti tvorjenk in pogostosti priponskih nizov

¹⁸ Tukaj upoštevamo samo *-ati* in *-iti* ne glede na to, ali je pred njima še kakšna obrazilna sestavina. Na enak način smo z iskalnim nizom *ati in *iti njuno številsko razmerje poiskali tudi v SSKJ 2.

Na prvi tvorbeni stopnji izmedmetnih nizov najpogosteje nastopajo glagolska priponska obrazila. In katera priponska obrazila jih najpogosteje zaključijo? Četrstopenjske tvorjenke zaključujeta glagolniški obrazili *-anje* s 3, *-enje* z 1 in pridevniški obrazili *-ast* in *-oč* prav tako s po 1 tvorjenko. Na tretji stopnji je nabor zaključnih obrazil v nizu večji, skupno jih je 21 različnih. Po številu tvorjenk iz BSSJ si sledijo *-anje* s 13 tvorjenkami, *-n* (9), *-ica* in *-ast* ter *-ava-ti* in *-lj-ati* s po 3 tvorjenkami. Preostalih 15 obrazil je potrjenih le z eno tvorjenko, in sicer samostalniška (*-ec*, *-ø*, *-a*, *-ača*, *-elj*, *-ina*, *-ost*, *-stvo* in *-ula*), pridevniška (*-av*, *-ev*, *-in*) in glagolska (*-ariti*, *-ati*, *-ova-ti*). Drugostopenjske nize zaključuje prav tako 21, vendar tokrat le samostalniških obrazil: *-anje* (12), *-ø* (4), *-a*, *-ač*, *-enje* (3), *-alec*, *-ot* (2); s po eno tvorjenko pa so zastopana obrazila (*-ača*, *-aj*, *-alo*, *-ar*, *-arija*, *-aš*, *-če*, *-e*, *-elj*, *-ež*, *-ica*, *-ika*, *-la*, *-ljaj*).

Stopnja tvorjenosti	Končno priponsko obrazilo	Št. tvorjenk
4	-anje	3
4	-ast	1
4	-enje	1
4	-oč	1
3	-anje	13
3	-n	9
3	-ast	3
3	-ava-ti	3
3	-ica	3
3	-lj-ati	3
3	-ec	2
3	-ø	1
3	-a	1
3	-ača	1
3	-ariti	1
3	-ati	1
3	-av	1
3	-elj	1
3	-ev	1
3	-in	1
3	-ina	1

3	-ost	1
3	-ova-ti	1
3	-stvo	1
3	-ula	1
2	-anje	12
2	-ø	4
2	-a	3
2	-ač	3
2	-enje	3
2	-alec	2
2	-ot	2
2	-ača	1
2	-aj	1
2	-alo	1
2	-ar	1
2	-arija	1
2	-aš	1
2	-če	1
2	-e	1
2	-elj	1
2	-ež	1
2	-ica	1
2	-ika	1
2	-la	1
2	-ljaj	1

PREGLEDNICA 10: Končna obrazila v izmedmetnih priponskih nizih in število tvorjenk

Pogostost s tvorjenkami potrjenih priponskih nizov je obratno sorazmerna s stopnjo tvorbe: nižja kot je stopnja tvorbe, več tvorjenk jo potrjuje. V dvopriponskih nizih (skupno 54) med produktivne vzorce sodijo *-ati + -anje* (7 tvorjenk), *-ati + -a*, *-ati + -ač* in *-kati + -anje* s po 3 tvorjenkami. S po 2 tvorjenkama pa niza *-ati + -ø* in *-ati + -n* (tudi prek glagolske sestavljenke), kar je praktično tudi predvidljivi del oblikoslovne paradigme tako predponskih kot nepredponskih glagolov. Preostalih 44 dvopriponskih nizov je enkratnic, potrjenih z 1 tvorjenko v BSSJ.

Triprisonske izmedmetne tvorjenke imajo 20 različnih nizov. Med njimi opazimo več skupnih obrazilnih nizov na prvi in drugi stopnji, npr. *-ati + -a* (4 priponski nizi); *-ati + -ika* (3 priponski nizi), *-ati + -av* (2 priponska niza), ki se na tretji stopnji obrazilno zelo različno razpršijo. Večinoma so potrjeni le z 1 tvorjenko, izjemi sta *-ati + -a + -ast* in *-ati + -a + -ica* s po dvema.

Štiripriponski izmedmetni niz tvorjenk je redek, v BSSJ potrjen zgolj z dvema kombinacijama priponskih obrazil, ki imajo po eno potrjeno tvorjenko (*bunčičast: -ati + -a + -ica + -ast; bučikarjenje: ati + -ika + -ariti + -enje*).

Št. priponskih obrazil	Priponski niz	Št. tvorjenk
4	<i>-ati + -a + -ica + -ast</i>	1
4	<i>-ati + -ika + -ariti + -enje</i>	1
3	<i>-ati + -a + -ast</i>	2
3	<i>-ati + -a + -ica</i>	1
3	Skl + <i>-ati + -lj-ati + -anje</i>	1
3	<i>-ati + -a + -ača</i>	1
3	<i>-ati + -a + -ula</i>	1
3	<i>-ati + -ač + -ev</i>	1
3	<i>-ati + -ar + -stvo</i>	1
3	<i>-ati + -av + -ec</i>	1
3	<i>-ati + -av + -ost</i>	1
3	<i>-ati + -ika + -ariti</i>	1
3	<i>-ati + -ika + -ast</i>	1
3	<i>-ati + -ika + -in</i>	1
3	<i>-ati + -lj-ati + -anje</i>	1
3	<i>-ati + -lj-ati + -av</i>	1
3	<i>-ati + po-Se + -ava-ti + -anje</i>	1
3	<i>-ati + po-Se + -ava-ti + -oč</i>	1
3	<i>-ati + po-Se + -ova-ti + -anje</i>	1
3	<i>-jati + -alen + -ica</i>	1
3	<i>-kati + -av + -ec</i>	1
3	Z- \emptyset + <i>-ski + -ina</i>	1
2	<i>-ati + -anje</i>	7
2	<i>-ati + za-Se + -anje</i>	4

2	-ati + -a	3
2	-ati + -ač	3
2	-kati + -anje	3
2	-niti + -enje	3
2	-ati + -ø	2
2	-ati + o-Se + -anje	2
2	-ati + o-Se + -n	2
2	-ati + za-Se + -n	2
2	-ati + -a + Z	1
2	-ati + -aj	1
2	-ati + -alec	1
2	-ati + -alo	1
2	-ati + -ar	1
2	-ati + -arija	1
2	-ati + -e	1
2	-ati + -elj	1
2	-ati + -ež	1
2	-ati + -ica	1
2	-ati + -ika	1
2	-ati + iz-Se + -n	1
2	-ati + -k-ati + po-Se	1
2	-ati + -ljaj	1
2	-ati + na-Se + -anje	1
2	-ati + na-Se + -n	1
2	-ati + od-Se + -anje	1
2	-ati + od-Se + -n	1
2	-ati + o-Se + -ava-ti	1
2	-ati + -ot	1
2	-ati + po-Se + -n	1
2	-ati + po-Se + -ova-ti	1
2	-ati + po-Se + -ava-ti	1
2	-ati + pre-Se + -anje	1
2	-ati + pre-Se + -n	1
2	-ati + pri-Se + -anje	1

2	-ati + pri-Se + -ava-ti	1
2	-ati + pri-Se + -lj-ati	1
2	-ati + raz-Se + -anje	1
2	-ati + za-Se + -lj-ati	1
2	-jati + -anje	1
2	-jati + -če	1
2	-kati + -ø	1
2	-kati + -la	1
2	-kati + -ot	1
2	-ketati + -ø	1
2	-ketati + -ača	1
2	-ketati + -alec	1
2	-ketati + -anje	1
2	-ketati + -aš	1
2	Skl + -ati + -anje	1
2	Skl + -ati + -lj-ati	1
2	Skl + -niti + -ø	1
2	Skl + -niti + -elj	1

PREGLEDNICA 11: Pogostejši priponski nizi izmedmetnih tvorjenk

6 SKLEP

Na podlagi statističnih analiz podatkov o izmedmetnih tvorjenkah v *Besedno-družinskem slovarju slovenskega jezika za iztočnice na b* ugotavljamo, da imajo izmedmetne tvorjenke največ različnih dvopriponskih nizov, število tri- in štiripriponskih je opazno manjše. Na prvi stopnji medmeti kot besedotvorno izhodišče najpogosteje generirajo glagolske izpeljanke, kar potrjujejo najpogostejši dvopriponski (I,V,S) in tripriponski (I,V,Ad,S; I,V,S,Ad; I,V,S,S) besedotvorni modeli, prvi omenjeni model ima tudi največ potrjenih tvorjenk v združenem korpusu mFida. Med prvostopenjskimi izglagolskimi izpeljankami najpogosteje nastopajo glagoli z začetnim obrazilom *-ati*, ki tvorijo tudi največ priponskih nizov s korpusno potrjenimi tvorjenkami. Druga začetna glagolska obrazila so nekoliko manj produktivna. Med končnimi obrazili priponskih nizov prevladujejo samostalniška s približno 70-odstotnim deležem, pridevniška (17-odstotni delež) in glagolska (manj kot 12-odstotni delež) so redkejša.

Priponski niz najpogosteje zaključujejo samostalniška obrazila *-nje*, *-ø*, *-ica*, *-a*, *-ač*. Najpogostejši pridevniški končni obrazili sta *-n*, *-ast*, med glagolskimi končnimi obrazili sta pogosti *-ava-ti*, *-lj-ati*. Pričakovana je tudi ugotovitev, da je pogostost korpusno potrjenih tvorjenk iz priponskih nizov obratnosorazmerna s stopnjo tvorbe – večino korpusno potrjenih tvorjenk imajo dvopriponski nizi.

REFERENCE

- Burkacka, Iwona. 2012. *Kombinatoryka sufiksalna w polskiej derywacji odrzeczownikowej*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.
- Erjavec, Tomaž, Pranjič, Marko, Kern, Boris, Pollak, Senja, Pelicon, Andraž, Stramljič Breznik, Irena. 2023. Automating derivational morphology for Slovenian. *Electronic lexicography in the 21st century (eLex 2023): proceedings of the eLex 2023 conference*. Brno: Lexical Computing. 449–465. <https://elex.link/ojs/index.php/elex/article/view/40>
- Fran. *Slovarji inštituta za slovenski jezik Frana Ramovša ZRC SAZU*. <https://fran.si/> (28. 10. 2023).
- Gigafida 2.0*. Referenčni korpus pisne standardne slovenščine. <https://viri.cjvt.si/gigafida/> (28. 10. 2023).
- Jadacka, Hanna (ur.). 2001. *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego*. Tom II: Gniazda odprzymiotnikowe. Kraków: Towarzystwo Wydawnicze »Historia Iagellonica«.
- Kern, Boris. 2017. *Stopenjsko besedotvorje (Na primeru glagolov čutnega zaznavanja)*. Ljubljana: Založba ZRC.
- Kern, Boris. 2020. Obrazilna kombinatorika v besedotvornih sestavih glagolov čutnega zaznavanja. M. Krajnc Ivič, A. Žele (ur.). *Pogled v jezik in iz jezika: Adi Vidovič Muha ob jubileju* (Mednarodna knjižna zbirka Zora, 133). Maribor: Univerzitetna založba. 67–79.
- Kern, Boris. 2023. Feminativi v izsamostalniških besedotvornih nizih. V: J. Jožef-Beg, M. Hočevar, N. Kočnik (ur.). *Naslavljanje raznolikosti v jeziku in književnosti* (Zbornik Slavističnega društva Slovenije, 33). Ljubljana: Zveza društev Slavistično društvo Slovenije. 197–205.
- Krvina, Domen, Žele, Andreja. 2017. O medmetih, zlasti o njihovih razločevalnih lastnosti: poudarjen slovarski vidik. *Slavistična revija* 65/2: 201–228.
- Lopatin, Vladimir V., Uluhanov, Igor S. 2016. *Словарь словообразовательных аффиксов современного русского языка*. Moskva: Издательский центр „Азбуковник“.
- Marvin Derganc, Tatjana. 2019. O glagolskem vidu pri izpeljavi samostalnikov za poimenovanje dejanja in vršilca dejanja v slovenščini. *Philological studies* 17/2: 251–261. <https://doi.org/10.17072/1857-6060-2019-17-2-251-267>
- Metafida 1.0*. *Slovenski združeni korpusi*. <https://www.clarin.si/ske/#dashboard?corpname=mfida10> (16. 1. 2024).
- Ološtiak, Martin, Ivanová, Martina (ur.). 2021. *Slovník slovotvorných prostředků v slovenštině*. Prešov: Prešovská univerzita v Prešove, Filozofická fakulta.
- Pleteršnik, Maks. 2014. *Slovensko-nemški slovar*. www.fran.si (16. 1. 2024).

- Pollak, Senja, Voršič, Ines, Kern, Boris, Ulčar, Matej. 2023. Novel Slovenian COVID-19 vocabulary from the perspective of naming possibilities and word formation. *Electronic lexicography in the 21st century (eLex 2023): proceedings of the eLex 2023 conference*. Brno: Lexical Computing. 419–438. <https://elex.link/ojs/index.php/elex/article/view/39>
- Saloni, Zygmunt idr. 2015. *Słownik gramatycznego języka polskiego*. Warszawa: Uniwersytet Warszawski.
- Skarżyński, Mirosław (ur.). 2004. *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego. Tom III: Gniazda odczasownikowe*. Kraków: Towarzystwo Wydawnicze »Historia Iagellonica«.
- Slovenski pravopis*. 2001. www.fran.si (28. 10. 2023).
- Snoj, Marko. 2015. *Slovenski etimološki slovar*. www.fran.si (16. 1. 2024).
- Snoj, Marko. 2022. Prispevki k slovnici slovenskih medmetov. *Slavistična revija* 70/3: 237–253. <https://doi.org/10.57589/srl.v70i3.4047>
- Snoj, Marko, Žele, Andreja. 2022. Merila za razločevanje med medmeti in členki. *Slavistična revija* 70/4: 557–572. <https://doi.org/10.57589/srl.v70i4.4082>
- Sokolová, Miloslava, Moško, Gustav, Šimon, František. 1999. *Morfematičský slovník slovenčiny*. Prešov: Náuka.
- SSKJ 1: Slovar slovenskega knjižnega jezika*. 1970–1991. www.fran.si (28. 10. 2023).
- SSKJ 2: Slovar slovenskega knjižnega jezika*. 2014. www.fran.si (28. 10. 2023).
- Stramljič Breznik, Irena. 2004. *Besednodružinski slovar slovenskega jezika: poskusni zvezek za iztočnice na B*. Maribor: Slavistično društvo.
- Stramljič Breznik, Irena. 2005. Kvantitativne lastnosti slovenskega tvorjenega besedja v poskusnem besednodružinskem slovarju za črko B. *Slavistična revija* 53/4: 505–520.
- Stramljič Breznik, Irena. 2012. Medmeti kot besedotvorni potencial. V: R. Dragičević (ur.). *Tvorba reči i njeni resursi u slovenskim jezicima: zbornik radova sa četrnaeste međunarodne naučne konferencije Komisije za tvorbu reči pri Međunarodnom komitetu slavista*. Beograd: Filološki fakultet. 341–352.
- Stramljič Breznik, Irena. 2013a. Besedotvorna produktivnost posnemovalnih medmetov za zvoke naprav v slovenščini. *Slavia Meridionalis* 13: 97–107.
- Stramljič Breznik, Irena. 2013b. Motivacijska moč slovenskih medmetov. *Jezikoslovni zapiski* 19/1: 173–185.
- Stramljič Breznik, Irena. 2013c. Medmeti kot sestavine frazemov. V: N. Jakop, M. Jemec Tomazin (ur.). *Frazeološka simfonija: sodobni pogledi na frazeologijo*. Ljubljana: Založba ZRC. 125–133. <https://doi.org/10.3986/9789610504047>
- Stramljič Breznik, Irena. 2014a. *Medmeti v slovenskem jeziku*. Maribor: Založba Pivec.
- Stramljič Breznik, Irena. 2014b. Medmet kot besedna vrsta v slovnicih opisih slovenskega jezika. *Slavistična revija* 62/1: 17–37.
- Stramljič Breznik, Irena. 2024. Značilnosti izmedmetih priponskih nizov v Besednodružinskem slovarju slovenskega jezika za iztočnice na b. *Slavia Centralis* 17/1: 1–17.

- Šircelj-Žnidaršič, Ivanka (ur.). 1998. *Besedišče slovenskega jezika z oblikoslovnimi podatki*. Ljubljana: Znanstvenoraziskovalni center SAZU, Založba ZRC. <http://bos.zrc-sazu.si/besedisce.html> (12. 1. 2024).
- Šiška, Zbyněk. 2005. *Bazový morfematičský slovník češtiny*. Olomouc: Univerzita Palackého.
- Toporišič, Jože. ⁴2000. *Slovenska slovnica*. Maribor: Obzorja.
- Vidovič Muha, Ada. 1988. *Slovensko skladiščno besedotvorje ob primerih zložen*. Ljubljana: Znanstveni inštitut Filozofske fakultete.
- Vidovič Muha, Ada. 2000. *Slovensko leksikalno pomenoslovje: govorica slovarja*. (Razprave Filozofske fakultete). Ljubljana: Znanstveni inštitut Filozofske fakultete.
- Vidovič Muha, Ada. ²2013. *Slovensko leksikalno pomenoslovje. Govorica slovarja*. Ljubljana: Znanstvena založba Filozofske fakultete, Partizanska knjiga.
- Vidovič Muha, Ada. ³2021. *Slovensko leksikalno pomenoslovje*. Ljubljana: Znanstvena založba Filozofske fakultete. <http://www.dlib.si/?URN=URN:NBN:SI:DOC-QSVADS5P> (28. 10. 2023).
- Vogelgesang, Teresa, Jadacka, Hanna (ur.). 2001. *Słownik gniazd słowotwórczych współczesnego języka ogólnopolskiego. Tom I: Gniazda odrzeczownikowe*. Kraków: Towarzystwo Wydawnicze »Historia Iagellonica«.

Prispelo januarja 2024, sprejeto aprila 2024.

Received January 2024, accepted April 2024.

ZAHVALE

Prispevek je nastal v okviru raziskovalnega projekta št. J6-3131 (Kombinatorika besedotvornih obrazil v slovenščini, vodja programa dr. Boris Kern), raziskovalnega programa št. P6-0156 (Slovensko jezikoslovje, književnost in poučevanje slovenščine, vodja prof. dr. Marko Jesenšek) in raziskovalnega programa št. P6-0038 (Slovenski jezik v sinhronem in diahronem razvoju, vodja dr. Kozma Ahačič), ki jih financira ARIS.

POVZETEK

KVANTITATIVNI PODATKI O BESEDOTVORNIH MODELIH IN PRIPONSKIH NIZIH IZMEDMETNIH TVORJENK V *BESEDNODRUŽINSKEM SLOVARJU SLOVENSKEGA JEZIKA ZA IZTOČNICE NA B*

Besednodružinski in stopenjski pristop sta v besedotvorju odprla pot morfotaktičnim raziskavam slovenskega jezika. Predstavitev kvantitativnih značilnosti izmedmetnih priponskih nizov v poskusnem besednodružinskem slovarju predstavlja eno izmed njih in prinaša naslednje ugotovitve. Izmedmetne tvorjenke imajo največ različnih dvopriponskih nizov, njihovo število pa opazno upada pri tri- in štiripriponskih, ki so

zelo redki. Medmeti kot besedotvorno izhodišče na prvi stopnji najpogosteje generirajo glagolske izpeljanke, in to potrjujejo najpogostejši dvopriponski (I,V,S), tropriponski (I,V,Ad,S; M,V,S,Ad; I,V,S,S) nizi, med štiripriponskimi nizi pa po enkrat nastopata le modela I,V,S,S,Ad in I,V,S,V,S. Dvopriponski niz (I,V,S) ima tudi največ potrjenih tvorjenk v korpusu mFida. Med prvostopenjskimi izglagolskimi izpeljankami v dvo-, tri- in štiripriponskih nizih najpogosteje nastopajo glagoli z obrazilom *-ati*, ki tvorijo največ priponskih nizov s korpusno potrjenimi tvorjenkami. Druga glagolska obrazila (*-kati*, *-ketati* in *-niti*) so tovrstno manj produktivna. Med končnimi obrazili priponskih nizov so pričakovano prevladujoča samostalniška obrazila, ki dosegajo 70-odstotni delež, medtem ko so pridevniška (17-odstotni delež) in glagolska (manj kot 12-odstotni delež) redka. Med samostalniškimi obrazili niz najpogosteje zaključijo obrazila: *-nje*, *-ø*, *-ica*, *-a*, *-ač*, med pridevniškimi *-n*, *-ast* in med glagolskimi *-ava-ti*, *-lj-ati*. Ugotoviti je mogoče tudi, da je pogostost korpusno potrjenih tvorjenk iz priponskih nizov v viru obratnosorazmerna s stopnjo tvorbe – nižja kot je stopnja, več tvorjenk je potrjenih.

QUANTITATIVE DATA ON THE WORD FORMATION MODELS AND THE WORD FORMATION SEQUENCES OF THE INTERJECTION IN THE *WORD FAMILY DICTIONARY OF THE SLOVENE LANGUAGE FOR ENTRIES BEGINNING ON B*

The word family and the stage approaches to word formation have paved the way for morphotactic research in Slovene. The present quantitative analysis of the characteristics of suffix chains formed by interjections in the word family dictionary is one such example. The key findings are that the majority of compounds formed by interjections are two-suffix chains, while the number of different word formation sequences decreases significantly for three- and four-suffix chains, which are very rare. Interjections as the starting point for word formation most often generate verbal derivations in the first stage, which is confirmed by the most frequent two-suffix (I,V,S) and three-suffix word formation sequences (I,V,Ad,S; I,V,S,Ad; I,V,S,Ad; I,V,S,S), while among the four-suffix word formation sequences the models I,V,S,S,Ad and I,V,S,V,S occur once each. The word formation sequences with two suffixes (I,V,S) also have the most confirmed compounds in the mFida corpus. Among the first-level verbal derivatives, verbs with the suffix *-ati* are the most frequent and form the most suffix chains with compounds attested in the corpus. Other verbal suffixes (*-kati*, *-ketati* and *-niti*) are less productive. As expected, noun suffixes predominate among the final suffixes of the suffix chains with a share of 70%, while adjectival (17%) and verbal suffixes (less than 12%) are less common. Among the noun suffixes, the suffixes *-nje*, *-ø*, *-ica*, *-a*, *-ač* are the most productive, while the adjectival suffixes *-n*, *-ast* and the

verbal suffixes *-ava-ti*, *-lj-ati* are the most frequent. It can also be observed that the frequency of confirmed compounds in the corpus is inversely proportional to the word formation level - the lower the level, the more compounds are confirmed.

Vrinda Subhalaxmi Chidambaram (ORCID: 0000-0002-2513-5417)

University of California, Riverside, USA

vrinda@ucr.edu

DOI: <https://doi.org/10.3986/16.1.04>

ON THE STRUCTURAL UNIFORMITY OF SUPERLATIVES: EVIDENCE FROM GEORGIAN, SLOVENE, AND TAMIL

This paper is an exploration of the structure of superlative degree adjectives and the syntactic constructions in which they occur across three very distinct languages. Slovenian, Georgian, and Tamil each represents a typological variant of the superlative construction. Slovenian has synthetic superlatives, built from the *naj-* prefix, while Georgian and Tamil have compositional superlatives, in which the meaning of the superlative is derived compositionally. In this paper, I show that these apparent differences between these constructions is superficial and that the underlying syntax is largely the same.

KEYWORDS: superlative, degree adjective, syntax, morphology, semantics, Tamil, Georgian

V prispevku obravnavam strukturo presežnika pridevnikov in skladenjskih zgradb, v katerih se presežniške zgradbe pojavljajo v treh različnih jezikih. V slovenščini, gruzinščini in tamilščini se pojavljajo tipološke različice presežniških zgradb. Slovenščina ima sintetične presežnike, zgrajene iz predpone *naj-*, v gruzinščini in tamilščini pa so v rabi sestavljeni kompozicionalni presežniki, v katerih je pomen presežnika zastavljen kompozicionalno. V prispevku dokazujem, da so razlike med zgradbami v obravnavanih jezikih samo površinske in da je globinska skladnja večinoma enaka.

Ključne besede: presežnik, presežna stopnja, superlativ, stopnjevalni pridevnik, skladnja, oblikoslovje, pomenoslovje, tamilščina, gruzinščina

1 INTRODUCTION

The morphological composition of superlative degree adjectives shows surprisingly limited cross-linguistic variation; indeed, it has been observed that despite differences in the surface realization of superlative adjective formation, there are cross-typological uniformities that lend themselves to universal generalizations, perhaps most strikingly that superlatives are morphologically derived from comparative adjectives (Bobaljik 2012). But the question that then inevitably arises

is how to reconcile a universal morphological principle, like that proposed by Bobaljik (2012) with the (limited but nevertheless significant) surface morpho-syntactic variation we find in the expression of comparatives and superlatives.

Superlative degree adjectives can be classified as either analytic/periphrastic (e.g., most sweet) or synthetic (e.g., sweetest). This does not appear to be a parametric difference, as languages do not necessarily conform to using just one or the other; rather many languages (e.g., English, Slovene) use some combination of the two. Although it is widely acknowledged that degree adjectives alternate between synthetic and analytic morphological realizations, little attention has been given to the *variability* across the construction of periphrastic superlatives. Overall, the syntax of superlatives has largely been ignored in the literature.

The syntactic properties of adjectives, including the syntax of comparative degree adjectives, have been the focus of a great deal of research (notably, for example, Cinque 2010, Aljović 2010, Bresnan 1973, Kennedy 2002). In addition, superlative degree adjectives have been explored extensively in terms of their semantics (e.g., Heim 1999, Gutiérrez-Rexach 2010, Sleeman 2010). And though there has been some recent interest in the syntax of superlative degree constructions (e.g., Loccioni 2020), the focus is centered on an analysis of those constructions that include an overt definiteness marker, as in Romance, to the exclusion of other varieties.

In this paper, I examine the relationship between the morphology of superlative degree adjectives, their semantic composition, and the syntactic environment in which they occur in order to propose a cross-typologically uniform syntax of superlatives.

1.1 BACKGROUND

Bobaljik (2012) illustrates that every superlative degree adjective properly contains the comparative degree adjective (*The Containment Hypothesis*). He offers a number of universally held generalizations to support this claim; perhaps most notably, he shows that in no language is there a comparative degree adjective with a suppletive stem whose superlative counterpart does not also bear a suppletive stem; i.e., the superlative cannot revert to the positive adjectival root once suppletion has occurred for the comparative.

In many languages, such as Slovene (predominantly synthetic) and Tamil (analytic), the compositional nature of the superlative is transparent. Bobaljik offers a compelling argument that even in languages where this embedding is less obvious (e.g., English), it nevertheless obtains. Assuming this is indeed

the case, the next step is to investigate the nature of the component parts of the comparative and superlative: does the form relate to the function, and if so, how? As an initial step toward answering this question, let's first examine the semantics of comparative and superlative degree adjectives.¹

1.2 THE DERIVED SEMANTICS OF SUPERLATIVES

The semantic value of the superlative degree adjective is built upon that of the comparative adjective. It is crucial to note that I will be looking strictly at the most prevalent use of the comparative, in which two distinct entities are undergoing comparison with respect to some feature. There are, of course, other ways in which comparatives are used (i.e. "The building is taller than it is wide." or the marginally acceptable distributed reading of "Tony and Nancy are taller than Andy and Kim", in which Andy may be taller than Nancy), but such uses go beyond the scope of this study.

In its typical usage, a comparative degree adjective denotes a function that selects one member of binary set, namely the one with the higher value with respect to some specified gradable feature F (denoted by the positive adjective). The member of the set selected by the degree adjective (A) has F greater than the other member (B), so ($F_A > F_B, \{A, B\}$). For a concrete example, let's consider the following sentence.

(1) Gaia is taller than Angelika.

The comparative applies to two entities, in this case Gaia and Angelika. Both Gaia and Angelika bear a height feature (F_{height}). The comparative selects the one entity whose F_{height} exceeds that of the other. In essence, the generic sentence model for a comparative in (2) can be paraphrased as (3).

(2) X is Z-er than Y

(3) "X exceeds Y in Z, where Z is a gradable property."² (Chidambaram 2008)

¹ Note that I will be looking strictly at the most prevalent use of the comparative, in which two distinct entities are undergoing comparison with respect to some feature. There are other ways in which comparatives are used (i.e. "The building is taller than it is wide." or the marginally acceptable distributed reading of "Tony and Nancy are taller than Andy and Kim", in which Andy may be taller than Nancy), but such uses go beyond the scope of this study.

² This follows directly from Christopher Kennedy's (1997) analysis of comparative degree adjective semantics.

The fundamental difference between a comparative adjective and a superlative adjective is the inclusion in the latter of either an explicit or implicit domain restriction as well as a universal quantifier taking scope over that domain. For example, in (4) we have an explicit domain restriction, which is the bracketed PP. And the meaning of the sentence is paraphrased in (5).

(4) Gaia is the silliest kid [_{PP} in the class].

(5) For every kid in the same class as Gaia but who is not Gaia, Gaia exceeds them in silliness.

Just as the definition of the comparative can be generalized, the definition of the superlative can also be generalized, as in (6).

(6) “For **every** Y, where Y represents any individual term **within the specified domain** that also includes X, and Y is not X, X exceeds Y in Z, where Z is a gradable property.” (Chidambaram 2008)

The relationship between (3) and (6) is quite plain to see. (6) in fact properly contains (3), but furthermore specifies the existence of a domain and includes a quantifier. In some languages, this relationship is grammatically transparent, as in Tamil. I will call this a *compositional superlative*.

2 COMPOSITIONAL SUPERLATIVES

As mentioned earlier, periphrasis of superlatives is not cross-linguistically uniform. While the majority of work on the topic has focused on the “most” variety of periphrastic superlatives, very little attention has been dedicated to examining other syntactically complex expressions of superlative degree. The compositional superlative is one such overlooked construction.

Compositional superlatives are transparently derived from comparatives, and thus adhere quite visibly to Bobaljik’s *Containment Hypothesis*. In addition to the inclusion of the comparative, compositional superlatives also include a quantifier and a domain restriction (which could be either explicit or implicit (Stanley, Szabó 2000)). Consider examples (7) and (8) from Tamil:

(9)

La ragazza più forte e Gaia.
 The girl more strong is Gaia.
 ‘The strongest girl is Gaia.’

In languages like Tamil, on the other hand, the focus seems to be the extent of the domain. We can call these *domain-extent compositional superlatives*. Domain-extent compositional superlatives are characterized by the presence of the universal quantifier; shown in Tamil example (10), *ella*: ‘all’:

(10)

Ra:mu vakuppile ella:ma:navarkalai viḍayum oyarama:navan
 Ramu_{NOM.SG.MASC.} class_{LOC.} all students_{ACC.PL.MASC.} than tall_{MASC.SG.PRED.ADJ.}
 ‘Ramu is the tallest student in the class.’

Georgian, too, forms superlatives in a similar manner, as illustrated in (11),³ in which the universal quantifier is *qvela* ‘all’.

(11)

Gaia qvelaze upro p’at’ara st’udent’ia.
 Gaia all_{LOC.} more small student-is.
 ‘Gaia is the smallest student.’

In this paper, I will restrict my investigation to the domain-extent compositional superlatives, as found in Tamil and Georgian (i.e., those who instantiate the superlative explicitly using a universal quantifier). That said, I do not rule out the possibility that an extension of the analysis I propose applies to Romance and other languages, as well.

³ It should be noted that Georgian additionally has a synthetic form of the superlative, which is formed with the circumfix *u- -es*: (1) p’at’ari– upro p’at’ar-i– u-p’at’ar-es-i →small– smaller– smallest. But from what I have understood from my Georgian informants, these are actually absolute superlatives, which contrast significantly from relative superlatives in terms of their semantics. Absolute superlatives have the meaning of a strong intensifier (e.g., Italian *bellissima*), whereas relative superlatives really choose the highest ranked member of an ordered set. It could be significant that one finds synthetic absolute superlatives in languages that have not only analytic but *compositional* relative superlatives (of the two different types). But this idea requires further research that goes beyond the scope of this paper.

3 THE SIGNIFICANCE OF THE UNIVERSAL QUANTIFIER

In many languages, we find the pattern that we find in Georgian and Tamil; they combine a universal quantifier with the comparative adjective to form the superlative. This inclusion of the universal quantifier has clear consequences on the semantics of the entire phrase. For one thing, it produces a downward entailing environment. Downward entailment is the property of a statement X , e.g. “I love all cheese”, such that X entails every statement that includes a subset of X , e.g. “I love all *French cheese*”. One interesting property of downward entailing environments is that they allow the inclusion of Negative Polarity Items (NPIs) (Ladusaw 1980), e.g. “I love all cheese that I *ever*_{NPI} tasted”. We see that NPIs are possible within superlative clauses in languages like Georgian, which overtly include a universal quantifier: both *k’i* and *odesme* are NPIs in Georgian (Chutkerashvili 2009) and are permitted in superlative clauses. Consider the Georgian example in (12):

- (12)
- | | | | | |
|---------------------------------|----------------------|-------------|---------------|--------------|
| Zaza | qvelaze | upro | k’et’ili | adamiani-a |
| Zaza | all _{LOC} | more | nice | person-is |
| (mat šoris) | visac | k’i | odesme | ševxvedrivar |
| (them _{DAT.PL.} among) | whom _{ACC.} | ever | ever | I-met |
- ‘Zaza is the nicest person whom I ever met.’

In general, NPIs are not licensed in relative clauses following positive adjectives nor in *than*-clauses following comparatives, as shown in (13) and (14).

(13) * Gaia is a sweet person that I ever knew.

(14) */? Gaia is a sweeter person than I ever knew.⁴

This suggests that there is something special about the superlative cross-linguistically that provides a downward entailing environment. In Georgian, this is unmistakably the work of the universal quantifier. The universal quantifier creates a downward entailing environment, licensing an NPI, otherwise

⁴ This sentence could be good if it is interpreted as containing some instantiation of the universal quantifier within the *than*-clause. If we were to interpret the sentence as truly comparative, as in “Gaia is a sweeter person than the one I ever knew”, it would be infelicitous if not outright ungrammatical.

unlicensed by an adjective (positive or comparative). In English and other languages, however, the source of the licensing is less evident, but nevertheless, we see that NPIs are indeed licensed in clauses within the scope of a superlative degree adjective, as demonstrated in (15), which supports the notion of a hidden universal quantifier in the superlative:

(15) Gaia is the sweetest person that I've ever known.

4 THE SLOVENE *KAR*-PHRASE

Slovene, unlike Tamil and Georgian, forms relative superlatives synthetically with the prefix *naj-*. The degree adjectival series is exemplified in (16):

(16)

lep	– lep-š-i	– naj-lep-š-i
beautiful	– more beautiful	– most beautiful

In this series, we see the transparent nesting of the comparative (and more deeply, positive) degree adjective within the superlative: a perfect illustration of Bobaljik's (2012) *Containment Hypothesis*. This manner of forming superlatives is far from unique; indeed, it is the most common way of forming superlatives in Slavic languages (Russian, Macedonian, and Bulgarian being exceptional).

It is natural at this point to consider the nature of the prefix *naj-*: what does it contribute semantically and where does it fit syntactically? The historical origins of this morpheme have been disputed, but recently Wandl (2022) provided a strong argument for the development of *naj-*, suggesting that it arose through composition of the excessive/intensifying adverbial prefix *na-* (occurring on verbs like *napiti se* 'to get drunk' formed from *piti* 'to drink') and *-j* which developed from an enclitic pronoun and appeared on adjectives in early Slavic to denote definiteness. For example in Old Church Slavonic, the morphemes *nova* 'new' + *ja* 'it_{NOM.SG.FEM.PRON.}' results in the word *novaja* 'the new (one_{NOM.SG.FEM.})' (Wandl 2022). This diachronic analysis of the superlative morpheme *naj-*, which includes a definiteness marker, closely resembles that of a type of superlative we have discussed previously: the individuating compositional superlative, which we see in some Romance languages (see example (9) for Italian).

There is evidence, however, that the synchronic analysis of *naj-*, i.e., the manner in which contemporary speakers interpret the morpheme, may have diverged somewhat from its historical origins. Specifically, there are some facts from contemporary Slovene that shed some light on the specialized function and properties of *naj-*. Before we examine those data, let's first consider some background on positive and comparative degree adjectives and the types of CP that follow them.

4.1 A BRIEF BUT NECESSARY DIGRESSION ON RELATIVE CLAUSES AND *THAN*-CLAUSES

Any NP can be modified by a relative clause, whether or not there is an adjectival modifier present.

(17) Jozef is the philosopher [_{RC} that I'm married to].

(18) Jozef is the [_{NP} [_{AP} **brilliant**] philosopher] [_{RC} that I'm married to].

The relative clauses in (17) and (18) are identical; the inclusion of an additional modifier in (18) has no consequence on either the interpretation or structure of the relative clause. But this is perhaps not the case when the modifying adjective is a comparative.

(19) #Jozef is the [_{NP} [_{compAP} more brilliant] philosopher] [_{RC} that I'm married to].

The meaning of sentence (19) can be paraphrased as follows: I am married to two different philosophers, of whom Jozef is the more brilliant. So the question is whether the phrase is truly interpreted as just a plain old relative clause. I will leave this as an open question for the moment and return to it in §4.3.

Typically, comparatives are followed not by relative clauses but by *than*-clauses, which occur as a function of the comparative itself. In Slovene, this can be instantiated phrasally (as a PP) or clausally (headed by the C⁰ *kot*) (Pancheva 2006). In both, the *than*-clause serves to denote the *comparand*, i.e. the lower-ranking member of the ordered set.

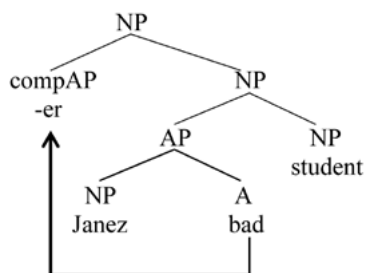
(20)

(a) Janez is a worse student [*than* Maja].

(b) Janez je slabši študent [_{PP} od Maje] / [_{CP} *kot* (je) Maja (slaba študentka)].

Janez is worse student [_{PP} from Maja] / [_{CP} *than* (is) Maja (bad student)].

The syntax of (20b) is complicated; there are several requirements that need to be met. One crucial consideration is that it be derived in a way that reflects the requisite morphological nesting of the positive adjective within the comparative. In Bobaljik (2012), this boils down to a direct selection relation between compA^0 and the AP, but I suggest that it must be more complex than that, simply because there is more to the compAP . On one hand, the structure should convey the functional similarity between a comparative degree adjective phrase and a positive degree adjective phrase: both are adjunct NP-modifiers. This can be achieved as in the following derivation, in which A^0 raises to adjoin to compA^0 to form the comparative adjective, and both AP and compAP are NP-adjuncts:



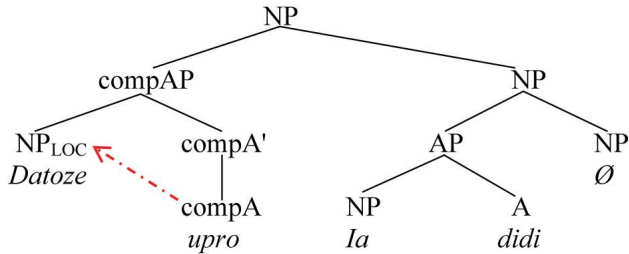
On the other hand, the structure must also reflect idiosyncratic properties of the comparative; notably, the close relationship between the comparative morpheme and the *than*-clause. Georgian provides an interesting clue as to how this condition may be satisfied.

(21)

Ia	Datoze	(upro) didi-a.
Ia _{NOM.SG.FEM.}	Dato _{LOC.SG.MASC.}	(more) big-is.

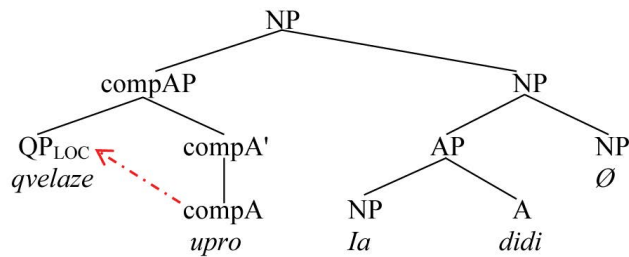
‘Ia is bigger than Dato.’

Here we see that the comparand in a phrasal comparative occurs in an oblique Case. In Georgian, there is no indication of a preposition assigning/licensing/checking Case, rather it seems that the comparative itself is responsible for the Case, thus suggesting the possibility of the following structure for the Comparative Adjective Phrase in (21):



The comparand is merged as the Specifier of compA, where its Case is checked by the compA head.

It is critical to note that both in Georgian and in Tamil, it is a quick (and superficially simple) step from the comparative to the superlative: it is achieved by replacing the DP-comparand with a universal quantifier. So eventually, one would hope the superlative would be derived as shown here, where *qvela*, the universal quantifier meaning ‘all’, occurs in the canonical position of the comparand:

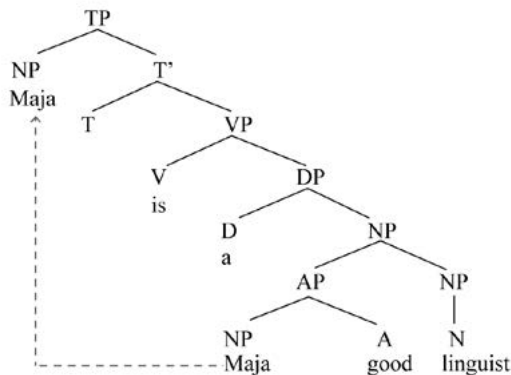


This structure represents only the final, surface/PF form; in SECTION 6, we address the operations that are necessary to derive it.

4.2 THE STRUCTURE OF SYNTHETIC COMPARATIVES

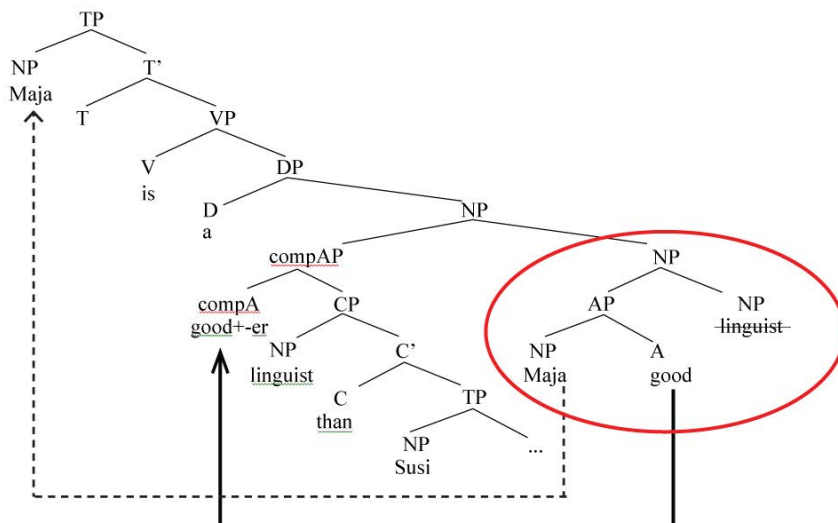
As a preliminary step toward understanding the syntax of superlatives, we must explore the structures of positive and comparative degree adjectival constructions.

A simplified derivation of the sentence *Maja is a good linguist* is shown below.



Note that the matrix subject originates as Specifier of the adjective, in accordance with a small clause analysis of predicate adjectives (Stowell 1980). It then raises to matrix subject, perhaps via intermediary cyclic steps (e.g. Spec-vP, in an expanded derivation).⁵

The syntax of a comparative includes the positive degree adjective but additionally includes movement of the A^0 to merge with the comparative morpheme *-er* and the *than*-CP comparand clause. Consider the following derivation of *Maja is a better linguist than Susi*.



⁵ In the absence of an adjective, the matrix subject would first merge into Spec-vP.

Note that the circled part of the derivation is identical to what we find in the derivation of *Maja is a good linguist*. Just as in the previous derivation, the Spec-AP *Maja* raises to matrix subject. In addition, the positive adjective A^0 raises to merge with the comparative morpheme. The morphological merger of the adjectival head *good* and *-er* triggers stem suppletion, giving rise to the form *better*. The final step is the deletion of the NP *linguist* on identity with the embedded Spec-CP. One could, alternatively, analyze this as a movement operation; that is, the NP *linguist* might move from its original position to the Specifier of the than-clause CP. This, however, would constitute a sideward movement, whose status in generative syntax is contentious.

4.3 BACK TO SUPERLATIVES IN SLOVENE

Now that we have considered the structure of a comparative, we are much better situated to understand the superlative. Let us recall the data on relative clauses adjoined to AP examples (17), (18), and (19), repeated here as (22), (23) and (24).

(22) Jozef is the philosopher [_{RC} that I'm married to].

(23) Jozef is the [_{NP} [_{AP} brilliant] philosopher] [_{RC} that I'm married to].

(24) #Jozef is the [_{NP} [_{compAP} more brilliant] philosopher] [_{RC} that I'm married to].

(25) #Jozef is the [_{NP} [_{supAP} most brilliant] philosopher] [_{RC} that I'm married to].

From the sentence in (25), it appears that I am married to ≥ 3 different philosophers. So, again, we must wonder whether the bracketed *RC* in this sentence is really a typical N^0 -modifying relative clause. If it were, we wouldn't expect it to render the sentence infelicitous. Data from Slovene suggest that this bracketed *RC* phrase in (25), which bears superficial similarity to a relative clause, perhaps is something quite different.

4.3.1 RELATIVE CLAUSES AND RESUMPTION IN SLOVENE

There are various ways to form a relative clause in Slovene, but one of the most frequently attested (especially when the relative head within the relative clause, or *internal relative head*, is assigned either Accusative or Dative Case) is a CP that is headed by the invariant complementizer *ki* and that contains a resumptive pronoun. The resumptive pronoun receives its case within the *RC*.

(26)

Maja je ta jezikoslovka, [_{RC} **ki** sem **jo** spoznala včeraj].
 Maja is the linguist_{NOM.SG.FEM.}, [_{RC} **that** AUX **her**_{ACC.SG.FEM.} met yesterday].
 ‘Maja is the linguist that I met yesterday.’

What we see in (26) closely resembles what we find in the clauses that follow superlatives; however, there are two significant points where the two clausal constructions diverge. First, and perhaps most noticeably, the complementizer itself is different: *kar*, rather than *ki*. And second, the resumptive pronoun does not necessarily match the relative head in number.

(27)

Oni so najpametnejši jezikoslovci, [_{CP} **kar** sem **jih** spoznala včeraj].
 They are the-smartest linguists_{NOM.PL.} that AUX **them**_{ACC.PL.} met yesterday.
 ‘They are the smartest linguists that I met yesterday.’

(28)

Maja je najpametnejša jezikoslovka, [_{CP} **kar** sem **jih** spoznala včeraj].
 Maja is the-smartest linguist_{NOM.SG.FEM.} that AUX **them**_{ACC.PL.} met yesterday.
 ‘Maja is the smartest linguist that I met yesterday.’

When the modified N⁰ is plural, as in (27) *jezikoslovci* ‘linguists’, the resumptive pronoun embedded in the *kar*-phrase will match it for number. But this apparent feature “matching” turns out to be an artefact of the invariant⁶ plurality of the *kar*-phrase resumptive, which is evident from (28), in which the modified N⁰ (*jezikoslovka* ‘linguist’), is singular, and the resumptive pronoun is nonetheless plural and obligatorily so.

The *kar*-phrase is certainly some kind of CP, but it differs from a typical *ki*-relative clause in 3 related ways: (1) it is not adjunct-like, (2) much like the *than*-clause, it seems to have a close relationship to a degree adjective, and (3) it serves to denote a quantifier domain restriction.

Let’s first consider the source of the plural number feature on the embedded clause internal pronoun. Within analyses of anaphora, resumptive pronouns

⁶ Nearly invariant. We will return to the point later in the section.

are always presented as something of an outlier. Whereas pronouns generally do not require any sort of sentence-internal antecedent, resumptive pronouns obligatorily have a local antecedent. And of course, it is a definitional requirement that the pronoun match its antecedent in all ϕ -features. The pronoun *jih* in (28) doesn't appear to have anything that could potentially serve as an antecedent. And indeed, if you ask any Slovenian what the *jih* stands for in (28), they will invariably respond *jezikoslovci* ('linguists'), which doesn't correspond to any overt lexical item in PF. But of course, that doesn't mean it isn't there in the syntax at all.

Interestingly, we can see parallels to this phenomenon cross-linguistically and cross-typologically. If we look back at Georgian, which is a Kartvelian language and has compositional superlatives (as opposed to Slovene synthetic superlatives), we see a striking similarity to the Slovene *kar*-phrase. Consider example (12), repeated here as (29):

- (29)
- | | | | | |
|---|----------------------|------|----------------|------------|
| Zaza | qvelaze | upro | k'e't'ili | adamiani-a |
| Zaza | all _{LOC} | more | nice | person-is |
| [_{CP} (mat šoris) | visac | k'i | ševxvedrivar]. | |
| [_{CP} (them _{DAT.PL.} among) | whom _{ACC.} | ever | I-met]. | |
- 'Zaza is the nicest person whom I ever met.'

The Postpositional Phrase [_{PP} *mat šoris*] contains a plural pronoun with no overt antecedent, although it is interpreted as "people".

This supports the notion that the CP following a superlative, regardless of the superlative's morphological structure, is in fact *not* a CP-adjunct of an N⁰ (i.e., a relative clause) but rather a specialized quantifier-domain-restricting clause that relates directly to the superlative adjective. The particular quantifier whose domain is being restricted is evident in Georgian: *qvela* 'all'. And there is evidence that it is the same in Slovene. As it turns out, the superlative degree adjective is not unique in its ability to license a *kar*-phrase.

- (30)
- | | | | |
|--|-----|------|-----------|
| Vsi jezikoslovci, [_{CP} kar sem jih včeraj spoznala], | so | bili | prijazni. |
| All linguists that AUX them yesterday met , | AUX | were | kind. |
- 'All the linguists I met yesterday were kind.'

I suggest that the plural feature of the resumptive pronoun is a direct consequence of the universal quantifier, not only in (30) but also in (28), where that universal quantifier is not overt but rather incorporated into the superlative degree adjective. This may seem like an unjustified theoretical leap; however, there is clear empirical evidence to support it. Let's consider the effect of quantification on mass nouns.

(31)

Vsa (različna) vina so bila izvrstna.
 All_{NOM.PL.NEUT.} (different) wines_{NOM.PL.NEUT.} AUX were excellent.
 'All the (different) wines [i.e. wine varieties] were excellent.'

(32)

Vse (*različno) vino je bilo pokvarjeno.
 All_{NOM.SG.NEUT.} (*different) wine_{NOM.SG.NEUT.} AUX was spoiled.
 'All the (*different) wine was spoiled.'

The contrast we find between (31) and (32) is restricted to mass nouns, as count nouns are obligatorily plural when they occur within the scope of universal quantification (this is also why we find obligatory plurality of the *kar*-phrase pronoun, i.e. due to its position within the scope of a universal quantifier). As expected, the quantified plural mass noun in (31) receives the *type*-reading (thereby allowing differentiation), whereas the quantified singular mass noun in (32) simply receives the collective reading (thereby precluding differentiation). Interestingly, we find a parallel to this in the *kar*-phrase.

The resumptive pronoun in a *kar*-phrase is obligatorily plural when the superlative modified N⁰ is a count noun, as in (33).

(33)

To je najhitrejši računalnik [kar jih /*ga lahko kupiš].
 That is fastest computer [that them/*it can you-buy].
 'That is the fastest computer that you can buy (them/*it).'

But an alternative emerges when the modified N⁰ is a mass noun, as in (34).

(34)

To je najkislejše pivo [kar sem jih / ga kdaj pila].
 That is sourest beer [that AUX them / it ever I-drunk].

- With *jih*: 'That is the sourest variety of beer I have ever drunk' (That particular beer may not be the sourest, but that variety (e.g., Sour Ale) is the sourest I ever drank.)

- With *ga*: 'That is the sourest beer that I have ever drunk' (I have never had any beer that is sourer.)

I submit this as further evidence that the superlative in Slovene contains universal quantifier, which is additionally supported by the licensing of an NPI in *kar*-phrases following superlatives (refer to SECTION 3 above).

(35)

Maja je najboljša jezikoslovka, kar sem jih kdaj spoznala.
 Maja is best linguist that AUX them ever met.

'Maja is the best linguist that I ever met.'

5 THE SYNTAX OF SUPERLATIVES

Thus far, I have argued that the following three and a half facts need to be represented in some way in the syntax of superlative degree adjectives, irrespective of the morphological composition (i.e., synthetic vs. compositional) of the superlative:

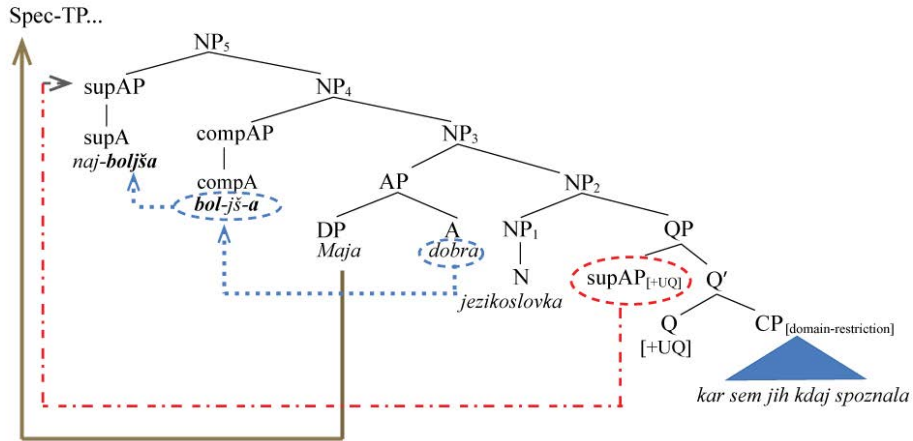
- (1) The superlative must contain the comparative (Bobaljik's (2012) Containment Hypothesis),
- (2) superlatives co-occur with domain restricting clauses whose semantics (and in some cases, whose surface syntactic structures) are unmistakably distinct from those of an ordinary NP-modifying relative clause, and
- (3) superlatives bear properties of universal quantifiers; in domain-extent compositional superlatives, this is morphologically transparent.
 - (a) In languages with transparent inclusion of the universal quantifier in the superlative, the UQ appears as the comparand.

Again, these requirements taken together are somewhat difficult to satisfy structurally, however, they can indeed be integrated, as illustrated in §5.1 and 5.2.

5.1 SLOVENE SUPERLATIVES

(36)

Maja je najboljša jezikoslovka, kar sem jih kdaj spoznala.
 Maja is best linguist that AUX them ever met.
 ‘Maja is the best linguist that I ever met.’



We begin with the Quantifier Phrase (QP), of which the Superlative Adjective Phrase (SupAP) is the Specifier, selected by the null universal quantifier in Q⁰. The SupAP is headed by the affix *naj-*, whose universal quantification [+UQ] feature percolates up to the phrasal node. In this Spec-QP position, the SupAP can check its ([+UQ]) feature against the null Q⁰.

The Q⁰ additionally takes a complement, namely the *kar*-phrase CP, which is the domain restriction of the universal quantifier. The entire QP is an adjunct of the NP₁, headed by N⁰ *jezikoslovka* ('linguist'). NP₂ is modified by an adjunct AP, which is a small clause, containing a subject (which eventually raises to matrix Subject position, i.e., Spec-TP) and the head adjective. NP₃ is modified by the Comparative Adjective Phrase (compAP), which is headed by the affix *-jš-*. This brings us to NP₄. The SupAP must raise to a position where it c-commands the comparative, so it adjoins to NP₄. This allows the compA⁰ *boljša*, which itself is derived from the raising of the A⁰ *dobra*, to raise and adjoin to the affix supA⁰ *naj-*, thereby deriving the synthetic morphological superlative, *najboljša*.

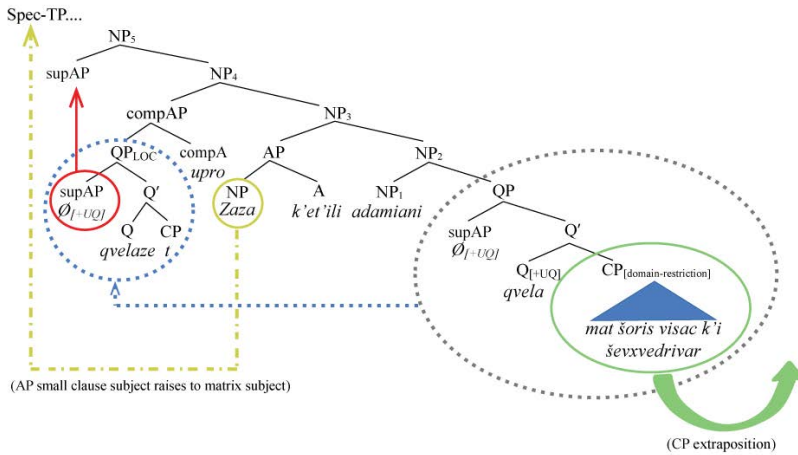
5.2 GEORGIAN SUPERLATIVES

(37)

Zaza qvelaze upro k'et'ili adamiani-a (**mat** **k'i** ševxvedrivar
šoris) visac

Zaza all_{LOC} more nice person-is (**them**_{DAT.PL.} **ever** I-met
among)whom_{ACC.}

‘Zaza is the nicest person whom I ever met.’



On the face of it, the derivation of (37) may seem considerably different from the Slovene one, but in fact, they are quite identical in their skeletal phrasal structure. It is the particulars of lexical items and movement operations that distinguish them. Here are the differences:

- (1) Georgian is an SOV language, and there is an abundance of evidence to show that Relative Clauses in SOV language undergo extraposition (they appear to the right of V). The same appears to be true for the domain restricting CP (i.e., it, too, appears to the right of V, so seemingly has undergone extraposition).
- (2) In Georgian, it is the supA⁰ that is null while the Q⁰ is overt (the opposite of what we find in Slovene). But crucially, they still are in a checking relation.
- (3) While in Slovene, only the SupAP raises, in Georgian, the entire QP raises. It raises into the Specifier of compAP, where it is assigned locative

case by the compA⁰. The locative case is realized on the only overt lexical item in the phrase, *qvela* ‘all’.⁷

(4) The null SupAP is sub-extracted from the QP to raise to a position c-commanding compAP (which is the landing site of the SupAP in Slovene, as well).

6 CONCLUSION

In this paper, I have explored a variety of the superlative construction that, though widely attested across languages, has largely been ignored in the literature: the compositional superlative; specifically, I provided an analysis of the domain-extent compositional superlative, which overtly contains a universal quantifier. This type of superlative appears in Tamil and Georgian. Subsequently, I illustrated that even in languages where a universal quantifier is not apparent, such as in Slovene, it may nevertheless be present: the two phenomena supporting this analysis are (1) the licensing of NPIs under the scope of superlatives and (2) the inclusion of a singular resumptive pronoun in *kar*-phrases modifying a mass noun, both of which are explained by the presence of a universal quantifier. Building on the analysis in which the universal quantifier is inherent in the superlative degree adjective, I suggest that both domain-extent compositional superlatives and synthetic superlatives select a domain-restricting clausal complement. In Slovene, this complement clause is overtly similar to what we find in domain-restricting clauses following bare universal quantifiers, i.e., the *kar*-phrase, which further supports the universal quantifier analysis. Having drawn this parallel between domain-extent compositional superlatives and synthetic superlatives, I suggest a syntactic frame that can be used to derive both.

While there are obvious superficial differences between languages with synthetic superlatives and those with compositional superlatives, I would suggest that the basic structures are the same, which given the semantic overlap, is

⁷ This QP-raising is easily the most objectionable aspect of this derivation, because what I am proposing here involves sideward movement. While there are analyses that make use of such movement (e.g., Nunes’s (2004) analysis of parasitic gapping), and while, quite frankly, I don’t know of compelling arguments against it, I recognize that any movement that fails to extend the structure is generally frowned upon. That said, given the 3.5 structural criteria listed at the start of this section, this derivation appears to be the most parsimonious possible.

entirely unsurprising. If we are to suggest that syntax and semantics bear a relationship, and if ambiguous semantics suggests distinct syntactic structures, then it should also be the case that synonymy (or near synonymy) reflects syntactic similarity. There are, of course, many outstanding questions with respect to the variety of uses and forms related to both comparative and superlative degree adjectives, but this is one small step toward a uniform analysis of superlatives.

REFERENCES

- Aljović, Nadira. 2010. Syntactic positions of attributive adjectives. In: P. Cabrero Hofherr, O. Matushansky (Eds). *Adjectives: Formal analysis in syntax and semantics*. The Netherlands: John Benjamins. <https://doi.org/10.1075/la.153.01alj>
- Bobaljik, Jonathan David. 2012. *Universals in Comparative Morphology*. Cambridge, MA: the MIT Press.
- Bresnan, Joan W. 1973. Syntax of the Comparative Clause Construction in English. *Linguistic Inquiry* 4: 275-343. <https://www.jstor.org/stable/4177775>
- Chidambaram, Vrinda. 2008. The Subcategorization and Quantification Properties of Superlatives in Slovene. In: J. Bailyn., C. Bethin, A. Antonenko (eds.). *Proceedings of Formal Approaches to Slavic Linguistics 16: the Stony Brook Meeting*. Ann Arbor, MI.: Michigan Slavic Publications. 84-101.
- Chutkerashvili, Ana 2009. On –c and ki particles in Georgian. In: P. Bosch, D. Gabelaia, J. Lang (eds.). *Logic, Language, and Computation: 7th International Tbilisi Symposium, Tbilisi*. Berlin: Springer-Verlag. 62-68. https://doi.org/10.1007/978-3-642-00665-4_6
- Cinque, Guglielmo. 2010. *The Syntax of Adjectives: A Comparative Study*. Cambridge, MA: the MIT Press.
- Farkas, Donka F., Kiss, Katalin E. 2000. On the comparative and absolute reading of superlatives. *Natural Language and Linguistic Theory* 18: 417-455.
- Gutiérrez-Rexach, Javier. 2010. Characterizing superlative quantifiers. In: P. Cabrero Hofherr, O. Matushansky (eds.). *Adjectives: Formal analysis in syntax and semantics*. The Netherlands: John Benjamins. <https://doi.org/10.1075/la.153.06gut>
- Heim, Irene. 1999. Notes on Superlatives. ms. MIT. <https://semanticsarchive.net/Archive/TI1MTlhZ/Superlative.pdf>
- Kennedy, Christopher. 2002. Comparative Deletion and Optimality in Syntax. *Natural Language and Linguistic Theory* 20/3: 553-621. <http://www.jstor.org/stable/4048056>
- Kennedy, Christopher. 1997. Antecedent Contained Deletion and the Syntax of Quantification. *Linguistic Inquiry* 28/4, 662-688. <https://www.jstor.org/stable/4178999>
- Kennedy, Christopher, Merchant, Jason. 1997. Attributive Comparatives and Bound Ellipsis. *Linguistics Research Center report 97/3*. University of California, Santa Cruz.
- Ladusaw, William A. 1980. *Polarity Sensitivity as Inherent Scope Relations*. New York: Garland.

- Loccioni, Nicoletta. 2020. The syntax of superlative phrases in Romance. In: I. Vogel. *Romance Languages and Linguistic Theory 16: Selected papers from the 47th Linguistic Symposium on Romance Languages, Newark, Delaware*. 171-186. <https://doi.org/10.1075/rllt.16.11loc>
- Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, MA.:The MIT Press.
- Pancheva, Roumyana. 2006. Phrasal and Clausal Comparatives in Slavic. In: J. Lavine, S. Franks, M. Tasseva-Kurkchieva and H. Filip (eds.) *Formal Approaches to Slavic Linguistics 14: The Princeton Meeting*. 236-257.
- Sleeman, Petra. 2010. Superlative adjectives and the licensing of non-modal infinitive subject relatives. In: P. Cabredo Hofherr, O. Matushansky (eds.). *Adjectives: Formal analysis in syntax and semantics*. The Netherlands: John Benjamins. 233–264. <https://doi.org/10.1075/la.153.07sle>
- Stanley, Jason, Szabó, Zoltan Gendler. 2000. On Quantifier Domain Restriction. *Mind and Language*. Malden, MA: Blackwell Publishing. 219-261.
- Wandl, Florian. 2022. Superlative morphology from syntax: Slavic Nai-/Naj- and internal definiteness marking in Old Lithuanian. *Transactions of the Philological Society* 120/1: 103-127. <https://doi.org/10.1111/1467-968X.12230>

Received February 2024, accepted June 2024.

Prispelo februarja 2024, sprejeto junija 2024.

ACKNOWLEDGEMENTS

My sincerest gratitude to my informants and those I have consulted for this paper; particularly, Tina Abesadze, E. Wayles Browne, Ekaterine Diasamidze-Graham, Alja Ferme, Tamuna Koshoridze, Hilda Koopman, Nicoletta Loccioni, Sandhya Sundaresan, and Sašo Živanović.

I would like very much to thank Florian Wandl for his many insightful comments on earlier versions of this paper, which served to make it much stronger. I would also like to thank all of my Slovenian language consultants, who patiently answered my many (and sometimes confusing) questions. Finally my sincerest gratitude to the two anonymous reviewers and the editing team at *Slovenski Jezik / Slovene Linguistics Studies* for their truly helpful and thoughtful advice and recommendations.

SUMMARY

ON THE STRUCTURAL UNIFORMITY OF SUPERLATIVES: EVIDENCE FROM GEORGIAN, SLOVENE, AND TAMIL

This article explores the structural basis for two morphosyntactically distinct expressions of the superlative: those that are semantically compositional and syntactically analytic and those that are synthetic. Slovenian falls into the latter category while Georgian and Tamil are placed the former. Through careful examination of superlative adjective data in all three languages, one finds evidence of a universal quantifier (UQ), essentially meaning ‘all’, in each of the three languages. In Tamil and Georgian, the UQ is overtly present, and in combination with the comparative degree adjective, it creates (compositionally) the semantics of the superlative. In this paper I argue, using data from Slovenian, that the same compositional structure exists in languages with synthetic superlative degree adjective, although the UQ is unpronounced. The paper aims to provide a single basic structure that allows for multiple surface realizations of superlativity.

○ STRUKTURNI ENAKOSTI PRIDEVNIŠKIH PRESEŽNIKOV: DOKAZI IZ GRUZINŠČINE, SLOVENŠČINE IN TAMILŠČINE

Prispevek obravnava strukturno osnovo dveh oblikoskladenjsko različnih zgradb pri tvorjenju presežnika: prve so pomensko kompozicionalne in skladenjsko analitične (sestavljene), druge pa so sintetične. V slovenščini so v rabi sintetične zgradbe, v gruzinščini in tamilščini pa sestavljene. Natančna analiza podatkov o rabi presežnika v treh jezikih kaže na obstoj univerzalnega kvantifikatorja (UK) s pomenom ‘vse’ v vseh treh jezikih. V tamilščini in gruzinščini je univerzalni kvantifikator izražen in v kombinaciji s primernikom pridevnikov izraža semantiko presežnika. V prispevku na podlagi podatkov iz slovenščine trdim, da enaka struktura obstaja tudi v jezikih s sintetičnim izražanjem presežnika, čeprav univerzalni kvantifikator ni površinsko izražen. Namen prispevka je podati enotno osnovno strukturo, ki združuje več površinsko različnih načinov tvorjenja presežnika.

Dejan Gabrovšek (ORCID: 0000-0002-4995-1814)

ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija

dejan.gabrovsek@zrc-sazu.si

DOI: <https://doi.org/10.3986/16.1.05>

BACKWARDS ANAPHORA AND BACKWARDS ELLIPSIS IN SLOVENIAN MULTI-CLAUSE SENTENCES

The article discusses backwards anaphora and backwards ellipsis in Slovenian multi-clause sentences. It demonstrates their characteristics on all types of Slovenian multi-clause sentences that allow the two phenomena. It turns out that backwards anaphora and backwards ellipsis are rare; backwards anaphora in the role of correlative is the most frequent, and backwards anaphora in coordination is the rarest. I have found individual instances of backwards anaphora and backwards ellipsis even in relationships where they were not envisaged in most studies to date. Backwards anaphora and backwards ellipsis can also have a stylistic effect. It is also possible to have more than one backwards anaphora or backwards ellipsis in one sentence. A semi-backwards anaphora and semi-backwards ellipsis are also shown as a special type of a backwards anaphora and backwards ellipsis.

KEYWORDS: syntax, sentence, cataphora, pronoun, correlative

Članek se ukvarja z vzvratno anaforo in vzvratno elipso v slovenski večstavčni povedi. Njune značilnosti prikazuje pri vseh tipih slovenske večstavčne povedi, ki ju v slovenščini omogočajo. Izkaže se, da sta vzvratna anafora in vzvratna elipsa redki: vzvratna anafora je najpogostejša v vlogi odnosnice, najredkejša pa v priredjih. Posamezne primere vzvratne anafore in vzvratne elipse smo našli tudi v razmerjih, pri katerih jih večina dosedanjih raziskav ni predvidevala. Vzvratna anafora in vzvratna elipsa imata lahko tudi stilistični učinek. V posameznih povedih je možna tudi več kot ena vzvratna anafora ali vzvratna elipsa. Kot poseben tip vzvratne anafore in vzvratne elipse sta obravnavni tudi polzvratna anafora in polzvratna elipsa.

KLJUČNE BESEDE: skladnja, poved, katafora, zaimek, odnosnica

1 INTRODUCTION

The article explores backwards anaphora and backwards ellipsis in the Slovenian multi-clause sentences,¹ which are rare options² of text building and are also limited as regards in which relationships they occur the most frequently, in which more rarely, in which only sporadically and in which at least one of the two is not possible, which does not apply to “ordinary” anaphora and ellipsis. Thus, it is sensible to explore their characteristics and the environments in which they occur separately. I have searched for examples in the *Gigafida 2.0* corpus of written standard Slovenian, using a CQL³ tool allowing me to find a number of examples belonging to a given type, which has made it easier to determine what backwards anaphora and backwards ellipsis have in common.

The aim of this article is to present and study all the types of backwards anaphora and backwards ellipsis that occur in Slovene multi-sentence sentences.

2 SHARED CHARACTERISTICS

Backwards anaphora and backwards ellipsis represent a text-building procedure where a lexical element (especially a noun or verb) or antecedent⁴ is first expressed with a pronoun (backwards anaphor), example (1) or left out

¹ There are also a backwards anaphora and backwards ellipsis within the clause:

Njen_i ginekolog bo zvezdnico_i morda spravi_l v zapor.
 her.nom.sg gynaecologist.nom.sg AUX.FUT star.acc.sg might put.lpt.m.sg in jail.acc.sg

‘Her_i gynaecologist might put the star_i in jail.’

Ljudje naj pred Ø_i in med vožnjo_i ne uživajo alkohola.
 people.nom.sg.pl should before and during driving.ins.sg not consume.ipfv.prs3pl alcohol.gen.sg

‘People should not consume alcohol before and during driving.’

Ellipsis of part of the words is also possible:

Na voljo so eno-, dvo- in tridnevne vozovnice.
 on will.acc.sg AUX one- two and three-day.nom.f.pl ticket.nom.pl

‘One-, two- and three-day tickets are available.’

² It is not possible to determine the frequency ratio between anaphora and ellipsis and backwards anaphora and backwards ellipsis, but a corpus search shows that it is significantly easier to find anaphora and ellipsis than backwards anaphora and backwards ellipsis. The literature review also shows that anaphora and ellipsis are discussed more frequently than backwards anaphora and backwards ellipsis.

³ Corpus Query Language.

⁴ In this case, the term is not particularly apt as the element in question follows the backwards anaphora, but I still use it due to established practice.

(backwards ellipsis), example (2) and only then expressed with a lexical word (Huddleston, Pullum 2002: 1455; Huang 2006: 231; Pavlič, Stepanov 2023: 294). The procedure with backwards ellipsis and backwards anaphora is thus the same. Backwards anaphora is also called cataphora or anticipatory anaphora (Huddleston, Pullum 2002: 1456). Backwards ellipsis is also termed cataphoric ellipsis (Vičar 2011: 96), catalipsis or anticipatory ellipsis (Huddleston, Pullum 2002: 1456). Anaphor is a pronoun in an anaphoric role.

(1)

Če jih_i še ne jemljete, je sedaj pravi čas, da začnete jemati [nosečniške vitamine]_i.

if he.gen.pl yet not take.ipfv.prs.2pl AUX now right.nom.m.sg time.nom.sg that start.pfv.prs.2pl take.ipfv.inf pregnancy.acc.m.pl vitamin.acc.pl

‘If you aren’t taking them yet, now is the right time to start taking pregnancy vitamins.’

(2)

Naloga ni Ø_i in ne bo lahka.

task.nom.sg AUX.NEG and not AUX.FUT easy.nom.f.sg

‘The task is not and will not be easy.’

In the article, I take backwards anaphora to mean any type of anaphora in which the pronoun refers to the following word or part of the text, even though some other linguistic theories would explain some relationships differently. The same applies to backwards ellipsis except that there is an empty space instead of a pronoun.

Backwards ellipsis and backwards anaphora are rarer than anaphora and ellipsis (Beaugrande, Dressler 1992: 50; Kazanina, Phillips 2010: 373; Trnavac, Taboada 2016: 72), even barely detectable in some types of Slovenian multi-clause sentences, though the thesis that there is no backwards anaphora or ellipsis in certain relationships, particularly coordination (Haspelmath 2004: 35), has not been confirmed, because I also found examples of backwards anaphora and backwards ellipsis in the coordination (SECTIONS 3.5 and 4.3).

The presence or possibility of backwards anaphora and backwards ellipsis are signs of approaching the subordinate pole: the inability of backwards anaphora and backwards ellipsis is one of the criteria for coordination (Gabrovšek 2021; Gabrovšek 2023a: 78-79), so they should not appear in typical compound

sentences,⁵ whereas subordination and relationships close to subordination have no such limitations, or such limitations are not as strict. As is shown further on, this is not quite true. The possibility of backwards anaphora and ellipsis is conditioned upon a closer connection (i.e. closer to subordination and they often appear in subordination and less common in other relationships) between clauses, which is the prerequisite for backwards anaphora and backwards ellipsis to be grammatical and understandable. This is also linked to a more pronounced hierarchisation of clauses, which facilitates the formation of more complex structures, which definitely include backwards anaphora and backwards ellipsis (Ross 1967: 121).

In subordination, backwards anaphor and backwards ellipsis can be placed in the dependent clause.⁶ In the correlative role (SECTIONS 3.1.1 and 4.2) they are in the main clause and are very common in this position. Only sporadically can the backwards anaphor which is not in the correlative role be placed in the main clause (SECTION 3.1). This means that backwards anaphor and backwards ellipsis can occur in a hierarchically lower clause when backwards anaphor and backwards ellipsis are not in the role of correlative, and in a hierarchically higher clause when they are in the role of the correlative.

In hierarchically equal clauses (i.e. in coordination), backwards anaphora and backwards ellipsis are limited to sporadic instances. This shows that there are structural differences between subordinate and coordinate relationships but that these boundaries are not clear-cut and that it is impossible to define subordination or coordination only based on individual criteria. The presence of backwards anaphora and backwards ellipsis in some coordinated sentences shows that even in these relationships, the clauses are not completely symmetrical; instead, there is at least a slight hierarchy between them, which allows presence of backwards anaphora and backwards ellipsis.

In principle, most cases of backwards anaphora and backwards ellipsis can be altered to create “regular” anaphora and ellipsis, often simply by switching the order of clauses, but the reverse is not always possible: this, too, shows that anaphora and ellipsis are prototypical, while backwards anaphora and ellipsis

⁵ This is an old thesis disproved as early as the 1980s (Mittwoch 1983). It would be better to say that there is no backwards anaphora and backwards ellipsis in the vast majority of coordination types; this thesis is also supported by the fact that I have only found rare examples of backwards anaphora and backwards ellipsis in coordination.

⁶ Because not all clauses in which backwards anaphora and backwards ellipsis occur are called a dependent clause sentence, a clause introduced by a conjunction is called a conjunctive clause.

are exceptions. This also points to at least some stylistic effect in all instances of backwards anaphora and backwards ellipsis. The only type where only backwards anaphora is possible is added proposition (SECTION 3.3) – reversing the order of sentences is impossible there.

Anaphora is the process of abstracting an antecedent, which is referenced with a pronoun instead of a lexical word, and the bond between them is kept mainly through proximity as they are normally located in the same sentence or adjacent sentences. Ellipsis is the next step in abstraction as the antecedent is expressed as a gap. This also applies to backwards anaphora and backwards ellipsis, which proves why they are rarer: it is easier to create a text going from the concrete towards abstraction (lexical word⁷ → pronoun (anaphor) → Ø) than vice versa. In backwards anaphora and backwards ellipsis, a text is created thus: Ø/pronoun (backwards anaphor) → lexical word.

The condition for backwards anaphora and backwards ellipsis is that at least one of the words is repeated in both clauses of a sentence. As anaphora and ellipsis are text-building processes, it is possible for the reference to stretch across multiple sentences.

Instances of backwards anaphor are coindexed with *i*, the position of backwards ellipsis is coindexed with the \emptyset_i symbol, and the antecedent of backwards anaphor or backwards ellipsis is coindexed with *i*. The second and third instances of backwards anaphora and backwards ellipsis in the same example are coindexed with *j* and *k*.

3 BACKWARDS ANAPHORA

This section discusses backwards anaphora. Backwards anaphora appears in a number of multi-clause relationships but is the most common in subordination. An anaphor is often located in the first clause of a sentence but relates to the previous sentence; such cases should not be considered backwards anaphora. To qualify as a backwards anaphora, it must refer only to a constituent in the following clause or the whole following clause. Sporadically, a backwards anaphor and its antecedent can be more than one clause apart. Considering the above, a backwards anaphor introduces new information, i.e. a new argument that has not yet been mentioned in the text.

⁷ I consider lexical words to include nouns, verbs, adjectives and adverbs.

As a rule, and this is especially apparent in the correlative role (SECTION 3.1.1) and in added proposition (SECTION 3.3), a backwards anaphor announces content with a pronoun:⁸ this content is thus as yet undetermined, unknown, and only made concrete by the following clause.

3.1 BACKWARDS ANAPHORA IN THE MAIN CLAUSE

Backwards anaphora in the main clause occurs only sporadically, and some discussions consider it impossible. We have found only one such example. Backwards anaphora as a correlative, on the other hand, is much more common (SECTION 3.1.1), although it also occurs in the main clause.

(3)

<i>Pri</i>	<i>njeni_i</i>	<i>zdravnici</i>	<i>sem</i>	<i>izvedela,</i>	<i>da je</i>	<i>hčerka_i</i>
at	her.loc.sg	doctor.loc.sg	AUX.PST	find out.lpt.f.sg	that	AUX.PST daughter.nom.sg
<i>bolezen</i>	<i>dobila</i>	<i>z</i>	<i>okužbo.</i>			
disease.acc.sg	get.lpt.f.sg	with	infection.ins.sg			

‘I found out from her doctor that my daughter had contracted the disease through an infection.’

3.1.1 BACKWARDS ANAPHORA AS CORRELATIVE

A dependent clause is introduced by a pronominal correlative, which is located in the main clause, reveals the function of the dependent clause and acts as a bridge between the main and dependent clauses (Ross 1967: 120; Petr 1986: 531; Žele 2016; 2017; Gabrovšek, Žele 2019; Gabrovšek 2023a: 61; 2023b). A correlative is typically located before the dependent clause, so it functions as a backwards anaphor (Jelovšek 2024: 228).⁹ It can also be left out; in which case it also functions as a backwards ellipsis (SECTION 4.2). A correlative in the backwards anaphora role is prototypical (examples (4-6)).¹⁰ A correlative announces the content, but the pronoun has too wide a meaning, so the dependent clause makes it more concrete (Halliday, Hasan 1976: 56). The same applies to

⁸ This sentence also contains a backwards anaphor.

⁹ Backwards anaphora can also be found in 16th century Slovenian texts (Jelovšek 2011: 421; 2024: 138, 143).

¹⁰ A correlative in the role of an anaphor is rarer:

[<i>Če</i>	<i>vam</i>	<i>je</i>	<i>to</i>	<i>všeč</i>],	<i>potem;</i>	<i>uživajte.</i>
if	you.dat.pl	AUX	this.nom.n.sg	like.adv	then.adv	enjoy.ipfv.imp.2pl

‘If you like this, then enjoy.’

a dependent clause preceding or inserted in the main clause. The position of the main clause thus has no (major) role in the placement of the correlative; there is a strong tendency to place it before the conjunction, i.e. before the dependent clause. A backwards anaphor refers to the whole dependent clause. The function of a correlative as a backwards anaphor is important because this is the only position where backwards anaphora is common and prototypical. In all other positions, backwards anaphora is an exception. A correlative plays the role of backwards anaphor anytime a dependent clause is located after or nested in the main clause, and often also where a dependent clause precedes the main clause.

(4)

Včasih naredim kaj, [kar ga malo razjezi].

sometimes do.pfv.prs.3sg what.acc.sg which.nom.sg.n.rel he.acc.sg little make angry.pfv.prs.3sg

‘Sometimes I do something that makes him a little angry.’

(5)

Najlepše je tu, [kjer živim].

beautiful.adv.sup AUX here where.adv.rel live.ipfv.prs.1sg

‘The most beautiful place is where I live.’

Backwards anaphora is mandatory in attributive clauses (Petr 1986: 524). However, the pronoun *ta* ‘this’ or *tisti* ‘that’ which stands at the beginning of a noun phrase is usually omitted. Such cases belong to the type 4.2 Backwards ellipsis as correlative.

(6)

Tisti, ljudje, [ki so uspešni v svojem

that.nom.m.pl people.nom.pl which AUX successful.nom.m.pl in your.adj.refl.loc

poklicu], so pogosto pod pritiskom.¹¹

profession.loc.sg AUX often under pressure.ins.sg

‘People who excel in their profession are often under pressure.’

3.2 BACKWARDS ANAPHORA IN DEPENDENT CLAUSE

Backwards anaphora is possible in referring to any sentence element. Among all dependent clauses, backwards anaphora in conditional clauses stands out

¹¹ [lemma="ta|tisti|tisti"] [tag="So.*"] [word=","] [word="ki|da"]

prominently as such clauses cover most of the examples found. Backwards anaphora in concessive clauses is also fairly common, whereas it is barely detectable in other dependent clauses. The dependent clause must always precede the main clause. Theoretically, it can be nested, though I have not found any such example. This makes backwards anaphora in a dependent clause significantly different from backwards anaphora as a correlative as the latter is not as bound to the placement of the dependent clause in the sentence.

(7)

Če jih_i še ne jemljete, je sedaj pravi čas, da začnete jemati [nosečniške vitamine].
 if he.gen.pl yet not take.ipfv.prs.2pl AUX now right.nom.m.sg time.nom.sg that start.pfv.prs.2pl take.ipfv.inf pregnancy.acc.m.pl vitamin.acc.pl

‘If you aren’t taking them yet, now is the right time to start taking pregnancy vitamins.’

(8)

Ker ga_i predsednik ni želel uporabljati, so Mehičani prodali letalo.
 because he.acc.sg president.nom.sg AUX.PST.NEG want.lpt.m.sg use.ipfv.inf AUX.PST Mexican.nom.pl sell.lpt.m.pl plane.acc.sg

‘Because the president refused to use it, the Mexicans have sold their jet.’

3.2.1 BACKWARDS ANAPHORA IN DIRECT SPEECH

As with other dependent clauses, the antecedent here is expressed in the non-initial clause, and the backwards anaphora is part of the literal quote.

(9)

Žalosti me_i, pravi strokovnjakinja, da ljudje tako malo hodijo peš.
 make sad.ipfv.prs.3sg I.acc.sg say.ipfv.prs.3sg expert.nom.f.sg that people.nom.pl that.adv little.adv walk.ipfv.prs.3pl on foot.adv

‘It makes me sad, says the expert, that people walk so rarely.’

3.2.2 BACKWARDS ANAPHORA REFERS TO THE WHOLE CLAUSE

The demonstrative pronoun *to* ‘this’ in the neuter gender and singular refers to the whole clause (Huddleston, Pullum 2002: 1458). This function of the neuter singular *to* is very common in the role of anaphora and backwards anaphora.

(10)

<i>Če</i>	<i>bodo</i>	<i>razmere</i>	<i>to</i> _i	<i>dopuščale,</i>
if	AUX.FUT	situation.nom.pl	this.acc.n.sg	allow.lpt.f.pl
<i>[bom</i>	<i>dal</i>	<i>več</i>	<i>priložnosti</i>	<i>mlajšim]</i> _i .
AUX.FUT	give.lpt.m.sg	more	opportunity.gen.pl	young.dat.pl.cmpr

‘If the situation allows it, I’ll give more opportunities to younger players.’

3.2.3 BACKWARDS ANAPHORA IS AN ADJECTIVAL PRONOUN

The backwards anaphora is an adjectival pronoun that is part of a noun phrase. The fact that the backwards anaphora and antecedent are different parts of speech does not decrease the understandability of the reference.

(11)

<i>Kljub</i>	<i>neprijetnim</i>	<i>značajskim</i>	<i>lastnostim,</i>	<i>kjer</i>
despite	disagreeable.dat.pl	character.dat.pl	trait.dat.pl	where
<i>vsekakor</i>	<i>vodi</i>	<i>njegova</i> _i	<i>pretirana</i>	<i>samovšečnost,</i>
definitely	lead.ipfv.prs.3sg	his.nom.f.sg	excessive.nom.f.sg	self-importance.nom.sg
<i>direktorju</i> _i	<i>priznavajo</i>	<i>izjemne</i>	<i>menedžerske</i>	<i>sposobnosti.</i>
director.dat.sg	recognize.ipfv.prs.3pl	outstanding.acc.f.pl	managerial.acc.f.pl	skill.acc.pl

‘Despite disagreeable character traits, especially his excessive self-importance, the director is hailed for his outstanding managerial skills.’

3.2.4 BACKWARDS ANAPHORA IS AN ADVERBIAL PRONOUN

The adverbial pronoun *tako* ‘so’ refers to an adverb, example (12), or a whole clause, example (13).

(12)

<i>Čeprav</i>	<i>se</i>	<i>ne</i>	<i>zdi</i>	<i>tako</i> _i ,	<i>je</i>	<i>[jasno]</i> _i ,
though	REFL.ACC	not	seem.ipfv.prs.3sg	so	AUX	clear.adv
<i>da</i>	<i>brez</i>	<i>avta</i>	<i>ne</i>	<i>moreš</i>	<i>nikamor</i> ¹²	
that	without	car.gen.sg	not	can.ipfv.prs.2sg	nowhere	

‘Though it may not seem so, it’s clear that you can’t get anywhere without a car.’

¹² [word="čeprav|čtudi"] []{0,3}[lemma="zdeti"][]{0,3} [word="tako"] [word=","] []{0,4} [tag="R.*"]

(13)

*Četudi se sprva ni zdelo tako, [so se
 though REFL.ACC at first AUX.PST.NEG seem.lpt.n.sg so AUX.PST REFL.ACC
 ljudje izvrstno zabavali].*
 people.nom.pl excellently fun.lpt.m.pl

‘Though it didn’t seem so at first, people had great fun.’

3.3 BACKWARDS ANAPHORA IN ADDED PROPOSITION

Added propositions are two-clause sentences punctuated with a colon or semicolon (Petr 1987: 535; Žele 2013: 21). The second clause usually refers to the last sentence element of the first clause. As a rule, a colon can be replaced by the conjunction *in sicer* ‘namely, that is’, though this conversion at least partly alters the meaning, as any conversion does. Direct speech is not part of this type. The second clause can concretise or explain the first one or list options according to the first clause (Petr 1987: 535; Belaj, Tanacković Faletar 2020: 403-404). In this type, the backwards anaphora has a strong linking role as the pronoun functioning as the backwards anaphora is the only word joining the two clauses to form a sentence – no conjunction is possible in this type. Backwards anaphora is mandatory in this type; reversing the order of the clauses would be ungrammatical or at least unusual (Huddleston, Pullum 2002: 1456).

(14)

*Navijači so doživeli marsikaj: [od
 fan.nom.m.pl AUX.PST experience.lpt.m.pl many things.acc.pl from
 proslavljanja napredovanja, napetosti zaradi preiskave
 celebrating.gen.sg promotion.gen.sg tension.gen.sg because investigation.gen.sg
 in zdaj še veliko razočaranje].¹³
 and now yet big.acc.n.sg disappointment.acc.sg*

‘The fans have experienced many a thing: celebrating the promotion, tensions due to the investigation and now great disappointment.’

(15)

*Prava težava je [nekje drugje]: [zahteve je
 real.nom.f.sg problem.nom.sg AUX somewhere else demand.acc.pl AUX*

¹³ [tag="Z.*"] [word="":"]

treba podpreti s prepričljivimi dokazi].

necessary support.pfv.inf with convincing.ins.m.pl evidence.ins.pl

‘The real problem lies elsewhere: demands need to be supported with real and convincing evidence.’

3.4 BACKWARDS ANAPHORA IN PARENTHESIS

Backwards anaphora is possible, though rare, in parentheses (Vičar 2011). Very often, backwards anaphora in parentheses behaves similarly in supplementary clauses (SECTION 5.1), the only difference being that the role of backwards anaphora is played by the conjunction and pronoun *kar* in supplementary clauses and by the pronoun *to* here, with *in* ‘and’ playing the role of conjunction. This separation of functions to two words also proves that *kar* intertwines both functions. It is characteristic of parenthesis to disrupt the flow of speech: information is not passed on linearly, unlike in typical coordination (where backwards anaphora is almost non-existent for exactly this reason); instead the order is altered as the parenthesis is inserted within a clause, turning an “ordinary” anaphora in a compound sentence (example (16a)) into a backwards anaphora (example (16)). Backwards anaphora mainly plays the role of additionally highlighting the content to follow; it is an announcement. The same applies to supplementary clauses.¹⁴ In this type, backwards anaphora has a strong linking, i.e. conjunctive role.

(16)

Naposled, in [to]_i je pomembna novost, [je
 in the end and this.nom.n.sg AUX important.nom.f.sg innovation.nom.sg AUX.PST
nastal tretji red – tisti, ki dela]_i.
 arise.lpt.m.sg third.nom.sg order.nom.sg that.nom.m.sg which work.ipfv.prs.3sg

‘In the end, as an important innovation, a third order arose – the working order.’

(16a)

Naposled, je [nastal tretji red – tisti, ki
 in the end AUX.PST arise.lpt.m.sg third.nom.sg order.nom.sg that.nom.m.sg which
dela]_i, in to_i je pomembna novost.
 work.ipfv.prs.3sg and this.nom.n.sg AUX important.nom.f.sg innovation.nom.sg

‘In the end, a third order arose – the working order, which is an important innovation.’

¹⁴ A parenthesis with a backwards anaphor can be converted into a supplementary clause and vice versa, which shows how similar these relationships are.

3.5 BACKWARDS ANAPHORA IN COORDINATION

In coordination, backwards anaphora is possible only in highly specific, often also stylistically marked examples: as a rule, arguments are first expressed with a lexical word before they can be left out (ellipsis) or expressed with a pronoun (anaphor) (Quirk 1985: 922; Orešnik 1992: 73; Haspelmath 2004: 35; Gabrovšek 2024: 78).¹⁵ Haspelmath provides the following example to demonstrate the ungrammaticality of backwards anaphora: **His_i wife worked/working at the mountain, (and) the old man_i tended the store* (Haspelmath 2004: 35).

An analysis of the materials has shown that very rarely, backwards anaphora does appear in coordination, usually for stylistic effect. Notably, I have not found backwards anaphora in copulative coordination, which is the most typical coordinate relationship, but I have found it in adversative coordination (Mittwoch 1983: 130-131) and in causal coordination. The most typical example of backwards anaphora in coordination is the construction *tega ne vedeti, ampak* ‘not know this but’, example (20).¹⁶ In this type, the first clause announces new information, which is typically unusual, unexpected. This shows that backwards anaphora is often subject to structural limitations: a number of specific conditions (like pronoun *ta* ‘this’, verb *vedeti* ‘know’, conjunction *ampak* ‘but’) must be met for it to appear, otherwise it is not possible.

It is worth noting that examples of backwards anaphora in coordination are so rare that sporadic examples are only found in big corpora, which contain more texts and thus have a higher chance of (very) rare structures appearing. My analysis shows that backwards anaphora is indeed absent from at least the most prototypical types of coordination, especially copulative coordination and disjunction. Adversativeness allows sporadic instances, probably in part because the clauses have antithetical content, and backwards anaphora further emphasises this opposition. As backwards anaphora appears in coordination only sporadically, I can confirm the thesis that in principle, backwards anaphora is not possible in coordination, but this is not an absolute rule. Example (19) is interesting because the antecedent is expressed only

¹⁵ The first clause of a compound sentence can include anaphoric pronouns, but as a rule, they relate to the previous sentence.

¹⁶ Similar example is in SECTION 4.3.4 Announcing unexpected information.

in the dependent clause following the main clause – this also proves that a dependent clause is part of its main clause as well as shows that backwards anaphora and its antecedent can be located more than one clause apart without this affecting the understandability of the sentence. Due to rarity, I cite more examples than with other types.

(17)

Morski pes se je večkrat zagnal vanjo, a najstnici je uspelo pobegniti.
 sea.nom.m.sg dog.nom.sg REFL.ACC AUX.PST repeatedly chase.lpt.m.sg in she.acc.sg but
 teenager.dat.f.sg AUX.PST manage.lpt.n.sg escape.pfv.inf

‘The shark snapped at her several times, but the teenager managed to escape.’

(18)

Dali so jima dva odstotka možnosti, da preživita, a [čudežni deklici] bosta kmalu praznovali sedmi rojstni dan.
 give.lpt.m.pl AUX.PST she.dat.du two.acc.m.du percent.acc.du chance.gen.sg
 that survive.pfv.prs.2du but miracle.nom.f.du girl.nom.du AUX.FUT
 soon celebrate.lpt.f.du seventh.acc.m.sg birth.acc.m.sg day.acc.sg

‘They gave them a two-percent chance of survival, but the miracle girls will soon celebrate their seventh birthday.’

(19)

V celico mu prinesejo šest istih obrokov, paznik pa izbere, katerega bo zapornik pojedel.
 in cell.acc.sg he.dat.sg bring.pfv.prs.3pl six.acc same.gen.pl meal.gen.pl
 guard.nom.sg and select.pfv.prs.3sg which.rel.acc.m.sg AUX.FUT prisoner.nom.sg eat.lpt.m.sg

‘They bring him six same meals to his cell, and the guard selects which one the prisoner is to eat.’

(20)

Morda tega še ne veste, a [sem hitre jeze].
 perhaps this.gen.n.sg yet not know.ipfv.prs.2pl but AUX quick.gen.f.sg anger.gen.sg

‘Perhaps you don’t know this yet, but I’m quick to anger.’

In sporadic examples, which have a significant semantic overlap with added proposition (SECTION 3.3), the second clause defines the pronoun located in the first clause in more detail.

(21)

<i>V</i>	<i>zadnjih</i>	<i>letih</i>	<i>je</i>	<i>avtobusna</i>	<i>postaja</i>
in	recent.loc.n.pl	year.loc.pl	AUX.PST	bus.nom.f.sg	station.nom.sg
<i>začela</i>	<i>sloveti</i>	<i>še</i>	<i>po</i>	<i>nečem_i</i>	<i>in</i>
start.lpt.f.sg	be famous.ipfv.inf	yet	after	something.loc.sg	and
<i>sicer</i>	<i>po</i>	<i>kriminalu_i</i> . ¹⁷			
otherwise	after	crime.loc.sg			

‘In recent years, the bus station has acquired a reputation of something else, namely crime.’

3.6 BACKWARDS ANAPHORA IN JUXTAPOSITION

Backwards anaphora is possible if there is a non-initial vocative in the sentence and a pronoun in the first clause refers to it. All examples have been found in dialogue – this is the typical environment for vocative use. This type is very close to exophora.

(22)

<i>Strinjam</i>	<i>se</i>	<i>z</i>	<i>vami_i</i>	<i>Ana_i</i> . ¹⁸
agree.ipfv.prs.1sg	REFL.ACC	with	you.ins.pl	Ana.nom.sg

‘I agree with you, Ana.’

3.7 BACKWARDS ANAPHORA REFERRING TO THE WHOLE TEXT

Particularly to attract interest in reading a whole newspaper or online article, the title only indicates the news content with a backwards anaphora that refers to the content of the whole article. This is an example of backwards anaphora reaching beyond the sentence in which it is located; the pronoun *to* ‘this’ can summarise the entire text that follows.¹⁹ This is also the only type of backwards anaphora systemically reaching beyond the limits of a sentence, though such cases are most likely possible in other types as well.

(23)

<i>To_i</i>	<i>po</i>	<i>drami</i>	<i>in</i>	<i>pritiskih</i>	<i>počne</i>	<i>novinarka</i> .
this.acc.n	after	drama.loc.sg	and	pressure.loc.pl	do.ipfv.prs.3sg	journalist.nom.sg

¹⁷ [tag="Zn.*"] [word=","] [word="in/to"] [word="sicer/je"] within <s/>

¹⁸ [tag="Gg.*"] []{0,2} [tag="Zod.*"] []{0,2} [word=","] [tag="P.*"]?[tag="S...i.*"]

¹⁹ Due to its length, this text is not cited here.

4 BACKWARDS ELLIPSIS

Backwards ellipsis behaves in a manner very similar to backwards anaphora; they differ mostly in the fact that backwards ellipsis is more common than backwards anaphora in coordination, though it is still very rare. Backwards ellipsis does not have such a strong linking role as it cannot function as a conjunction or referent, and supplementary clauses, added propositions and parentheses need at least one linking element, i.e. a conjunction or referent or both. This is also why there are more types of backwards anaphora than types of backwards ellipsis, but on the other hand, there are more types of backwards ellipsis than of backwards anaphora in coordination. This is because compound sentences do not need as many linking elements (which definitely include pronouns playing the role of backwards anaphor) as complex sentences do, and ellipsis is also more characteristic of coordination, so backwards ellipsis, too, can be somewhat more common in coordination.

4.1 BACKWARDS ELLIPSIS IN SUBORDINATION

Backwards ellipsis is somewhat more common in subordination and more systemic than backwards anaphora. In backwards ellipsis, conditional clauses feature very common and well-established constructions. It seems that even where it is possible, backwards ellipsis is very rare and limited to certain more or less established constructions. Conditional clauses are a prominent exception; in other dependent clauses, I have found only sporadic instances of backwards ellipsis, which shows that the relationships where backwards anaphora and backwards ellipsis are more common overlap significantly. In examples (25-26), the subject is left out in the dependent clause: this is normal in Slovenian,²⁰ so these are examples of backwards ellipsis.²¹

(24)

Če kdaj \emptyset_i , [si nagrado zaslužijo]_i, letos.
 if when REFL.DAT award.acc.sg deserve.pfv.prs.3pl this year
 ‘If ever, they deserve to win the award this year.’

²⁰ Slovenian is a pro-drop language.

²¹ If a pronoun was inserted in the dependent clause in the example (26): *Odkar je ona, rodila, se je vplivnica, zelo spremenila*, the subject would be emphasised, and this would be an example of backwards anaphora. It would also be perfectly normal for the word *vplivnica* ‘influencer’ to appear only in the second clause. *Odkar je vplivnica, rodila, se je \emptyset_i zelo spremenila*.

(25)

<i>Čeprav</i> Ø _i	<i>ne</i>	<i>spada</i>	<i>med</i>	<i>najbolj</i>
though	not	belong.ipfv.prs.3sg	among	most.adv.sup
<i>mondene</i>	<i>predele</i>	<i>Sredozemlja,</i>	<i>postaja</i>	<i>Jadran</i> _i
cosmopolitan.acc.m.pl	part.acc.pl	Mediterranean.gen.sg	become.ipfv.prs.3sg	Adriatic.nom.sg
<i>zanimiv</i>	<i>tudi</i>	<i>za svetovne</i>	<i>bogataše.</i>	
interesting.nom.m.sg	also	for world.acc.m.pl	rich people.acc.pl	

‘Though it isn’t among the most cosmopolitan parts of the Mediterranean, the Adriatic is becoming more and more interesting to rich people of the world.’

(26)

<i>Odkar</i>	<i>je</i> Ø _i	<i>rodila,</i>	<i>se</i>	<i>je</i>	<i>vplivnica</i> _i	<i>zelo</i>	<i>spremenila.</i>
Since	AUX.PST	give birth.lpt.f.sg	REFL.ACC	AUX.PST	influencer.nom.f.sg	very	change.lpt.f.sg

‘Since she gave birth, the influencer has changed significantly.’

I have not found any examples with backwards ellipsis in the main clause, which is confirmed by scholarly literature (Kehler 2018: 318).

4.2 BACKWARDS ELLIPSIS AS CORRELATIVE

Since I claim a correlative is mandatory in subordination (Gabrovšek 2023a: 61; 2023b) but not always expressed, cases of unexpressed correlatives are instances of ellipsis. If a correlative precedes a dependent clause, this is a case of backwards ellipsis.²² Because the correlative is left out, it is not always unequivocally clear which kind of ellipsis there is as the correlative could be inserted before or after the dependent clause. There is definitely backwards ellipsis if the dependent clause follows the main clause; see example (27). The pronoun *to* ‘this’ is elided. The same type with the pronoun is in the SECTION 3.1.1.

(27)

<i>Mislím</i> Ø _i ,	[<i>da nam</i>	<i>bo</i>	<i>uspelo</i>] _i .	
think.ipfv.prs.1sg	that	I.dat.pl	AUX.FUT	succeed.lpt.n.sg

‘I think we’ll make it.’

4.3 BACKWARDS ELLIPSIS IN COORDINATION

In coordination, backwards ellipsis is very rare but still more common than backwards anaphora, so the thesis that backwards ellipsis is not possible in

²² As noted in SECTION 3.1.1, an expressed correlative is a backwards anaphor.

coordination (Žele, Krajnc Ivič 2020: 232) has not been confirmed (Mihaljevič 2024: 315). Backwards ellipsis in coordination is either stylistically marked or a special construction with a very specific meaning. Unlike backwards anaphora, which is only a sporadic exception, backwards ellipsis has a strong semantic role. The reason for the rarity of backwards ellipsis lies in the structural equality of clauses, which is one of the criteria for coordination (Gabrovšek 2021; 2023a: 74; 2024: 78). As the clauses are structurally equal, they can complement each other, and repeated arguments can be left out, which they usually are. Ellipsis in coordination is (almost) always anaphoric. Arguments must be expressed first (with the exception of the subject, which is evident from the finite verb form, in Slovenian) before being left out.

The basic type of backwards ellipsis is the same as in backwards anaphora in coordination (SECTION 3.5). Because the pronoun (*oni* ‘they’) in the first clause is in the nominative case and functions as the subject, it is left out.

(28)

<i>Sicer</i>	<i>so</i> Ø _i	<i>imeli</i>	<i>dobre</i>	<i>namene,</i>	<i>a</i>
while	AUX.PST	have.lpt.m.pl	good.acc.m.pl	intention.acc.pl	but
<i>je</i>	<i>stranka</i> _i	<i>krepko</i>	<i>zgrešila,</i>	<i>ko</i>	<i>je</i>
AUX.PST	party.nom.sg	strongly	miss.lpt.f.sg	when	AUX.PST
<i>na</i>	<i>Twitterju</i>	<i>objavila</i>	<i>napačno</i>	<i>zastavo.</i>	
on	Twitter.loc.sg	post.lpt.f.sg	wrong.acc.f.sg	flag.acc.sg	

‘While it had good intentions, the Party made quite a flop when it posted the wrong flag on Twitter.’

4.3.1 EMPHASISED STYLISTIC ROLE

This section includes examples where the role of backwards ellipsis is purely stylistic, i.e. to express particular information in a more prominent manner – this proves that backwards ellipsis is an exception in coordination, but useful to emphasise particular information for this exact reason. All examples could be expressed equally with no backwards ellipsis. The following types of coordination are also stylistically marked, but categorised into specific groups based on shared characteristics.

(29)

Še dobro se ni iztekel en Ø_i,
 yet good.adv REFL.ACC AUX.PST.NEG end.lpt.m.sg one.nom.m.sg
že se obeta nov [jazzovski festival]_i.
 already REFL.ACC promise.ipfv.prs.3sg new.nom.m.sg jazz.nom.m.sg festival.nom.sg
 ‘While one festival has not even quite ended, a new jazz festival is in the works.’

4.3.2 EMPHASISED ADDITION WITH BACKWARDS ELLIPSIS

One piece of information is emphasised with backwards ellipsis, which seems marked due to its great rarity. This type is structurally limited to the same verb in two different forms: these can differ in tense, or one of them is negated (Quirk 1985: 976).

(30)

On je bil Ø_i in še je brez dvoma legenda_i.
 he.nom.sg AUX.PST be.lpt.m.sg and yet AUX without doubt.gen.sg legend.nom.sg
 ‘He was and undoubtedly still is a legend.’

(31)

Naloga ni Ø_i in ne bo lahka_i.
 task.nom.sg AUX.NEG and not AUX.FUT easy.nom.f.sg
 ‘The task is not and will not be easy.’

4.3.3 SUBSTITUTION EXPRESSED WITH BACKWARDS ELLIPSIS

The first clause conveys a first alternative, and the second clause a second one, which is better, more appropriate. The verb appears only in the second clause, so there is backwards ellipsis of the verb. The same meaning of the whole sentence could also be expressed without backwards ellipsis, but such a substitution is more emphasised, so the construction has an expressive effect. This is adversative coordination, which has also been found to be the most common type of coordination with backwards anaphora. The first clause is negated, and the second one is affirmative.

(32)

Ne Ø_i Sloveniji, tokrat spodrsuje_i Evropi.
 not Slovenia.dat.sg this time slip.ipfv.prs.3sg Europe.dat.sg
 ‘Not Slovenia, this time Europe is failing.’

(33)

Ne Ø_i *veter,* *odnos* *do igre bo odločil*_i
 not wind.nom.sg attitude.nom.sg to play.gen.sg be.ind.fut.3sg determine.lpt.m.sg
*o uspehu.*²³
 about success.loc.sg

‘It isn’t the wind but rather the attitude to playing that will determine who succeeds.’

4.3.4 ANNOUNCING UNEXPECTED INFORMATION

The first clause in the form *ne boste verjeli, ampak* ‘you won’t believe [this] but’ announces the second one; example (20) in SECTION 3.5 has the same effect. A pronoun which functions as argument can be inserted in the first clause of each example, which points to similarity in the behaviours of backwards anaphora and backwards ellipsis.

(34)

Ne boste verjeli Ø_i, *ampak* [še vedno imam *tremo*]_i.²⁴
 not AUX.FUT believe.lpt.m.pl but yet always have.ipfv.prs.1sg fright.acc.sg

‘You won’t believe this but I still have stage fright.’

4.3.5 BACKWARDS ELLIPSIS IN ADDITIVE CORRELATION

Backwards ellipsis in additive correlation introduced by the conjunctive phrase *ne samo – ampak tudi* ‘not only – but also’ (with synonymous variants) behaves in a manner similar to substitution with systemic ellipsis in the first clause. The possibility of backwards ellipsis in additive correlation is most likely enabled by its similarity to adversative coordination as some conjunctions are the same.

(35)

Ne Ø_i *le izplačilo nagrade, želi* *še dodaten milijon*
 not just payment.acc.sg bonus.gen.sg want.ipfv.prs.3sg yet additional.acc.m.sg million.acc.sg
*in pol.*²⁵
 and half.acc.sg

‘He wants not only the bonus but an additional million and a half.’

²³ [word="\." [word="Kaj|Ne"] [tag="S.*|Z.*"] [word=","] [tag!="Vd"]

²⁴ [word="ne"][] {,3} [lemma="verjeti"] [] {,3} [word=","] [word="ampak|toda|a"]

²⁵ [word="ne"] [word="samo|le"] [] {1,5} [word="ampak|temveč|marveč|pač"]

(36)

Ne Ø_i *le* *zvečer*, *s* *pojavom* *tigrastih* *komarjev*
 not just evening with advent.ins.sg tiger.gen.m.pl mosquitoes.gen.pl
 [*smo* *na* *pikanje* *obsojeni*]_i *ves* *dan*.
 AUX on biting.acc.sg doomed.nom.m.pl all.acc.m.sg day.acc.sg

‘With the advent of tiger mosquitoes, we are doomed to being bitten not only in the evening, but throughout the day.’

4.3.6 BACKWARDS ELLIPSIS WITH TWO SUBORDINATE CONJUNCTIONS

In examples where two dependent clauses differ only in the conjunction whereas the other words are the same, the sentences can be joined by coordinating the subordinating conjunctions, with the content of the dependent clause following (Ha 2008: 123). As a rule, only conjunctions close in meaning (e.g. temporal and conditional conjunctions) can be joined. This possibility is a very good way to concisely convey information but is also systemically very limited and thus quite rare. Example (37) is a shortened form of example (37a). Predominantly, it is the conjunctions *ko* ‘when’ and *če* ‘if’ that are joined. The most common combinations are: *ko in če* ‘when and if’, *če in ko* ‘if and when’, *ko oziroma če* ‘when or if’, *če oziroma ko* ‘if or when’, *če in kadar* ‘if and when’. The backwards ellipsis results in coordination of two subordinate conjunctions.

(37)

Znate *se* *zavzeti*, *kadar* Ø_i *in* *kjer* [*je* *treba*]_i.²⁶
 know.ipfv.prs.2pl REFL.ACC commit.pfv.inf when and where AUX necessary

‘You know how to make a stand when and where needed.’

(37a)

Znate *se* *zavzeti*, *kadar* *je* *treba* *in* *kjer* *je* *treba*.
 know.ipfv.prs.2pl REFL.ACC commit.pfv.inf when AUX necessary and where AUX necessary

‘You know how to make a stand when needed and where needed.’

(38)

Do *sprememb* *bo* *prišlo* *ko* Ø_i *in* *če*
 to change.gen.pl AUX.FUT come.lpt.n.sg when and if
 [*bo* *gospodarstvo* *presodilo*, *da* *mu* *koristijo*]_i.
 AUX.FUT economy.nom.sg deem.lpt.n.sg that it.dat.n.sg benefit.ipfv.prs.3pl

‘The changes will take place when and if the economy deems them useful.’

²⁶ [tag="Vd"] [tag="Vp"] [tag="Vd"]

5 SEMI-BACKWARDS ANAPHORA AND SEMI-BACKWARDS ELLIPSIS

Sporadically, in addition to backwards anaphora and backwards ellipsis, semi-backwards anaphora and semi-backwards ellipsis are also possible. Semi-backwards anaphora is only possible where the dependent clause is embedded in the main clause and the neuter pronoun *to* ('this') refers to the entire main clause surrounding the dependent: part of the content referred to by the pronoun is before the pronoun and part is after the pronoun, hence the proposed term semi-backwards anaphora, examples (39-40). The same is true of semi-backwards ellipsis, where the pronoun *to* is absent, example (41). Both possibilities are rare.²⁷

(39)

[*Do oltarja se jima*]_{i/2} *kljub temu, da si*
 to altar.gen.sg REFL.ACC they.acc.du despite of this.dat.n.sg. that REFL.DAT
to_i močno želita, [*še ni uspelo sprehoditi*]_{i/2}.²⁸
 this.acc.n.sg strongly wish.ipfv.prs.2du yet AUX.PST.NEG succeed.lpt.n.sg walk.pfv.inf

'They have not yet managed to get married, despite their strong desire to do so.'

(40)

[*Zaposleni so*]_{i/2}, *če je to_i bilo mogoče,*
 employee.nom.pl AUX.PST if AUX.PST this.nom.n.sg be.lpt.n.sg possible
*[zapustili delovna mesta ter odhiteli domov]*_{i/2}.
 left.lpt.m.pl working.acc.n.pl place.acc.pl and rush.lpt.m.pl home.

'Employees left their jobs and rushed home whenever possible.'

(41)

[*Vsak grafolog bi lahko*]_{i/2}, *če bi*
 every.nom.m.sg graphologist.nom.sg COND can.adv if COND
bilo *Ø_i treba,* [*dokazal, da je to*
 be.lpt.n.sg necessary prove.lpt.m.sg that AUX.PST this.nom.n.sg
njegov podpis]_{i/2}.
 his.nom.m.sg signature.nom.sg

'Any graphologist could, if necessary, prove that it is his signature.'

²⁷ The proposed Slovenian equivalent terms are *polvzvrtna anafora* and *polvzvrtna elipsa*.

²⁸ [tag="[SG].*"] [] {,2} [word=","] [tag="Vd"] [] {,3} [lemma="ta"] [tag!="[PS].*"] [] {,3} [word=","] [] {,2} [tag="[SG].*"] within <s/> neg filter [word="ki"] in [tag="Vp"]

5.1 SEMI-BACKWARDS ANAPHORA IN SUPPLEMENTARY CLAUSES

The second, structurally similar possibility is represented by supplementary clauses. These are clauses that are introduced by originally subordinating conjunctions and express a typical coordinate relationship (Gabrovšek 2019; 2023a). In English, they partially overlap with non-restrictive attributive clauses (Huddleston, Pullum 2002: 1035). Generally speaking, supplementary clauses structurally²⁹ behave the same as instances of subordination or very similarly, so the presence of backwards anaphora is expected. Despite structural similarities, supplementary clauses have a special type of semi-backwards anaphora as the role of semi-backwards anaphora is not played by the pronoun *to*, but rather by *kar* ‘which’, which is both a conjunction and relative pronoun. *Kar* summarises an entire clause and introduces a new one, so it is usually in a non-initial position, but is sporadically also embedded, in which case it functions as a semi-backwards anaphor. In this type, semi-backwards anaphora has a strong conjunctive role. Similar constructions, but without relative conjunction, can be found in SECTION 3.4.

(42)

[Šprint	bo] _{i/2} ,	kar _i	že	dolgo	vem,
sprint	AUX.FUT	which.nom.sg.n.rel	already	long.adv	know.ipfv.prs.1sg
[v	klasični	tehniki] _{i/2} .			
in	classic.loc.f.sg	technique.loc.sg			

‘As I’ve known for a long time, the sprint will be done in classic technique.’

(43)

[Malček	nam	je] _{i/2} ,	kar _i	smo	na	teh
little boy.nom.sg	I.dat.pl	AUX.PST	which.nom.sg.n.rel	AUX.PST	on	this.loc.pl
straneh	že	zapisali,	[resnično	zlezel	pod	kožo] _{i/2} .
page.loc.pl	already	write.lpt.m.pl	really	crawl.lpt.m.sg	under	skin.acc.sg

‘As we’ve already said on these pages, the little car really grew on us.’

6 COMBINATION OF BACKWARDS ANAPHORA AND BACKWARDS ELLIPSIS

In the example (44), there are two backwards anaphors and two backwards ellipses. This example is the title of an article and in the next two sentences the participants are expressed by nouns. In this example, three participants are

²⁹ The differences are mainly semantic and significant enough for these elements to form their own group.

expressed by one backwards anaphora and two backwards ellipses, making it difficult to understand without further context.³⁰

(44)

Po 24 letih zakona jo_i je Ø_j prevaral, tega
 after 24 year.gen.pl marriage.gen.sg she.acc.sg AUX.PST cheat.lpt.m.sg this.gen.n.sg
mu_j Ø_k nikoli ni odpustila.³¹
 he.dat.sg never AUX.PST.NEG forgive.lpt.f.sg.

‘After 24 years of marriage, he cheated on her and she never forgave him.’

Slovenian and translated further context:

[Rowan Atkinson]_j se je leta 2015 po 24 letih zakona ločil od [svoje žene Sunetre Sastry]_i, s katero ima dva otroka: 30-letnega sina Benjamina in hčerko Lily_k. Lily_k nikoli ni odpustila očetu zaradi njegove prevare mame.

[Rowan Atkinson]_j divorced his wife of 24 years, [Sunetra Sastry]_i, in 2015, with whom he has two children, 30-year-old son Benjamin and daughter Lily_k. Lily_k has never forgiven her father for cheating on her mother.

The example (45) is an example of a backwards ellipsis and a backwards anaphora in the same sentence. The backwards anaphora is also present in the first clause of the second sentence, and the antecedent is only present in the third sentence.

(45)

Vodstvo je_i ni želelo videti, toda vseeno je Ø_i
 leadership.nom.sg her.gen.sg AUX.PST.NEG want.lpt.n.sg see.ipfv.inf but anyway AUX.PST
postala izbranka ljudskih src.
 become.lpt.f.sg favourite.nom.sg people.adj.gen.n.pl heart.gen.pl

‘The leadership did not want to see her, but she became the favourite of the people’s hearts anyway.’

Slovenian and translated further context:

Na valu vsesplošnega antisemitizma ji_i je predsednik želel preprečiti nastop na Evroviziji. Toda [Eden Golan]_i je prišla in postala ljubljenska evropskega občinstva.

³⁰ That is one of the reasons why we quote it here.

³¹ This example also includes ‘regular’ anaphors, but they are not coindexed.

On the wave of widespread anti-Semitism, the President wanted to prevent her from performing at Eurovision. But Eden Golan came and became the darling of the European public.

7 CONCLUSION

Backwards anaphora and backwards ellipsis have proved to be important but rather rare text-building procedures. The claim that they are impossible in certain relationships, particularly coordination, has proved false; it is better to say they are rare in such relationships. A larger quantity of corpus materials has enabled me to find even such rarer examples. Backwards anaphora and backwards ellipsis have been shown to be one of the indicators of hierarchy between clauses: in coordination, therefore, the chances of forming backwards anaphora and backwards ellipsis are almost zero, whereas in subordination they are more frequent and sometimes even predominate.

ABBREVIATIONS

ACC – accusative; ADJ – adjective; ADV – adverb; CMPR – comparative; COND – conditional mood; DAT – dative; DU – dual; F – feminine gender; FUT – future tense; GEN – genitive; IMP – imperative mood; IND – indicative mood; INF – infinitive; INS – instrumental; IPFV – imperfective aspect; LOC – locative; LPT – L participle; M – masculine gender; N – neuter gender; NEG – negation; NOM – nominative case; PFV – perfective aspect; PL – plural; PRS – present; REFL – reflexive pronoun; REL – relative; SG – singular; SUP – superlative; 1 – first person; 2 – second person; 3 – third person

REFERENCES

- Fran*: Slovarji Inštituta za slovenski jezik Franba Ramovša. www.fran.si
- Gigafida 2.0: Corpus of Written Standard Slovene*. <https://viri.cjvt.si/gigafida/>
- Gigafida corpus in the NoSketch Engine tool: https://www.clarin.si/noske/run.cgi/first_form?corpname=gfida20_dedup;align=
- Belaj, Branimir, Tanacković Faletar, Goran. 2020. *Kognitivna gramatika hrvatskoga jezika. Sintaksa složene rečenice*. Zagreb: Disput.
- De Beaugrande, Robert Alain, Dressler, Wolfgang Ulrich. 1992. *Uvod v besediloslovje*. Ljubljana: Park.
- Gabrovšek, Dejan. 2019. Tipologija nestavčnočlenskih nematičnih dopolnil. *Jezikoslovni zapiski* 25/2: 83-96. <https://doi.org/10.3986/JZ.25.2.5>

- Gabrovšek, Dejan. 2021. Termina priredje in podredje z vidika stopenj odvisnosti. In: N. Kočnik et al. (eds.). *9. simpozij mladih slavistov. Mejniki in prelomnice v slovanskih jezikih in literaturah*. Ljubljana: Študentska sekcija Zveze društev Slavistično društvo Slovenije. 259-272.
- Gabrovšek, Dejan. 2023a. *Slovenska zložena poved z vidika stopenj odvisnosti*. PhD Thesis. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Gabrovšek, Dejan. 2023b. Vloga odnosnice v slovenski podredno zloženi povedi. *Jezik in slovstvo* 68/3: 43-55. <https://doi.org/10.4312/jis.68.3.43-55>
- Gabrovšek, Dejan. 2024. Criteria for Coordination in Slovenian Multiclausal Sentences. *Slavistica Vilnensis* 69/1: 67-86. [https://doi.org/10.15388/SlavViln.2024.69\(1\).5](https://doi.org/10.15388/SlavViln.2024.69(1).5)
- Gabrovšek, Dejan, Žele, Andreja. 2019. Tipologija stavčnočlenskih odvisnikov v slovenščini. *Slavistična revija* 67/3: 487-507. <https://srl.si/ojs/srl/article/view/2019-3-1-5>
- Ha, Seungwan. 2008. Backwards Ellipsis is Right Node Raising. *Berkeley Linguistics Society* 34/1: 121-132. <https://doi.org/10.3765/bls.v34i1.3562>
- Halliday, Michael, Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.
- Haspelmath, Martin. 2004. Coordinating constructions: An overview. In: M. Haspelmath (ed.). *Coordinating constructions*. Amsterdam: Benjamins. 3-39.
- Huang, Yan. 2006. Anaphora, Cataphora, Exophora, Logophoricity. *Encyclopedia of Language & Linguistics*. Reading: University of Reading. 231-237.
- Huddleston, Rodney, Pullum, Geoffrey (eds.). 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Jelovšek, Alenka. 2011. Neosebna raba oblike ono v jeziku slovenskih protestantskih piscev 16. stoletja. *Slavistična revija* 59/4: 415-435. https://srl.si/ojs/srl/article/view/COBISS_ID-33760557
- Jelovšek, Alenka. 2024. *Osebnostni zaimki in nastajajoča slovenska knjižna norma*. Ljubljana: Založba ZRC. <https://doi.org/10.3986/9789610507628>
- Kazanina, Nina, Phillips, Colin. 2010. Differential effects of constraints in the processing of Russian cataphora. *The Quarterly Journal of Experimental Psychology*. 63/2: 371-400.
- Kehler, Andrew. 2018. Ellipsis and Discourse. In: J. Craenenbroeck, T. Temmerman (eds.). *The Oxford Handbook of Ellipsis*. Oxford: Oxford University Press. 314-341.
- Mihaljević, Milan. 2024. Coordination and Subordination in Slavic Languages. In: D. Šipka, W. Browne (eds.). *The Cambridge Handbook of Slavic Linguistics*. Cambridge: Cambridge University Press. 304-337.
- Mittwoch, Anita. 1983. Backward anaphora and discourse structure. *Journal of Pragmatics* 7: 129-139.
- Orešnik, Janez. 1992. *Udeleženske vloge v slovenščini*. Ljubljana: SAZU.
- Pavlič, Matic, Stepanov, Arthur. 2023. Number mismatch effect and processing cataphora in a pro-drop language: The case of Slovenian. In: *Advances in formal Slavic linguistics 2021*. Language Science Press. 293-314. <https://doi.org/10.5281/zenodo.10123649>
- Petr, Jan, et al. 1987. *Mluvnice češtiny 3*. Prague: Academia.

- Ross, John Robert. 1967. *Constraints on variables in syntax*. PhD Thesis. Cambridge: Massachusetts Institute of Technology.
- Quirk, Randolph, et al. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.
- Trnavac, Radoslava, Taboada, Maite. 2016. Cataphora, backgrounding and accessibility in discourse. *Journal of Pragmatics* 93: 68-84. doi.org/10.1016/j.pragma.2015.12.008
- Vičar, Branislava. 2011. *Parenteza v novinarskem in parlamentarnem diskurzu*. Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti, Filozofska fakulteta.
- Žele, Andreja. 2013. *Slovenska besedilna skladnja z jezikovnosistemskega vidika – temeljni pojmi*. Self-published.
- Žele, Andreja. 2016. Odvisniki v slovenščini: vsebinski odvisniki in nepravi prislovnodoločilni odvisniki. *Slavistična revija* 64/2: 81-94. https://srl.si/ojs/srl/article/view/COBISS_ID-61135714
- Žele, Andreja. 2017. Razmerje osebek proti osebkov odvisnik v slovenskih povedih. *Slavistična revija* 65/1: 81-97. https://srl.si/ojs/srl/article/view/COBISS_ID-63781986
- Žele, Andreja, Krajnc Ivič, Mira. 2020. *Sodobna slovenska skladnja: diskurzni in slovnični vidik*. Maribor: Univerzitetna založba Univerze.

Received November 2023, accepted January 2024.

Prispelo novembra 2023, sprejeto januarja 2024.

ACKNOWLEDGEMENTS

This article has been produced as part of the P6-0038 programme, The Slovenian Language in Synchronic and Diachronic Development, financed by the Slovenian Research and Innovation Agency.

SUMMARY

BACKWARDS ANAPHORA AND BACKWARDS ELLIPSIS IN SLOVENIAN MULTI-CLAUSE SENTENCES

The article discusses backwards anaphora and backwards ellipsis in Slovenian multi-clause sentences. What both phenomena have in common is that a pronoun (backwards anaphor) or ellipsis (backwards ellipsis) is followed by the lexical element (especially noun or verb) that the pronoun or ellipsis refers to and not the other way round, unlike in anaphora and ellipsis, which are more common ways of text building. The second part of the article first examines all types of backwards anaphora and then of backwards ellipsis that are possible in Slovenian multi-clause sentences,

using corpus examples. Backwards anaphora is possible in all types of multi-clause sentences and is the rarest, only sporadic, in coordination. The most common role it plays is that of a correlative. In most types, backwards anaphora has a strong linking role. Backwards ellipsis is possible in subordination and coordination but in no other types of Slovenian multi-clause sentences; in coordination, it appears more often than backwards anaphora and has a strong semantic extension role. A special type of backwards ellipsis joins two subordinate clauses into one. A semi-backwards anaphora and semi-backwards ellipsis are also shown as a special type of a backwards anaphora and backwards ellipsis.

VZVRATNA ANAFORA IN VZVRATNA ELIPSA V SLOVENSKI VEČSTAVČNI POVEDI

Članek se ukvarja z vzvratno anaforo in vzvratno elipso v slovenski večstavčni povedi. Njuna skupna značilnost je, da zaimku (vzvratni anafori) oziroma elipsi (vzvratni elipsi) sledi polnopomenski element (zlasti samostalnik, glagol), na katerega se zaimek ali elipsa nanaša, in ne obratno, kot to velja za anaforo in elipso, ki sta pogostejša načina tvorbe besedila. Drugi del članka na korpusnih zgledih proučuje vse tipe najprej vzvratne anafore in nato še vzvratne elipse, ki so v slovenski večstavčni povedi možni. Vzvrtna anafora je možna v vseh tipih večstavčne povedi, najredkejša, zgolj sporadična, pa je v priredjih. Najpogostejša je v vlogi odnosnice. V večini tipov ima vzvratna anafora močno povezovalno vlogo. Vzvrtna elipsa je možna v podredjih in priredjih, v ostalih tipih slovenske večstavčne povedi pa ne, s tem da se v priredjih pojavlja pogosteje kot vzvratna anafora in ima tudi močno pomenotvorno vlogo. Poseben tip vzvratne elipse je povezava dveh podrednih stavkov v enega. Kot poseben tip vzvratne anafore in vzvratne elipse sta obravnavni tudi polvzvratna anafora in polvzvratna elipsa.

Danko Šipka (ORCID: 0000-0002-9144-8155)

Arizona State University, Arizona, USA

danko.sipka@asu.edu

Wayles Browne (ORCID 0000-0001-5843-6497)

Cornell University, New York, USA

ewb2@cornell.edu

DOI: <https://doi.org/10.3986/16.1.06>

SLAVIC REFERENCE WORKS AND THE REPRESENTATION OF SLOVENE: A CASE STUDY

The present paper uses the example of the Cambridge Handbook of Slavic Linguistics to elucidate the process of documenting Slavic languages in reference works: the rationale for the project and its contextualization amid other reference works for Slavic languages; the execution of the project, from the selection of the topics and authors to providing editors' feedback to refereeing the submissions using external reviewers, and the content of the volume. Special attention is devoted to how the Slovene language is represented in this volume.

KEYWORDS: reference works, Slavic studies, Slovene language, Slavic linguistics, Slavic languages

Avtorji se na primeru projekta Cambridge Handbook of Slavic Linguistics ukvarjajo s procesom dokumentiranja slovanskih jezikov v referenčnih delih: utemeljitvijo projekta in njegovo umeščenostjo med druga referenčna dela s področja slovanskih jezikov; izbiro tem in avtorjev, posredovanjem povratnih informacij urednikov in recenziranjem prispevkov z uporabo zunanjih recenzentov in vsebino zvezka. Posebna pozornost je namenjena predstavitvi zastopanosti slovenskega jezika v tem zvezku.

KLJUČNE BESEDE: referenčna dela, slavistika, slovenski jezik, slovansko jezikoslovje, slovanski jeziki

INTRODUCTION

The present paper discusses the process of documenting Slavic linguistics in reference works with special emphasis on the representation of the Slovene language in that process. The case study used here is the *Cambridge Handbook*

of *Slavic Linguistics* (Šipka, Browne 2024). We will first address the rationale for the project and contextualize it amid other reference works for Slavic languages. We will then proceed to discuss the execution of the project, from the selection of the topics and authors, to providing editors' feedback, to refereeing the submissions using external reviewers. Next, we will address the content of the volume. Finally, we will show how the Slovene language is represented in this volume and more broadly in similar reference works.

CONTEXTUALIZATION

At the time this volume (Šipka, Browne 2024) was conceived, there were various other general surveys of Slavic languages or linguistics, but a handbook of Slavic linguistics, a compact book that could be used in various scenarios in teaching and research, was conspicuously absent. An account of those related resources that are still most relevant and useful looks as follows. First, there are two general introductions to Slavic languages: Sussex and Cubberley (2006) and Comrie and Corbett (1993). However, these are introductions to the languages per se, not to linguistic research about them. Second, there is an unfolding online *Encyclopedia of Slavic Languages and Linguistics* (Greenberg 2022), which is a large encyclopedic work rather than a handbook. Then, there are collected volumes that offer a much narrower list of topics in Slavic linguistics, for example: *Aspects of Slavic Linguistics: Formal Grammar, Lexicon and Communication* (Mueller-Reichau, Guhl 2017), which actually contains papers by University of Leipzig linguists and their friends, reflecting just their own interests, or *New Insights into Slavic Linguistics* (Witkoś, Jaworski 2014), which is a conference volume. There is, furthermore, *Die slavischen Sprachen: ein internationales Handbuch zu ihrer Struktur, ihrer Geschichte und ihrer Erforschung* (Kempgen et al. 2009) which is very thorough, but most texts are in German, and, as such, not accessible to most English speakers. It is furthermore encyclopedic in its nature, so it is meant to be a library resource, rather than something used in classes and as a handbook. The present handbook thus complements existing resources. It is worth noting that the time of its publication roughly coincides with two related complementary offerings – *The Balkan Languages* (Friedman, Joseph 2025), a survey of Slavic and non-Slavic languages of the Balkan Sprachbund ('Linguistic League'), and *Water, Whiskey, and Vodka: A Story of Slavic Languages* (Šipka 2023), an introduction to Slavic languages for a general audience.

The present volume is a handbook. The Oxford English Dictionary (OUP 2023: s.v.) defines handbook as follows:

Originally: a book small enough to be easily portable and intended to be kept close to hand, typically one containing a collection of passages important for reference or a compendium of information on a particular subject, esp. a book of religious instruction (now historical). Later also more generally: any book (usually but not necessarily concise) giving information such as facts on a particular subject, guidance in some art or occupation, instructions for operating a machine, or information for tourists.

In our case, the relevant part of the definition is that it is a book giving information such as facts on a particular subject, a book that is concise. The concise nature is apparent even from the term, calqued from Latin, *manuāle* with the same meaning, ultimately from *manuālis* ‘suitable to be held by hand’. The subject at hand was Slavic linguistics. The intention then was to provide a succinct compendium of main topics in Slavic linguistics. This orientation is what differentiates the present volume from all aforementioned compendia, which focus on Slavic languages rather than linguistics or strive for comprehensiveness rather than succinctness. With that in mind, we will now proceed to discussing the process of compiling this volume.

EXECUTION

The Cambridge Handbook of Slavic Linguistics sailed through the turbulent waters of global events, between the Scylla of COVID-19 and the Charybdis of the War in Ukraine. It now reaches its destination primarily due to the resilience and collegiality of all its collaborators, starting from the authors of its 35 chapters, through the graduate and undergraduate student researchers who helped with various aspects of the project, to the colleagues who served as reviewers for each chapter, selflessly contributing their time and expertise.

The Cambridge Handbook of Slavic Linguistics seeks to provide a systematic review of relevant topics and research about them in Slavic linguistics. The review is approach-neutral and involves synchronic and diachronic perspectives. The goal of each chapter is to identify and review the following: (a) the linguistic features pertinent to Slavic languages, (b) the development of these features from Proto-Slavic to the present-day Slavic languages (to the degree

appropriate for the topic of the chapter), (c) the main findings in historical and ongoing research devoted to these features, and (d) a summary of what the state of the art in the field is and what the directions of further research will be.

While remaining accessible to a broad circle of scholars and students in the fields of linguistics and Slavic studies, the present volume caters in particular to the following three readerships. First, it offers a review of main areas of inquiry in Slavic studies to the current and prospective students of Slavic linguistics. The brisk introductions to the field provided in each chapter are thus meant to be teasers that would help these students to select the field or fields of their specialization. Second, nowadays Slavic linguists typically specialize in one or several rather narrow areas of inquiry. The chapters discussing the fields other than one's own offer to these scholars an accessible introduction and a chance to broaden their horizons in Slavic studies. Finally, in contrast to the previous group, non-Slavic linguists may be interested in the chapters of their specialization. For example, a scholar of inflection in Baltic or Germanic languages may want to get some introductory information about that field in Slavic languages.

Given its intent and target readership, chapters are approach-neutral. Similarly, they deploy broadly known terminology. This makes the text accessible to every student of Slavic studies and linguistics, no matter what theoretical background they may have. Just like any field of human inquiry, the field of Slavic linguistics is multifaceted, with lines of investigation meandering through countless problems and topics. This made the task of selecting the topics to be included in a volume of a limited length difficult if not impossible. Cognizant of the aforementioned complexity of the problem, we attempted to include the most prominent problems and topics that are of relevance for the intended readership of this volume. Needless to say, other editors might have selected a different set of topics.

Authors for each chapter have been selected, without fear or favor, as established experts in their particular fields. It is important that in addition to inclusivity in the subject matter (which comes with the territory) the volume proves to be temporally and geographically inclusive. Thus it includes early, mid, and late career scholars of Slavic studies. In a way, it gives insight into the past, present, and future of Slavic linguistics. The range of places from which the authors come testifies to the latter. While, as is usual in publications in English, authors from countries like the United States, Canada, and

the United Kingdom constitute the core of the contributors, there are authors from Germany and Austria, where Slavic studies are exceptionally strong, but also from other European countries, such as Belgium and Spain, and from Japan. Last, but definitely not least, there are authors from the Slavic-speaking countries – from Croatia, Czech Republic, Montenegro, Poland, Russia, Slovenia, and Serbia. This is particularly valuable given that, despite recent rapprochement processes, research traditions in English-speaking countries and those around the Slavic world still seem to live separate lives with some notable exceptions. For example, there is a prominent research group at Tromsø, Norway, led by an American, Laura Janda, and a Norwegian, Tore Nessel. Various exchange programs, most notably Fulbright, and more recently Erasmus Plus, have also been instrumental in cross-pollination of research. One should furthermore mention regular meetings that have brought together scholars internationally on a regular basis, e.g., the Biennial Balkan and South Slavic Linguistics, Literature and Folklore; the annual meetings of the Slavic Linguistics Society, which alternate between sides of the Atlantic, as well as the smaller, but influential SCLA conferences (Slavic Cognitive Linguistics Association). There were also individual scholars, such as Joseph Schallert, one author in this volume, who collaborated with the Moscow Accentological School. All that being said, mainstream research still remains largely divided by the Atlantic Ocean.

The writing and production process looked as follows. First, the authors for the chapters were selected and confirmed. The next step involved soliciting 500-word abstracts from the authors. The authors were asked to address main issues of their respective topics, provide relevant research in the field, and, where relevant outline historical developments. Both editors provided feedback to the authors and asked them to commence with designing chapters with that feedback in mind.

Once the first draft of each chapter had been received, the two editors provided their feedback, and asked the authors to incorporate it in a second draft. The new draft was then distributed to at least two external reviewers (many chapters received three and even four external reviews). The resulting feedback was reviewed by the editors and *mutatis mutandis* forwarded to the authors, who then produced the third draft. Finally, a professional editor from Cambridge University Press provided her feedback, which was reviewed by the editors. The editors then shared this feedback. The fourth and final draft

was reviewed by the authors in the production process. One can see that each chapter has undergone a most rigorous procedure of multiple checks and balances between the authors' and editors' visions. One should also note the most collegial atmosphere around this project and the reviewers' selfless contribution of time and expertise. Indeed, professional courtesy and friendliness within this project was exemplary and we hope for such an atmosphere in all further projects of this kind.

It is particularly worth noting that in this intricate and time-consuming process we have enjoyed support of various institutions – from multiple undergraduate student research grants of the Melikian Center at Arizona State University, to funds from Cornell University, to an advance from Cambridge University Press. A hallmark of this project, perhaps a formula to be replicated in similar undertakings, was inclusion of undergraduate and graduate students in the project activities at various stages. McKenna Kellar, an undergraduate student at the time, was instrumental in corresponding with the authors in the initial stages when they were selected. Mary Murphy, another undergraduate, was in charge of corresponding with reviewers and authors after they had been selected. Gina Scarpete Walters, a graduate student, created the index for the volume. This was a win-win solution. The students were included in the research process as a kind of service learning, and became acquainted with both the landscape of Slavic linguistics and with important stages in preparing linguistic publications. The editors received tangible help with some time-consuming activities in the process of producing this manual. The process we have outlined has led to the volume of the content that will be discussed next.

STRUCTURE

The volume has the following structure. The first six chapters are devoted to phonology. Draga Zec, Cornell University, discusses *Word Stress*. Irena Sawicka (Copernicus University, Toruń) presents *Vocalism*, and Peter Jurgec (University of Toronto) discusses *Consonantism*. Alexei Kochetov of the same University is next with his analysis of the *Syllable Structure*. Next, Darya Kavitskaya, of University of California Berkeley, addresses *Phonologically Conditioned Alternations*. The final chapter in this section, penned by Radek Šimík (Charles University, Prague), is devoted to *Prosodic Reflexes of Information Structure*.

The next section comprises four chapters focusing on inflectional morphology and two exploring lexical morphology. Marek Majer (University of Łódź) discusses *Declensions*, while Ian Press (University of St. Andrews) presents *Conjugations*. Hagen Pitsch (Leipzig) writes about *Tense and Mood Forms*. In a final chapter, Stephen Dickey (University of Kansas) explores *Aspect in Verbs*. In the first chapter devoted to lexical morphology, Frank Y. Gladney (University of Illinois Champaign-Urbana) discusses *Lexical Derivation*. The section on morphology is concluded by the chapter on *Lexical Composition* written by Mate Kapović (University of Zagreb).

A cohort of chapters treating syntax is next. Jana Willer-Gold (University College London) discusses *Agreement*. In the next chapter, Barbara Citko (University of Washington) analyses *Wh-Dependencies*. Milan Mihaljević (University of Zagreb) is next with his presentation of *Coordination and Subordination*. In the following chapter, Steven Franks (Indiana University) discusses *Numerals and Quantity Expressions*. Then, there is a chapter by Franc Lanko Marušič, Petra Mišmaš, and Rok Žaucer (University of Nova Gorica) about the *Placement and Ordering of the Enclitics*. Nerea Madariaga (University of the Basque Country) analyses *Secondary Predication* in the next chapter. Luka Szucsich (Humboldt University, Berlin) is next with his discussion of *Polarity*. Then Alina Israeli (American University) presents *Null Subjects*. Jasmina Milićević (Dalhousie University) presents *Voice* in the next chapter. Finally, Elena Titov (University College London) discusses *Morphosyntactic Reflexes of Information Structure*.

The next section presents three chapters devoted to the lexicon. Valentina Apresjan (Nazarbaev University/Dartmouth College) and Alexei Shmelev (Moscow State Pedagogical University) discuss the *Structure of the Lexicon*. Rajna Dragičević (University of Belgrade) and Milan Ivanović (University of Montenegro) analyze *Lexical Semantics*. Dragičević and Danko Šipka (Arizona State University) are next with the presentation of *Lexical Borrowing*.

The next section includes chapters addressing sociolinguistics, broadly understood, and geographical approaches. Serge Sharoff (University of Leeds) and Nenad Ivanović (Institute for the Serbian Language of SASA) discuss *Sociolinguistic Variation*. Danko Šipka (Arizona State University) and Mladen Uhlik (University of Ljubljana, ZRC SAZU) consider *False Cognates*. Joseph Schallert (University of Toronto) is next with his presentation of *Dialectal Fragmentation*. In the following chapter Dieter Stern (Ghent University)

discusses *Language Contacts*. Next, Motoki Nomachi (Hokkaido University, Sapporo) addresses *Slavic Micro-Languages*. Oksana Laleko (State University of New York, New Paltz) takes up *Heritage Language Forms*. There follow two chapters on epilinguistic issues of relevance in Slavic languages, both penned by Daniel Bunčić (University of Cologne). The first discusses *Scripts*, the second *Orthographies*.

The final section is devoted to prominent applied linguistic fields in Slavic linguistics. In the first chapter, Tania Ivanova-Sullivan (University of California, Los Angeles) and Irina A. Sekerina (City University of New York) discuss *Psycholinguistics and Language Acquisition*. In the second, Tomaž Erjavec (Jožef Štefan Institute, Ljubljana) addresses *Natural Language Processing*.

It is of particular importance that the coverage of the topics features a balance between various fields of Slavic linguistics. Thus, for example, there are syntactic and phonological topics that, since the Chomskian turn, have dominated Slavic linguistics in North America, but there are also topics in lexicology and morphology, which is traditionally strong on the other side of the Pond. Furthermore, in addition to more theoretically-minded topics (for example those addressing information structure or psycholinguistics), there are those that tackle more applied issues (e.g., computational linguistics). One can say that the choice of topics represented in this volume displays the rich variety of topics and approaches in current Slavic linguistics.

The volume with the contents as outlined above has already received accolades from many notable colleagues in the field. As the editors of this volume, we can only hope that the following words by Lenore Grenoble of the University of Chicago, featured on the publisher's web page of this Handbook, will become prophetic: "This monumental volume brings together cutting-edge research in the field of Slavic linguistics by a collection of leading scholars, tracing both the historical foundations and providing state-of-the art theoretical contributions to the field. More than a handbook, it sets the stage for future directions in Slavic linguistics."

SLOVENE REPRESENTATION

The Slovene language is represented very well in this and various other complementary publications. In the two volumes devoted to Slavic languages, Sussex and Cubberley (2006) and Comrie and Corbett (1993), Slovene is considered

one of 13 major Slavic languages, and is covered quite well. Similarly, in Šipka (2023), Slovene is mentioned on 31 (out of 200 pages), there is a section about the development of literacy in Slovenia and the index testifies to references to various Slovene men of letters, from Trubar to Prešeren to Ahačič. The coverage is equally strong in the Brill encyclopedia (Greenberg 2022-), which comes as no surprise given that the editor-in-chief is a Slovenianist.

The present volume continues the tradition. Slovene is amply represented. The index mentions the name of the language 153 times. The abbreviation Sln. (used for examples from Slovene) has a frequency of 150. In addition, some topics that are exemplified primarily using Slovene data are well represented. For example, “number, dual” is mentioned 12 times. The representation of Slovene is also well-documented in the appropriate article in Greenberg (2022-): <https://bibliographies.brill.com/BSLO/>. One should not understate the role of the Seminar of the Slovene Language, Literature, and Culture (Seminar slovenskega jezika, literature in kulture, annual since 1965) in making Slovene linguistic issues prominent in a broader research community. Multiple authors of this volume and its editors have participated in this seminar.

Last but not least, Slovenia is embodied by Slovene authors. There are five authors from Slovenia (Erjavec, Marušič, Mišmaš, Uhlik, and Žaucer) and one Slovene from Canada (Jurgec).

REFERENCES

- Corbett, Greville, Comrie, Bernard (eds.). 1993. *The Slavonic Languages*. London: Routledge. <https://doi.org/10.4324/9781136861376>
- Friedman, Victor, Joseph, Brian. 2025. *The Balkan Languages* (Cambridge Language Surveys series). Cambridge: Cambridge University Press.
- Greenberg, Marc (ed.). 2022-. *Encyclopedia of Slavic Languages and Linguistics Online*. Leiden: Brill. <https://brill.com/view/db/eslo>
- Kempgen, Sebastian, Gutschmidt, Karl, Berger, Tilman, Kosta, Peter (eds.). 2009-. *Die slavischen Sprachen: ein internationales Handbuch zu ihrer Struktur, ihrer Geschichte und ihrer Erforschung*. Berlin: DeGruyter. <https://doi.org/10.1515/9783110215472>
- Mueller-Reichau, Olav, Guhl, Marcel. 2017. *Aspects of Slavic Linguistics: Formal Grammar, Lexicon and Communication*. Berlin: DeGruyter. <https://doi.org/10.1515/9783110517873>
- Oxford University Press. 2023. Handbook. In *Oxford English dictionary*. <https://www.oed.com> (December 19, 2023).

- Sussex, Ronald, Cubberley, Paul (eds.). 2006. *The Slavic Languages*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486807>
- Šipka, Danko. 2023. *Water, Whiskey, and Vodka: A Story of Slavic Languages*. Washington, DC: Georgetown University Press.
- Šipka, Danko, Browne, Wayles (eds.). 2024. *The Cambridge Handbook of Slavic Linguistics* (Cambridge Handbooks in Language and Linguistics). Cambridge: Cambridge University Press.
- Witkoś, Jacek, Jaworski, Sylwester. 2014. *New Insights into Slavic Linguistics*. Berlin: Peter Lang. <https://doi.org/10.3726/978-3-653-04359-4>

Received December 2023, accepted January 2024.

Prispelo decembra 2023, sprejeto januarja 2024.

SUMMARY

SLAVIC REFERENCE WORKS AND THE REPRESENTATION OF SLOVENE: A CASE STUDY

The present paper uses the example of the *Cambridge Handbook of Slavic Linguistics* to elucidate the process of documenting Slavic languages in reference works. The first section addresses the rationale for the project and contextualizes it amid other reference works for Slavic languages. The second section is devoted to the execution of the project, from the selection of the topics and authors to providing editors' feedback to refereeing the submissions using external reviewers. The third section discusses the content of the volume. Special attention is devoted to how the Slovene language is represented in this volume and more broadly in similar reference works, all of which is the topic of the final section of this paper.

SLAVISTIČNA REFERENČNA DELA IN ZASTOPANOST SLOVENŠČINE: ŠTUDIJA PRIMERA

Avtorji se na primeru projekta *Cambridge Handbook of Slavic Linguistics* ukvarjajo s procesom dokumentiranja slovanskih jezikov v referenčnih delih. Prvi del obravnava utemeljitev projekta in ga umešča med druga referenčna dela slovanskih jezikov. Drugi del je posvečen izvedbi projekta: izbiri tem in avtorjev, posredovanju povratnih informacij urednikov in recenziranju prispevkov z uporabo zunanjih recenzentov. Tretji del obravnava vsebino zvezka, pri čemer je posebna pozornost namenjena temu, kako je v tem zvezku zastopan slovenski jezik in širše v podobnih referenčnih delih.

Domen Krvina (ORCID 0000-0002-2276-1156)

ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
domen.krvina@zrc-sazu.si

Špela Petric Žižić (ORCID 0000-0001-7451-4264)

ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija
spela.petric@zrc-sazu.si

DOI: <https://doi.org/10.3986/16.1.07>

THE RELATION BETWEEN THE COMPOSITION OF CORPORA (GENRE BALANCE AND REPRESENTATIVENESS) AND THEIR RELIABILITY IN COMPILING GENERAL EXPLANATORY DICTIONARY

This paper aims to examine the genre composition of certain Slovenian corpora as sources for lexicographic analysis (especially when compiling dictionaries such as eSSKJ, the general explanatory dictionary), particularly of the largest corpus, Gigafida 2.0 (divided into two sub-corpora: a sub-corpus of non-fiction and literary texts and a sub-corpus of journalistic texts), the Corpus of Slovenian School Texts, the Corpus of Scientific Texts of Contemporary Slovenian, as well as the KRES corpus. We argue that corpora with major discrepancy in the proportions between different text genres used as lexicographic resources do not reflect the proportions between meanings which originate in semantic extension processes. Thus, one of the largest corpora available for Slovene, Gigafida (in both versions, 1.0 and 2.0, updated in 2019), could hardly be regarded as a reference source of data for a general explanatory dictionary. This is because various journalistic texts and web texts are predominant in Gigafida, while the share of non-fiction and literary texts does not exceed 10% in total. We suggest that a corpus should be at least approximately balanced, which could in turn provide its representativeness.

KEYWORDS: Corpora, Dictionaries, Reference corpus, Representativeness, Balance, Meanings Proportion, Lexicology, Lexicography, Slovene

Namen prispevka je proučiti žanrsko sestavo nekaterih slovenskih korpusov kot virov za leksikografsko analizo (zlasti za slovarje, kot je eSSKJ, torej splošni razlagalni slovar), posebej največjega korpusa Gigafida 2.0 (razdeljenega v dva podkorpusa: podkorpus neumetnostnih in leposlovnih besedil ter podkorpus publicistike), Korpusa šolskih besedil slovenskega jezika, Korpusa znanstvenih besedil sodobne slovenščine ter korpusa KRES. V prispevku korpus obravnavamo predvsem kot vir gradiva za izdelavo slovarjev

in sorodnih referenčnih del. Trdimo, da korpusi z večjimi odstopanji v razmerju med različnimi besedilnimi vrstami kot leksikografski viri ne odražajo razmerij med pomeni, ki so rezultat pomenotvornih procesov. Zato bi enega večjih korpusov, ki je na voljo za slovenščino, Gigafido (v obeh različicah, 1.0 in 2.0, posodobljeni leta 2019) le stežka obravnavali kot referenčni vir za splošni razlagalni slovar. V njem namreč prevladujejo različna publicistična besedila in spletna besedila, medtem ko skupni delež neumetnostnih in leposlovnih besedil ne presega 10 %. Poudarjamo, da bi korpus moral biti vsaj približno uravnotežen, kar bi posledično lahko zagotovilo tudi njegovo reprezentativnost.

KLJUČNE BESEDE: korpusi, slovarji, referenčni korpus, reprezentativnost, uravnoteženost, razmerje med pomeni, leksikologija, leksikografija, slovenščina

1 INTRODUCTION

The aim of this paper is to examine the genre composition of certain Slovenian corpora, particularly of the largest corpus, Gigafida (divided into a sub-corpus of non-fiction and literary texts (STVL) and a sub-corpus of journalistic texts (PUBL)) – in its current version Gigafida 2.0, and previous one, 1.0, –, the Corpus of Slovenian School Texts (*Korpus šolskih besedil slovenskega jezika*, KŠBSJ), which was made and is used especially as a source for the School Dictionary of the Slovenian Language (*Šolski slovar slovenskega jezika*), the Corpus of Scientific Texts of Contemporary Slovenian (*Korpus znanstvenih besedil sodobne slovenščine*, KZB), and the KRES corpus, as sources for lexicographic analysis. We aim to demonstrate in practice that corpora with major discrepancy in the proportions between different text genres used as lexicographic resources do not reflect the proportions between meanings within an analysed lexeme which are derived from the knowledge of lexicological theory about meaning development and semantic extension processes (cf. Atkins and Rundell 2008: 130–150, 263–309; Vidovič Muha 2013: 217; Novak 2004; Snoj 2004: 32, 77). This knowledge is (as a rule) implemented in reference lexicographic works, which are based on established lexicographic practice – as also shown in entries in authoritative explanatory dictionaries such as the Dictionary of the Slovenian Standard Language (*Slovar slovenskega knjižnega jezika*; SSKJ). In this regard, we can hardly recognize Gigafida corpus as a reference (albeit readily available) source of data for a general explanatory dictionary. This is because various journalistic texts and web texts (among them forum news comments) are distinctly predominant in Gigafida, whereas the share of non-fiction and literary texts is lower than 10% in total. In KŠBSJ

(much smaller in size and aimed at compiling a school dictionary), on the other hand, non-fiction (especially textbook) texts and texts from children's and young adult literature are predominant, which means it is a polar opposite of sorts to Gigafida in terms of its genre composition.

The markedness of journalistic texts – particularly those which exhibit also advertising characteristics,¹ which are not uncommon in the corpus Gigafida – due to the conative function taking precedence over the referential function (cf. Jakobson 1996) is a well-known fact that has been addressed in a relatively detailed manner in scholarly literature (cf. Korošec 2005). A lexical description based foremost on (recent) journalistic texts goes against general (long-term) semantic extension trends in the lexical system as reflected in descriptions of meaning in existing reference works (dictionaries such as SSKJ). Stylistic labelling relying particularly on journalistic texts (as well as genre labelling, due the lack of other genres) would be highly problematic, too, as the conative function of such texts makes them inherently stylistically marked.

We are interested in how differences in the genre composition affect the usability of a corpus in creating a relevant lexical description; which text genres are more suitable for that purpose (and what the interrelations between these genres are); how the distinct predominance of one genre, especially journalism, can affect the frequency balance of meanings within semantic relations if a lexicographer relies only on a corpus where journalism is the predominant genre, as well as how it affects the perception what is stylistically neutral and what is at least partially marked; how relying on a corpus imbalanced in terms of genre can affect dictionary descriptions; why it would make sense to pursue at least a relatively equal distribution of shares of individual genres in future corpus updates and in compiling new corpora – especially when one is not entirely sure which genre is, could or should be the most representative of the central standard register of language.²

¹ Cf. also Centa Strahovnik (2023: 24–25) on a trend in modern advertising discourse, *zaslužiš si* 'you deserve', and Gregorčič 2023 on the view of J. Habermas, who discusses the asymmetry of the operation of the media in modern society in terms of the limited engagement of users in creating or influencing content. Cf. also Vodičar 2023 about the (real) authority in the digital world, strongly influenced mostly by various marketing strategies.

² Cf. also Górski, Łaziński (2012: 26): Representativeness refers to a reality that exists outside the corpus. Balancing, on the other hand, is taking care to build the corpus in such a way that no component dominates the others at any level [...] The first possible rule of corpus building is not to set any criteria of representativeness, but to concentrate on building as large and diverse a corpus of randomly selected texts as possible [...].

In this context, we draw attention to the term reference corpus, which suggests that a corpus plays (or can play) the role of a reliable reference in terms of data representativeness.³ In this regard, we highlight that a corpus not exhibiting a tendency to be at least approximately balanced, should probably not be called a reference corpus. It is worth emphasising that in this paper we analyse corpora foremost as a source of materials for compiling dictionaries (such as general explanatory dictionary eSSKJ: Dictionary of the Slovenian Standard Language 2016–) and related reference works,⁴ which can indeed serve as references – under the precondition that the materials are balanced.

2 THE METHODOLOGY OF BUILDING A SOURCE OF MATERIALS FOR COMPILING LEXICOGRAPHIC WORKS

This section presents the fundamental methodological starting points for materials, which are based especially on the reflection how to provide as diverse and balanced quality material as possible so that the corpus will be as representative as possible and can thus be effectively used in dictionary compilation. A general explanatory dictionary is the core reference work in these reflections, so this overview starts with findings of Stane Suhadolnik, who led and directed the work on the Dictionary of the Slovenian Standard Language (*Slovar slovenskega knjižnega jezika*; SSKJ), an authoritative explanatory dictionary made in 1970–1991 – a recent, almost contemporary period when corpora were mostly unavailable in electronic form but were certainly judiciously structured to pursue a primary objective: to make a reference dictionary that reflects, as reliably as possible, semantic relations within lexemes and thus provides a comprehensive description of Slovenian lexical system.

2.1 SUHADOLNIK ON THE SSKJ CORPUS

On the material sources constituting the core of the corpus for compiling the dictionary in the early 1960s, Stane Suhadolnik notes in the concept of the SSKJ that this core represents approximately 2,200,000 cards that can be used in compiling a dictionary of the contemporary standard language. Among these,

³ This term is actually used by Suhadolnik (1963: 929) as early as the 1960s to describe a dictionary, referring to what we would term the “reference status” of a reference work.

⁴ “The corpus is not a replacement for any linguistic reference books” (<https://korpus.sk/en/about-corpora/corpora/>).

there were 700,000 reliable cards with data on ordinary words and phrases from classics and selected essayistic and popular scientific works, 400,000 good transcripts from the most recent journalism, from magazines and daily newspapers, over 100,000 cards with terminological data⁵ and 900,000 cards with quotes of rarer words or phrases. All this amounted to a lexical material that made it possible to start a trial draft according to internationally valid standards (Suhadolnik 1963: 929; see also Suhadolnik 1968: 220).

To summarise, in the underlying corpus for making SSKJ, the share of non-fiction and literature was approx. 32%; the share of journalism (in the sense of anything published in recent years) was approx. 18%; the share of terminology was approx. 4.5%, and 41% of the materials were not yet categorised, consisting mostly of words occurring less frequently. The aim of the materials collected in this manner was to comprehensively demonstrate the richness of the standard language of the last half a century: the dictionary should present the vocabulary and language use of the last 60–70 years as they are reflected in the card materials and take into account the vocabulary of the classics of the previous century, modern technical terms in a secondary-school scope, as well as dialectal lexis, colloquial and jargon elements to the extent they are attested in written standard language (Suhadolnik 1968: 220). By design and size, the dictionary should serve its purpose for a number of decades and as objectively as possible present the entire central (i.e. core) vocabulary; it should not excessively fragment and map meanings and their nuances that do not actually exist in general use or represent only an emotional and sociological side of words or professional usage (Suhadolnik 1968: 223).

In a paper on language registers, Suhadolnik highlights the following with regard to dictionary descriptions, in particular:

The central group is the most important for our study [...] This group of living, neutral words is typical for each language and the most interesting to each linguist and stylist but is sadly highly neglected because it is so widely known that nobody notices it, which is the reason even dictionaries do not address it in a satisfactory manner (Suhadolnik, Janežič 1962: 47).

⁵ According to the paper, terminological lexis was selected based on pre-prepared glossaries for each field made in collaboration with experts.

It is thus apparent that Suhadolnik was aware of importance of describing the core vocabulary, first its basic meanings and only then its figurative meanings.

2.2 VIEWS ON BUILDING BALANCED CORPORA

The importance of genre diversity and, as far as possible, balance among the genres is also often highlighted by the compilers of various corpora, among them the compilers of the KRES corpus – this represented the basic motivation to judiciously design and build the KRES corpus as a partial adaptation of the underlying billion-word Gigafida corpus, which is distinctly imbalanced in terms of genre:

KRES is a balanced sub-corpus sampled from Gigafida. It is key for corpora that represent the comprehensive image of a language to be large and diverse in terms of text genre. While Gigafida is such a reference corpus, it would be hard to say it is balanced as 77% of its words come from periodicals (newspapers, magazines) and only slightly over 6% from books (literature, non-fiction), for example. This composition of Gigafida is foremost the result of it including the entire FidaPLUS and everything we obtained anew under copyright contracts. This is why we have planned the 100-million KRES from the outset as Gigafida's balanced sub-corpus. (Erjavec, Logar Berginc 2012; <http://www.korpus-kres.net/Support/About>)

Logar Berginc et al. (2012) present the genre composition of KRES in more detail. The compilers agreed on including 17 million words from literature and 18 million words from non-fiction in KRES. To reach the agreed 18 million words, it sufficed to include 35.72% of all non-fiction texts; this share was obtained through random sampling from each title. Every newspaper out of 53 daily, weekly and free newspapers in the National Readership Survey (NRB) 2010 survey chart were also included in KRES. Magazines contribute 255,271,089 words to Gigafida, but only 20 million in KRES, which is less than 8% of the total. 20-percent share was allocated to web texts in KRES, which amounts to 20 million words, of which 8 million were allocated to texts from news portals, and 12 million to websites of institutions and enterprises (80–81).

As far as the genre composition of Gigafida is concerned, a detailed description is given for each text genre and the difficulty of obtaining such texts is highlighted. The authors note that the shares in the Gigafida taxonomy were ultimately the subjective choice of the corpus compilers. In collecting texts for

Gigafida, the data from the NRB for 2006, 2007, 2008, 2009 and 2010 was used. The most-read were some popular (free) newspapers and magazines.⁶ The compilers also relied upon data from the MOSS survey (Measuring visits to websites), which was commissioned by the Slovenian Advertising Chamber.⁷ In Gigafida the newspaper texts account for more than half, followed by magazines with 21%. Periodicals in total hold a 77-percent share in Gigafida. Books hold a 6-percent share in Gigafida, of which 2 percentage points of words come from literature, and 4 come from non-fiction. The authors point out that the desire to achieve 15–35% of book materials was too optimistic: approximately ten times too little literature and non-fiction was obtained to achieve 20% and 30% of the corpus, respectively. Such a deviation in the final shares can be attributed to two facts: (a) the monthly, weekly or daily production of periodicals is inherently more extensive than book production, and (b) the authors and publishers of literary works as well as everything labelled as non-fiction in the corpus are much more careful in transferring copyright than media outlets are, and approaching each individual author is more time-consuming considering the yield. Gigafida thus includes all materials the copyright was obtained and acquired for. The authors posit that if reference corpora aim to have a higher share of book materials in the future, its makers will have to be more convincing in contacting book publishers and authors; the alternative is a lower share of newspapers and magazines. Works that carry the non-fiction tag are comprised mostly of available secondary- and primary-school textbooks, manuals and guides, i.e. popular scientific and professional texts, with scientific monographs, however, appearing only sporadically. In selecting news websites, the key criterion was the number of visits. The frequency of crawling a particular website was determined intuitively; websites periodically posting listings and news about events were crawled more often, while relatively static websites were crawled less frequently (Logar Berginc et al., 2012: 21–48).

The following paragraphs present the views on how to achieve the representativeness of a corpus in the process of planning, balancing, or compiling various text corpora. Stefanowitsch notes that with large corpora composed of a broader range of web-accessible text their size is the only argument in their

⁶ Such as *Žurnal*, *Nedeljski dnevnik*, *Dobro jutro*, *Slovenske novice* etc. and (magazines) *Lady*, *Ognjišče*, *Motorevija*, *Zdravje*.

⁷ The 10 most visited websites included: 24.ur.com, najdi.si, siol.net, rtvslo.si, bolha.com, zurnal24, avto.net, itis.si, zadovoljna.si and enaa.com.

favour, as their creators and their users must give up pretense that they are dealing with a representative corpus. On the one hand, corpus size correlates with representativeness only to the extent that we take corpus diversity into account. On the other hand, assuming that language structure and use are not infinitely variable, size will correlate with the representativeness of a corpus at least to some extent (Stefanowitsch, 2020: 37–38). This view is shared by *Corpas Pastor* and Seghiri, who also point out that corpus-based studies should rely on the quality and representativeness of each corpus as foundation for producing valid results. This entails deciding on valid external and internal criteria for corpus design and compilation. A basic tenet is that corpus representativeness determines the kinds of research questions that can be addressed and the generalizability of the results obtained. Thus, the representativeness is a crucial point in the creation of a corpus, but also one of the most controversial aspects among specialists. As for the quality of the texts that are included, a system for gauging the quality of digital information through adopting an evaluation protocol should be applied to all the documents – a vast collection of texts itself is usually not sufficient as point is reached when the addition of more documents will not in practice bring anything new to the collection (*Corpas Pastor*, Seghiri, 2010: 111–121). This means that when compiling a corpus, one should be selective in choosing texts, so that the quality or content of the data takes equal or more precedence over issues of quantity. Corpus representativeness can be obtained by establishing coherent limits and carefully selecting textual genres for inclusion. These could be considered as external selection criteria to be established from the outset in order to ensure corpus representativeness and quality (*Corpas Pastor*, Seghiri, 2010: 122–135).

To summarise briefly: the readership factor should not be reduced to the mere question what is most popular.⁸ In compiling a general-purpose corpus, the main principle should be the quality⁹ of obtained texts and their diversity in terms of fields (and thematic diversity within fields). This should be prioritised

⁸ R. L. Górski advocates that a reference corpus should reflect the readership of persons who graduated institutions of tertiary education, because these people read much more than the rest of the society. He also states that nobody knows what is the amount of text in a given newspaper which is read by an average reader (Górski 2008: 122–123).

⁹ This refers both to language quality (grammatical correctness, coherent and clear syntax etc.) as well as to content quality (absence of factual errors in non-fiction texts; also avoiding advertising tendencies wherever possible).

over corpus size itself.¹⁰ A high-quality, genre-diverse, balanced and, as far as possible, representative corpus provides a good basis to describe the core, fundamental vocabulary in its main meanings – something that, as noted by Suhadolnik, is often overlooked.

3 THE COMPOSITION OF ANALYSED TEXT CORPORA

What follows is a presentation of the genre composition of, first, the Slovenian text corpora under consideration (Gigafida 2.0, KRES, KŠBSJ) and then some Slavic ones, especially Central European, which are the closest to the Slovenian language, linguistic and cultural area. We pay particular attention to the proportions between the shares of non-fiction, literature and journalism (which, in certain corpora, also encompasses most of what is defined as “web texts” in terms of their medium).

3.1 GIGAFIDA, KRES, CORPUS OF SLOVENIAN SCHOOL TEXTS

Corpus	Gigafida 2.0	KRES	KŠBSJ	Difference Gigafida:KRES	Difference Gigafida:KŠBSJ
Size [words]	1,1 x 10 ⁹	100 x 10 ⁶	3,9 x 10 ⁶	1 x 10 ⁹	1,096 x 10⁹
Non-fiction	3.8%	18%	44.1%	-14.2 pp	-40.3 pp
Literature	3.5%	17%	50.7%	-13.5 pp	-47.2 pp
Journalistic texts	64.3%	40%	-	24.3 pp	-
Web texts	28%	20%	-	8 pp	-
Pupils' own texts	-		5.2%		-

TABLE 1: Genre distribution of texts in Gigafida 2.0 compared to KRES and KŠBSJ (shares among all texts)

The Corpus of Slovenian School Texts (*Korpus šolskih besedil slovenskega jezika*; KŠBSJ)¹¹ consists mainly of children's and young adult literature and of equally represented subject fields from textbook materials – which is highly relevant both to its primary use, compilation of the School Dictionary of the Slovenian Language (*Šolski slovar slovenskega jezika*; ŠSSJ),¹² and, to some extent, to a semantic analysis of texts in general. Editing work for ŠSSJ has

¹⁰ When compiling a corpus for e.g. natural language processing (and not necessarily lexicographic analysis), its size may be more important than genre balance.

¹¹ On the corpus, see Ledinek et al. 2022.

¹² On the dictionary, see Petric Žižić 2020 and Petric Žižić 2022.

shown that the basic meaning (denotative, i.e. non-figurative meaning) is usually represented well both in terms of frequency and relevant collocations. Such a proportion of meanings in KŠBSJ is influenced by a large share of textbook texts, which ensures that ŠSSJ properly presents terms encountered by pupils in class. Due to the large share of literature, established figurative meanings¹³ are also represented well in and proportionally to the basic meaning. Literature (especially select authors) arguably has an important impact on the semantic competence of pupils¹⁴ – by reading such texts, they expand their own abilities to understand and form figurative meanings with regards to basic ones. Texts produced by pupils themselves have a smaller representation in the corpus,¹⁵ but they do contribute to the presence of certain meanings typical of spoken language.

As will be shown in the SECTION 4, in Gigafida – unlike in KŠBSJ –, the basic meanings of most words, at least as far as it can be judged from the analysed sample (and can be in general supported by lexicographic experience), are often underrepresented in terms of frequency, collocations and syntactic structures. We assume this is due to Gigafida's disproportionate share of texts from particular thematic areas (e.g. sport, police blotter, automotive, food, health) of the journalistic genre, in particular. On the other hand, some meanings that are actually marginal from the point of view of general semantic relations (usually figurative meanings) but are thematically tied to journalistic areas are overrepresented – this can make it seem that the basic meaning is hardly present in use. The same can be said for some seemingly absent terminological meanings students encounter in primary school or secondary school at the latest. We can thus already note that what a corpus with a disproportionate prevalence of journalistic texts reflects is more topical, popular usage (i.e. what is written, read about; what sells) in a particular interval of time. A corpus with the predominance of non-fiction on the other hand, generally – as far as the knowledge of lexicological theory about semantic extension processes

¹³ This is because children's and young adult literature are mostly not particularly experimental in linguistic terms.

¹⁴ The evaluation of the quality of a literary work is to a certain extent intuitive, so there are no uniform criteria. However, distinguished literary works, particularly those catering to a youthful audience, are commonly acknowledged as such when the equilibrium of their tripartite functions – cognitive, ethical, and aesthetic – is maintained (Svetina 2009: 67–68).

¹⁵ Open-access requirements resulted in legal issues, particularly relating to personal data protection (see Ledinek et al. 2022: 134).

and reference lexicographic works which follow such lexicological principles are concerned – reflects more proportionate relations between the basic and derived meanings of a given lexeme as they appear (and remain valid) over a longer interval of time.

3.2 CERTAIN OTHER SLAVIC (CZECH, RUSSIAN, POLISH, SLOVAK) AND ENGLISH (BNC) CORPORA

Corpus	Russian National	Czech National (Syn2020) ¹⁶	Polish National	Slovak National	British National ¹⁷	Average across the 5 corpora	Gigafida 2.0	Difference Gigafida: Average across the 5 corpora
Size [words]	2,1 x 10 ⁹	100 x 10 ⁶	1,8 x 10 ⁹	1,4 x 10 ⁹	100 x 10 ⁶	1.1 x 10 ⁹	1,1 x 10 ⁹	0
Non-fiction	33.5%	33.6%	~30%	~12.2%	~50%	37.2%	3.8%	-33.4 pp
Literature	40.5%	34.7%	~20%	16.8%	16.6%	25.6%	3.5%	-22.1 pp
Journalistic texts	36%	33.3%	~50%	71%	18.4%	41.6%	64.3% (+ web texts 28%)	22.7 pp (50.7 pp)

TABLE 2: Genre distribution of texts in national corpora of Russian, Czech, Polish, Slovak and English compared to Gigafida 2.0 (shares among all texts)

Russian National Corpus:¹⁸

Non-literary texts hold the largest share in this corpus: 59.5% (administrative 4.6%, art and culture 10.6%, history 7.2%, science and technology 9.6%, politics and society 14.2%, journalism 60.5%, everyday life 14.7%). The share of journalism in the corpus as a whole is thus 59.5% × 60.5% ~ 36%. Literary texts hold a large share: 40.5% (documentary prose 8.2%, historical prose 9.4%, fantasy 6.3%, detective stories 5%, hard-to-define prose 56.1%).

¹⁶ This is the largest corpus in the “syn” (synchronic) group with the label “representative” (<https://www.korpus.cz/kontext/corpora/corplist>).

¹⁷ In the context of corpora comparison, the shares of the parts covering the spoken language and miscellaneous are not shown here.

¹⁸ <https://ruscorpora.ru/stats>

Czech National Corpus:¹⁹

Literary texts hold the largest share: 34.7% (short prose 16%, memorial prose, autobiographic prose 12 %, other prose 72%); they are closely followed by non-fiction: 33.6% (administrative 0.01%, popular scientific 39.2%, scientific 20% and scholarly literature 28%); with the share of journalism not far behind: 33.3% (leisure journalism 40%, traditional journalism 60%).

Slovak National Corpus:²⁰

This is a corpus (prim-10.0-public-all) of a relatively small language in which the share of journalism was first around 60% and grew to roughly 70% with upgrades. Journalism thus accounts for 71%, literature for 16.8%, and non-fiction and other texts for 12.2%. A balanced sub-corpus has already been made (prim-10.0-public-vyv), with the genres above represented equally in thirds.

Polish National Corpus:²¹

In this corpus, journalism accounts for around 50% – but, by design, does not exceed this share, as noted by the Górski, Łaziński (2012: 29–31) – followed by non-fiction (also including administrative texts) with roughly 30%, and the share of literature is around 20%.²²

British National Corpus:²³

This corpus, the original version of which emerged around 1990 under the Oxford University Press publisher, has a prevalent, about 50% share of non-fiction (which includes scientific and popular scientific texts as well as other, e.g. administrative, essayistic, religious texts), followed by journalistic texts (newspapers and magazines) with over 18% and literature (divided into prose, poetry and drama) with 16.6%.

¹⁹ <https://www.korpus.cz/kontext/query?corpname=syn2020>; <https://wiki.korpus.cz/doku.php/cnk:syn2020>

²⁰ <https://korpus.sk/en/corpora-and-databases/snc-corpora/publicly-available-snc-corpora/structure-of-the-corpus-prim-10-0/>

²¹ <https://nkjp.pl/poliqarp/>

²² The classics, which are (also) read in schools, encompass the period after 1900 – i.e. the period when the average Polish secondary school graduate does not need a dictionary to read them.

²³ <https://www.english-corpora.org/bnc/>

Calculating the average of fundamental genres across the 5 studied corpora reveals that non-fiction averages around 37.2%; literature amounts to 25.6% (both were also targets of the Gigafida compiler, cf. Logar Berginc et al. 2012: 33), and journalism to 41.6%. Journalism thus mostly has a moderate prevalence (or is balanced with the other genres), with the exception of the British National Corpus, but is not distinctly predominant at the expense of other genres. In fact, it would be worth pursuing such an average or proportion between the shares of fundamental genres in compiling a corpus in the first place (with consideration for the quality of the included texts). For example, the KRES corpus falls short for about 20 percentage points in non-fiction and less than 10 percentage points in literature, whereas in journalism (if not considering web texts, which are also mostly journalistic) it comes quite close.

4 ANALYSIS OF THE INDIVIDUAL LEXEMES IN THE COMPARED CORPORA

This section presents the semantic relations within selected lexemes (mostly everyday, not specialised words) as revealed through an analysis of the corpora:

Gigafida 1.0, divided in two subcorpora – the larger PUBL²⁴ (consisting of texts with tags “časopisi” (newspapers), “revije” (magazines, journals), “internet” (web), “ostalo” (other)) and much smaller STVL (consisting of texts with tags “literatura” (fiction) and “stvarna besedila” (non-fiction) as well as, shared with PUBL, “ostalo” (other)) –, KRES, KŠBSJ and KZB.

The methodology is as follows:

Alongside concordance analysis we use so-called word sketches (Krek, Kilgarriff 2006) with the large PUBL subcorpus of Gigafida 1.0 and comparable context search with the corpus KRES as their larger size makes context search viable. For the subcorpus STVL and the corpora KŠBSJ, KZB – due to their smaller size – we use concordance analysis only. For the lexemes with greater frequencies the calculations are based on random sample of 300 shuffled concordances. The results obtained in such a way serve as an overall lexicographic assessment of the proportions of meanings within the lexeme. The same applies to the calculated percentages of individual meanings within the lexeme in each corpus.

²⁴ Due to overrepresentation (90+%) of the journalistic texts using the entire Gigafida corpus for analysis instead of 90+% share in PUBL subcorpus would not yield much different results.

4.1 PROPORTIONS BETWEEN MEANINGS WITHIN A LEXEME IN DIFFERENT CORPORA

Zadek ‘abdomen, buttocks, rear’: in the word sketch of the larger PUBL subcorpus of Gigafida, the figurative meaning ‘rear of vehicle’ is distinctly predominant, whereas the basic meanings ‘rear part of body in an animal, especially an insect’ and ‘buttocks’ are almost imperceptible (0.2%), occurring in sporadic examples (e. g. *pajkov zadek* ‘spider abdomen’) from popular science journals. They are much more perceptible (29,4%) in the KRES corpus (especially in coordinated structures, such *glavoprjsje in zadek* ‘cephalothorax and abdomen’). In KŠBSJ and STVL sub-corpus, the basic meanings are predominant, but the figurative vehicle-related meaning can also be found (9–10 %). KZB, on the other hand, displays a distinct prevalence (96%) of the basic meaning related to animals, especially insects, though there are some occurrences of the figurative meaning.

Zadetek ‘hit, goal, prize’: in the Gigafida PUBL subcorpus as well as KRES, the meaning of a point in sports is prevalent, whereas a web search hit *iskalni zadetek* (which is now very common) or lottery prize is barely perceptible (0.1–0.6%). KŠBSJ has a relatively equal representation for the meanings of web search hit, lottery prize or sports point, whereas physical hit is not as perceptible, which also applies to KZB, where the web-search hit stands out (96%) – something not completely unexpected taking into account the type of texts included in the corpus. In the STVL sub-corpus, on the other hand, physical hit (e.g. *v oko* ‘to the eye’, *s topom* ‘with a cannon’) is more perceptible (40%), and lottery prize is less perceptible.

Konjič ‘small horse, wheels, horsepower, toy horse’: in Gigafida PUBL subcorpus, there is a distinct prevalence (96%) of the figurative meaning ‘car’ or ‘horsepower’. Also notable is typical of car-related texts expressive use of collocates²⁵, which primarily refer to an animal: *isker* ‘lively’, *rezgetati* ‘to neigh’ (*Voznik s pritiskom stopalke za plin do pločevine mobilizira 207 iskrih konjičev, ki glasno rezgetajo pri 6000 vrtljajih* ‘By putting the pedal to the metal, the driver mobilises 207 lively horses, which neigh loudly at 6000 rpm’), whereas the basic meaning ‘animal’ and the figurative meaning ‘toy’ are perceptible very poorly (3.5%); they are represented much better in KRES (56%). In

²⁵ The by-far most frequent phrase (Gigafida 1.0) *srebrnogrivi konjič* ‘little silver-maned horse’ refers to the proper name of an animated series and is the result of crawling frequently refreshed websites, such as TV listings, so it is irrelevant to the dictionary description. In *lisica* ‘fox’, a similar example is the phrase *zlata lisica* ‘golden fox’.

KŠBSJ, KZB²⁶ and the STVL sub-corpus, the ‘animal’ meaning is predominant (82–86%), with the ‘toy’ meaning also perceptible. Such a difference between Gigafida and all other corpora, where the car-related meaning (and its collocates which should primarily refer to an animal) is practically imperceptible, raises the question whether to even consider this meaning (characteristic only of automotive texts) when drafting a dictionary entry in general explanatory dictionary, such as eSSKJ.

Similar examples include *lisica* ‘fox, handcuff’ (a distinct prevalence of the figurative meaning ‘restraint device’, especially from crime news, in Gigafida PUBL sub-corpus) and *oven* ‘ram, Aries’ (a distinct prevalence of the horoscope meaning in Gigafida PUBL sub-corpus, KRES, even the STVL sub-corpus²⁷).

Predlog ‘proposal, suggestion, preposition’: in Gigafida PUBL sub-corpus as well as KRES, there is a distinct prevalence of political and administrative context, whereas the word is perceptible rather poorly (10% in Gigafida PUBL, 19% in KRES) in everyday contexts (e.g. KŠBSJ: *Odličen predlog! Kar takoj se ga lotiva*. ‘Great suggestion! Let’s do it right away.’) or in the ‘preposition’ meaning (0.3–0,7%). On the other hand, the contexts in KŠBSJ, KZB and the STVL sub-corpus are more balanced (from everyday to more formal ones, including political and administrative), and the grammatical meaning is perceptible (from 3% in STVL to 10% in KŠBSJ, and up to 25% in KZB), too.

Koš ‘basket, bin’: in Gigafida PUBL sub-corpus, the sports meaning is distinctly prevalent (it also has distinctive collocations, e.g. *napolniti/polniti koš*, literally ‘to fill the basket’, i.e. ‘to score points’: *uspešno morajo polniti koš*: *tako z natančnimi meti z razdalje kot s prodori pod koš* ‘they have to be successful at scoring: both by precise distance throws and breakaways’), while the meanings of an instrument to collect rubbish, laundry or carry cargo (12%) are limited mainly to individual phrases (*pletен koš* ‘woven basket’, *koš za smeti/odpadke* ‘rubbish bin’). Those meanings are somewhat more prominent in KRES (32%) and the STVL sub-corpus (24%). The situation in KŠBSJ and KZB is similar to that in KRES and STVL, though the sports meaning only holds a minor share (3–5%).

²⁶ In KŠBSJ as well as KZB a significant portion of the occurrences of this lemma come from (children’s) songs. Such examples show that specialised corpora might serve only as a complementary source when compiling a general explanatory dictionary.

²⁷ This is the result of frequently categorising non-scientific, non-professional and usually fringe fields, such as astrology, as non-fiction.

A similar example is *vilice* ‘fork’: in Gigafida PUBL sub-corpus, there is a distinct prevalence of the meaning referring to vehicle parts or tools (which are also advertised), especially in phrases such as *nihajne vilice* ‘swinging arm’, *(hidravlične) teleskopske vilice* ‘(hydraulic) telescopic fork’, *paletne vilice* ‘pallet fork’; as a utensil, it is used not so much as cutlery in eating (10% in Gigafida PUBL, 34% in KRES), but especially for preparing food, as influenced by numerous texts containing recipes. These technical element-related meanings are less distinct in KRES, KŠBSJ, KZB and STVL, where these meanings have a minor share (4–7%), with the basic meaning of cutlery prevalent – especially for eating and less frequently for preparing food. An example of distinct occurrence in an advertising context is *koža* ‘skin’ (in Gigafida PUBL sub-corpus and mostly KRES as well, there is a prevalence of collocates such as *negovati* ‘to care’, *pomirjati* ‘to soothe’, *vlažiti* ‘to moisturise’, *ščititi* ‘to protect’, *obnavljati* ‘to repair’, *gladiti* ‘to smooth’, *napeti* ‘to tighten’, *učvrstiti* ‘to firm’; *občutljiv* ‘sensitive’, *suh* ‘dry’, *masten* ‘greasy’, *razdražen* ‘irritated’, *trd* ‘hard’, *razpokan* ‘cracked’). This context of *koža* is substantially less (4%) present in the STVL sub-corpus and practically absent from KŠBSJ and KZB.

*Prisegati*_{ipt}/*priseči*_{pt} ‘to swear’: in Gigafida PUBL sub-corpus and mostly KRES as well, the figurative (imperfective, thus present only in *prisegati*) meaning ‘to value, have a very good opinion of; to like to use’ (which, in SSKJ, for example, is only noted in the phraseology section as a non-standard meaning) with collocates such as *kreator* ‘creator’, *ljubitelj* ‘fan’, *zvezdnica* ‘star’, *voditeljica* ‘presenter’, *navdušenec* ‘enthusiast’, *poznavalec* ‘connoisseur’ (e.g. *Zvezdnice prisegajo na najrazličnejše odtenke rdeče šminke, ki odlično pristojijo njihovi polti, barvi las in oblačilom* ‘Stars swear by various shades of red lipstick, which splendidly suit their skin tone, hair colour and clothes) with the preposition *na* (e.g. *na klasiko* ‘by a classic’, *znamko* ‘trademark’, *tradicijo* ‘tradition’, *udobje* ‘comfort’, *kozmetiko* ‘cosmetics’, *eleganco* ‘elegance’, *videz* ‘appearance’, *slog* ‘style’, *lepoto* ‘beauty’) stands out so much (from 68% in KRES up to 91% in Gigafida PUBL) that it seems that *prisegati*, despite its aspectual correlation²⁸ with *priseči*, shares almost no collocates with the latter (only, for example, with the object *ljubezen* ‘love’, *zvestobo* ‘loyalty’; with the subject *predsednik* ‘president’, *domobranec* ‘Home Guard’; *pred bogom*, *pred*

²⁸ It is clear, though, that one-to-one correspondence in all collocations and especially in their frequency cannot be expected as the difference in non-categorical semantic features in an aspectual correlation is close to 0 but not actually 0. Cf. Krvina 2018.

predsednikom ‘before God, before the president’). It seems, for example, that a *predsednica* ‘president [f]’, *premierka* ‘prime minister [f]’, *sodnica* ‘judge [f]’, *vitez* ‘knight’ only *priseže*_{pf} ‘is sworn in [perfective aspect]’, never *prisega*_{ipf} ‘is sworn in [progressive aspect]’. In subcorpus STVL and corpora KŠBSJ, KZB and however, this figurative meaning in *prisegati* is proportionate (19–25%) to the basic meanings ‘to affirm that something is the truth’ and ‘to officially take up a position’.

A similar example is *računati*_{ipf}/*izračunati*_{pf} ‘to calculate, count’, where there is a distinct prevalence (90%) of the figurative (again, imperfective, present only in *računati*) meaning ‘to count on’ (*resno* ‘really’, *potihem* ‘secretly’, *upravičeno* ‘legitimately’, *trdno* ‘reliably’; with the subject *selektor* ‘selector’, *trener* ‘coach’, *strateg* ‘strategist’, *priređitelj* ‘organiser’; *na podporo* ‘on support’, *pomoč* ‘help’, *uvrstitev* ‘placing’, *zmago* ‘win’, *uspeh* ‘success’, *rezultat* ‘result’, *denar* ‘money’, *uslugo* ‘favour’; *igralce* ‘players’, *kupce* ‘buyers’) in Gigafida PUBL sub-corpus and mostly (82%) KRES, too. The basic meaning is represented much more poorly (< 10–18%), so it seems that *računati* ‘to calculate [progressive aspect]’ and *izračunati* ‘to calculate [perfective aspect]’ only share the collocates (with object) *koren* ‘root’, *razdaljo* ‘distance’, *vrednost* ‘value’, *obresti* ‘interest’, *indeks* ‘index’, *porabo* ‘consumption’, *povprečje* ‘average’; (with adverb) *pravilno* ‘correctly’, *približno* ‘approximately’, *natančno* ‘precisely’, whereas the objects *razmerje* ‘ratio’, *koncentracija* ‘concentration’, *površino* ‘area’, *hitrost* ‘velocity’, *oddaljenost* ‘distance’, *verjetnost* ‘probability’; *znesek* ‘amount’, *dohodnino* ‘income tax’ and the adjuncts *po postopku* ‘according to the procedure’, *metodi* ‘method’, *standardu* ‘standard’, *metodologiji* ‘methodology’ etc. only collocate with *izračunati*. The situation is different in KŠBSJ, KZB and STVL, where the basic meaning, with diverse collocations (e.g. *računati ulomke* ‘to calculate fractions’; *merijo dolžine in računajo ploščine* ‘they measure lengths and calculate areas’; *koeficiente povezanosti za pojave s številčnim podatki računamo s pomočjo linearne korelacije* ‘correlation coefficients for phenomena with numerical data are calculated with a linear correlation’), is represented well, and the figurative meaning is proportionate (58–60%) in relation to it, while it is much rarer (13%) in the KŠBSJ.

Čvrst ‘solid, firm’: in Gigafida and mostly KRES as well, the meaning ‘compact, elastic’, which appears in advertising texts, particularly in relation to body care (*koža* ‘skin’, *prsi* ‘breasts’, *zadnjica/ritka* ‘buttocks’, *trebuh* ‘stomach’,

mišice ‘muscles’, *stegna* ‘thighs’, *nohti* ‘nails’, *kosti* ‘bones’), the automotive field (*vzmetenje* ‘suspension’, *karoserija* ‘bodywork’, *podvozje* ‘undercarriage’) and in recipes (*sneg* ‘whipped egg white’, *hruška* ‘pear’, *tofu* ‘tofu’), is prominent (45–55%). The basic or derived meaning ‘full of force, strong, decisive’ appears practically only in texts on sports (*čvrsta obramba* ‘solid defence’: *V napadu igrajo zelo hitro, računam pa, da jih bomo onemogočili z našo čvrsto obrambo* ‘They play a very fast offence, but I’m counting on disabling them with our solid defence’). In KŠBSJ, KZB and STVL, advertising style is imperceptible ($\rightarrow 0$) even in phrases relating to body parts and food, and the basic meanings have a stronger presence, e.g. *bila je še čvrsta in gibčna* ‘she was still vigorous and flexible’, *stisk njegove roke je bil proti pričakovanju čvrst* ‘unexpectedly, his handshake was firm’.

To summarise briefly: search results in the STVL sub-corpus are closer to search results in KŠBSJ and KZB, whereas searching the KRES yields results somewhere between STVL/KŠBSJ/KZB and Gigafida PUBL. In KŠBSJ, KZB and STVL, the basic meaning is always perceptible well, while the representation of derived meanings depends on the inclusion of texts from a particular field in the corpus, but such meanings are mostly perceptible. If a lexicographic description for a general explanatory dictionary were to be made based purely on materials with a distinct prevalence of (mostly) advertising journalistic texts, the presentation of semantic relations within a lexeme would be more or less inadequate. Figurative meanings (e.g. to describe sports activities, automotive and other products, healthy food, body care etc.) would be distinctly prevalent.

5 DISCUSSION

The above analysis shows that Slovene lacks a reference corpus which would serve as a source for comprehensive linguistic research and compiling a general explanatory dictionary, grammar and normative guide based on thorough description of the Slovenian language system.

In defining semantic relations within a lexeme, a certain theoretical basis is needed. We find the knowledge of lexicological theory about meaning development, semantic extension processes (cf. Atkins and Rundell 2008: 130–150, 263–309; Vidovič Muha 2013: 217; Novak 2004; Snoj 2004: 32, 77) to be a solid foundation for defining semantic relations within a lexeme. This knowledge

is (as a rule) implemented in reference lexicographic works which are based on established lexicographic practice. As already noted by Suhadolnik (1968), it seems that a material offering insight into processes of semantic extension (from the basic meaning into figurative meanings) is the point of reference that is desirable in lexicographic description of a lexical system.

It would therefore be worth reflecting on compiling as balanced corpus as possible which would improve the textual proportions.²⁹ It should be sizeable enough and while the share of periodicals (particularly weightier journalistic discussions, interviews etc., if possible) would probably be relatively large, an effort should be made that it is not overrepresented. In preserving such proportions, editorial interventions are desirable, if not necessary, to ensure texts are included according to pre-defined criteria not determined mostly by the accessibility of texts itself.

Well-defined criteria should improve current genre tagging which is often overly simplistic in terms of uniting very different text types (e.g. newspapers, journals, magazine web sites as well as some popular science publications) under a common tag with no further distinction. As far as non-fiction texts are concerned, one possibility would be to follow the example of KŠBSJ in principle: the core of non-fiction texts would consist of reviewed, mainly professional texts at the secondary school level, complemented by scientific texts (such as those collected in the KZB corpus), presumably in fundamental fields of science. Apart from this, popular science texts, instructions, various handbooks etc. should be taken into account.

Literature would be represented by quality Slovenian and foreign (semi-) literary texts. The criteria of quality – which are, as already said, always intuitive to certain extent, but generally prefer distinguished literary works which manage to maintain the equilibrium of cognitive, ethical, and aesthetic function – and readership should be both taken into account. Various lists of prize-winning literature, librarians' and similar lists of recent literature could provide some insight; world classics should probably not be omitted as well. Journalistic texts would also be selected based on the criterion of quality (daily news – reporting, discussions, articles, columns; sections marked by advertising and other texts with marketing patterns should be minimised)

²⁹ This includes taking into account the legal ramifications, i.e. anticipating potential legal issues and devising way to overcome them as effectively as possible (cf. Ledinek et al. 2022: 131–132).

and – regardless of their bigger production and relative ease of access – should not be overrepresented.

Such a balance between individual genres is necessary when compiling general explanatory dictionary, such as eSSKJ, as it contributes to the representativeness of the corpus data – to the extent it can be ensured when one cannot be quite sure about the actual linguistic and language-formation influence (especially in the core of language, i.e. standard language) of a genre (cf. Logar et al. 2023: 88–89). In no case should the size of the corpus take precedence over planning and compiling the corpus in a way described above. On the other hand, an excessive share of topical materials whose linguistic quality can be rather poor, provides more language innovations that are only emerging – these are represented well in the Trendi corpus (cf. Kosem et al. 2023), a valuable source of materials for the Growing Dictionary of the Slovenian Language (*Sprotni slovar slovenskega jezika*), for instance, which actually aims to capture new lexical trends (and possibly innovations).³⁰

Suggestions by respondents to improve the Gigafida 2.0 corpus, obtained in a survey on the use of this corpus conducted in 2021 as part of the RSDO project, also point to the necessity of updating (and probably also compiling new) general corpora, with texts as diverse as possible in terms of genre (Logar et al. 2023: 86, 88–89). The suggestion to increase genre diversity took first place among all suggestions (Logar et al. 2023: 82).³¹

Extensive, billion-word corpora such as Gigafida (from version 1.0 onward) may be suitable for machine rather than manual analysis. However, due to their seeming primary goal to achieve the desired size, their genre composition can be seriously deficient (as noted, for example, by Górski), with a disproportionate prevalence of journalistic texts. Due to their extensive production and accessibility, such texts do enable achieving the goal of a billion-word corpus but they obscure the semantic relations within lexemes in the process (cf. Rundell, Atkins 2013: 1339). This is also reflected in the results provided by word-sketch machine analysis. In addition, due to poorer precision (and a

³⁰ Cf. also Krvina 2022.

³¹ Judging by the overall average of ratings, ~4, the lowest-rated (~3.6) statement “the corpus offers appropriate search options” could point to this as well. The statement probably does not only refer to the range of complex (enough) search options, but very likely also to the quality of materials being searched – a user can interpret an excessive occurrence of particular search results as the inadequacy of search tools or complex search options even though the issue actually lies in the problematic genre composition of materials originating in the already mentioned distinct imbalance.

more sensational nature) of reporting, both linguistic and factual errors can occur, especially in non-peer-reviewed and thematically often relatively peripheral non-fiction texts but especially journalistic texts.³²

If one were to pursue an effort to balance genres (it is impossible to rely only on readership data), this would involve designing the corpus (especially for compiling general explanatory dictionary) in advance so that:

- 1) different or at least fundamental genres (literature, non-fiction, journalism) are represented as equally as possible, despite the fact that some can be accessible in a much larger quantity than others;
- 2) the included texts are of the highest possible quality from the perspective of linguistic and factual correctness;
- 3) the texts are also represented as equally as possible in terms of the years of their creation.

This enables detecting linguistic phenomena and possible trends over a longer period; in addition, this avoids an excessively current focus – which, to a certain extent, understandably characterises dictionaries such as the *Growing Dictionary of the Slovenian Language* – by reporting on popular themes in a given year. All this entails that the design, compiling and implementation of a corpus – within technological and legal limits – must be managed by a group of editors. To obtain a sufficient number of texts, especially literary texts, it would be worth simultaneously conducting a promotion campaign to raise awareness of the importance of such texts in compiling a corpus, which would increase the willingness of both authors and publishers to provide their texts for such purposes. In any case, a high-quality composition of a corpus should be the first criterion. When considering the use of the corpus for compiling general explanatory dictionary, in particular, its size should not be too large if that entails disrupting the proportions between fields and genres. Even though it is probably harder to achieve balance in smaller languages (as shown by Slovak, where the share of journalism in the national corpus is 60–70%, though literature still amounts to about 18%) than in large ones, it is worth pursuing exactly that – rather than mere size.

³² E.g. *Nega matere in otroka*: Epiduralni anestetik **ohromi hrbtenjačo** in tako onemogoči zaznavanje bolečine ‘An epidural anaesthetic **paralyses the spinal cord** and prevents feeling pain’; *Slovenske novice*: *24 ur na preži*: Ta del obale skriva še eno zanimivost, dno pokrivajo goste preproge **pozejdonke**, temno zelenih **travnatih alg** ‘This part of the coast reveals another interesting fact; the sea bed is covered by thick carpets of **Neptune grass**, dark green grass-like algae’ (Neptune grass is not a species of algae).

The Russian and Czech national corpora feature well-balanced proportions between fields and genres. As regards language quality, a methodological commentary on the Russian National Corpus on its website highlights the importance of the adequacy of the written standard language in literary and related texts. The share of literary texts in the Czech National Corpus also amounts close to a third of the corpus; this share amounts to slightly less than a third in the Polish National Corpus (and was only a little lower in the SSKJ card index). The share of journalism, on the other hand, does not exceed a third in the Russian National Corpus, is similar in the Czech corpus (where it is further divided into “traditional”, which probably means a somewhat higher level of quality, and “leisure” with a lower share) and does not exceed half of all the texts in the Polish National Corpus.

As regards the term reference corpus it seems to be sometimes used in relation to corpora, such as Gigafida, which lack balance and could hardly be considered representative – which is important for a reference corpus (cf. Atkins, Clear, Ostler 1992, Biber 1993, Gorjanc 2005). When that is the case,³³ we suggest to stop using the term reference corpus to ambiguously refer to a corpus as if it was balanced and representative.³⁴ The use of the phrase reference corpus is not meaningful if used every time to refer to a corpus (it seems devalued, an epithet of sorts), especially when it does not seem justified.

Górski and Łaziński note the FIDA(Plus) corpus as an outlier for its noticeably disproportionate share of journalism (and very low shares of non-fiction and literature) and the temporal limitation of texts, which are tied to the period after 1990 and especially 2000 with leaps between particular years.³⁵ The project

³³ “The starting point here is a clear awareness of the limited familiarity with relations between discourse and texts in a particular speech community, so analyses of these relations focus on their complexity and attempt to base criteria for the balanced nature of a corpus on this complexity” (Stefanowitsch 2020 in Logar Berginc et al. 2023: 88–89). For corpus GIGAFIDA see *Corpus Compilation: Specifications*, pp 3–7.

³⁴ Cf., for example, “Unless language structure and language use are infinitely variable (which, at a given point in time, they are clearly not), increasing the diversity of the sample will increase representativeness even if the corpus design is not strictly proportional to the incidence of text varieties or types of speakers found in the speech community. It is important to acknowledge that this does not mean that diversity and representativeness are the same thing, but given that representative corpora are practically (and perhaps theoretically) impossible to create, diversity is a workable and justifiable proxy” (Stefanowitsch 2020: 35).

³⁵ “One would notice that FIDA, the corpus of a small community such as Slovenes, is very poorly balanced, simply because in a country with a population of 2 million there are not enough books being written to easily create a corpus of 100 million that would not show a large imbalance between texts published in the press and in books. However, even in the case of larger communities, it may happen that certain types of texts are not very numerous, such as scientific literature in particular fields” (Górski, Łaziński 2012: 25–26).

of compiling Gigafida attempted to rectify the situation, but it seems that it pursued first and foremost a billion-word size as its seeming main objective, doing so with the accessible text production – this means mass-produced daily news and, in its tone or even altogether, advertising journalism, which has resulted in disproportions in semantic relations within a lexeme (i.e. *vilice* ‘fork’ first as machine part, not cutlery; *koš* ‘basket’ first for sports, not as a (home) accessory; *konjič* ‘horse’ first as a car, not an animal or toy; *nalagati* ‘to load, place; to impose’ first as a figurative description of obligations according to rules and regulations etc., not a concrete verb of putting something somewhere). In addition, with the domination of the conative function over the referential one (identifying objects), such mass journalism has made it difficult to judge the (un)markedness of lexis as the user gets the impression that every text has at least traces of expressive effect resulting from the conative function. Such a distinctly disproportionate prevalence of journalism with its own models of verbalisation also – not as much, but still – affects the diversity and representativeness of syntactic structures or syntactic modes of expression in general.³⁶

As for the analysis of neologisms (new words and new meanings in existing words), the Trendi corpus is fairly suitable, as confirmed by the experience of its use in compiling the *Growing Dictionary of the Slovenian Language*. Due to its design, its usefulness in analysing new words surpasses that of Gigafida – even in its latest version. As regards the widest possible analyses which – if that is their purpose – encompass the whole temporal range of linguistic phenomena, there is the Metafida corpus collection, which Gigafida is part of and to the size of which it materially contributes.

In light of the experience of using the 100-million KRES, it would be worth trying whether it is possible to compile a corpus with about 150–350 million words and a balance at least close to the Polish National Corpus (the Czech and Russian corpora seem unattainable), so that the share of journalism (of the highest possible quality, consisting of daily news, reports, discussions, columns) does not exceed half of the corpus. The second half would feature a quality literature (children’s, young adult, adult; Slovenian and translated) and especially adequately checked (reviewed, proofread) non-fiction primarily in the fundamental sciences,

³⁶ Cf. predicative modifiers in corpus Gigafida – the most common predicative modifiers are for instance *pijan* ‘drunk’, *mrtev* ‘dead’, *(ne)poškodovan* ‘(un)injured’ (from the police blotter); *gol* ‘naked’, *oblečen* ‘dressed’ (from advertising texts); *oslabljen* ‘weakened’ (from sports); *vroč* ‘hot’, *ohlajen* ‘cooled’ (from recipes) (cf. Gabrovšek 2023, 118–122).

with which most people are familiarised in primary and secondary school (and later in narrower specialised fields). These two factors – an equal representation of quality texts of different genres and a size that enables the proper operation of word sketches but is also not unmanageable for manual concordance analysis – seems essential in compiling a versatile, representative and thus reference text corpus as a resource for making quality reference works (especially general explanatory dictionaries) and for independent linguistic analysis.

REFERENCES

- eSSKJ: Slovar slovenskega knjižnega jezika 2016–, www.fran.si (1. 1.–31. 5. 2024).
- Gigafida 2.0: Korpus pisne standardne slovenščine. <https://viri.cjvt.si/gigafida/> (subcorpora PUBL and STVL available within search options: https://www.clarin.si/noske/run.cgi/first_form?corpname=gfida20_dedup;align=)
- Gigafida 2.0. Corpus Compilation: Specifications. https://www.cjvt.si/gigafida/wp-content/uploads/sites/10/2019/06/Gigafida2.0_specifikacije.pdf
- KRES. <http://www.korpus-kres.net/> (October 2024)
- Korpus šolskih besedil slovenskega jezika (KŠBSJ). Internal materials.
- Korpus znanstvenih besedil (KZB). <https://www.clarin.si/ske/#dashboard?corpname=kzb10>
- British National Corpus. <https://www.english-corpora.org/bnc/>
- Czech National Corpus. <https://www.korpus.cz/kontext/query?corpname=syn2020> (October 2024); <https://wiki.korpus.cz/doku.php/cnk:syn2020> (October 2024)
- Polish National Corpus. <https://nkjp.pl/poliqarp/>
- Russian National Corpus. <https://ruscorpora.ru/stats>
- Slovak National Corpus. <https://korpus.sk/en/corpora-and-databases/snc-corpora/publicly-available-snc-corpora/structure-of-the-corpus-prim-10-0/>

LITERATURE CITED

- Atkins, Sue, Clear, Jeremy, Ostler, Nicholas. 1992. Corpus Design Criteria. *Literary and Linguistic Computing* 7/1: 1–16.
- Biber, Douglas. 1993. Representativeness in Corpus Design. *Literary and Linguistic Computing* 8/4: 243–257.
- Centa Strahovnik, Mateja. 2023. *Čustva, človekova odnosnost in doseganje dobrega življenja*. Ljubljana: Teološka fakulteta.
- Corpas Pastor, Gloria, Seghiri, Miriam. 2010. Size Matters: A Quantitative Approach to Corpus Representativeness. In: R. Rabadán, M. Fernández López, and T. Guzmán González (ed.). *Lengua, traducción, recepción en honor de Julio César Santoyo*. León: Universidad de León Área de Publicaciones: 111–145. <http://hdl.handle.net/2436/622560>

- Gabrovšek, Dejan. 2023. Povedkov prilastek v slovenščini. *Slavistična revija* 71/2: 113–128. <https://doi.org/10.57589/srl.v71i2.4108>
- Gorjanc, Vojko. 2005. *Uvod v korpusno jezikoslovje*. Domžale: Izolit.
- Górski, Rafał L. 2008. Representativeness of a written part of a Polish general-reference corpus. Primary notes. In: B. Lewandowska-Tomaszczyk (ed.). *Corpus Linguistics, Computer Tools, and Applications – State of the Art*, Frankfurt am Main: Peter Lang. 119–123. http://nkjp.pl/settings/papers/representativeness_primary_notes.pdf
- Górski, Rafał L., Łaziński, Marek. 2012. Reprezentatywność i zrównoważenie korpusu. In: A. Przepiórkowski, M. Bańko, R. L. Górski, B. Lewandowska-Tomaszczyk (ed.). *Narodowy korpus języka polskiego*. Warszawa: Wydawnictwo naukowe PWN. 25–36.
- Gregorčič, Rok. 2023. Tehnološki razvoj v luči Habermasove etike diskurza. *Bogoslovni vestnik* 83/4: 911–922.
- Jakobson, Roman. 1996. *Lingvistični in drugi spisi*. Ljubljana: Inštitut za humanistične študije.
- Korošec, Tomo. 2005. *Jezik in stil oglaševanja*. Ljubljana: Fakulteta za družbene vede.
- Kosem, Iztok, Čibej, Jaka, Dobrovoljc, Kaja, Kuzman, Taja, Ljubešić, Nikola. 2023. Spremljevalni korpus Trendi in avtomatska kategorizacija. *Slovenščina 2.0: Empirične, Aplikativne in Interdisciplinarne Raziskave* 11/1: 161–188. <https://doi.org/10.4312/slo2.0.2023.1.161-188>
- About KRES. <http://www.korpus-kres.net/Support/About> (November 2023).
- Krek, Kilgariff. 2006. Slovene Word Sketches. *Proceedings of 5th Slovenian/First International Languages Technology Conference*. Ljubljana. <https://www.kilgariff.co.uk/Publications/2006-KrekKilg-Ljub-SloveneWS.pdf>
- Krvina, Domen. 2018. *Glagolski vid v sodobni slovenščini 1. Besedotvorje in pomen*. Ljubljana: Založba ZRC. <https://doi.org/10.3986/9789610500742>
- Krvina, Domen. 2022. The Growing Dictionary of the Slovenian Language (2014–) and Slovenian Neologisms: Study on Types of Data and Their Use. *Slovenski jezik / Slovene Linguistic Studies* 14: 117–151. <https://doi.org/10.3986/sjls.14.1.05>
- Ledinek, Nina, Jemec Tomazin, Mateja, Trojar, Mitja, Perdih, Andrej, Ježovnik, Janoš, Romih, Miro, Erjavec, Tomaž. 2022. Korpus šolskih besedil slovenskega jezika: zasnova in gradnja. *Jezikoslovni zapiski* 28/1: 123–137. <https://doi.org/10.3986/JZ.28.1.07>
- Logar Berginc, Nataša, Grčar, Miha, Brakus, Marko, Erjavec, Tomaž, Arhar Holdt, Špela, Krek, Simon. 2012. *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko, in Fakulteta za družbene vede.
- Logar Berginc, Nataša, Gorjanc, Vojko, Arhar Holdt, Špela. 2023. Korpus Gigafida 2.0: Mnenje uporabnikov. *Jezik in Slovtvo* 68/2: 75–91.
- Novak, France. 2004. *Samostalniška večpomenskost v jeziku slovenskih protestantskih piscev 16. stoletja*. Ljubljana: Založba ZRC
- Petric Žižić, Špela. 2020. Tipologija razlag v Šolskem slovarju slovenskega jezika. *Slavistična revija* 68/3: 391–409. <https://srl.si/ojs/srl/article/view/3875>
- Petric Žižić, Špela (tran.). 2022. School Dictionary of the Slovenian Language on the Franček Web Portal. *Slavistica Vilnensis*, 67/2: 126–140. <https://orcid.org/0000-0001-7451-4264>

- Rundell, Michael, Atkins, Sue. 2013. Criteria for the design of corpora for monolingual lexicography. In: R. H. Gouws, U. Heid, W. Schweickard, H. E. Wiegand (eds.). *Dictionaries. An International Encyclopedia of Lexicography*. Berlin/Boston: De Gruyter Mouton. 1336–1343.
- Snoj, Jerica. 2004. *Tipologija slovarske večpomenskosti slovenskih samostalnikov*. Ljubljana: Založba ZRC. <https://doi.org/10.3986/9616500309>
- Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology (Textbooks in Language Sciences 7)*. Berlin: Language Science Press.
- Suhadolnik, Stane. 1963. Problemi slovenske leksikografije. *Sodobnost* 11/10: 926–934.
- Suhadolnik, Stane, Janežič, Marija. 1962. Plasti in pogostnost leksike. *Jezik in slovstvo* 8/1–2: 45–49.
- Suhadolnik, Stane. 1968. Koncept novega slovarja slovenskega knjižnega jezika. *Jezik in slovstvo* 13/7: 219–224.
- Svetina, Peter. 2009. Kaj naj beremo z otroki? In: Livija Knaflič, N. Bucik (ed.). *Branje za znanje in branje za zabavo: priročnik za spodbujanje družinske pismenosti*. Ljubljana: Andragoški center Slovenije. 67–69. https://arhiv.acs.si/publikacije/Branje_za_znaje_in_branje_za_zabavo-prirocnik.pdf
- Vidovič Muha, Ada. 2013. *Slovensko leksikalno pomenoslovje*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Vodičar, Janez. 2023. Avtoriteta na področju vzgoje in verovanja v digitalni dobi. *Bogoslovni vestnik* 83/4: 1035–1047.
- About ZgoSLiP Project: <https://www.zgoslip.si/> (November 2023).

Received November 2023, accepted December 2023.

Prispelo novembra 2023, sprejeto decembra 2023.

ACKNOWLEDGMENTS

The publication of article was made possible by programme Slovenski jezik v sinhronem in diahronem razvoju (P6-0038 (A)), which is financially supported by the Slovenian Research Agency (ARIS).

SUMMARY

THE RELATION BETWEEN CORPORA COMPOSITION (GENRE BALANCE AND REPRESENTATIVENESS) AND THEIR RELIABILITY IN COMPILING GENERAL EXPLANATORY DICTIONARY

The general explanatory dictionary is expected to describe the lexical system as comprehensively as possible. The knowledge of lexicological theory about meaning

development, semantic extension processes which is (as a rule) implemented in reference lexicographic works represent a sound theoretical basis for such a description. Regarding the suitable corpus for such a description to be possible, we posit that it should have adequate shares of non-fiction and literature of the highest possible quality. The share of periodicals (particularly journalistic discussions, interviews etc.) would still be relatively large; it should, however, probably not represent more than half of the corpus (as is the case, for example, in the Polish National Corpus). The corpus should be sizeable enough for (semi-)automatic analysis – but not by distorting the genre proportions through expansion. To preserve such proportions, editorial interventions are desirable, if not necessary, to ensure texts are included according to appropriate criteria, and not determined mostly by the accessibility of texts itself. The core of non-fiction would consist of reviewed, mainly professional texts at the secondary school level, complemented by scientific texts (such as those collected in the KZB corpus), preferably in fundamental fields of science. Literature would be represented by quality both Slovenian and foreign (potentially based on data collected in the ZgoSLiP project) (semi-)literary texts. Journalistic texts would also be selected based on the criterion of quality (daily news – reporting, discussions, articles, columns; sections marked by advertising and other texts with marketing patterns should be minimised) and – in spite of their bigger production and relative ease of access – should not exceed half of the corpora. Such a balance would increase the representativeness of the corpus data, even when one cannot be quite sure about the actual influence of a genre. Compiling the corpus in this way should take precedence over its size with stronger effect on the representation of long-time, representative language trends.

RAZMERJE MED SESTAVO KORPUSOV (ŽANRSKA URAVNOTEŽENOST IN
REPREZENTATIVNOST) IN NJIHOVO ZANESLJIVOSTJO PRI IZDELAVI SPLOŠNEGA
RAZLAGALNEGA SLOVARJA

Od splošnega razlagalnega slovarja se pričakuje čim bolj celovit opis leksikalnega sistema. Poznavanje leksikološke teorije o pomenskem razvoju in pomenotvornih procesih, ki se (praviloma) odraža v referenčnih leksikografskih delih, je dobra teoretična podlaga za tak opis. Za tak opis bi bilo smiselno zgraditi korpus z ustreznim deležem stvarne literature in leposlovja najvišje možne kakovosti. Delež periodike (zlasti novinarskih razprav, intervjujev itd.) bi bil še vedno razmeroma velik, vendar ne bi smel predstavljati več kot polovice korpusa (kot je to na primer v Poljskem narodnem korpusu). Korpus bi moral biti dovolj velik za (pol)avtomatsko analizo – vendar ne na račun izkrivljenosti žanrskih razmerij. Za ohranitev takšnih razmerij so

zaželeni, če ne celo nujni, uredniški posegi, ki zagotavljajo, da so besedila vključena v skladu z ustreznimi merili in da jih ne določa zlasti dostopnost besedil. Jedro stvarne literature bi sestavljala preverjena, predvsem strokovna besedila srednješolske ravni, dopolnjevala pa bi jih strokovna besedila (kot npr. ta, ki so zbrana v korpusu KZB), po možnosti s temeljnih področij znanosti. Leposlovje bi bilo zastopano s kakovostnimi slovenskimi in tujimi (potencialno na podlagi podatkov, zbranih v projektu ZgoSLiP) (pol)literarnimi besedili. Tudi novinarska besedila bi izbrali na podlagi merila kakovosti (dnevne novice – poročanje, razprave, članki, kolumne; besedila z oglaševalskimi prvinami v čim manjšem deležu). Kljub veliki produkciji in relativno enostavnemu dostopu publicistika ne bi smela presegati polovice korpusa. Takšno ravnovesje bi povečalo reprezentativnost korpusnih podatkov, in to kljub temu, da o dejanskem vplivu posamezne zvrsti ne moremo biti povsem gotovi. Tovrstno oblikovanje korpusa bi moralo za zastopanost reprezentativnih jezikovnih trendov imeti prednost pred velikostjo.

Alenka Vrbinc (ORCID 0000-0002-7330-4158)

Univerza v Ljubljani, Ekonomska fakulteta, Slovenija
alenka.vrbinc@ef.uni-lj.si

Donna Farina (ORCID 0000-0001-5695-7782)

New Jersey City University, Department of Multicultural Education, ZDA
dfarina@njcu.edu

Marjeta Vrbinc (ORCID 0000-0002-6866-6023)

Univerza v Ljubljani, Filozofska fakulteta, Slovenija
marjeta.vrbinc@ff.uni-lj.si
DOI: <https://doi.org/10.3986/16.1.08>

HOMONIMIJA IN POLISEMIJA V DVOJEZIČNIH SLOVARJIH ZA SLOVENSKE PRISELJENCE V ZDA V ZAČETKU 20. STOLETJA

V metaleksikografski literaturi ni veliko znanstvenih prispevkov, ki bi obravnavali slovarje za priseljence. Slovarji za priseljence so klasifikacijsko gledano del pedagoške leksikografije. V prispevku obravnavamo štiri slovensko-angleške oz. angleško-slovenske slovarje za priseljence (Kubelka 1904, Kubelka 1912, Košutnik 1912 in Kern 1919), ki so izšli na začetku 20. stoletja, to je v času največjega izseljevanja Slovencev v ZDA. V analizi se osredotočamo na prikaz informacij s posebnim poudarkom na obravnavi homonimije, polisemije in konverzije ter vključevanju usmerjevalnih besed, zapisanih v oklepajih. Iščemo tudi odgovor na vprašanje, ali uporabnikom – priseljencem olajšamo razumevanje in/ali rabo jezika novega okolja, če v slovarju uporabimo makrostrukturni pristop oz. če uporabimo mikrostrukturnega. Rezultati naše analize kažejo, da je še posebej pomembno, kako so v slovarjih za priseljence prikazane informacije v mikrostrukturi. Zaključimo lahko, da je treba v tovrstnih slovarjih dati prednost prikazu pomena oz. pomenov iztočnice, vse ostale informacije, ki so tipično vključene v slovarje, pa so za priseljence drugotnega pomena.

KLJUČNE BESEDE: slovarji za priseljence, homonimija, polisemija, konverzija, usmerjevalne besede v oklepaju

Regrettably, scholarly attention has been limited concerning dictionaries designed for immigrants, what could be called *immigrant dictionaries*. These books, which aim to address the immediate requirements of new immigrants, are part of the broader sphere of learner lexicography. This article scrutinizes four such immigrant dictionaries, Slo-

venian–English and English–Slovenian works published in the early twentieth century during the peak of Slovenian immigration to the United States (Kubelka 1904, Kubelka 1912, Košutnik 1912, and Kern 1919). The analysis focuses on the presentation of information, examining the dictionaries' treatments of homonymy, polysemy, conversion, and parenthetical information. The article discusses implications of adopting a macrostructure-oriented approach or a microstructure-oriented one, in terms of which best facilitates the comprehension of immigrant users. The research findings highlight the significance of lexicographic presentation in the microstructure of immigrant dictionaries. The recommendation is to prioritize the presentation of meaning over other elements typically covered in dictionaries, when a dictionary is intended for immigrants.

KEYWORDS: immigrant dictionaries, homonymy, polysemy, conversion, parenthetical information

1 UVOD

V devetnajstem stoletju so imeli enojezični slovarji v večini ameriških gospodinjstev, še posebej na podeželju, zelo pomembno mesto (Béjoint 2010: 90–91), kajti angleško govoreči prebivalci so v slovarjih videli pripomoček, s pomočjo katerega bi se lahko povzpeli do boljšega družbenega položaja. Sestavljavci enojezičnih slovarjev so skušali pomagati deprivilegiranim prebivalcem pri vključevanju v družbo, niso pa imeli namena spreminjati družbe (Béjoint 2010: 93). Na prelomu devetnajstega in dvajsetega stoletja se je povečevalo število priseljencev v ZDA, ki so govorili različne jezike in so potrebovali dvojezične slovarje, da bi lahko razumeli jezik novega okolja in se seznanili z ameriško kulturo. V tem obdobju so sestavljali dvojezične slovarje in druge priročnike tudi priseljenci, ki so želeli pomagati rojakom v novi deželi. Frank Javh-Kern, avtor enega od slovarjev, ki ga bomo obravnavali v nadaljevanju, je v svojih spominih zapisal, da je bila dobrobit Slovencev v ZDA in v stari domovini glavni cilj njegove javne dejavnosti (Javh-Kern 1937).

Drugi val množičnega priseljevanja na ozemlje ZDA (1860–1915) se je začel v Avstro-Ogrski (Granatir Alexander 2009: 20) in kar 80 % ljudi, ki so se izselili iz Avstro-Ogrske, je odšlo v ZDA (Kalc in sod. 2020: 45). J. Granatir Alexander (2009: 20) navaja, da je v 70. letih devetnajstega stoletja v Ameriko prišlo manj kot 73.000 prebivalcev Avstro-Ogrske, to število se je v 90. letih devetnajstega stoletja povečalo kar za sedemkrat, in sicer na 593.000, v prvem desetletju dvajsetega stoletja pa je doseglo več kot 2,1 milijon ljudi, kar pomeni, da se je v obdobju med letoma 1880 in 1915 v Ameriko izselilo več kot 3,9

milijona prebivalcev Avstro-Ogrske. V 90. letih devetnajstega stoletja so v ZDA začeli prihajati tudi Slovenci s področja Kranjske, ki so jih ameriški uradniki za priseljevanje le s težavo ločevali od Hrvatov ali drugih etničnih skupin, zaradi česar točno število izseljenih Slovencev ni znano. Glede na ocene je pred 1. svetovno vojno v ZDA prišlo od 250.000 do 350.000 Slovencev (Granatir Alexander 2009: 21; Kalc in sod. 2020: 25).

Novi priseljenci so se morali naučiti angleško in se prilagoditi novemu kulturnemu okolju. Pri tem so si pomagali s priročniki, izdanimi v ZDA in na Slovenskem, v katerih so dobili informacije o angleščini, ameriški ustavi in vladi, ameriškem denarju in gospodarstvu (Stanonik 1996). Med temi publikacijami najdemo slovnice, slovarje, učbenike, priročnike s koristnimi stalnimi večbesednimi izrazi iz vsakdanjega življenja in otroške abecednike, številne med njimi so bile večnamenske (npr. Kubelka 1904 in Kubelka 1912, ki ju bomo analizirali v nadaljevanju).

1.1 SLOVARJI ZA PRISELJENCE

V prispevku bomo obravnavali slovensko-angleške in angleško-slovenske slovarje, namenjene slovenskim priseljencem v ZDA v začetku dvajsetega stoletja. V ZDA so v tem obdobju obstajali tudi številni drugi dvojezični slovarji z angleščino, ki so bili namenjeni priseljencem iz drugih držav. Slovarji za priseljenca nastajajo tudi v današnjem času, še posebej v državah, v katere se priseljuje veliko ljudi iz različnih jezikovnih in kulturnih okolij. Ti slovarji so po obliki in vsebini podobni slovarjem z začetka dvajsetega stoletja, so pa pri njihovem sestavljanju v veliko pomoč sodobna jezikovnotehnološka orodja in viri, ki jih sestavljavci slovarjev pred več kot sto leti niso imeli na voljo. Kljub temu Tarp (2004: 228) ugotavlja, da je večina sodobnih slovarjev za priseljenca slabe kakovosti (kot izjemo navaja švedski Lexin), zaradi česar bi bilo k sestavljanju tovrstnih slovarjev treba pritegniti usposobljene leksikografe.

V metaleksikografski literaturi ni veliko virov, ki bi obravnavali slovarje za priseljenca, čeprav so pomembni za zgodovino slovaropisja (Fjeld 2012). R. V. Fjeld opisuje slovar norveščine, ki je izšel v Minneapolisu v Minnesoti leta 1897, to je ob približno enakem času kot slovarji, ki jih analiziramo v našem prispevku, pa tudi informacije, vključene v ta slovar in omenjene v uvodu, ki je napisan v norveščini, so podobne informacijam, vključenim v slovarje, ki smo jih preučevali v naši raziskavi (npr. pravopisna pravila, pravila za pravilno rabo ločil, družbenopolitični sistem in druge informacije, ki lahko koristijo priseljencem).

O. Karpova in M. A. Kulagina (2009: 169, 174) raziskujeta proces integracije priseljencev, ki se v novem družbenem okolju soočajo z drugačnim načinom življenja, z drugačnimi navadami, običaji in tradicijo ter se spopadajo z različnimi izzivi v komunikaciji z lokalnim prebivalstvom ter z uradniki v uradih za priseljence in v socialnih službah. Pomanjkljivo razumevanje in neuspešna komunikacija nista samo posledica slabega znanja jezika novega okolja, ampak tudi pomanjkanja znanja o kulturnih posebnostih novega okolja. Tudi A. Vacalopoulou in sod. (2011: 274), ki so preučevali slovarske vire za priseljence v Grčiji, so prišli do podobnih ugotovitev. Ti viri naj bi priseljencem pomagali pri komunikaciji z oblastmi, pri komunikaciji v različnih družbenih okoljih in situacijah, pri reševanju vsakodnevnih problemov, pri učenju jezika v formalnem in neformalnem okolju in pri seznanjanju s splošnim kulturnim in družbenim kontekstom.

Priseljenci v specifičnih okoljih uspešno delujejo le, če imajo na voljo priročnike z osnovnimi izrazi, ki se uporabljajo v tem okolju, kar O. Karpova in M. A. Kulagina (2009: 169) poimenujeta jezik za potrebe priseljske politike. O. Grønvik (2016: 23) poudarja, da morajo slovarji za priseljence, ki se učijo jezik novega okolja, poleg ustreznikov v ciljnem jeziku vključevati tudi razlage besed in konceptov, specifičnih za novo deželo in kulturo. Vključevanje kulturnospecifičnih informacij v slovarje za priseljence je izredno pomembno, ker se priseljenci tudi s pomočjo slovarjev seznanjajo z navadami, običaji in kulturo nove dežele (Karpova in Kulagina 2009: 174). A. Vacalopoulou in sod. (2011: 276) omenjajo nekaj segmentov besedišča, ki morajo biti vključeni v slovarje za priseljence: lastna imena, kot so zemljepisna imena (npr. imena mest, pokrajin, rek, otokov), imena uradnih organov (npr. ministrstva in druge vladne in nevladne organizacije) in imena geopolitičnih entitet skupaj z njihovimi akronimi.

O. Karpova in M. A. Kulagina (2009: 171) ugotavljata, da v številnih državah, še posebej v ZDA, nastajajo slovarji za priseljence, ki vključujejo zgoraj navedene segmente besedišča. Ti slovarji so namenjeni ljudem, ki se morajo hitro integrirati v novo okolje, na primer učenci, dijaki, študenti in vsi ostali, ki morajo obvladati jezik novega okolja na različnih nivojih (Karpova 2019). Tovrstno besedje je vključeno tudi v slovarje, ki nastajajo v okviru projekta Lexin, s katerim so v poznih 70. letih dvajsetega stoletja začeli na Švedskem, namen pa je bil sestaviti dvojezične slovarje, ki vključujejo švedščino in matere jezike priseljencev. V 90. letih dvajsetega stoletja se je

projekt razširil tudi v druge skandinavske države, tako da so danes na voljo številni dvojezični slovarji s skandinavskimi jeziki in jeziki priseljencev, še posebej številni so slovarji s švedščino in norveščino kot izhodiščnim jezikom (gl. tudi Grønvik 2016: 23; Karpova in Kulagina 2009: 171; Malmgren 1999).

1.2 SLOVARJI ZA SLOVENSKE PRISELJENCE V ZDA

V prispevku bomo podrobneje obravnavali štiri slovarje, ki so izšli v prvih dveh desetletjih dvajsetega stoletja, to je v času največjega izseljevanja Slovencev v ZDA:

Victor J. Kubelka. 1904. *Slovensko-angleški žepni rečnik v olajšavo naučenja obeh jezikov, Slovenian-English pocket dictionary to facilitate the study of both languages* (v nadaljevanju Kubelka 1904),

Viktor/Victor J. Kubelka. 1912. *Slovensko-angleška slovnica, tolmač, spisovnik in navodilo za naturalizacijo. Angleško-slovenski in slovensko-angleški slovar. Slovenian-English grammar, interpreter, letterwriter and information on naturalization. English-Slovenian and Slovenian-English dictionary* (v nadaljevanju Kubelka 1912),

Silvester Košutnik. 1912. *Ročni slovensko-angleški in angleško slovenski slovar: Zlasti namenjen izseljencem v Ameriko* (v nadaljevanju Košutnik 1912),

F(rank) J(avh) Kern. 1919. *A complete pronouncing dictionary of the English and Slovene languages for general use, Popoln angleško-slovenski besednjak z angleško izgovarjavo* (v nadaljevanju Kern 1919).

Oba slovarja Viktorja Kubelke sta izšla v New Yorku in sta tipična primera slovarjev za priseljence, saj poleg samega slovarja vključujeta še različne informacije, ki so jih potrebovali priseljenci. Tako najdemo v Kubelki 1904 poleg slovensko-angleškega slovarja še razdelek s pogovori »Easy dialogus [sic.]« oz. »Lahki pogovori«, teme pogovorov pa so: Pozdravi, V družbi, Povprašanja, Življenju potrebne stvari in Za delom. Poleg tega so vključene še druge informacije o ZDA, kot so imena ameriških zveznih držav ter denar ZDA. Pod naslovom Imena so navedena slovenska osebna imena z ustrezniki v angleščini. Kubelka 1912 ima poleg angleško-slovenskega in slovensko-angleškega slovarja še štiri dele: Angleška slovnica, Vsakdanji razgovori, Slovensko-angleški spisovnik in Navodila za naturalizacijo. V Košutniku 1912, ki je edini od štirih preučevanih slovarjev, ki je izšel v Ljubljani, ni dodatnih informacij, enako pa velja tudi za Kerna 1919, ki je izšel v Clevelandu.

Kubelka, Košutnik in Kern¹ so se s sestavljanjem slovarjev ukvarjali ljubiteljsko, nihče od njih ni bil jezikoslovec ali leksikograf in z veliko verjetnostjo lahko trdimo, da njihovo delo ni temeljilo na jezikoslovni, leksikološki ali leksikografski teoriji, kar je razvidno tudi iz različnih pristopov, ki so jih uporabljali (npr. pri izbiranju iztočnic ali pri prikazu pomenov).

1.3 CILJ RAZISKAVE

V raziskavi smo se osredotočili na prikaz informacij v omenjenih štirih slovarjih za slovenske priseljence v ZDA s poudarkom na homonimiji, konverziji, polisemiji in usmerjevalnih besedah, zapisanih v oklepajih. Glede na to, da o slovarjih za priseljence do sedaj ni bilo veliko napisanega, skušamo z našo raziskavo prispevati k metaleksikografskemu razumevanju tovrstnih slovarjev. Preučevanje zgodnejših slovarjev za priseljence, težave, ki so jih imeli njihovi sestavljavci, in rešitve, ki so jih bolj ali manj dosledno uporabljali v svojih slovarjih, so lahko v pomoč tudi pri sestavljanju sodobnih slovarjev za priseljence. Našega dela pa ne motivira samo leksikografski, ampak tudi družbeni vidik, saj so migracije tudi del sodobnega sveta, zaradi česar potreba po kakovostnih slovarskih virih ni nič manjša kot v preteklosti. Priseljenci se morajo tudi danes naučiti jezik nove dežele, prav tako se morajo prilagoditi novemu kulturnemu okolju, pri čemer jim lahko družba v novi deželi priskoči na pomoč s priročniki in ustreznimi slovarji.

2 HOMONIMIJA, KONVERZIJA IN POLISEMIJA V SLOVARJIH ZA PRISELJENCE

V dvojezičnih slovarjih se uporabljajo različni pristopi k obravnavi homonimije in polisemije, kar je odvisno od specifičnega jezikovnega para, ki ga slovar obravnava, in od ciljnih uporabnikov slovarja (informacije, ki jih potrebujejo, intuicija pri iskanju določene informacije v slovarju, vrsta slovarja (Svensén 2009: 101)). Homonimijo (pomenska neprekrivnost) se obravnava v povezavi s polisemijo (pomenska prekrivnost), pri čemer lahko natančno definiramo

¹ Victor John Kubelka: krščen kot Viktor Maria Johann Nep. Kubelka, rojen 29. 4. 1879 v Ljubljani, umrl 24. 9. 1961 v Loomisu v zvezni državi New York.

Silvester Košutnik: krščen kot Silvester Podergajs, rojen 20. 12. 1874 v Škednju, umrl 30. 6. 1944 v Šentilju v Slovenskih Goricah.

Frank J. Kern: krščen kot Franz Javh, rojen 18. 3. 1887 v Breznici pod Lubnikom, umrl 4. 10. 1979 v Clevelandu v zvezni državi Ohio.

Več biografskih podatkov o treh avtorjih, gl. Vrbinc, Farina, Vrbinc 2024.

pomen homonimije le, če jo jasno razlikujemo od polisemije (Halas Popović 2021: 38). Pri razlikovanju homonimije in polisemije se uporabljajo različni formalni in semantični kriteriji ter testi, na splošno pa lahko te kriterije delimo na tri glavne tipe: etimološki, slovnični in semantični (prav tam: 39–41; Svensén 2009: 96–100). Kljub upoštevanju vseh teh kriterijev je zelo težko razlikovati med polisemijo in homonimijo, saj ne gre za dva popolnoma nasprotna koncepta; med njima obstajajo različne stopnje formalne in semantične prekrivnosti, zato ju lahko umestimo na dva skrajna konca lestvice, ki sta med seboj povezana (Lipka 2002: 157). Dvojezični slovarji obravnavajo homonimijo na različne načine – v okviru enega slovarskega sestavka ali v okviru dveh ali več slovarskih sestavkov; podobno velja tudi za konverzne lekseme. Iz slovarskih ustreznikov so razvidni različni pomeni večbesedne iztočnice, slovarski ustrezniki so lahko tudi sopomenke, zaradi česar je v dvojezični slovar smiselno vključiti sopomenske ustreznike na drugačen način kot ustreznike z različnimi pomeni.

Ker sta za homonime značilni popolna izrazna prekrivnost in popolna vsebinska oz. pomenska neprekrivnost in so homonimna razmerja možna le znotraj iste besedne vrste, torej le med slovničnokategorialno povezanimi leksemi, se moramo vprašati, ali je konverzija lahko tudi vrsta homonimije (Žele 2015: 74). Za angleščino so v veliko večji meri kot za slovenščino² značilni konverzni leksemi, ki pripadajo dvema ali več različnim besednim vrstam in ohranjajo medsebojno pomensko povezanost. Kljub temu, da osnovna oblika ostane nespremenjena, pa se pomen spremeni sorazmerno s spremenjeno skladenjsko vlogo in slovničnimi lastnostmi motivirajočega leksema (Žele 2015: 65). Za A. Legan Ravnika (2021: 178) predstavlja konverzija izziv za sestavljalce slovarjev, ki morajo najprej poiskati odgovor na leksikološko vprašanje, ali gre zaradi prehajanja iz enega stavčnega člena v drugega, nove skladenjske rabe in prehajanja med besednimi vrstami za enega, dva ali več leksemov, nato pa še na leksikografsko vprašanje, ali naj homografe vključijo v slovar kot nove iztočnice v samostojnih slovarskih sestavkih ali kot podiztočnice v istih slovarskih sestavkih z izhodiščnim leksemom (gl. tudi Žele 2015: 75).

Uporabniki, ki iščejo v slovarju rešitve za komunikacijske težave, potrebujejo nazorne razlage, če gre za enojezični slovar, oziroma ustreznike, če

² A. Žele (2015: 65) omenja povečano število konverzij v *Slovarju novejšega besedja slovenskega jezika* zaradi povečanega prevzemanja besedja.

gre za dvojezični slovar, to pa jim omogoča, da razumejo besedo ali frazem v konkretnem sobesedilu; za te uporabnike možni drugi pomeni v drugih sobesedilih niso pomembni, prav tako jih ne zanimajo homonimni ali polisemični odnosi med različnimi pomeni (Tarp 2009: 297). To vsekakor velja tudi za uporabnike slovarjev, ki smo jih preučevali v naši raziskavi.

2.1 KUBELKA 1904: SLOVENSKO-ANGLEŠKI SLOVAR

V Kubelki 1904 enako pisane iztočnice niso označene s številko zadaj nad iztočnico, kot je to značilno za SSKJ2,³ ampak so samo navedene ena za drugo. Prav tako ni označeno mesto naglasa ali kvantiteta in kvaliteta naglašene samoglasnika.⁴ To je razvidno iz prikaza samostalnika *svet* (SLIKA 1), ki je obravnavan kot štiri iztočnice, ki ustrezajo štirim različnim pomenom v angleščini. V primerih obsežnejšega homonimnega niza je možno, da ima uporabnik težave z iskanjem ustreznega pomena, ki se skriva pod eno od navedenih iztočnic (gl. Gantar 2015: 356). Samo pri prvi iztočnici je uporabnikom pri izbiri ustreznega pomena v angleščini v pomoč usmerjevalna beseda *zemlja*, ki je zapisana v oklepaju. Na splošno je v tem slovarju zelo malo usmerjevalnih besed ali dodatnih informacij, s katerimi bi si uporabnik lahko pomagal; tudi v primeru samostalnika *svet* bi bile informacije v oklepaju koristne za uporabnika.

svet (zemlja)	world	<i>vorl</i>
svet	counsel	<i>kaunsl</i>
“	advice	<i>advajs</i>
svet	saint	<i>sejnt</i>

SLIKA 1: Kubelka 1904: 103

Raba ponavljanja [“] v Kubelki 1904 ni razložena in uporabniku verjetno ni najbolj razumljiva. Pri samostalniku *svet* (SLIKA 1) je Kubelka s ponavljanjem verjetno označil polisemijo. Za označevanje polisemije je Kubelka poleg ponavljanja uporabljal tudi zavite oklepaje ({}) med prvim in drugim stolpcem. Zaviti oklepaji včasih pomenijo, da ima slovenska iztočnica več ustreznikov v angleščini, torej imajo oklepaji funkcijo označevanja polisemije; zaviti oklepaj povezuje ustreznike (1904: 26):

³ Naveden je SSKJ2 kot najnovejši že zaključeni splošni razlagalni slovar za slovenski jezik.

⁴ S tem bi verjetno slovenski priseljenci v ZDA imeli težave glede na relativno nizko stopnjo izobrazbe večine priseljenecv (gl. Cvrček 2020).

barva { colour
paint

Če bi bila tovrstna raba zavitega oklepaja dosledna, bi bila za uporabnika lahko koristna, vendar pa Kubelka uporablja zaviti oklepaj tudi v druge namene. Z zavitim oklepajem v primeru na str. 35

črnoglav } black-headed
črnolas }

najverjetneje označuje sopomenskost slovenskih besed *črnoglav* in *črnolas*, saj zaviti oklepaj povezuje slovenski iztočnici. Problematičen je angleški ustreznik *black-headed*, ki se glede na *Oxford English Dictionary* uporablja predvsem kot sestavni del imen živali in bi se na podlagi tega lahko zdelo, da ni pravi ustreznik slovenskih iztočnic. Glede na časopisne vire iz obdobja, ko je bil objavljen Kubelkov slovar iz leta 1904, lahko trdimo, da se *black-headed* v tistem času ni uporabljal samo v imenih živali. Npr. *a black-headed doll* (*The Charlotte News*, 20. 12. 1904); *black-headed girls* (*The Baltimore Sun*, 31. 8. 1904); *a little black-headed boy* (*The Indianapolis News*, 5. 1. 1904); *black-headed men* (*The Toronto Star*, 26. 8. 1904); *the little black-headed woman* (*The Gazette*, 7. 10. 1910). Nenavadno je, da je Kubelka vključil informacije o sopomenskosti slovenskih besed v slovar, ki naj bi uporabniku pomagal pri enkodiranju. Vprašanje je seveda tudi, kako so uporabniki slovarja na začetku 20. stoletja razumeli in interpretirali rabo zavitega oklepaja, ki je bil v slovarju obrnjen enkrat levo in drugič desno.

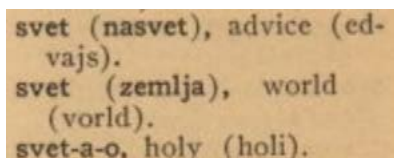
Kubelka v slovarju iz leta 1904 uporablja še tretji način za prikaz polisemije, to je ponavljanje iztočnice (na primer na str. 54):

jezik tongue
jezik language

Kubelka vključuje enake informacije na različne načine (ponavljanje, zaviti oklepaj, ponavljanje iztočnice), ki jih nikjer ne pojasni, zato je le malo verjetno, da so jih uporabniki slovarja – priseljenci v ZDA pravilno razumeli in interpretirali.

2.2 KUBELKA 1912: SLOVENSKO-ANGLEŠKI IN ANGLEŠKO-SLOVENSKI SLOVAR

Kubelka tudi v slovensko-angleškem delu slovarja iz leta 1912 ne označuje mesta naglasa ali kvantitete in kvalitete naglašene samoglasnika. Če pogledamo isto iztočnico kot v Kubelki 1904, to je *svet* (SLIKA 2), vidimo, da so v tem slovarju tri iztočnice: dve samostalniški za pomena ‘nasvet’ in ‘zemlja’ in pridevniška; besedne vrste iztočnic niso zapisane.⁵ Glede na odsotnost naglasa so te tri iztočnice vključene kot homonimi, kar pa ciljnim uporabnikom verjetno ni predstavljalo problema, ker so v slovarju iskali angleške ustreznike slovenskih iztočnic. Razlikovanju pomena samostalniških iztočnic služita usmerjevalni besedi ‘nasvet’ in ‘zemlja’, je pa usmerjevalnih besed v tem delu slovarja zelo malo. Pridevniku *svet* sledita obliki za ženski in srednji spol, kar je praksa tudi v SSKJ2, to pa je značilnost, po kateri se ta iztočnica razlikuje od samostalniških iztočnic. Na ta dva načina je Kubelka jasno prikazal razlike med iztočnicami in uporabnikom olajšal razumevanje pomenov večbesednih iztočnic.



SLIKA 2: Kubelka 1912: 405

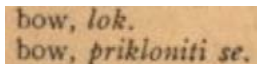
V angleško-slovenskem delu Kubelka iztočnico *lie* s tremi slovenskimi ustrezniki ‘laž, lagati, ležati’, ki jih ločujejo vejice, obravnava v enem slovarskem sestavku (1912: 253), čeprav gre v angleščini za tri homonime in dve besedni vrsti. Tudi v tem delu slovarja besedne vrste iztočnic niso navedene, iz oblike slovenskih ustreznikov pa je bilo slovenskim izseljencem jasno, kateri ustreznik naj izberejo. Glede na to, da je ta del slovarja namenjen dekodiranju, takšno vključevanje ustreznikov ni problematično. Seveda uporabniki niso izvedeli, da je *lie*, rabljen kot glagol, pravilen v pomenu ‘lagati’ in nepravilen v pomenu ‘ležati’, vendar pa ta podatek za dekodiranje ni potreben.

Čeprav v angleško-slovenskem delu izgovor angleških iztočnic ni naveden,⁶ je Kubelka homografa *bow* obravnaval v dveh slovarskih sestavkih glede na

⁵ V nobenem od preučevanih slovarjev besedna vrsta iztočnice ni zapisana. Enako velja tudi dvojezične slovarje, ki so izšli mnogo kasneje (npr. Grad 1971 in Škerlj, Komac 1991).

⁶ V slovarju za dekodiranje izgovor ni primarnega pomena, ga pa sodobni angleško-slovenski slovarji navajajo.

različen izgovor (SLIKA 3): /bou/ (samostalnik s pomenom 'lok') in /bau/ (glagol s pomenom 'prikloniti se').⁷ Na ta način je Kubelka jasno prikazal razliko v pomenu, uporabniki so ustrezni pomen z lahkoto poiskali, čeprav verjetno niso vedeli, zakaj sta v slovarju dva sestavka.



SLIKA 3: Kubelka 1912: 220

V nasprotju s slovarjem iz leta 1904 Kubelka v slovensko-angleškem delu slovarja iz leta 1912 za prikaz polisemije uporablja vejico, s katero ločuje ustreznike večpomenske iztočnice. Na primer:

goba, mushroom, sponge. (1912: 314)

Ta način vključevanja ustreznikov je problematičen pri enkodiranju, ki mu je slovensko-angleški del namenjen, kajti ponujena ustreznika 'mushroom' in 'sponge' se ne uporabljata v istem sobesedilu, uporabniki pa tega ne morejo predvideti oziroma tega ne vedo. Uporaba vejice za razlikovanje polisemičnih besed ni najbolj posrečena rešitev, saj uporabniki ne dobijo vseh informacij o polisemiji; zaradi tega Kubelka v redkih primerih uporablja tudi usmerjevalno besedilo. Na primer:

kazalec, forefinger, hand [of clock]. (1912: 327)

Ustreznika 'forefinger' in 'hand' nista sopomenki, poleg tega je samostalnik 'hand' v angleščini polisemična beseda, tako da je za uporabnika usmerjevalno besedilo v oklepaju koristno, saj izve, v katerem sobesedilu lahko uporabi ustreznik 'hand'.

V slovensko-angleškem delu Kubelke 1912 vejica ne ločuje samo polisemov, ampak tudi sopomenke. Na primer:

kupiti, to buy, to purchase. (1912: 333)

Glede na to, da Kubelka uporablja vejico tako v primeru polisemije kot tudi v primeru sinonimije, lahko zaključimo, da njegov prikaz pomenov iztočnic ni najprimernejši, uporabnost slovensko-angleškega dela slovarja pa je zaradi

⁷ Gl. tudi Tarp (2009: 302), ki pravi, da je homografe, ki niso homofoni, bolje obravnavati v različnih sestavkih.

tega okrnjena. Za prikaz polisemije resda uporablja tudi usmerjevalna besedila v oklepaju, vendar le zelo poredko in nikakor ne dosledno.

Večpomenske ustrezniške Kubelka 1912 tudi v angleško-slovenskem delu slovarja ločuje z vejico:

convince, *dokazati, prepričati*. (1912: 228)

V slovarskem sestavku iztočnice *hope*, ki je v angleščini tako samostalnik kot glagol, Kubelka kombinira konverzijo⁸ in sinonimijo:

hope, upanje, nada, upati, pričakovati. (1912: 245)

Dvema samostalniškima ustreznikoma sledita dva glagolska ustreznika, za ločevanje ustreznikov uporablja samo vejico, podpičja v tem slovarju ne uporablja. Besedne vrste niso označene, vendar lahko predvidevamo, da uporabniki – slovenski priseljenci v ZDA s tem niso imeli težav, ker so že iz oblike ustreznika vedeli, katerega morajo izbrati. V slovarskem sestavku iztočnice *cold*, ki je v angleščini tako samostalnik kot pridevnik, Kubelka kombinira konverzijo in polisemijo:

cold, prehlajenje, mraz, mrzel. (1912: 226)

2.3 KOŠUTNIK 1912: SLOVENSKO-ANGLEŠKI IN ANGLEŠKO-SLOVENSKI SLOVAR

Tako kot Kubelka tudi Košutnik v slovensko-angleškem delu slovarja ne označuje mesta naglasa ali kvantitete in kvalitete naglašene samoglasnika slovenske iztočnice. Za besedo *svet* (SLIKA 4) ima tri iztočnice: dve samostalniški in eno pridevniško. Pri prvi samostalniški iztočnici v oklepaju navede še glagol *svetovati*, ustreznika 'advice' in 'counsel' sta v angleščini lahko samostalnika ali glagola. Čeprav je navajanje glagola v slovarskem sestavku za samostalnik neobičajna praksa, pa to verjetno slovenskim priseljencem v ZDA ni povzročalo težav pri izbiri ustreznika. Poleg tega imamo lahko glagol *svetovati* za usmerjevalno besedo, s pomočjo katere Košutnik razlikuje obe samostalniški iztočnici. Pri drugi samostalniški iztočnici so trije ustrezniki ločeni z vejico. Pri pridevniški iztočnici Košutnik za razliko od Kubelke 1912 ne navaja oblik za ženski in srednji spol, kar bi to iztočnico že na prvi pogled razlikovalo od predhodnih samostalniških iztočnic. Tudi za Košutnikov slovar velja ugotovitev, da bi bilo večje število usmerjevalnih besedil koristno za uporabnike.

⁸ Tudi v kasnejših slovarjih (npr. Grad 1971 in Škerlj, Komac 1991) so primeri konverzije obravnavani znotraj enega geselskega članka.

**svet (svetovati) - advice,
counsel.
svet - world, universe, people.
svet - holy, sacred, inviolable.**

SLIKA 4: Košutnik 1912: 52

V angleško-slovenskem delu ima Košutnik dve iztočnici za *lie* (1912: 111), ki ju razlikuje glede na pomen in ne glede na besedno vrsto (samostalnik, glagol):

lie, laj – laž, lug; lagati.

lie, laj – ležati, spati.

Pri prvi iztočnici najprej navaja samostalniška ustreznika ‘laž’ (pravilni ustreznik) in ‘lug’ (nepravilni ustreznik); pri ustrezniku ‘lug’ gre za napako, ker je angleška beseda *lye* ‘lug’ homofon besede *lie*. Samostalniška ustreznika sta s podpičjem ločena od glagolskega ustreznika ‘lagati’. Pri drugi iztočnici besede *lie* navaja dva glagolska ustreznika. Z delitvijo na dve iztočnici Košutnik jasno pokaže, da ima *lie* različne med seboj nepovezane pomene, vendar bi bilo to razliko mogoče pokazati tudi v okviru enega slovarskega sestavka. Problematične so napake pri izboru ustreznikov, ki smo jih opazili pri obeh zgoraj navedenih primerih.

Pri prikazu iztočnice *bow* Košutnik navaja samo glagolski pomen ‘vpogniti, vpogniti se’ (1912: 72), ne pa tudi ustreznika ‘prikloniti se’, ki ga navaja Kubelka 1912, in bi ga bilo tudi smiselno vključiti v slovar. Ne glede na izbiro ustreznikov lahko zaključimo, da je prikaz iztočnice *bow* v Kubelki 1912 ustrežnejši, ker Košutnik ne obravnava obeh homografov, ampak samo *bow* z izgovorom /bau/.

Podobno kot Kubelka 1912 tudi Košutnik sopomenske ustreznike ločuje z vejico (*colour* in *hue* v primeru spodaj), vendar včasih z vejico ločuje tudi večpomenske ustreznike (*paint* in *printer’s ink* v primeru spodaj sta večpomenski v razmerju drug do drugega kot tudi v razmerju do *colour* in *hue*). Glede na to, da občasno za ločevanje polisemije uporablja tudi podpičje, bi bilo v primeru spodaj verjetno bolje uporabiti podpičje pred ustreznikom *paint* in pred ustreznikom *printer’s ink*:

barva – colour, hue, paint, printer’s ink. (1912: 5)

Pri prikazu iztočnic *breskva* in *figa* se prvi ustreznik nanaša na sadež, drugi pa na drevo, vendar je Košutnik izbral različne načine prikaza:

breskva – peach, peach-tree. (1912: 7)

figa – fig; (drevo) – fig-tree. (1912: 12)

Pri iztočnici *breskva* ustreznika, ločena z vejico, nista sopomenki; podobno velja tudi za iztočnico *figa*, vendar pri tej iztočnici Košutnik z uporabo podpičja jasno pokaže, da sta ustreznika polisemična. Poleg tega uporabi še usmerjevalno besedo 'drevo', s pomočjo katere pri enkodiranju uporabnik lažje izbere primeren ustreznik. Pri iztočnici *figa* Košutnik uporablja pomišljaj, ki sledi usmerjevalni besedi, podobno opazimo tudi pri iztočnici *breg*, kjer pomišljaj sledi usmerjevalnemu besedilu pred obema ustreznikoma:

breg (pri reki) – bank; (breg pri morju) – shore. (1912: 7)

Košutnik z uporabo pomišljaja in podpičja pokaže, da se pomeni angleških ustreznikov razlikujejo in da ne gre za sopomenke. V primeru iztočnice *breg* usmerjevalno besedilo v oklepajih uporabnika opozarja na določen pomen iztočnice; iztočnica *breg* je del usmerjevalnega besedila v drugem oklepaju, v prvem pa ne. Košutnik pri navajanju cele iztočnice v usmerjevalnem besedilu ni dosleden, kar je razvidno tudi iz prikaza iztočnice *brazda*, ki jo v usmerjevalnem besedilu okrajša z b.:

brazda – furrow; (b. v obrazu) – wrinkle. (1912: 7)

Polisemija v slovensko-angleškem delu je v Košutniku 1912 obravnavana veliko bolj nazorno kot v Kubelki 1904 in 1912, saj Košutnik za razlikovanje različnih pomenov uporablja vejice, podpičja, pomišljaje in usmerjevalno besedilo.

Iz prikaza iztočnice *labour* v angleško-slovenskem delu Košutnika 1912 vidimo, kako obravnava polisemijo, konverzijo in sinonimijo:

labour, le'ber – delo, trud, napor delati, truditi se, v sili biti, obdelovati, izdelovati (1912: 109)

V angleščini je *labour* konverzni leksem, ki pripada dvema različnima besednima vrstama, kar je razvidno tudi iz ustreznikov (prva dva ustreznika sta samostalnika, ostali ustrezniki so glagoli). Košutnik vse ustreznike ločuje z vejico ne glede na besedno vrsto in na pomen. Samostalniška ustreznika 'delo' in 'trud' sta večpomenki, glagolska ustreznika 'napor delati' in 'truditi se' sta sopomenki, vsi ostali glagolski ustrezniki pa so večpomenke v razmerju drug

do drugega in tudi v razmerju do sopomenskih ustreznikov 'napor delati' in 'truditi se'. Za konverzne lekseme je značilno, da je ohranjena medsebojna pomenska povezanost (npr. med samostalniškim ustreznikom 'trud' in glagolskima ustreznikoma 'napor delati' in 'truditi se'). Košutnik z uporabo različnih ustreznikov zajame celotni pomen iztočnice. V angleško-slovenskem delu, ki je slovenskim priseljencem v ZDA pomagal pri razumevanju angleščine, so različni ustrezniki (večpomenske) iztočnice *labour* omogočali uporabnikom izbiro ustreznika, ki je bil v danem sobesedilu najustreznejši, z veliko verjetnostjo pa lahko trdimo, da jih niso zanimala kompleksna razmerja med ustrezniki, ki bi jih bilo treba reševati na ustrezen leksikografski način (npr. dva slovarska sestavka, uporaba podpičja za ločevanje polisemije ipd.).

2.4 KERN 1919: ANGLEŠKO-SLOVENSKI SLOVAR

Kern v angleško-slovenskem slovarju iz leta 1919 obravnava iztočnico *lie* (SLIKA 5) v dveh slovarskih sestavkih, besednih vrst iztočnice pa ne navaja:

lie, (laj), (lay, lain), ležati; položiti.
lie, (laj), laž; lagati.

SLIKA 5: Kern 1919: 139

V prvem slovarskem sestavku za nepravilni glagol *lie* ponudi ustreznika 'ležati' in 'položiti', ki sta ločena s podpičjem. Običajna praksa v sodobnih dvojezičnih angleško-slovenskih slovarjih je, da sta dovršni in nedovršni glagol navedena kot ustreznika in ločena z vejico, tako da je Kernova raba podpičja v tem primeru z današnje perspektive nenavadna. Problematičen je tudi ustreznik 'položiti', ki ga sodobni angleško-slovenski slovarji ne navajajo kot ustreznik glagola *lie*, kajti *lie* je v angleščini pomensko gledano nedovršen. V drugem slovarskem sestavku za *lie* sta samostalniški ustreznik 'laž' in glagolski ustreznik 'lagati' ločena s podpičjem. Prikaz dveh različnih besednih vrst v okviru enega sestavka je ustrezen za uporabnike, ki jih ne zanima vprašanje homonimije in polisemije, z rabo podpičja pa so vseeno opozorjeni na razliko med obema ustreznikoma.

Kern v angleško-slovenskem slovarju vključuje izgovor angleških iztočnic, kar je še posebej pomembno pri homografih. Homograf *bow* (SLIKA 6) Kern obravnava v enem slovarskem sestavku, za razlikovanje pomenov pa uporablja izgovor in podpičje. Izgovor je prilagojen tedanjemu ciljnemu uporabniku, Kern ga skuša zapisati na čim lažji način brez uporabe posebnih znakov,

značilnih za sodobni zapis izgovora angleških besed. Prvemu izgovoru (v sodobnem fonetičnem zapisu /bau/) sledi glagolski ustreznik, ki je s podpičjem ločen od dveh sopomenskih samostalniških ustreznikov, ta dva ustreznika pa sta s podpičjem ločena od samostalniškega ustreznika ‘ospredje ladje’, ki je večpomenka in je pravzaprav opisni ustreznik.⁹ Glagolski ustreznik ‘upogniti [se]’ je enak kot v Košutniku 1912, le da ga Košutnik zapiše z *v* in ne *u*. Samostalniška ustreznika ‘priklon, poklon’ sta pomensko povezana z glagolom ‘prikloniti se’, ki ga kot ustreznik navaja Kubelka, ne pa tudi z ustreznikom ‘upogniti [se]’, ki ga najdemo v Kernovem slovarju.

bow, (baw), upogniti [se]; priklon, poklon; ospredje ladje; (bō)— lok; zanjka; pentlja.

SLIKA 6: Kern 1919: 26

V istem slovarskem sestavku navaja Kern še drugi izgovor iztočnice *bow* (v sodobnem fonetičnem zapisu /bou/) s tremi polisemičnimi ustrezniki, ki so med seboj ločeni s podpičjem. ‘Lok’ (navaja ga tudi Kubelka 1912) in ‘pentlja’ sta pravilna ustreznika, ‘zanjka’ pa ne ustreza pomenu angleške iztočnice.

Kot je razvidno iz slovarskega sestavka za *bow*, uporablja Kern podpičje za razlikovanje besednih vrst (glagol, samostalnik), za polisemične pomene samostalnika in za razlikovanje homografov. Ne glede na to, kaj jezikoslovci razumejo pod homonimijo, homografijo ali polisemijo, je bila Kernova uporaba podpičja za razlikovanje različnih pomenov iztočnice za tedanjega uporabnika najverjetneje ustrezna.

Kern za razliko od Kubelke 1912 in Košutnika 1912 pogosto in relativno dosledno uporablja podpičje za prikaz polisemije:

abuse, (ebjūz’), zlorabiti; prevariti; grditi; (ēbjūs’) – zloraba; napad; razžaljenje. (1919: 2)

V slovarskem sestavku iztočnice *abuse* Kern s podpičjem ne ločuje samo polisemov, ampak tudi izgovor konverznih leksemov (glagol, samostalnik). Samostalniški ustreznik ‘zloraba’ je pomensko povezan z glagolskim ustreznikom ‘zlorabiti’. ‘Prevariti’ v sodobni angleščini pomensko ni povezan z

⁹ Možno je, da v času sestavljanja tega slovarja strokovnega izraza *premec* še ni bilo; gl. Snój 2015, ki navaja kot čas nastanka termina *premec* 20. stoletje.

drugimi ustrezniki, ker gre za zastarelo rabo glagola *abuse*. Ustreznika ‘napad’ v sodobnih angleško-slovenskih slovarjih ne najdemo, ustreznik ‘razžaljenje’ pa je pomensko povezan z glagolom ‘grditi’ v pomenu ‘grdo, slabo govoriti o kom’; manjka ustreznik ‘razžaliti’, ki je pomensko že na prvi pogled povezan z ustreznikom ‘razžaljenje’. Pri Kernu najdemo tako pare glagolov in samostalnikov, kjer je ohranjena medsebojna pomenska povezanost (npr. ‘zlorabiti’ in ‘zloraba’), kot tudi večpomenske glagolske in samostalniške ustreznike. Med analiziranimi slovarji se Kernov slovar odlikuje po prikazu pomenov iztočnic in po številu ustreznikov, zaradi česar lahko s precejšnjo verjetnostjo rečemo, da je bil za ciljne uporabnike najbolj uporaben, čeprav mogoče niso vedno razumeli, zakaj je Kern določeno informacijo vključil v slovar na določen način.

3 RAZPRAVA

Preden se lotimo sestavljanja slovarja, moramo določiti način vključevanja enako pisanih iztočnic. Ali jih bomo vključili kot eno ali kot dve iztočnici? Ali moramo iztočnico obravnavati kot dve različni besedi? V tem primeru gre za homonimijo, ki sodi v makrostrukturo slovarja (samostojne iztočnice). Ali ima iztočnica različne pomenne in torej tudi različne ustreznike v ciljnem jeziku? V tem primeru gre za polisemijo, ki jo obravnavamo v mikrostrukturi slovarja (več pomenov ene iztočnice). Obstaja več možnih načinov reševanja problema polisemije in homonimije: zgodovinski, semantični, morfosemantični in formalnoslovnični pristop (Svensén 2009: 94–101). Svensén (2009: 101) se ne zavzema za tradicionalni zgodovinski pristop do homonimije in polisemije, ker uporabniki ne poznajo izvora besed, zaradi česar ne razumejo informacij, ki so v slovar vključene na zgodovinski osnovi, prav tako pa se leksikografska interpretacija, pri kateri je etimologija merilo razločevanja med homonimijo in polisemijo, velikokrat razlikuje od sinhrono govorne predstave (Gantar 2015: 256). Tudi v Uvodu v SSKJ2 (2014: 28, § 17) je razloženo, da se homonimna postavitve iztočnic ne opira na etimologijo (torej ne gre za zgodovinski pristop), ampak na stanje v sodobnem jeziku; težnja v SSKJ2 je, da je homonimnih iztočnic čim manj. Ne glede na pristop, uporabljen v določenem slovarju, je za Svenséna (2009) reševanje problema homonimije in polisemije pomembno le za sestavljalce slovarja, za uporabnika pa je popolnoma irelevantno (gl. tudi Gantar 2015: 356). Na osnovi analize štirih slovarjev za priseljence iz prvih dveh desetletij dvajsetega stoletja delimo Svensénovo mnenje, saj slovarji za

priseljence ne bi smeli biti osnovani na teoretičnih jezikoslovnih in leksikografskih premisah, ki so uporabnikom popolnoma nepoznane. Svensén (2009: 101) zagovarja vključevanje homonimov glede na semantične razlike, saj meni, da bo slovar, v katerem je mikrostruktura preprostejša, makrostruktura pa bogatejša, prispeval k boljši preglednosti in bo tudi pedagoško učinkovitejši. Glede na analizo slovarjev v naši raziskavi bi bilo primerneje, da v geslovník ne bi vključevali enako pisanih iztočnic, ampak bi semantične informacije obravnavali v okviru enega slovarskega sestavka.

Kubelka, Košutnik in Kern so bili ljubiteljski sestavljavci dvojezičnih slovarjev, ki o jezikoslovnih in leksikografskih obravnavah homonimije in polisemije niso vedeli veliko ali pa sploh nič. Slovarje so najverjetneje poznali le kot uporabniki, vendar o tem nimamo verodostojnih podatkov. V skladu s Svensénom (2009) smo štiri slovarje analizirali na podlagi tega, ali vključujejo več enako pisanih iztočnic (homonimični pristop) in imajo torej bogatejšo makrostrukturo, ali pa vključujejo informacije o pomene iztočnice v okviru enega slovarskega sestavka (polisemični pristop) in imajo torej kompleksno mikrostrukturo. Poudariti moramo, da nismo ocenjevali dejanske obravnave homonimije ali polisemije v konkretnem slovarju. Osredotočili smo se na to, ali so homonimne iztočnice rabljene dosledno ali ne in kako bi uporabnik interpretiral oz. razumel informacije, vključene v slovar, pri čemer smo se skušali postaviti v vlogo takratnega uporabnika. Tu gre zgolj za ugibanja, ker ne razpolagamo z rezultati študij rabe slovarjev v tistem času, saj take študije ne obstajajo, tako da je zelo težko ugotavljati, česa tedanji uporabniki dejansko niso razumeli, s čim so imeli težave in kaj se jim je zdelo v slovarju nejasno. Poleg tega smo preučevali tudi kompleksne iztočnice s homonimi, različnimi pomeni in sopomenkami, pri čemer smo se spraševali, ali bi povprečni slovenski priseljenec v ZDA prišel do potrebnih informacij.

3.1 SLOVENSKO-ANGLEŠKI SLOVARJI

Kubelka je v slovar iz leta 1904 za razliko od slovarja iz leta 1912 in tudi za razliko od Košutnika 1912 pogosto vključeval enako pisane iztočnice (makrostrukturni pristop). Za razlikovanje enako pisanih iztočnic je uporabljal različne načine (ponavljanje, zaviti oklepaji, ponavljanje iztočnice), pa tudi namen vključevanja enako pisanih iztočnic je bil različen (prikaz »tradicionalnih« homonimov, prikaz polisemije). Težko je verjeti, da so ciljni uporabniki tega slovarja dejansko razumeli Kubelkov makrostrukturni pristop.

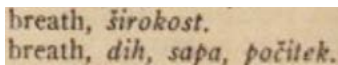
V slovarju iz leta 1912 je Kubelka uporabljal drugačen način prikaza iztočnic, saj je obravnava usmerjena v mikrostrukturo, makrostruktura pa je manj bogata. Čeprav tudi v ta slovar vključuje enako pisane iztočnice, jih je manj kot v Kubelki 1904, prav tako jih ne uporablja za prikaz polisemije kot v Kubelki 1904. Verjetno je tak mikrostrukturni pristop v slovensko-angleškem delu Kubelke 1912 za uporabnike lažje razumljiv kot makrostrukturni pristop v Kubelki 1904.

Podobno kot Kubelka 1912 tudi Košutnik v slovarju iz leta 1912 uporablja v glavnem mikrostrukturni pristop. Vključuje le malo enako pisanih iztočnic in v slovarskem sestavku uporablja na pomenu temelječ pristop k prikazu informacij o polisemiji in/ali sinonimiji ustreznikov, kar je uporabniku prijazen način vključevanja informacij.

3.2 ANGLEŠKO-SLOVENSKI SLOVARJI

Pri vključevanju enako pisanih iztočnic v tri analizirane slovarje je praktično nemogoče ugotoviti, po katerih kriterijih so jih avtorji vključevali, obstajajo pa razlike med slovarji v številu tovrstnih iztočnic.

V Kubelki 1912 je vključevanje enako pisanih iztočnic izredno redko, vendar je bolj ali manj dosledno in temelji v glavnem na izgovoru. Mikrostrukturni pristop je opaznejši v angleško-slovenskem delu Kubelke 1912 kot v slovensko-angleškem delu. Kot smo videli, Kubelka deli *bow* na dve iztočnici glede na izgovor (v slovarju ni zapisan) ali pa glede na besedno vrsto (v slovarju ni zapisana). Podobno velja tudi za prikaz dveh iztočnic za *sow* z različnim izgovorom samostalnika in glagola. Pri delitvi *lead* na dve iztočnici je kriterij izgovor in ne besedna vrsta, dokaz za to so samostalniški in glagolski ustrezniki pri iztočnici z izgovorom /li:d/; podobno velja tudi za prikaz dveh iztočnic za *row*. V Kubelki 1912 smo samo pri samostalniku *breath* našli delitev na dve enako pisani iztočnici, ki temelji na dejanski homonimiji (SLIKA 7).



SLIKA 7: Kubelka 1912: 220

Tu sploh ne gre za homonimijo, ker je Kubelka pri zapisu prve iztočnice naredil pravopisno napako, saj bi moral napisati *breadth* in ne *breath*.

Kljub temu, da je v Kubelki 1912 izgovor relativno dosledno uporabljen kriterij za sicer redko rabo več enako pisanih iztočnic, pa ta kriterij uporabniku

ni jasen, ker izgovor v slovarju ni naveden. Verjetno bi bilo za ciljne uporabnike razumljiveje in lažje, če bi bil pomen prikazan znotraj enega slovarskega sestavka.

V Košutniku 1912 je več enako pisanih iztočnic kot v Kubelki 1912; Košutnik uporablja tako mikro- kot tudi makrostrukturni pristop. V angleško-slovenskem delu Košutnika 1912 je več enako pisanih iztočnic kot v slovensko-angleškem delu, kar je razumljivo, če upoštevamo dejstvo, da je zaradi konverzije določanje besedne vrste v angleščini težje kot v slovenščini. Košutnik pri delitvi na več enako pisanih iztočnic ni dosleden, delitev največkrat temelji na besedni vrsti, ki v slovarju ni navedena (npr. samostalnik in glagol *bank* in *brook*; pridevnik in glagol *blank*, *blind* in *lean*). Raziskave rabe slovarjev so pokazale, da imajo tudi dandanes celo izobraženi uporabniki težave s prepoznavanjem besedne vrste (Farina, Vrbinc, Vrbinc 2019, Vrbinc, Farina, Vrbinc 2022 in Vrbinc, Vrbinc, Farina 2023); z veliko gotovostjo lahko predpostavljamo, da so imeli podobne težave tudi (večinoma slabo izobraženi) priseljenci v Košutnikovem času. Delitev na več iztočnic ne temelji samo na besedni vrsti. Pri delitvi *lie* (gl. 2.3) na dve iztočnici gre za homonimijo, kar ni pogosto razlog za več iztočnic. Tako kot v Kubelki 1912 je tudi v Košutniku 1912 pri delitvi *lead* na dve iztočnici kriterij izgovor, kar dokazuje vključevanje samostalniških in glagolskih ustreznikov pri eni od iztočnic. Zaključimo lahko, da so uporabniki v Košutnikovem slovarju težje našli iskane informacije zaradi delitve na več enako pisanih iztočnic brez razumljivih in doslednih kriterijev, ki niso nikjer razloženi.

Kern od vseh avtorjev preučevanih slovarjev uporablja najbolj makrostrukturni pristop in v svoj slovar vključuje največ enako pisanih iztočnic, vendar tudi on uporablja različne in ne povsem jasne kriterije za vključevanje več enako pisanih iztočnic. Pri delitvi *lie* na dve iztočnici uporablja kot kriterij homonimijo (gl. 2.4), vendar je ta kriterij le redko rabljen. Podobno kot v Košutniku 1912 sta tudi v Kernovem slovarju najbolj pogosta kriterija za delitev na več enako pisanih iztočnic besedna vrsta, ki v slovarju ni navedena, in izgovor.

Glede na to, da je v Kernov slovar vključen izgovor, bi bil to lahko kriterij za vključevanje več enako pisanih iztočnic, vendar pa je Kern za več enako pisanih iztočnic uporabljal različne pristope, ki niso nikjer razloženi. Večini ciljnih uporabnikov tega slovarja verjetno ni bilo jasno, zakaj se določena iztočnica pojavi v geslovniku večkrat. Vključevanje več enako pisanih iztočnic brez (doslednih in pojasnjenih) kriterijev zagotovo ni ena od prednosti Kernovega

slovarja, je pa med analiziranimi slovarji Kernov sestavljen najbolj profesionalno. Kot je bilo že omenjeno (gl. 2.4), je Kern v okviru slovarskega sestavka dobro prikazal pomen(e) iztočnice, vključil pa je tudi več ustreznikov kot Kubelka in Košutnik. Možno je, da je posnemal enojezične slovarje za rojene govorce angleščine,¹⁰ ki jih je poznal in uporabljal, vendar v svojem dvojezičnem slovarju ni upošteval težav, ki so jih imeli ciljni uporabniki – slovenski priseljenci z razumevanjem strukture in sestavnih delov slovarja.

4 ZAKLJUČEK

V angleških enojezičnih slovarjih za tujejezične govorce namenjajo veliko pozornost vključevanju in prikazu različnih informacij, ki uporabniku pomagajo pri razumevanju in rabi tujega jezika. To je še toliko pomembneje za uporabnike, ki prihajajo iz kulturnih okolij, ki se (zelo) razlikujejo od kulturnega okolja jezika, ki se ga učijo (gl. Tarp 2004). Poleg tega angleški enojezični slovarji za tujejezične govorce ne posvečajo več toliko pozornosti elementom, kot so oznake besednih vrst, homonimija ipd., ampak je v središču pozornosti prikaz pomena oz. pomenov iztočnice. Béjoint (2010: 178, 183) omenja prvo izdajo slovarja COBUILD iz leta 1987, kjer je edini kriterij za razvrščanje pomenov pogostnost in ne besedna vrsta; homonimi so obravnavani v okviru istega slovarskega sestavka. Landau (2001: 101) poudarja, da je v leksikografski praksi, še posebej v komercialni leksikografiji, razlikovanje med homonimijo in polisemijo povsem nepomembno, saj je v središču zanimanja predstavitev jezika na način, ki se bo zdel uporabnikom slovarja smiseln in razumljiv, podobno stališče pa zagovarjajo tudi Tarp (2009) ter S. Atkins in Rundell (2008: 281). S. Atkins in Rundell (2008) pravita, da je homonimija v sodobni leksikografiji postala nepomembna pri organizaciji makrostrukture, kar še posebej velja za enojezične slovarje za tujejezične govorce, ki uporabljajo polisemični pristop. R. Moon (1987: 88) meni, da v slovarje za tujejezične govorce ni smiselno vključevati homonimov, saj uporabniki iščejo iztočnice glede na pravopis besede, o homonimiji, ki temelji na etimologiji, pa ne vedo nič. Tudi P. Gantar (2015: 346) ugotavlja, da je homonimija primarno slovarski problem, sestavljavci slovarja pa se morajo odločiti, kako bodo prikazali izrazno enoto v odnosu do

¹⁰ Kern je edini, ki v slovenskem uvodu omenja, da je njegov dvojezični slovar primeren za priseljence, ki se učijo angleščino, vsem tistim, ki angleščino že bolje obvladajo, pa svetuje, naj raje uporabljajo *New Standard Dictionary* (prvič izšel leta 1913) ali »Webster«.

njene vsebine v slovarju in kako bodo povezali posamezne besedne oblike s posameznimi pomeni, pri čemer se predvsem slovarji za tujejezične govorce in slovarji za določene uporabniške skupine, kamor lahko prištevamo tudi slovarje za priseljence, nagibajo k reševanju uporabniških problemov in ne k prenosu teoretičnih izhodišč v slovarsko prakso.

Enojezični slovarji za tujejezične govorce in dvojezični slovarji so namenjeni istim ciljnim uporabnikom, med katere sodijo tudi priseljenci. Kot je bilo že omenjeno (1.3), ne obstaja veliko znanstvenih prispevkov, ki bi obravnavali slovarje za priseljence, zaradi česar tudi ne moremo govoriti o optimalni zasnovi tovrstnih slovarjev. Štirje slovarji, ki smo jih preučevali v naši raziskavi in ki so nastali pred več kot 100 leti, dobro ponazarjajo težave z zasnovo in prikazom pomena oz. pomenov iztočnice. Z bolj ali manj podobnimi težavami se spopadajo tudi sestavljalci sodobnih slovarjev za priseljence, čeprav uporabljajo različne tehnike prikaza pomena oz. pomenov (gl. Tarp 2004: 27–28). Pri preučevanju štirih slovarjev smo ugotavljali, da elementi prikaza, ki ne prispevajo k neposrednemu razumevanju pomena iztočnice, predstavljajo velik problem za uporabnike. Zaključimo lahko, da se morajo (eno- in dvojezični) slovarji za priseljence v prvi vrsti osredotočiti na pomen, kar pomeni, da mora biti mikrostrukturni prikaz podroben, natančen in dosleden, slovarski sestavek pa mora vključevati več pomenov in/ali več sopomenk.

ZAHVALE

Projekt, katerega izsledki so predstavljeni v pričujočem članku, je finančno podprla Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije, in sicer v okviru bilateralnih projektov »Stare besede, nove besede, novi svet: Življenje slovenskih leksikografov – priseljencev v ZDA« (BI-US/22-24-030) ter »Slovarji in izkušnje Slovencev – emigrantov v Združenih državah Amerike (konec 19. in začetek 20. stoletja)« (BI-US/22-24-042). Posebna zahvala gre dvema anonimnima recenzentoma, ki sta natančno prebrala rokopis članka in predlagala izboljšave, ki so znatno prispevale h končni podobi članka.

REFERENCE

Med viri je Kern naveden pod Kern ali pod Javh-Kern glede na to, s katerim priimkom je bil naveden v viru.

SLOVARJI

Kern, Frank J. 1919. *A complete pronouncing dictionary of the English and Slovene languages for general use. Popoln angleško-slovenski besednjak z angleško izgovarjavo*. Cleveland, Ohio: Tisk ameriške domovine.

Košutnik, Silvester. 1912. *Ročni slovensko-angleški in angleško slovenski slovar: Zlasti namenjen izseljencem v Ameriko*. Ljubljana: Anton Turk.

Kubelka, Victor J. 1904. *Slovensko-angleški žepni rečnik v olajšavo naučenja obeh jezikov. Slovenian-English pocket dictionary to facilitate the study of both languages*. New York. [https://babel.hathitrust.org/cgi/pt?id=loc.ark:/13960/t4qj8p026&view=1up&seq\(20.12.2023\)](https://babel.hathitrust.org/cgi/pt?id=loc.ark:/13960/t4qj8p026&view=1up&seq(20.12.2023)).

Kubelka, Viktor J. 1912. *Slovensko-angleška slovnica, tolmač, spisovnik in navodilo za naturalizacijo. Angleško-slovenski in slovensko-angleški slovar. Slovenian-English grammar, interpreter, letterwriter and information on naturalization. English-Slovenian and Slovenian-English dictionary*. New York: Viktor J. Kubelka.

Slovar slovenskega knjižnega jezika. Druga, dopolnjena in deloma prenovljena izdaja (SSKJ2). 2014. www.fran.si (18. 12. 2023).

Snoj, Marko. 2015. *Slovenski etimološki slovar*. Ljubljana: Založba ZRC.

DRUGI VIRI

Atkins, B. T. Sue, Rundell, Michael. 2008. *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.

Béjoint, Henri. 2010. *The lexicography of English*. Oxford: Oxford University Press.

Cvrček, Tomáš. 2020. Schooling under control: The origins of public education in imperial Austria 1769-1869. Tübingen: Mohr Siebeck.

Farina, Donna M. T. Cr., Vrbinc, Marjeta, Vrbinc, Alenka. 2019. Problems in online dictionary use for advanced Slovenian learners of English. *International Journal of Lexicography* 32/4: 458–479. <https://doi.org/10.1093/ijl/ecz017>

Fjeld, Ruth Vatvedt. 2012. Et forsømt kapittel i leksikografiens historie: Leachman's Norsk-Engelsk Lomme-Ordbog/Norwegian-English pocket dictionary. *Norsk Lingvistisk Tidsskrift* 30/2: 151–164.

Gantar, Polona. 2015. Homonimija in večpomenskost: od teorije do slovarja. V: V. Gorjanc, P. Gantar, I. Kosem, S. Krek, M. Bratanić, W. Browne, V. Cvrček (ur.). *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete. 340–357.

Grad, Anton. 1971. *Angleško-slovenski slovar*. Maribor: Obzorja.

- Granatir Alexander, June. 2009. *Daily life in immigrant America, 1870–1920: How the second great wave of immigrants made their way in America*. Chicago: Ivan R. Dee.
- Grønvik, Oddrun. 2016. The lexicography of Norwegian. V: P. Hanks, G.-M. de Schryver (ur.). *International handbook of modern lexis and lexicography*. Berlin, Heidelberg: Springer-Verlag. 1–34. https://doi.org/10.1007/978-3-642-45369-4_44-1 (10. 11. 2023).
- Halas Popović, Ana V. 2021. Homonymy in English monolingual general-purpose dictionaries. *Nasleđe* 48: 37–51.
- Javh-Kern, Frank. 1937. *Memoirs on the occasion of the thirtieth anniversary of arrival in America*. [Govor na Slovenskem mednarodnem rodoslovnem združenju]
- Kalc, Aleksej, Milharčič Hladnik, Mirjam, Žitnik Serafin, Janja. 2020. *Doba velikih migracij na Slovenskem*. Založba ZRC.
- Karpova, Olga. 2019. Modern English dictionaries. A foreign user's view. *Respectus philologicus* 36/41: 11–18.
- Karpova, Olga M., Kulagina, Margarita A. 2009. What type of LSP dictionary do immigrants need in a global world? V: O. Karpova, F. Kartashkova (ur.). *Lexicography and terminology: A worldwide outlook*. Newcastle upon Tyne: Cambridge Scholars Publishing. 169–179.
- Komac, Daša, Škerlj, Ružena. 1991. *Angleško-slovenski in slovensko-angleški moderni slovar*. Ljubljana: Cankarjeva založba.
- Landau, Sidney I. 2001. *Dictionaries: The art and craft of lexicography*. Cambridge: Cambridge University Press.
- Legan Ravnikar, Andreja. 2021. Konverzija v slovnici in slovarju: izhodišča za prepoznavanje konverzije v leksikografski praksi. *Slavia Centralis* 14/2: 177–194.
- Lexin. <https://lexin.oslomet.no/#/> (22. 7. 2023).
- Lipka, Leonhard. 2002. *English lexicology*. Tübingen: Gunter Narr Verlag.
- Malmgren, Sven-Goran. 1999. De svenska LEXIN-ordböckerna ur ett (ställföreträdande) användarperspektiv. *LexicoNordica* 6: 79–90.
- Moon, Rosamund E. 1987. The analysis of meaning. V: J. M. Sinclair (ur.). *Looking up: An account of the COBUILD project in lexical computing*. London: Collins. 86–103.
- Stanonik, Janez. 1996. Slovenci v Združenih državah: obdobje 1848–1891. *Dve domovini* 7: 113–129.
- Svensén, Bo. 2009. *A handbook of lexicography: The theory and practice of dictionary-making*. Cambridge: Cambridge University Press.
- Tarp, Sven. 2004. Basic problems of learner's lexicography. *Lexikos* 14: 222–252.
- Tarp, Sven. 2009. Homonymy and polysemy in a lexicographical perspective. *Zeitschrift für Anglistik und Amerikanistik* 57/3: 289–305.
- Vacalopoulou, Anna, Gioulu, Voula, Giagkou, Maria, Efthimiou, Eleni. 2011. Online dictionaries for immigrants in Greece: Overcoming the communication barriers. V: I. Kosem, K. Kosem (ur.). *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 2011, 10–12 November 2011, Bled, Slovenia*. Ljubljana: Trojina, Institute for Applied Slovene Studies. 274–279.

- Vrbinc, Marjeta, Farina, Donna M. T. Cr., Vrbinc, Alenka. 2022. Who is sharpest at looking up sharp?: Comparing two parallel groups of dictionary users. *Lexikos* 32: 368–391.
- Vrbinc, Marjeta, Farina, Donna M. T. Cr., Vrbinc, Alenka. 2024. Oris slovarjev in priročnikov za učenje angleščine slovenskih izseljencev v ZDA v obdobju 1896–1919. *Dve domovini/Two Homelands* 59: 179–202.
- Vrbinc, Marjeta, Vrbinc, Alenka, Farina, Donna M. T. Cr. 2023. The lexicographer's dream audience: Dictionary use among English majors at a Slovenian university. *International Journal of Lexicography* 36/2: 149–172.
- Žele, Andreja. 2015. Konverzija v slovenščini. *Jezik in slovstvo* 60/ 2: 65–77.

Prispelo januarja 2024, sprejeto marca 2024.

Received January 2024, accepted March 2024.

POVZETEK

HOMONIMIJA IN POLISEMIJA V DVOJEZIČNIH SLOVARJIH ZA SLOVENSKE PRISELJENCE V ZDA V ZAČETKU 20. STOLETJA

V članku obravnavamo slovarje, ki so jih imeli na voljo slovenski priseljenci v ZDA v času množičnega izseljevanja od leta 1860 do leta 1915. Priseljenci so se morali naučiti angleško in se prilagoditi novemu kulturnemu okolju, pri čemer so si pomagali s priročniki, izdanimi v ZDA in na Slovenskem, v katerih niso dobili samo informacij o angleščini, temveč tudi informacije o ameriški ustavi in vladi, denarju in gospodarstvu. Med temi publikacijami najdemo slovnice, slovarje, učbenike, priročnike s koristnimi stalnimi večbesednimi izrazi iz vsakdanjega življenja in otroške abecednike. Številne tovrstne publikacije sodijo v okvir zgodovine (pedagoške) leksikografije, ki je trenutno še relativno slabo raziskana. V prispevku bomo podrobneje obravnavali štiri slovensko-angleške oz. angleško-slovenske slovarje (Kubelka 1904, Kubelka 1912, Košutnik 1912 in Kern 1919), ki so izšli v prvih dveh desetletjih dvajsetega stoletja, to je v obdobju, ko se je v ZDA izselilo največ Slovencev. V raziskavi smo se osredotočili na prikaz informacij v omenjenih štirih slovarjih s poudarkom na homonimiji, polisemiji, konverziji in usmerjevalnih besedah, zapisanih v oklepajih.

Slovensko-angleški slovarji: Kubelka vključuje v slovar iz leta 1904 več enako pisanih iztočnic, ki jih prikaže z uporabo ponavljaja, zavitega oklepaja in s ponavljanjem iztočnice. V slovarju iz leta 1912 Kubelka uporablja drugačen način prikaza iztočnic, saj je obravnava usmerjena v mikrostrukturo, makrostruktura pa je manj bogata. Tudi v ta slovar vključuje enako pisane iztočnice, vendar jih je številčno manj kot v slovarju iz leta 1904. V Košutniku 1912 je v glavnem uporabljen mikrostrukturni

pristop, enako pisanih iztočnic je malo, v slovarskem sestavku pa Košutnik uporablja na pomenu temelječ pristop prikaza informacij o polisemiji in/ali sinonimiji ustreznikov.

Angleško-slovenski slovarji: V Kubelki 1912 je vključevanje enako pisanih iztočnic izredno redko, temelji pa v glavnem na izgovoru. Košutnik 1912 ima več enako pisanih iztočnic kot Kubelka 1912, za delitev na več enako pisanih iztočnic pa se najpogosteje odloča na podlagi besedne vrste, ki v slovarju ni navedena. Košutnik za prikaz pomena oz. pomenov iztočnice uporablja tako mikro- kot tudi makrostrukturni pristop. Kern 1919 od vseh preučevanih slovarjev najpogosteje vključuje enako pisane iztočnice in uporablja najbolj makrostrukturni pristop. Dejstvo je, da Kern makrostrukturnega pristopa ne uporablja dosledno, se pa temu še najbolj približa.

Ugotovitve, do katerih smo prišli v naši raziskavi, kažejo, da lahko makro- oz. mikrostrukturni pristop vpliva na razumevanje uporabnika – priseljence. V vseh štirih preučevanih slovarjih so očitne težave z zasnovo in prikazom pomena, to pa so težave, ki so še vedno prisotne v slovarjih za priseljence, čeprav so tehnike prikaza sedaj povsem drugačne. Pri preučevanju teh slovarjev smo ugotovili tudi, da elementi prikaza, ki ne prispevajo k neposrednemu razumevanju pomena iztočnice, predstavljajo velik problem za neizkušene uporabnike. Zaključimo lahko, da se morajo (eno- in dvojezični) slovarji za priseljence v prvi vrsti osredotočiti na pomen.

HOMONYMY AND POLYSEMY IN BILINGUAL DICTIONARIES FOR SLOVENIAN IMMIGRANTS TO THE US AT THE BEGINNING OF THE 20TH CENTURY

This article examines what can be called *immigrant dictionaries*, available to Slovenian immigrants who arrived in the U.S. during the wave of mass immigration from 1860 to 1915. These books, designed to assist in learning English and acculturating to the U.S., were published on both sides of the Atlantic; their aim was to provide information on English, the U.S. constitution and government, on American money, and economy. Among these reference books were grammars, dictionaries, handbooks or textbooks, phrasebooks, children's alphabet books, and other. The books are part of the less-explored history of (learner) lexicography. This article examines four such immigrant dictionaries, Slovenian–English and English–Slovenian (Kubelka 1904, Kubelka 1912, Košutnik 1912, and Kern 1919), published in the early twentieth century at a time when the Slovenian immigration to the United States was at its height. The presentation of information in these dictionaries is investigated with a focus on the treatment of homonymy, polysemy, conversion, and parenthetical information.

Slovenian–English Dictionaries: In Kubelka 1904, the main approach was to use multiple lemmata. Kubelka used ditto marks, curly brackets, and other formal markers

to differentiate multiple lemmata. A transition to a less macrostructure-rich and more microstructure-oriented presentation style can be observed in Kubelka 1912. Here, while there are some multiple lemmata as in Kubelka 1904, they are not relied upon nearly as often. On the other hand, Košutnik 1912 uses a mostly microstructure-oriented presentation; there are few multiple lemmata and within a single lemma he uses a meaning-based approach that presents information about polysemy and/or synonymy of equivalents.

English-Slovenian Dictionaries: In Kubelka 1912, the very infrequent division of a meaning presentation into multiple lemmata appears to be based primarily on considerations of pronunciation. Košutnik 1912 divides his presentation of meaning into multiple lemmata much more frequently than Kubelka 1912. He most often bases his divisions on part of speech, even though part of speech is not marked in the dictionary. We could characterize Košutnik's presentation style as in between a microstructure and a macrostructure orientation. Of all the authors, Kern 1919 makes the most use of multiple entries and is the most macrostructure-oriented. While it would not be accurate to say that Kern has a macrostructure-oriented approach overall, his dictionary is the closest to one.

Our findings address the implications of a macrostructure- versus a microstructure-oriented approach for immigrant user comprehension. The four target Slovenian and English dictionaries of this study illustrate aspects of the problem of design, of presentation of meaning — a problem that remains relevant in immigrant lexicography today even as presentational techniques are now different. These four immigrant dictionaries from the past demonstrate well how elements of presentation that do not contribute to the direct understanding of word meaning can create obstacles for the inexperienced user. Our takeaway should be that dictionaries for immigrants, whether monolingual or bilingual, must provide their special group of users with presentations that put meaning above all else.

Jera Oman (ORCID 0009-0006-2659-6166)

ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Slovenija

jera.oman@zrc-sazu.si

DOI: <https://doi.org/10.3986/16.1.09>

PRIMERJAVA GLAGOLA BITI V SLOVENŠČINI IN RUŠČINI

V prispevku je obravnavan glagol *biti* v slovenščini in ruščini, in sicer z vidika njunih pomensko-skladenjskih zmožnosti. Glede na sorodnost obeh jezikov je pričakovati, da bo raba glagola *biti* v slovenščini in *быть* v ruščini podobna. Raziskava je zato osredotočena na iskanje podobnosti pa tudi razlik, ki so posledica drugačnih jezikovnih zakonitosti. Glagol je v obeh jezikih večpomenski in se uresničuje v treh glavnih vlogah; nastopa lahko kot polnopomenski glagol, opravlja vlogo slovnične vezi ali pa je sestavina zložene glagolske oblike, vse to pa prispeva k njegovi razširjenosti tako v slovenščini kot v ruščini. V prispevku je predstavljen pomenski opis glagolov *biti* in *быть* na primeru prevajanja nekaterih konstrukcij iz ruščine v slovenščino, skladenjsko-pomenski preplet med glagolom *biti* in glagolom *imeti* ter vloga slovnične vezi, ki jo opravlja glagol *biti*, s poudarkom na vprašanju ničte vezi v ruščini.

KLJUČNE BESEDE: glagol biti, slovanski jeziki, pomenoslovje, skladnja, vez

This article explores the verb *biti* in Slovenian and Russian languages from the perspective of their semantic and syntactic characteristics. In light of the connectedness of the two languages, it is reasonable to expect that the use of the verb *biti* in Slovenian and *быть* in Russian is similar. With regard to that, the study is focused on the search for similarities and differences resulting from different linguistic norms. The verb is multidimensional in both languages and as such, acts in three cardinal capacities; as a full verb, as a grammatical conjunction or as an element of a compound form. This multidimensionality advances its widespread use in Slovenian and Russian languages. The semantic account of the verbs *biti* and *быть* presented in this article is based on the example of translating certain constructions from Russian to Slovenian language and the syntactic and semantic intertwining of the verbs *biti* and *imeti*. The account also analyses the role of a conjunction as exercised by the verb *biti*, with particular attention dedicated to the issue of zero copula in Russian language.

KEY WORDS: verb biti, Slavic languages, semantics, syntax, copula

UVOD

Glagol¹ *biti* (tako kot tudi glagola *delati* in *imeti*) sodi v skupino t. i. primarnih glagolov ali glagolskih primitivov. Zanje je značilno, da je njihovo vezljivostno polje načeloma hkrati tudi baza za vse ostale glagole (Žele 2001: 138). A. Vidovič Muha glagolske primitive v sistemu uvrščevalnih pomenskih sestavin (UPS) postavlja povsem na vrh pomenske piramide. UPS na vrhu piramide je tista, ki s svojimi kategorialnimi pomenskimi sestavinami določa vse druge UPS. Vrh pomenske piramide je lahko zaimek (pronomeni) ali primitiv (proverb). Zanje velja, da se kot podpomenka ne morejo pojavljati, saj nastopajo le kot zadnja, najvišja, najsplošnejša nadpomenka. Glagol *biti* (z ostalimi glagolskimi primitivi) je pomensko najširši in pokriva celotno največje pomensko polje (Vidovič Muha 2013: 68). *Biti* je pomensko najbolj posplošujoč glagol, ki ga ni mogoče uvrstiti v širše pomensko polje. Žele (2011: 138) glagolski primitiv *biti* s širokim pomenskim poljem 'bivanje, obstajanje' skupaj še s primarnim glagolom *imeti* s pomenskim poljem 'razmerja, odnosi' označuje kot nadpomenko za vse temeljne glagole stanja. Tako torej *biti* pokriva stanje pa tudi netvorne procese/dogajanja/dejanja.²

Najbolj celovito je glagol *biti* predstavil Toporišič (1980) v svojem članku *O strukturalnem določanju besednih pomenov (ob glagolu biti)*, ki vključuje vidik strukturalnega pomenoslovja. Prispevek je služil kot kritika slovarskega sestavka za glagol *biti* v *Slovarju slovenskega knjižnega jezika I* (1970). V njej je avtor predstavil izčrpno obravnavo glagola, ob pregledu pomenov je zagovarjal strukturalno določanje pomenov, pri čemer naj bi posamezen pomen potrjevala sama jezikovna struktura, ne pa kakršne koli druge vanjo vnesene prvine. Pomene je preverjal s pomočjo sopomenskega izrazja (če to obstaja), s čimer je želel ugotoviti tudi pravilnost ustrezne uvrstitve ponazoritev oz. zgledov za posamezne pomene. Članek natančno predstavlja glagol *biti*, na kar smo se oprli v tem prispevku, ob strani pa smo pustili Toporišičevo kritiko, ki odpira vprašanja, povezana s konceptom SSKJ.

¹ Prispevek je nastal na osnovi magistrskega dela z naslovom Skladenjsko-pomenske zmožnosti glagola BITI v slovensščini in ruščini (2021).

² Glagol *biti* v slovenskem knjižnem jeziku 16. stoletja je z vidika skladenjsko-pomenskih ter oblikoslovnih značilnosti obravnavala Majda Merše (2017). Avtorica ugotavlja, da je glagol *biti* zaradi svoje dvojne funkcionalnosti (lahko nastopa kot polnopomenski in kot pomožni glagol) najpogosteje rabljeni glagol v slovenskem knjižnem jeziku 16. stoletja (Merše 2017: 211).

Slovarski sestavek za glagol *biti* je v SSKJ2³ razdeljen na pet odstavkov, označenih z rimskimi številkami: **I.** ‘v osebni rabi’, **II.** ‘v brezosebni rabi ali brezosebno v pomenih kakor pod I.’, **III.** ‘v osebni ali brezosebni rabi, navadno s prislovno rabljenim samostalnikom v pomenih kakor pod I.’, **IV.** ‘eliptično navadno za poživitev pripovedi’, **V.** ‘kot pomožni glagol, z opisnim ali trpnim deležnikom za tvorbo časov, naklonov ali načinov’.

Toporišič je pri **I.** razdelku menil, da slovarskemu sestavku v razlagi primanjkuje splošna stopnja sopomenk, zato sam predlaga: obstajati (namesto izraza navzočnost), nahajati se (namesto ‘izraža navzočnost v prostoru in času’), (imeti), dogajati se (tu ima slovar deloma zgoditi se; Toporišič opozori na zgled omenjena nesreča je bila včeraj, s stransko razlago se je zgodila). Kjer predlagani izrazi niso variantna sredstva,⁴ je zanje smiselno odpreti poseben pomen, in sicer Toporišič predlaga: I. 1. a) obstajati, b) obstajati/živeti, 2. a) nahajati se, b) nahajati se/imeti, c) nahajati se/stati, č) nahajati se/živeti, bivati, d) nahajati se/udeležiti se, e) nahajati se/biti pri (pregledu), 3. dogajati se. Pomene obstajati, nahajati se, imeti družijo skupna oblikovna lastnost: ob zanikanem glagolu *biti* se nosilec stanja ali poteka iz osebkevega imenovalnika prestavi v rodilnik smiselnega osebka (npr. *Ali je kak izhod? – Ali ni nobenega izhoda?, Tukaj je jama – Tukaj ni jame, Tedaj je bila vojna – Tedaj ni bilo vojne*). Toporišič poudarja, da nam prehod osebka iz imenovalnika v rodilnik ob zanikanem glagolu pomaga preverjati predmetnopomensko vlogo glagola *biti* v primerjavi s slovnično, ki prestavljanja v rodilnik ne pozna.⁵ *Mož je poštenjak – Mož ni poštenjak* (Toporišič 1980: 152–154).

³ Geselski sestavek za glagol *biti* se v drugi izdaji slovarja (glede na prvo) sicer ni bistveno spremenil.

⁴ Toporišič ugotavlja, kdaj in kako jih zaznamuje t. i. količek ali dva količka, s katerim(a) »v okviru večjih istopomenskih odstavkov skušamo sopomensko izraziti tudi pomenske premike« (1980: 151). Gre za dve poševni črti pred razlago (/), ki uvajata podpomen, in eno poševno črto (/), ki loči pomenske odtenke, stalne zveze in stilizme (Uvod v SSKJ).

⁵ Ob pregledu zgledov je Toporišič ugotovil, da je samo primer *meso je brez kosti* tak, da osebkevo imenovalnik lahko prestavimo v rodilnik in se pomen pri tem spremeni: *meso ni brez kosti* ‘ima kosti’, *ni mesa brez kosti* ‘ne obstaja meso brez kosti’ (155).

POMENI GLAGOLA *BITI* V SLOVENŠČINI IN RUŠČINI

Predstavitev pomenskega polja za glagol *biti* ~ *быть* se opira na njun geselski članek v Slovarju slovenskega knjižnega jezika (SSKJ2) ter v Aktivnem slovarju ruskega jezika⁶ (v nadaljevanju AS) (ur. Jurij D. Apresjan).

AS ima pri glagolu *быть* šestnajst pomenskih kategorij, iz njih pa je mogoče izluščiti šest glavnih z njihovimi podpomeni: **БЫТЬ 1.1** ‘являться’ sln. ‘biti (vezno)’, **БЫТЬ 1.2** ‘быть тождественным’ sln. ‘biti istoveten’, **БЫТЬ 2.1** ‘находиться’ sln. ‘nahajati se’, **БЫТЬ 2.2** ‘прибывать куда-л.’ sln. ‘prihajati kam’, **БЫТЬ 3.1** ‘иметься’ sln. ‘imeti’, **БЫТЬ 3.2** ‘иметь возраст’ sln. ‘biti (star koliko)’, **БЫТЬ 4.1** ‘существовать’ sln. ‘obstajati’, **БЫТЬ 4.2** ‘наступить’ [о времени суток] sln. ‘nastopiti (čas dneva)’, **БЫТЬ 4.3** ‘иметь место, состояться’ [о событии, мероприятии] sln. ‘dogajati se’ (dogodek, prireditev), **БЫТЬ 4.4** ‘происходить, случаться’ sln. ‘dogajati se’, **БЫТЬ 5.1** ‘уверенность в неизбежности события’ sln. ‘prepričanost o neizbežnosti dogodka’, **БЫТЬ 5.2** ‘уверенность в неизбежности плохого’ sln. ‘prepričanost o neizbežnosti slabega’, **БЫТЬ 5.3** ‘надо прекратить ситуацию’ sln. ‘treba je končati situacijo’, **БЫТЬ 5.4** ‘надо прекратить действие’ sln. ‘treba je končati dejanje’, **БЫТЬ 6.1** в составе формы БУД sln. del oblike prihodnjega časa nekega glagola, **БЫТЬ 6.2** в составе формы СТРАД sln. del oblike trpnega deležnika.

Razporeditev pomenov je obrazložena v konceptu Aktivnega slovarja ruskega jezika (*Перспектив активное словаря русского языка*, ur. Apresjan). Osnovni pomen (zato na prvem mestu), ki je pripisan glagolu *быть*, je vezni (*быть* 1.1). Kot podpomen je podan še vezni pomen enakosti oz. identifikacije (*быть* 1.2). Drugi pomen združuje lokativne pomene, najprej lokativni pomen ‘nahajati se’ (*быть* 2.1) in razširitev s ‘prispeti’, ‘začeti nekje bivati’ (*быть* 2.2). Lokativni pomeni sledijo veznim zaradi posebne povezanosti teh dveh pomenov, ki v ruščini obstaja še pri nekaterih drugih glagolih npr. *стать мастером спорта – стать у стены* (sln. *postati športni prvak – stati pri steni*).

⁶ Glagol v navedenih slovarjih ni obravnavan enakovredno, ker imamo opraviti z dvema različnima tipoma slovarjev. SSKJ2 je splošnorazlagalni enojezični slovar srednjega obsega, vsebuje približno sto tisoč gesel in predstavlja opis jezika druge polovice 20. stoletja. AS pa vključuje samo besede, ki jih naravni govorniki ruščine uporabljajo najpogosteje (zato jih je precej manj kot v SSKJ2; le nekaj tisoč). Vsako geslo dopušča natančnejšo obdelavo, geslo za glagol *biti* je sploh eno od obsežnejših. Vira sta torej neenakovredna, se je pa med raziskavo izkazalo, da sta za izbrana jezika najbolj informativna, saj oba ponujata celovito predstavitev glagola v obeh jezikih, ponazarjata njune skladenjsko-pomenske zmožnosti ter primere rabe.

Med veznimi in lokativnimi pomeni obstaja izrazita pomenska bližina. V nekaterih primerih imajo lahko glagoli z osnovnim lokativnim pomenom tudi vezni pomen, npr. *находиться в столице – находиться под наблюдением врачей* (sln. *nahajati se v prestolnici – biti pod zdravniškim nadzorom*). Avtorji slovarja menijo, da je pod tretji pomen smiselno uvrstiti pomene posedovanja (*быть* 3.1–*быть* 3.3), ki se pomensko tudi približujejo veznim, npr. *Она обладала умом и красотой – Она была умна и красива* (sln. *Bila je pametna in lepa*). Tem pomenom, zadnji z leksikalnimi pomeni glagola kot vezi, sledijo pomeni obstoja (*быть* 4.1–*быть* 4.4). Pod petim pomenom so predstavljeni modalni pomeni glagola (*быть* 5.1–*быть* 5.3), šesti pomen pa združuje dva slovnična pomena.

быть 1.1 ‘являться’: Мой отец был архитектором. (sln. Moj oče je bil arhitekt.)

Prvi pomen AS ustreza 3. pomenu v SSKJ2 ‘izraža lastnost ali stanje osebk’, in sicer pod a) s pregibno besedo: mož je poštenjak, prijatelj je učitelj; otrok je bolan, nadarjen; pšenica še ni zrela; knjiga je moja; klobuk je očetov; ura bo tri / dolžen mu je veliko vsoto; njegov načrt je bil, da bi napadli; on je odgovoren za tovariše; uspehi so rezultat trdega dela / zdaj je čas, da poveste svoje mnenje; ni bila moja dolžnost, naloga popravljati napake.

быть 1.2 ‘быть тождественным’: Это был Иван. (sln. To je bil Ivan.)

V SSKJ2 ustreznico najdemo pod 3. pomenom ‘izraža lastnost ali stanje osebk’ kot podpomen d) s primerjavo // ‘izraža istost, enakost’: dvakrat dve je štiri; ena in ena je dve; evro je sto centov; kolikokrat je dve v osem / pedagogika je nauk o vzgoji / pride jutri, to je v torek. V AS je primer Сколько будет два плюс два? (prihodnjik 3. osebe ednine) sln. Koliko je dva plus dva? uvrščen pod prvi pomen, torej 1.1 ‘являться’, z razlago, da ima v odgovoru na to vprašanje pomen sedanjika, pri čemer se pomen glagola pomakne v smer ‘biti ekvivalenten ali enak’.

быть 2.1 ‘находиться’: Дети были на озере. (sln. Otroci so bili na (pri) jezeru.)

V SSKJ2 je to 2. pomen ‘izraža navzočnost v prostoru ali času’: tukaj je jama, v kateri je raslo drevo; hiša je sredi polja; drva so na dvorišču; v

vsakem človeku je kaj dobrega; v knjigi so napake; v vodi je kisik; za gozdom je travnik ... pa tudi proslava bo v nedeljo. Toporišič (1980: 153) govori o sopomenki ‘nahajati se’. V AS pod tem pomenom ni časovne dimenzije oz. t. i. ‘navzočnosti v času’.

БЫТЬ 2.2 ‘прибывать куда-л.’: Он сегодня непременно будет < уже был в два часа>. (sln. Danes zagotovo pride <je že bil ob dveh>.)

Pomen je omejen na sopomenko ‘prispeti’, povezuje se tako s prostorom kot časom. Slovar AS navaja, da se pogosto pojavlja ob členkih še in že (еще, уже). Ob pregledu možnih rešitev v SSKJ2 omenjenega ni mogoče najti. Primer, ki bi ustrezal, je v II. razdelku pod a) očeta še ni, kar je mogoče razumeti kot oče še ni prišel/prispel. Vendar drugi primeri, navedeni v nadaljevanju: jutri ne bo šole; pri tej hiši ni miru; ni ga človeka, ki bi mi pomagal, ne ustrezajo pomenu, kot ga opredeljuje AS.

БЫТЬ 3.1 ‘иметься’: У него была прекрасная библиотека. (sln. Imel je čudovito knjižnico.)⁷

V SSKJ2 ustreznico najdemo pod 3. pomenom ‘izraža lastnost ali stanje osebk’, in sicer pod a) s pregibno besedo knjiga je moja, klobuk je očetov; pa tudi pod c) s predložno besedo meso je brez kosti (meso nima kosti), biti brez denarja ne imeti ga, brez težav.

БЫТЬ 3.2 ‘иметь возраст’: Ему было двадцать лет. (sln. Star je bil dvajset let.)⁸

V SSKJ2 v II. razdelku pod b) s smiselnim osebkom v odvisnem sklonu njej je bilo dvajset let bila je stara. V slovenščini se raba tovrstnih konstrukcij z glagolom *biti* v pomenu ‘starost’ zdi najpogostejša za preteklik, čeprav morda nekoliko arhaična (bilo mu je pet mesecev) – česar SSKJ2 sicer ne potrjuje, precej vprašljiva pa za prihodnjik (februarja mu bo eno leto) ter za sedanjik (tri leta so mu).

⁷ Poved bi v slovenščino najverjetneje prevedli z glagolom *imeti*, lahko pa bi jo preoblikovali tudi v: *V njegovi lasti je bila čudovita knjižnica.*

⁸ *Bilo mu je dvajset let.*

быть 4.1 ‘существовать’: Есть еще добрые люди на свете! (sln. Še so na svetu dobri ljudje!)

SSKJ2 pomen obstajanja opredeljuje 1. ‘izraža materialno ali duhovno navzočnost v stvarnosti’: še so stvari, ki jih ne poznamo; ali je kak izhod? red mora biti, poudarjeno sò znamenja, da se obrača na boljše, bolnik ne bo več dolgo; v pravljicah bil je kralj, ki je imel tri sinove. Če primerjamo zglede, ki so podani v obeh slovarjih, bi pod 1. pomen v SSKJ2 sodila primera: Есть еще добрые люди на свете! (sln. Še so na svetu dobri ljudje!) in Есть разные теории на этот счет (sln. Na to temo so različne teorije). Primeri, kot so npr.: В Архангельской и Вологодской областях будут грозы (sln. V Arhangelski in Vologdski oblasti bodo nevihte); Осенью было много свадеб (sln. Jeseni je bilo veliko porok), На Курилах есть горячие источники (sln. Na Kurilskem otočju so termalni vrelci) (AS), pa bi se uvrstili pod drugi pomen v SSKJ2 ‘izraža navzočnost v prostoru in času’. Toporišič (1980: 153, 154) za ta pomen predlaga sopomenki ‘nahajati se’ in ‘dogajati se’. V primerih V Arhangelski in Vologdski oblasti bodo nevihte ter Jeseni je bilo veliko porok gre za vprašanje dogajanja, ki bi ga bilo mogoče opredeliti kot nekakšen podtip obstoja.

быть 4.2 ‘наступить’ [о времени суток]: Было пять часов. (sln. Ura je bila pet.)

V SSKJ2 najdemo zgled pod 3. pomenom ‘izraža lastnost ali stanje osebk’, a) ura bo tri in pa v II. razdelku pod e) s prislovno rabljenim izrazom v povedku: skoraj bo poldne.

быть 4.3 ‘иметь место, состояться’ [о событии, мероприятии]: Был дождь; Была воздушная тревога; Обед был в два часа. (sln. Deževalo je; Bila je grožnja zračnega napada; Kosilo je bilo ob dveh.)

V SSKJ2 so to primeri pod 2. pomenom ‘izraža navzočnost v prostoru in času’, in sicer tisti, ki jih Toporišič (1980: 154) označuje s sopomenko ‘dogajati se’: tedaj je bila vojna, v Kamniku je bil semenj; proslava bo v nedeljo / omenjena nesreča je bila včeraj se je zgodila; to je bilo lani.

быть 4.4 ‘происходить, случаться’: С ним несчастье; Что с тобой? (sln. Nesrečo ima (nesrečen je); Kaj je s tabo?)

V SSKJ2 (II. razdelek) pod b) s smiselnim osebkom v odvisnem sklonu: z njim je slabo ter s kvalifikatorjem ekspresivno kaj bo s teboj se bo zgodilo.

БЫТЬ 5.1 ‘уверенность в неизбежности события’: Быть грозе. (sln. Nevihta bo.)

Za konstrukcijo je značilno, da nedoločnik glagola БЫТЬ stoji v prepoziciji samostalnika v dajalniku, slovar jo označuje s kvalifikatorjema уходящ. (izginja) ali книжн. (knjižno). Pomen je ‘prepričanost o neizbežnosti nekega dogodka’ (AS).

БЫТЬ 5.2 ‘уверенность в неизбежности плохого’: Нам теперь крышка. (sln. Konec je z nami (Po nas je).)

V tem primeru gre za pomen ‘prepričanost o neizbežnosti česa slabega’. V SSKJ2 pomen najdemo v t. i. frazeološkem gnezdu s kvalifikatorjem ekspresivno, po njem je ‘izgubljen je, mrtev je’.

БЫТЬ 5.3 ‘надо прекратить ситуацию’: Будет с тебя. (sln. Dovolj je (Dovolj imaš).)

Konstrukcija obstaja samo v 3. osebi ednine prihodnjika in izginja, označuje pa pomen ‘treba je končati situacijo’ (AS).

БЫТЬ 5.4 ‘надо прекратить действие’: Будет тебе плакать. (sln. Dovolj je jokanja.)

Konstrukcija obstaja samo v 3. osebi ednine prihodnjika, v pomenu ‘treba je končati dejanje’, lahko gre za govorno ali fizično dejanje, vedenje ali stanje (AS).

БЫТЬ 6.1 в составе формы БУД: Не буду вам мешать; Чай будешь? (sln. Ne bom vas motil; Boš čaj?)

Glagol *biti* nastopa kot del oblike prihodnjika. Gre za pomožni glagol, ki s svojimi osebnimi prihodnjiškimi oblikami skupaj s polnopomenskim glagolom v nedoločniku nedovršnega vida tvori analitično obliko prihodnjega časa tega polnopomenskega glagola, npr. Я буду работать (sln. Delal bom). V pogovornem jeziku glagol nastopa skupaj s samostalniki (v 4. sklonu) v pomenu

hrane, pijače, cigaret ipd. in označuje željo ali namero človeka to uporabiti (zaužiti): Кофе <хлеб с ветчиной> будешь? = Будешь пить кофе <есть хлеб с ветчиной> ? (sln. Boš kavo <kruh s šunko>? = Boš pil kavo <jedel kruh s šunko?>); Что будешь на завтрак? (sln. Kaj boš za zajtrk?) (AS). V SSKJ2 so taki primeri pod IV. razdelkom eliptično, navadno za poživitev pripovedi a) vsak en kozarček ga bova. V teh primerih v obeh jezikih opažamo elipso glagolov jesti in piti.

БЫТЬ 6.2 в составе формы СТРАД: Проект был <будет> закончен в декабре. (sln. Projekt je bil <bo> končan decembra.)

Glagol kot del trpnega deležnika – pomožni glagol, ki skupaj s poln-pomenskim glagolom v dovršnem trpnem deležniku preteklega časa tvori analitične oblike trpnega načina tega glagola, npr. Работа начата только что (sln. Delo je ravno začeto) (AS). SSKJ2 v V. razdelku obravnava glagol *biti* kot pomožni glagol, z opisnim ali trpnim deležnikom za tvorbo časov, naklonov ali načinov: bo delal, je delal, je bil izdelal; bi delal, bi bil delal; je izdelan / v ekspresivni rabi stoji pomožni glagol tudi na začetku stavka sem si kar mislil, da je nekaj narobe.

VEZNA VLOGA GLAGOLA

V slovenščini ima glagol *biti* lahko poln-pomensko vlogo ali pa zgolj slovnično oz. vezno. Toporišič v *Slovenski slovnici* (2004: 588, 589) te glagole ločuje glede na njihovo sporočilno oz. obvestilno samozadostnost; merilo je možnost razstavljanja na dve enoti in presojanje predmetnopomenskih ali pomožnih glagolov, npr. *hoditi bos* – *hoditi* + *biti bos*; *vrniti se kot zmagovalec* – *vrniti se* + *biti zmagovalec* (kot primer predmetnopomenskih) ter *biti star* (primer pomožnega glagola). Zadnje zveze ni mogoče razstaviti, ker ne bi bila več sporočilno samozadostna; v primeru razstavitve bi glagol dobil drugačen pomen, npr. ‘obstajanja, nahajanja, dogajanja’, kar ne ustreza pomenu zveze *biti star*. V prvem primeru gre za zvezo predmetnopomenskega glagola in povedkovega prilastka, v drugem pa za pomožni ali vezni glagol s povedkovim določilom. Teh zvez ne moremo razstaviti tako, da dobimo dve pomensko samozadostni enoti. Glavni pomožni glagol v slovenščini je ravno *biti*, čeprav ob njem obstajajo še drugi (npr. nevezavna *ostati*, *postati*, vezavni *imeti*).

Svoj sestavek o glagolu *biti* tako Toporišič deli na dva dela. V I. je predstavljen kot polnopomenski glagol, v II. pa kot vez s povedkovim določilom: a) v dvodelnih stavkih in b) v enodelnih stavkih (1980: 163,164).

Tudi glagolu *быть* v ruščini lahko pripišemo različne vloge. Asya Pereltsvaig (2007: 4) (po Chvany) glagol deli v dva sintaktična tipa; prvi je glagol *biti* s t. i. eksistencialno vlogo, drugi pa glagol *biti* kot vez. Navaja kriterije za ločevanje med obema tipoma: Pri prvem tipu se stavčna negacija izraža z *нет*: *Его нет дома* (*Ni ga doma*), pri drugem (*biti* v vlogi kopule) pa z *не* + *быть*: *Он не был писателем* (*Ni bil pisatelj*). Sklon osebkov v zanikani konstrukciji pri prvem tipu je roditeljnik, pri drugem pa imenovalnik. Kadar glagol *быть* nastopa v vlogi vezi, dovoljuje zamenjavo glagola tako, da osebek konstrukcije z glagolom *быть* postane predmet nove konstrukcije *Я считаю Ваню гением* (*Vanja se mi zdi genij*); kopula dovoljuje tudi zamenjavo glagola tako, da osebek nove konstrukcije ostaja enak stari *Ваня кажется нерыхой* (*Vanja je videti malomaren*); v prvem sintaktičnem tipu glagol *biti* tega ne dopušča. Iz bivanjskega glagola *biti* je mogoče tvoriti izpeljanke, npr. *бытие* (*obstoj*), lahko pa mu poiščemo tudi sinonim, ki je sintaktično ustrezen, npr. *существовать* (*obstajati*). Glagol *biti* v vlogi vezi ne dopušča ničesar od tega.

V ruščini obstaja poseben pojav opuščanja slovnične vezi, imenovan ničta vez (*нулевая связка*),⁹ ki je poseben podtip vezi oz. veznih elementov. A. B. Letučij (2018) ga opisuje kot pojav, pri katerem v povedi glagol ni izražen, na mestu povedka pa se lahko pojavlja t. i. neglagolska skupina, torej samostalniška beseda (*Серёжа — студент*, sln. *Serjoža je študent*), pridevniška beseda (*Ты уже не молод*, sln. *Nisi več mlad*; *Она красивая*, sln. *Ona je lepa*), prislov (*Концерт — завтра*, sln. *Koncert je jutri*), predikativ (*Мне холодно*, sln. *Zebe me*). Konstrukcije lahko vsebujejo kanonični osebek (*Иван — студент*, sln. *Ivan je študent*) ali pa so brez osebkov (*Здесь интересно*, sln. *Tukaj je zanimivo*). Izraz ničta vez se uporablja predvsem takrat, ko se ničta oblika povedka s spremembo časa in naklona zamenja z neničto obliko. V ruščini ničta vez zapolnjuje povedni naklon glagola *быть* v sedanjiku. Pri tem ne gre za primer elipse, saj ničta vez ne nadomešča opuščene prvine, izražene v predhodnem ali sledečem odlomku besedila. Pogoji, ki morajo biti izpolnjeni, da lahko govorimo o ničti vezi: (v povedi) glagol ni izražen, predstavljeno je

⁹ Vez se v ruščini, če ne gre za poudarek, v sedanjiku ne izraža.

realno stanje, ki se dogaja v sedanjem času (zajema tudi trenutek ubeseditve dejstva), v vseh drugih časih in naklonih je glagol *быть* izražen (obstaja nekaj konstrukcij, ki temu pogoju ne ustrezajo, vendar se v skupino vključujejo zaradi semantične podobnosti standardnim konstrukcijam). Ničta vez je v ruščini prisotna samo v sedanjiku povednega naklona, v drugih časih in naklonih pa se na njenem mestu uporabljajo izražene oblike glagola *быть*, pri čemer mora imenski del predikativa stati v orodniku. *Серёжа – студент* (sln. *Serjoža je študent*), *Серёжа был студентом* (sln. *Serjoža je bil študent*), *Серёжа был бы студентом* (sln. *Serjoža bi bil študent*).

Sodobna ruščina za glagol *biti* v sedanjiku pozna naslednje oblike: ničto vez (нулевая связка), *есть* in *суть*.¹⁰ Zimmerling (2021: 278) pojasnjuje, da prvi dve obliki ne ločujeta med glagolsko osebo in številom, medtem ko oblika *суть* del naravnih govorcev uporablja le kot vez za tretjo osebo množine v znanstvenih in publicističnih besedilih. V primerih, ki zahtevajo uporabo ničte vezi, oblika *есть* pa ni mogoča, tam prav tako ni mogoča oblika *суть*. Avtor (s sklicevanjem na Apresjana 1996) (ibid.) razlaga, da se sicer glagolska oblika *суть* v sodobni ruščini domnevno ohranja le v vlogi vezi (ne pa kot polnopomenski glagol, o čemer je govoril tudi Apresjan), vendar je nabor kontekstov, ki dovoljujejo rabo te oblike, manjši kot za ničto vez ali *есть*.

BITI IN IMETI

Med glagoloma *biti* in *imeti* obstaja pomensko-skladenjska povezanost. Povezanost glagolov *biti* in *imeti* je Vidovič Muha (1998: 293) raziskovala preko izražanja prostorske umeščenosti; *biti* je prostorsko odprt in nedoločen, *imeti* pa prostorskost veže na posedovalca oz. posesorja. Pojasnjuje, da sta glagola v nad- oz. podpomenskem razmerju. *Biti* je temeljni glagolski primitiv, ki ne predvideva nadaljnje pomenske členitve. Slovanska razlaga zanj nima uvrščevalne pomenske sestavine, kar je razvidno iz tipa razlag v SSKJ (npr. *izraža navzočnost v stvarnosti*). Za pomen je značilna enovezljivost, na paradigmatski ravni pa se vzpostavlja sopomenskost z glagolom *obstajati*. Glagol *biti* lahko brez dodatnih opredelitev izraža zavzemanje prostora: *Nekdo/Nekaj je*. Gledano z vidika pomenskega okvira je torej primitiv *imeti* primitivu *biti* podrejen. Tudi ta izraža zavzemanje oz. zasedanje prostora; kaj/kdo zavzema prostor, kar je

¹⁰ *Есть* se navadno uporablja za izražanje poudarka, *суть* pa v določenih besedilih.

možno tudi kot glede na koga/kaj: *kaj/kdo* je od *koga/česa*, *kaj/kdo* je pri *kom/čem* oz. *kdo ima kaj* – ne nujno kot svojo lastnino (Vidovič Muha 1998: 316).

Globoka skladdenjsko-pomenska prepletenost med omenjenima glagoloma (*biti*, *imeti*) se je pokazala tudi pri pregledu pomenov v *Aktivnem slovarju ruskega jezika*, natančneje pri pomenu *быть* 3.1 ‘иметься’.¹¹ Povezanost se je pokazala pri prevodu primerov v slovenščino, npr.: *У Гегеля была большая библиотека* (sln. *Hegel je imel veliko knjižnico*); *У нее две квартиры в Москве* <антикварная мебель> (sln. *V Moskvi ima dve stanovanji* <starinsko pohištvo>); *У них пластиковые лыжи* <надувная лодка, мотороллер> (sln. *Imajo plastične smuči* <napihljiv čoln, skuter>); *У директора будет* <было> *для меня ровно пять минут* (sln. *Direktor bo imel* <*je imel*> *zame natanko pet minut*) ... Skoraj vse primere, navedene pod tem pomenom, ki se v ruščini izražajo z glagolom *biti*, bi v slovenščino nevtravno prevajali z glagolom *imeti*. Nekateri primeri pa so taki, da dopuščajo obe možnosti, npr.

У нее две квартиры в Москве → *V Moskvi ima dve stanovanji* ali *V njeni lasti sta dve stanovanji v Moskvi*

У нее впереди несколько бессонных ночей → *Pred sabo ima nekaj neprespanih noči* ali *Pred njo je nekaj neprespanih noči*

У наших детей было беззаботное детство → *Naši otroci so imeli brezskrbno otroštvo* ali *Za našimi otroki je brezskrbno otroštvo*.

VIDIK PREVAJANJA GLAGOLA *BITI* V LEPOSLOVNEM GRADIVU

Naša raziskava temelji na omejenem gradivu, na naključno izbranih odlomkih z glagolom *biti* iz romana *Pomladni dan* Cirila Kosmača¹² in njegovega prevoda v ruščino. Na primeru 47 krajših odlomkov smo poskušali ugotoviti, kaj stoji namesto glagola *biti* v primerih, kjer je ta uporabljen v zgolj enem jeziku. Pregledali smo pojavitve glagolov v izbranih odlomkih in njihovih prevodih. Izkazalo se je, da je v 87 % (celoto predstavljajo pojavitve glagolov *biti/быть* v pregledanih primerih) za glagol *biti* v slovenščini tudi v prevodu glagol *быть*, kar je mogoče pripisati sorodnostim med jeziki. V 66 % (tu celoto

¹¹ Pomen ‘imeti’ se v ruščini nevtravno izraža prav s konstrukcijo *есть у кого* (»*biti u koga*«).

¹² Ciril Kosmač, *Pomladni dan* (DZS, 2001) ter Ciril Kosmač, *Избранное*, (Moskva: Прогресс, 1976), prevod Jevgenija Ivanovna Rjabova.

predstavljajo zgolj primeri, kjer se glagol *biti/быть* pojavlja samo v enem od obeh jezikov) je glagol *biti* v izvirniku, v ruskem prevodu pa ne. V ruščini se za take primere pojavljajo naslednje prevodne rešitve:

* *je grobar – он гробы делаем* (glagol *delati* + samostalnik); glagol *biti* je v ruščini zamenjan s polnopomenskim glagolom *delati* v zvezi s samostalnikom, za izbiro je mogoče reči, da je posledica stilističnega posega prevajalke in precej vpliva na pomen. V slovenščini (*je grobar*) je prikazana stalnost poklica; medtem ko zveza v ruščini deluje nekoliko bolj priložnostno. Prevajalka se je za prevodno rešitev morda odločila zato, ker *могильщик* (grobar) v prenesenem pomenu v ruščini poimenuje tistega, ki 'ima uničujoč učinek na nekaj', nečemu »prinaša smrt«.

* *steza je ozka in prečka strmino – узенькая тропинка взбирается по крутому склону; rob je pameten – умный парень* (zveza pridevnika in samostalnika); v tem primeru predikacijo (v slovenščini) v ruščini nadomešča zveza z levim prilastkom. Tudi tu je mogoče reči, da se pomen nekoliko spremeni. V ruskem prevodu je morda lahko lastnost razumeti bolj priložnostno kot v slovenščini.

* *bom že – я сама* (pomožni glagol *biti* samo v slovenščini); pri zamolku je glagol *biti* pomožni in nastopa kot del zložene glagolske oblike ob odsotnosti polnopomenskega glagola. Ruski prevod pa prav tako napoveduje izpust polnopomenskega glagola (npr. *я сама могу* (zmorem sama)).

* *v njem je nekaj sanjskega, pretresujoče lepega in vabljivega – видение родины, мечтательное, неотразимо прекрасное и манящее* (sprememba diateze in precejšen pomenski poseg).

* *za pamet je zmeraj čas – для умного время всегда найдется* (uporaba povratnega glagola *найтись*); v ruskem prevodu je uporabljen povratni glagol *найтись* z oslabljenim pomenom; prevajalka je *pamet* spremenila v *pametnega* (človeka).

Za najbolj sistemsko prevodno rešitev bi lahko označili uporabo glagolskega para *biti – imeti* – glede na njuno skladijsko-pomensko povezavo in ostale prevode slovarskih zgledov je bila tudi najbolj pričakovana. Glagol *быть* se v določenih konstrukcijah v slovenščino prevaja z glagolom *imeti*. To so tiste, ki v sodobni ruščini izražajo svojilnost, prvotno pa so pomenile 'nahajanje v bližini', torej v pomenu 'biti v bližini'. Zaradi svojilnosti jih v slovenščino največkrat prevajamo z glagolom *imeti*. Kljub sorodnosti gre pri slovenščini in ruščini za dva različna jezikovna sistema, z različnimi izraznimi možnostmi,

kar se odraža v drugačnih jezikovnih konstrukcijah. Nekatere razlike, ki se pojavljajo, je težje natančno opredeliti, ker so posledica prevajalkine odločitve oz. stilistike, ki jo prevajalec izrablja, da v ciljnem jeziku doseže podoben jezikovni učinek kot v izvirniku. Zato lahko pride tudi do večjih pomenskih posegov.

SKLEP

Slovenski glagol *biti* in ruski glagol *быть* sta večpomenska, zaradi svoje elementarnosti pa sta v jezikovni rabi zelo pogosta v obeh jezikih. Imata tri glavne vloge: polnopomensko, vlogo slovnične vezi in vlogo dela zložene glagolske oblike. S skladijsko-pomenskega vidika je glagolu *biti* najbližji *imeti*, kar je razvidno iz prevajanja nekaterih konstrukcij, predvsem *У кого (есть) что*, iz ruščine v slovenščino. Na analiziranem primerku leposlovnega gradiva so preverjene vse glavne vloge glagola *biti*: polnopomenska, vezna in vloga pomožnega glagola. Raziskava je razkrila načine uresničitve glagola v obeh jezikih, njune skladijske in pomenske lastnosti ter primere rabe. Pokazale so se nekatere bistvene podobnosti in tudi razlike; SSKJ2 na prvem mestu oz. kot osnovni pomen glagola *biti* navaja polnopomenski 'obstajati', AS pa ima za osnovni pomen vezni (ker je glede na rabo najpogostejši), pomen 'obstoja' pa šele na četrtem mestu. Oba jezika imata *biti* in *быть* kot slovnično vez, vendar v ruščini ni vedno izražena, pozna namreč pojav t. i. ničte vezi, ki se pojavlja le v povednem naklonu sedanjika, v drugih časih in naklonih pa se premenjuje z izraženo obliko glagola *быть*. Naj tu omenimo še tvorbo preteklika v ruščini, ki se tvori brez pomožnega glagola *biti*, zato zahteva osebni zaimek oz. osebek.

REFERENCE

- Апресян, Юрий Д. (ur.) 2014. *Активный словарь русского языка (AS)*. Том 1. А-Б. Москва: Языки славянской культуры. [Aprėsjan, Jurij D. (ur.) 2014. *Aktivnyj slovar' russkogo jazyka (AS)*. Том 1. А-В. Moskva: Jazyki slavjanskoj kul'tury.]
- Апресян, Юрий Д. (ред.) 2010. *Перспект активного словаря русского языка*. Москва: Институт русского языка им. В. В. Виноградова РАН. [Aprėsjan, Jurij D. (red.) 2010. *Prospekt aktivnogo slovarja russkogo jazyka*. Moskva: Institut russkogo jazyka im. V. V. Vinogradova RAN.]

- Циммерлинг, Антон В. 2021. Суть и есть. V: И. М. Ладыженский, М. А. Пузина (ur.). *Sub specie aeternitatis. Сборник научных статей к 60-летию Вадима Борисовича Крысько*. Москва: Азбуковник. 278–291. [Zimmerling, Anton V. 2021. Sut' i est'. V: I. M. Ladyženskij, M. A. Puzina (ur.) *Sub specie aeternitatis. Sbornik naučnyh stat'ej k 60-letiju Vadima Borisoviča Krysko*. Moskva: Azbukovnik. 278–291.]
- Kosmač, Ciril. 2001. *Pomladni dan*. Ljubljana: DZS.
- Космач, Цирил. 1976. *Избранное*. Москва: Прогресс. [Kosmač, Ciril. 1976. *Izbrannoe*. Moskva: Progress.]
- Летучий, Александр Б. Нулевая связка. Материалы для проекта корпусного описания русской грамматики. На правах рукописи. М., 2018. [Letučij, Aleksandr B. Nulevaja svjazka. Materialy dlja proëkta korpusnogo opisanija russoj grammatiki. Na pravah rukopisi. M., 2018.] <http://www.rusgram.ru/>
- Merše, Majda. 2017. Raba glagola biti sem v slovenskem knjižnem jeziku 16. stoletja. *Jezikoslovni zapiski* 23/2: 211–227.
- Pereltsvaig, Asya. 2007. *Copular sentences in russian: A Theory of Intra-Clausal Relations*. Dordrecht: Springer.
- Sitar, Jera. 2021. Skladijsko-pomenske zmožnosti glagola biti v slovenščini in ruščini. Magistrsko delo. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja. (SSKJ2) 2014. www.fran.si
- Toporišič, Jože. 1980. O strukturalnem določanju besednih pomenov (ob glagolu biti). *Linguistica* 20/1: 151–167.
- Toporišič, Jože. 2004. *Slovenska slovnica*. Maribor: Obzorja.
- Vidovič Muha, Ada. 1998. Pomenski preplet glagolov *imeti* in *biti* – njuna jezikovnosistemska stilistika. *Slavistična revija* 46/4: 294–323.
- Vidovič Muha, Ada. 2013. *Slovensko leksikalno pomenoslovje*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Žele, Andreja. 2001. Vezljivost kot pomensko- in strukturoskladijski pojav. *Slavistična revija* 49/3: 138–161.

Prispelo januarja 2023, sprejeto februarja 2024.

Received January 2023, accepted February 2024.

ZAHVALE

Prispevek je nastal v okviru programa P6-0038, ki ga financira ARIS.

POVZETEK

PRIMERJAVA GLAGOLA BITI V SLOVENŠČINI IN RUŠČINI

V prispevku obravnavamo glagol *biti* v slovenščini in glagol *быть* v ruščini. Posvečamo se njenemu pomenskemu opisu, kot ga je mogoče razbrati iz Slovarja slovenskega knjižnega jezika in Aktivnega slovarja ruskega jezika. Skladenjsko-pomenski preplet glagolov *biti* in *imeti* je predstavljen na osnovi primerjav med določenimi jezikovnimi konstrukcijami v ruščini in slovenščini. Vloga glagola kot slovnične vezi nas zanima predvsem z vidika ničte vezi, pojava v ruščini, ko se v povednem naklonu sedanjika vez opušča, v drugih časih in naklonih pa se premenjuje z izraženo obliko glagola *быть*. Na koncu prispevka predstavljamo izsledke analize posameznih odlomkov iz leposlovnega gradiva – romana *Pomladni dan* v slovenščini in njegovega prevoda v ruščino. Večinoma se izvirnik in prevod ujemata: slovenska konstrukcija vsebuje glagol *biti*, ruski prevod pa glagol *быть*. Opazili smo nekaj neujemanj, ki so na eni strani posledica jezikovnih zakonitosti posameznega jezika, na drugi pa prevodnih rešitev, ki so prevajalčeva odločitev.

COMPARISON OF THE VERB BITI (TO BE) IN SLOVENIAN AND RUSSIAN LANGUAGES

This article explores the verb *biti* in Slovenian and the verb *быть* in Russian. It focuses on their semantic descriptions as derived from the Dictionary of Standard Slovenian Language and the Active Dictionary of Russian. Syntactic-semantic interactions of the verbs *biti* and *imeti* are illustrated on the basis of a comparison of certain linguistic constructions in Russian and Slovenian. The role of the verb as a grammatical link is of particular interest to us in the context of the zero copula, a phenomenon in Russian in which the link is omitted in the indicative mood in present tense, but is replaced by the expressed form of the verb *быть* in other tenses and moods. The final part of the article presents the results of the analysis of some examples from literary fiction, the novel *Pomladni dan* in Slovenian and in its Russian translation. In most cases, there is a correspondence between the Slovenian construction (containing the verb *biti*) and the Russian translation (containing the verb *быть*). We have identified some discrepancies which arise, on the one hand, due to the linguistic characteristics of the respective languages, and on the other, due to the solutions chosen by the translator.

In memórium

Marc L. Greenberg (ORCID 0000-0001-8419-8779)

University of Kansas, Department of Slavic, German & Eurasian Studies, Kansas, USA
mlg@ku.edu

DOI: <https://doi.org/10.3986/16.1.10>

SLAVIST WILLEM R. VERMEER (1947–2024)

Slavic linguistics has lost a major scholar with the departure of Dutch Slavist Willem R. Vermeer, who passed away after a long illness in September 2024. Willem's interests were not limited to Slavic linguistics, but he had a wide range of interests, including visual art, music, and literature. He was an accomplished and generous teacher and mentor, a devoted husband to Trille and father to their daughter, Wotienke, both gifted artists. He will be greatly missed in the Slavic field for erudition and insight.

I first became aware of Willem's name when I was studying South Slavic dialectology and Slavic word prosody at UCLA in 1986, during two intensive graduate seminars given by Academician Pavle Ivić, then a visiting Fulbright scholar. The term papers he had assigned sent me to the University Research Library, where I found Willem's publications and those of other Dutch Slavists, who were among the few scholars writing in English on South Slavic topics. More than that, I was astonished to learn that Willem and other Dutch Slavists had conducted dialectological fieldwork in Yugoslavia (and elsewhere), rather than limit their analyses to dialect data only from published sources. There had been a common understanding in those days that that sort of thing was reserved for native speakers, and it was the lot of outsiders like me to do our dialect work in the library. In consultation with Academician Ivić I discussed my hopes to work directly with dialect speakers and he suggested that the right interlocutor was Willem,¹ for whom he had great praise as both a scholar and fieldworker, who also had close familiarity with Slovene. (My local role model was Ronelle Alexander, who had collected field data for her dissertation

¹ From the beginning Willem insisted that I call him by his first name and that we dispense with academic titles. In the continuation of this text, I invoke his egalitarian spirit by referring to him this way.

on Torlak dialects, whom Professor Ivić also held in high regard, but Torlak presented a completely different set of structural problems.) By then Willem had published his 1982 study on development of the vowel system, which gave a thorough structural account of the dialectalization of Slovene, building on earlier work by Fran Ramovš and Jakob Rigler. A letter of introduction from Academician Ivić to Professor (then Universitair Docent in the Dutch system) Vermeer helped to open a lively correspondence between Willem and me, which was to begin a fruitful mentorship and later also friendship.

When my Fulbright Fellowship to conduct research in the former Yugoslavia came through in 1987, Willem invited me to stay at his house in Amsterdam on the way to my fieldwork destinations. That short visit was a formative experience for me. I benefitted from rich discussions about linguistic topics from Willem, who then also introduced me to the experienced Dutch Slavist-fieldworkers Janneke Kalsbeek and Peter Houtzagers, as well as Han Steenwijk, who was then preparing for his work on the dialect of Resia. Willem's daily routine included sifting through dialect description at the breakfast table; he had filled hundreds of handwritten notebooks with forms that awaited explanation and integration into his analyses. He also set me straight then about the place of dialectologists-cum-historical-comparativists in the firmament of Slavic linguistics: "We're the lunatic fringe." I had found my tribe.

Being a mere graduate student, I was honored and fortunate to have found such a generous and helpful mentor as Willem. He was more than just a caring and patient mentor, however. His writings proved to be a continued source of inspiration and insight in the decades hence.

Willem's many Slavic research interests ranged from South Slavic historical dialectology to the East Slavic Birchbark letters and many related topics in between. The main thrust of his work grew from the Dutch School of Slavic accentology, which dates back to Nicolaas van Wijk (1880–1942), and developed in the post-Stangian vein through the work of Carl Ebeling (1924–2017) and his PhD student Frederik Kortlandt (b. 1946), Willem being the latter's first PhD dissertator. Through this lineage Willem was an indirect heir to Jakobsonian linguistics (RJ having been Ebeling's mentor), and through dialogue to the Moscow Accentological School (MAS). From these influences one can discern leitmotifs in Willems work, such as close attention to systemic structure, concern with the spatial dynamics of innovations, and focus on prosodic phenomena in inflection. It is this latter point that may arguably have

been the foremost impetus for the Dutch School researchers to venture into the field, because full accented paradigms were then, and remain today, rare sightings in the published dialectological literature. Willem's work tackled both individual matters of detail, thorough analysis of single (village) dialect systems, as well as large, complex topics. His work was always insightful and original, characteristically providing a thoughtful critique of previous work, proceeding logically and with disciplined attention to the comparative method, along with pertinent exemplification.

A remarkable feature of Willem's writing, whether it was in the form of lengthy personal letters or substantive articles, was that they seamlessly bridged pedagogy and research. As his daughter Wotienke aptly put it, explaining why Willem had saved his correspondence with me, as well as with other scholars, "Writing with people has always been a way to order his thoughts." Among his publications in the "thinking-aloud" vein is his 64-page article critiquing two shorter articles, totaling 36 pages, by D. J. L. Johnson (Vermeer 1984). This extended discourse with Johnson is a brilliant exegesis on the insights afforded to the development of the (Balto-)Slavic accent system by Stang 1957 and its followers in Moscow, Amsterdam, and Leiden. As a graduate student trying to crack the code of the esoteric subfield of Slavic accentology, I spent many intensive weeks studying Willem's 1984 paper and thinking about not only how much more was to be mined from the existing descriptive literature, but how new field data could enrich the collective endeavor to comprehend the labyrinthine paths of accent developments in Slavic dialects.

Willem's close attention to detail could also illuminate and historically contextualize an existing work in such a way as to add considerable value to it, such as to Jedvaj's classic description of the Bednja kajkavian² dialect (Jedvaj 1956), as is the case with Vermeer 1979. This sophisticated and insightful work, more than three-fifths the length of Jedvaj's description, is among the earliest of Willem's publications, the first one on his vita being his 1975 essay on Susak čakavian, where he had himself conducted fieldwork. Today Jedvaj's description cannot be read responsibly without considering Willem's critical analysis of 1979.

² In most of Willem's English-language publications he does not capitalize the labels kajkavian, čakavian, and štokavian, and I follow this practice here for consistency. In my writing I prefer using capitals, Kajkavian, etc., in accord with the current mainstream English practice, which bestows quasi-ethnonym status on the labels and thus comports with native usage.

Not one to shy away from tackling the most complex developments in the Western South Slavic sphere, Willem undertook a thorough and original reanalysis of the development of Resian vowel systems, one of most perplexing in all of Slavic dialectology, pointing out shortcomings in previous analyses and providing a reasoned structural account in 11 synchronic cross-cuts, beginning with the first South Slavic innovations (Vermeer 1987).

Focus on detailed dialect systems did not prevent Willem from engaging with big-picture matters – far from it. A great insight, in my view, is contained in his article “The rise and fall of the kajkavian vowel system” (Vermeer 1983), which reconciles two broad tendencies to explain a peculiarity of kajkavian, i.e., the merger of the reflex of Proto-Slavic jat (*ě) with the reflex of the jers (*ь/ъ): first, the tendency towards the raising of Common Slavic *ě, a general Slavic trend; second, the tendency in Western South Slavic to merge the jers as a low vowel. In his 1983 article, Willem discusses possible scenarios for this merger, preferring the explanation that this merger occurred when both the jat reflex and the merged-jers reflex were low vowels and, subsequently, they raised together in accord with the general (i.e., all-Slavic) tendency toward raising the jat reflex (448–451). This possibility, hitherto not entertained in the literature, provides a natural explanation for how this “defective” — as Willem was fond of characterizing such developments — outcome fits in with contiguous dialects, without assuming a radical systemic divergence. Willem’s vision could extend also beyond the South Slavic data to “see through” to the substratum languages: in Vermeer 1989, in the *Gedenkschrift* for Jakob Rigler, he demonstrated that two types of Romance substratal vowel systems underlay Western South Slavic, a western one, characteristic of modern Friulian and Dalmatian (Vegliote), corresponding to Slovene and kajkavian; and an eastern one, characteristic of Romanian, corresponding to čakavian and štokavian.

One could continue for a long time describing the many virtues that shone in Willem’s writing, but I will limit myself to just one further observation. In a subfield that lends itself to dense, dry, even mathematics-like writing, Willem managed to imbue his prose with a storyteller’s gift for lively narrative. Take, for example, his treatise on the continuity of vowel-quantity oppositions from Proto-Indo-European to Slavic (Vermeer 1992a), in which he describes successive vowel systems through time. On pages 128–129 we come to a crescendo of pent-up tension: “In pretonic syllables the laryngealized vowels

had been eliminated at an earlier stage [...]. So in that position either quantity was redundant or the new timbre contrasts were. Something was bound to happen.” And then comes the reveal: “What happened was that length was lost.” I like to think that this narrative technique comes from one of Willem’s eclectic reading choices, which, included — when he was “feeling low,” he once confided — Raymond Chandler’s crime novels (set in my native Los Angeles, where my grandfather had been a police reporter in the 1920s). But I really don’t know.

In the early days of Willem’s mentorship in the mid-1980s, he advised me to steer clear of Yugoslav politics, emphasizing rightly that outsiders like us would inevitably miss nuances and could thus get ourselves into trouble. He departed from his own advice when the killing began, publishing on the Yugoslav wars both in the Dutch press (e.g., Vermeer 1991) and in academic journals. Regardless of the audience, his writing remained on a high scholarly level, building arguments with verifiable facts and with a view to illuminating the problems to a readership that would otherwise have trouble identifying — let alone differentiating — Croats, Serbs, Bosnians, Montenegrins, Macedonians, Slovenes, or Kosovar Albanians. The bewildered reader could hardly do better to consult Willem’s scholarly explainer on the Serbian-Albanian conflict than his article Vermeer 1992b, which in a mere 23 pages sorts out not just the conflict, ranging from the prehistoric Illyrians to Serbia in 1878 in the wake of the Congress of Berlin to the infamous SANU Memorandum of 1991.

My sadness on the loss of Willem to our field is compounded by a particular regret: I could not manage to convince him to write an entry for the *Encyclopedia of Slavic Languages and Linguistics* (Brill), which I have been editing since 2016. While he expressed that he was honored to have been invited, he felt uncomfortable having the last word on a topic. This attitude reflects the nature of his engagement with the field: he was continually in dialogue with its ideas and refining them through deep analysis and systematic discovery. For him, I believe, there was no obvious end point to inquiry. He was also a consummate perfectionist, as is evidenced by his annotations and corrections of errata to his earlier publications, republished in *samizdat* form — with scrupulous attention to original pagination — on the Academia.edu website (leidenuniv.academia.edu/WillemVermeer). Let this be an invitation to read and engage with Willem’s incandescent and important writings, which richly deserve to be studied by future scholars.

REFERENCES

- Jedvaj, Josip. 1957. Bednjanski govor. *Hrvatski dijalektološki zbornik* 1. Zagreb: JAZU. 279–330.
- Stang, Christian S. 1957. *Slavonic Accentuation*. Oslo: i komisjon hos H. Aschehoug & Co. / W. Nygaard.
- Vermeer, Willem R. 1975. Problems in the synchronic and diachronic phonology of Susak čakavian. *Zbornik za filologiju i lingvistiku* 18/2: 139–159.
- Vermeer, Willem R. 1979. Innovations in the kajkavian dialect of Bednja. In: Jan M. Meijer (ed.), *Dutch Contributions to the Eighth International Congress of Slavists*. Lisse: Peter de Ridder. 347–381.
- Vermeer, Willem R. 1982. Raising of *ě and loss of the nasal feature in Slovene. *Zbornik za filologiju i lingvistiku* 25/1: 97–120.
- Vermeer, Willem R. 1983. The rise and fall of the kajkavian vowel system. *Studies in Slavic and General Linguistics* 3. Amsterdam, Atlanta: Rodopi. 439–477.
- Vermeer, Willem R. 1984. On clarifying some points of Slavonic accentology: the quantity of the thematic vowel in the present tense and related issues. *Folia Linguistica Historica* 5/2: 331–395.
- Vermeer, Willem R. 1987. Rekonstruiranje razvoja samoglasniških sestavov v rezijanskih govorih. *Slavistična Revija* 35/3: 237–257.
- Vermeer, Willem R. 1989. Traces of an early Romance isogloss in Western Balkan Slavic. *Slavistična Revija* 37/1–3 (= *Riglerjev zbornik*): 15–30.
- Vermeer, Willem R. 1991. De genocide van hoofdman Ante Pavelic en zijn Ustase op Serviërs. *NRC Handelsblad*, 19 september 1991 (archived). <https://www.nrc.nl/nieuws/1991/09/19/de-genocide-van-hoofdman-ante-pavelic-en-zijn-ustase-6980792-a323447> (21. 10. 2024).
- Vermeer, Willem R. 1992a. *In the beginning was the lengthened grade: On the continuity of Proto-Indo-European vowel quantity in Slavic*. In: R. Beekes, A. Lubotsky, J. Weitenberg (eds.). *Rekonstruktion und relative Chronologie: Akten der VIII. Fachtagung der Indogermanischen Gesellschaft, Leiden, 31. August – 4. September 1987*. Innsbruck (= *Innsbrucker Beiträge zur Sprachwissenschaft* 65). 115–136.
- Vermeer, Willem R. 1992b. Albanians and Serbs in Yugoslavia. *Yearbook of European Studies* 5. *The disintegration of Yugoslavia*. 101–124.