

Korpusna analiza slovenskega deležja v različnih besedilnih tipih

Tamara Mikolič Južnič

Oddelek za prevajalstvo, Filozofska fakulteta, Univerza v Ljubljani
Aškerčeva 2, Ljubljana
tamara.mikolic@guest.arnes.si

Povzetek

Prispevek se osredotoča na raziskavo rabe deležja v različnih besedilnih tipih v slovenskem jeziku. O splošno uveljavljenem mnenju o redkosti slovenskega deležja govorijo številni avtorji, vendar so dejanske, s konkretnimi podatki podprte raziskave na tem področju izjemno redke. V tem prispevku želimo ugotoviti, kako pogosto se deležje pojavlja v nekaterih slovenskih neliterarnih besedilnih tipih, kot so znanstvena besedila (humanistično-družboslovna ter naravoslovna) in strokovna pravna besedila (zakoni) ter v korpusu splošne slovenščine, pa tudi v leposlovnih besedilih. Pri zbiranju podatkov so bili uporabljeni različni razpoložljivi korpusi, kot je korpus JOS 100K, korpus Spook in ad hoc korpus neliterarnih besedil. Ob zbiranju podatkov smo preverili tudi uspešnost oblikoskladenjskega označevanja projekta JOS glede deležij.

Corpus Analysis of the Slovene Participle in various text types

The paper focuses on a study of the use of the Slovene participle in various text types. Several authors have pointed out that the participle is rarely used, but actual research based on concrete data is extremely rare in this field. Here we want to find out how frequent the participle is in certain Slovene non-literary text types such as scientific texts (from the fields of the arts and humanities and the natural sciences) and legal texts (laws) and in a reference corpus of Slovene, as well as in literary texts. Several different corpora were used in collecting data (JOS 100K, Spook and an ad hoc corpus of non-literary texts). When collecting the data, we also verified how successful the morphosyntactic tagging system of the JOS project is with respect to participles.

1. Uvod

Slovensko deležje, tako kot nekatere druge slovenske neosebne glagolske oblike, nima dobrega slovesa med uporabniki jezika: Žagar (1998: 41) ugotavlja, da njihova raba peša, Schlamberger Brezar (2005: 127) pa vidi rabo deležja kot oddaljeno od naravne, zlasti govorne rabe, ter jo povezuje s pridihom arhaičnosti. Dejansko pa prepričanje, da se deležje dandanes ne uporablja (naj ne bi uporabljalo?), zaenkrat še ni dobilo dokončne znanstvene potrditve, saj se obstoječe študije osredotočajo bodisi na zgodovinski vidik (Jesenšek 1998) bodisi na posamezne besedilne tipe (prim. naslednji razdelek).

Z vidika zastopanosti določenega jezikovnega elementa v jeziku nasploh ali v posameznih žanrih oz. področjih, se korpusi samoumevno ponujajo kot odlično raziskovalno orodje. Slovenščina je v tem smislu privilegirana, saj obstaja že cela vrsta prosto dostopnih korpusov in orodij za gradnjo lastnih korpusov, ki raziskovalcem omogočajo relativno hitro in enostavno pregledovanje velikih količin podatkov.

Tako želimo v pričujoči raziskavi na podlagi različnih korpusov preveriti, koliko in kako se deležje pojavlja v slovenskih pisnih besedilih, in sicer zlasti v določenih neleposlovnih žanrih v primerjavi s splošnim jezikom oz. z leposlovjem. Pričakujemo, da se deležje v nobenem besedilnem tipu ne bo pojavljalo pogosto; po drugi strani pa pričakujemo tudi, da se bodo med različnimi besedilnimi tipi pojavljale konkretne razlike, saj predhodne raziskave (prim. Mikolič Južnič, v tisku) nakazujejo, da je pojavljanje deležja do določene mere odvisno od besedilnega tipa. Poleg tega skozi raziskavo nameravamo opazovati, kako natančno so korpusi oblikoskladenjsko označeni oz. ali so v izbranih korpusih, označenih večinoma z avtomatskimi orodji, deležja pravilno kategorizirana.

2. Slovenska deležja

Deležje ima v slovenskem jeziku dvojno naravo: je glagolska oblika, ki meji na prislov. V skladu s Toporišičem (1991: 339), deležja na *-č* in *-aje* izražajo »glagolsko dejanje, ki se vrši hkrati s kakim drugim dejanjem«, deležje na *-e* pa se »večinoma rabi kot prislov načina«. Deležijski polstavek, ki deležje vsebuje, »izraža glagolsko dejanje, ki poteka hkrati z drugim dejanjem« (Žagar 2001: 291). Jesenšek (1998: 41) navaja, da deležja najdemo zlasti v umetnostnih in strokovnih besedilih, vendar pa da njihova raba upada, če pa se pojavljajo, izražajo istodobnost dogajanja z glagolskim dogodkom v glavnem stavku. Po drugi strani Mezeg (2011: 323) ugotavlja, da se deležja na *-č* pojavljajo v nezanimljivem številu tako v publicističnih kot tudi (še pogosteje) v literarnih prevodih iz francoščine, Schlamberger Brezar (2005) pa ugotavlja povišano prisotnost deležijskih stavkov v prevodih političnih besedil iz francoščine.

Deležja prepoznamo po značilnih obrazilih *-oč* (npr. *čakajoč*, primer 1), *-eč* (*hodeč*, primer 2), *-aje* (*upoštevaje*, primer 3) in *-(v)ši* (*začenši*, primer 4, *vštevši*, primer 5), poleg teh pa tudi *-e* (*miže*, *smeje*, *sede*, primeri 6, 7, 8), vendar so taka deležja pogosto kategorizirana kot prislovi z glagolsko etimologijo.¹

- (1) *Čakajoč*, da obiskovalci opravijo svoj ritual ...
- (2) Modno pisto sem zapustil *hodeč* nazaj ...
- (3) *Upoštevaje* okoliščine, je bil jedilni kot ...
- (4) ... bil pred tem, *začenši* z letom 1961, sekretar ...
- (5) ... Jackson pa je, *vštevši* še chicaške uspehe, nanizal ...
- (6) Časopis bi lahko delila *miže* ...
- (7) ... žensko, ki je *smeje* se zapuščala prizorišče ...

¹ Navedeni primeri so bili naključno izbrani iz korpusa Fidaplus.

- (8) ...je življenje *sedé*, brez hoje in brez dela ali športa nezdravo, ...

V pričujočo raziskavo so vključene vse navedene oblike, saj nas po eni strani zanima celovit pregled rabe deležja v obravnavanih besedilnih tipih, po drugi strani pa je dejansko praktično vse primere takih 'glagolskih' prislovov mogoče pretvoriti v osebne glagolske oblike, tako kot to lahko naredimo za 'prava' deležja, kar je dodaten dokaz, da gre v bistvu za glagolske oblike in ne prislove.²

Iz omenjenih raziskav lahko sklepamo, da je raba deležja in njegova pogostnost do določene mere odvisna od besedilnega tipa, s katerim imamo opravka, zato želimo v pričujoči raziskavi s pomočjo korpusov, opisanih v naslednjem razdelku, ter s primerjavo s podatki, pridobljenimi v predhodni raziskavi (Mikolič Južnič, v tisku), ugotoviti, ali naši podatki dejansko potrjujejo to odvisnost.

3. Korpusi in metode

3.1. Korpusi

Ker želimo v raziskavo vključiti različne besedilne tipe ter predstaviti, kako se raba deležja spreminja v nekaterih različicah slovenskega jezika, se kot samoumevni vir ponuja korpus Fidaplus, ki je uravnotežen, izredno zajeten referenčni korpus (621 milijonov besed) in do neke mere omogoča izbiro prenosnika in zvrsti besedil, ki jih vključuje. Vendar pa ta korpus prinaša določene težave, ki jih v okviru te raziskave ni bilo mogoče rešiti: po eni strani so podatki o zvrstnosti besedil oz. s tem povezane možnosti izbire podkorpusov preveč omejene, da bi bile dejansko uporabne za natančnejše analize; po drugi strani v korpusu deležja niso označena s svojo lastno kodo MSD (prim. Arhar 2007), kar bistveno oteži luščenje primerov, saj druge metode (podobne tistim, opisanim v naslednjem razdelku) prinašajo preveliko število šuma (oz. nezazelenih zadetkov) in s tem za ročno analizo neobvladljivo število primerov.

Zaradi navedenih razlogov smo se odločili, da uporabimo drug korpus (ali bolje korpus). Kot kontrolni korpus splošne slovenščine smo uporabili JOS 100K, korpus s 100.000 pojavnici, ki vsebuje vzorce referenčnega korpusa Fidaplus s podrobno ročno preverjenimi jezikoslovnimi oznakami. Poleg tega smo uporabili specializirani korpus znanstvenih in strokovnih besedil, ki je nastal v sklopu predhodne raziskave o pojavljanju nominalizacije v izbranih slovenskih besedilnih tipih (Mikolič Južnič 2011). Korpus skupno šteje približno 870.000 pojavnici ter združuje znanstvena besedila s področja humanistike in družboslovja na eni ter naravoslovja na drugi strani in strokovna besedila s področja slovenske zakonodaje. Podrobna sestava korpusa je navedena v Tabeli 1. Korpus je tudi jezikoslovno

² Izjema je beseda *glede* (npr. *Prečne pregrade razporedimo glede na izvedbo notranjosti vlaka*; vir: Fidaplus). Čeprav bi pogosto v primerih, ki vsebujejo *glede*, lahko besedno zvezo prav tako pretvorili v odvisni stavek, to ni vedno mogoče oz. smiselno. Poleg tega ima beseda izjemno veliko pogostnost (preko 100.000 primerov v Fidaplus) in bi izkrivila sliko o splošni rabi deležja v jeziku.

označen, in sicer s pomočjo spletnega servisa projekta JOS (Erjavec idr. 2010) oz. orodjem ToTaLe.

Podatke, pridobljene po metodah, opisanih v nadaljevanju, smo primerjali tudi s podatki iz predhodne raziskave (Mikolič Južnič, v tisku), opravljene na osnovi korpusa Spook (Vintar 2009, v tisku), ki združuje izvirna ter prevedena slovenska leposlovna besedila ter izvirnike danih prevodov v štirih jezikih (angleščini, francoščini, nemščini in italijanščini).³ Na ta način smo želeli omogočiti širši vpogled v rabo deležja v slovenskem jeziku.

3.2. Metode

Po izboru korpusov je bil prvi korak analize luščenje primerov rabe deležja. Korpus JOS 100K omogoča ustvarjanje konkordanc v lastnem spletnem iskalniku, za brskanje po specializiranem korpusu znanstvenih in strokovnih besedil pa smo uporabili Sketch Engine (Krek in Kilgarriff 2006), pri čemer smo predhodno korpus oblikoskladenjsko označili, kot smo omenili zgoraj.

Ker sta izbrana korpusa torej označena z jezikovnimi informacijami (npr. glede besedne vrste), v okviru katerih imajo deležja svojo specifično kodo, smo najprej izluščili primere z iskanjem te kode. A ker pri avtomatskem označevanju včasih lahko prihaja tudi do napak (narobe označenih besed, in sicer bodisi besed, ki jim je koda za deležje napačno pripisana, bodisi takih, ki deležja so, vendar niso označena kot taka), smo podatke želeli preveriti tudi s kombinirano avtomatsko in ročno metodo, oz. z iskanjem po obrazilih. Tako smo preverili prisotnost pripon *-eč*, *-oč*, *-aje*, *-ši*, poleg tega pa tudi *-če*, *-de*, *-te* in *-eje*.⁴

Ko smo avtomatsko pridobili vse primere, v katerih se pojavljajo dane pripone, je bilo treba odstraniti šum, tj. vse tiste primere, kjer se naključne besede končajo z identičnimi črkami kot naša obrazila, vendar ne gre za deležja (npr. *več* za *-eč*). Tega dela smo se lotili v dveh fazah: v prvi smo odstranili vse naključno podobne besede, nato pa posebej razločili še med deležji in deležniki, saj imata slednji dve kategoriji pogosto identični obliki, ki seveda nista uporabljena v enakih okoliščinah. Na ta način smo dobili seznam praktično vseh deležij, ki se pojavljajo v danih besedilih.⁵

Dobljene rezultate smo končno primerjali tudi s predhodno raziskavo o deležjih v literarnih (izvirnih in prevedenih) besedilih in podatki, ki so bili pridobljeni s

³ Dejansko so bili v dani raziskavi bili uporabljeni samo prevodi iz italijanskega jezika.

⁴ Čeprav se določena obrazila v obravnavanih korpusih pravzaprav ne pojavijo niti enkrat (prim. razdelek 4.), jih vseeno navajamo, in sicer iz dveh razlogov: a) ker s tvorbenega in slovarskega vidika kot možne oblike obstajajo in smo njihovo prisotnost zato preverjali; in b) ker smo želeli doseči čim večjo primerljivost z rezultati, pridobljenimi iz literarnih besedil (prim. razdelek 4.3.).

⁵ Mogoče je, da se je kako posamezno deležje izmuznilo, saj zaradi neizmerno velikega šuma, ki bi se pojavil ob iskanju z obrazilom *-e*, nismo pregledali vseh mogočih kombinacij deležij na *-e*, temveč le določene, ki so se predhodno izkazale za bolj pogoste, vendar s precejšnjo mero gotovosti lahko trdimo, da smo jih zajeli zelo veliko večino (tudi z ozirom na dejstvo, da med deležji, ki smo jih izluščili s pomočjo kod MSD, praktično ni takih, ki jih ne bi zajeli z obrazili).

Vrsta besedila	Izbrana področja (in viri)	Število pojavníc
Znanstvena besedila	Humanistika in družboslovje <i>Slavistična revija</i> <i>Sodobna pedagogika</i> <i>Revija za kriminalistiko in kriminologijo</i> <i>Geografski vestnik</i> <i>Etnolog</i>	415.000 81.399 74.355 109.788 53.611 95.847
	Naravoslovje <i>Zobozdravstveni vestnik</i> <i>Medicinski razgledi</i> <i>Acta biologica slovenica</i> <i>Geodetski vestnik</i> <i>Farmaceutski vestnik</i>	304.744 55.456 72.178 52.780 55.364 68.966
Strokovna besedila	Slovenska zakonodaja <i>Stanovanjski zakon SZ-1</i> <i>Zakon o društvih ZDRu-1</i> <i>Zakon o verski svobodi ZVS</i> <i>Zakon o volilni in referendumski kampanji ZVRK</i> <i>Zakon o javnih zbiranjih ZJZ</i> <i>Zakon o splošnem upravnem postopku ZUP</i> <i>Zakon o upravnih taksah ZUT</i> <i>Kazenski zakonik KZ-1</i>	154.088
	SKUPAJ POJAVNIC	873.832

Tabela 1: Podrobna zgradba specializiranega korpusa znanstvenih in strokovnih besedil ter število pojavníc v posameznem podkorpusu (povzeto po Mikolič Južnič 2011: 323).

pomočjo korpusa Spook, kot smo že omenili (Mikolič Južnič, v tisku). Ker je bila ta raziskava izvedena z enako metodologijo in smo rezultate normalizirali na 500.000 pojavníc, so rezultati obeh analiz primerljivi (čeprav je seveda normaliziranje navzgor pri tako majhnih številkah do neke mere tvegano. Sicer pa gre v Mikolič Južnič (v tisku) za širšo analizo, ki upošteva tudi določene aspekte z vidika prevajanja, zato smo povzeli le del podatkov, ki je relevanten za pričujočo analizo (navedeni so v razdelku 4.3.)

4. Rezultati in diskusija

4.1. Iskanje s kodami MSD

Tako spletni konkordančnik korpusa JOS 100K kot tudi Sketch Engine omogočata iskanje po oblikoskladenjskih oznakah besed v korpusih. Ker je bila pri korpusu JOS 100K uporabljena tabela oznak JOS in je bil specializirani korpus, ki smo ga uporabili, označen z orodjem ToTaLe,

Korpus	Absolutno št. deležij	Št. deležij na 500.000 pojavníc
JOS 100K	18	90
Znanstvena besedila	58	40,3
	<i>Humanistika/družboslovje</i> 43	29,9
	<i>Naravoslovje</i> 15	10,4
Strokovna besedila	Zakonodaja 4	13,0

Tabela 2: Pojavitve deležja v treh obravnavanih besedilnih tipih (iskanje po kodah MSD)

je bilo iskanje izvedeno z enakim iskalnim pogojem. Rezultati te analize so povzeti v Tabeli 2.

Podatki o pojavitvah deležij, zbrani po kodah MSD, so zaradi različne velikosti korpusov in lažje primerljivosti preračunani na 500.000 pojavníc. Na prvi pogled lahko opazimo, da so razlike med posameznimi korpusi precejšnje, vendar se bomo temu podrobneje posvetili v nadaljevanju, ko bomo pregledali pojavitve tudi po posameznih obrazilih. Pred tem nas je zanimalo tudi, katera obrazila se pojavljajo pri zgoraj navedenih deležjih: ti podatki so zbrani v Tabeli 3, v kateri smo prav tako v oklepajih navedli normalizirane podatke na 500.000 pojavníc.

	JOS 100K	Znanstvena besedila		Strokovna besedila
		Humanistika/družboslovje	Naravoslovje	Zakonodaja
-eč	2 (10)	5 (3,5)	3 (2,1)	1 (3,2)
-oč	9 (45)	31 (21,5)	10 (6,9)	3 (9,7)
-aje	2 (10)	3 (2,1)	0	0
-ši	1 (5)	3 (2,1)	2 (1,4)	0
-če	3 (15)	0	0	0
-te	0	1 (0,7)	0	0
-de	1 (5)	0	0	0
-eje	0	0	0	0

Tabela 3: Pojavitve posameznih obrazil ob iskanju s kodami MSD (podatki, preračunani na 500.000 pojavníc, v oklepaju)

Preden se lotimo interpretiranja podatkov, ki smo jih pridobili, si oglejmo še rezultate iskanja s posameznimi obrazili.

4.2. Iskanje z obrazili

Iskanje z obrazili je seveda malo bolj zamudno zaradi nujnega ročnega pregledovanja primerov, vendar so rezultati pogosto vredni dodatnega truda.

Preverili smo vsa zgoraj navedena obrazila (-eč, -oč, -aje, -ši, -če, -de, -te in -eje), končne rezultate pa strnili v tabeli 4, v kateri so v zadnjem stolpcu navedeni tudi rezultati, preračunani na 500.000 pojavníc.

Korpus		Absolutno št. deležij	Št. deležij na 500.000 pojavníc
JOS 100K		23	115
Znanstvena besedila		83	57,7
	Humanistika/družboslovje	61	42,4
	Naravoslovje	22	15,3
Strokovna besedila	Zakonodaja	15	46,7

Tabela 4: Pojavitve deležja v treh obravnavanih besedilnih tipih (iskanje po obrazilih)

Če primerjamo Tabelo 4 s Tabelo 2, takoj opazimo, da podatki ne sovpadajo popolnoma, oziroma da so številke v Tabeli 4 višje. Pred natančnejšo diskusijo si lahko ogledamo še podrobnejšo sliko za posamezna obrazila, ki jo navajamo v Tabeli 5, in sicer v absolutnih ter normaliziranih številkah, preračunanih na 500.000 pojavníc (v oklepaju).

	JOS 100K	Znanstvena besedila		Strokovna besedila
		Humanistika/družboslovje	Naravoslovje	Zakonodaja
-eč	3 (15)	6 (4,2)	3 (2,1)	1 (3,2)
-oč	9 (45)	43 (29,8)	15 (10,4)	5 (16,2)
-aje	2 (10)	3 (2,1)	0	1 (3,2)
-ši	1 (5)	2 (1,4)	2 (1,4)	0
-če	8 (40)	4 (2,8)	2 (1,4)	8 (26,0)
-te	0	3 (2,1)	0	0
-de	1 (5)	0	0	0
-eje	0	0	0	0

Tabela 5: Pojavitve posameznih obrazil ob iskanju z obrazili (podatki, preračunani na 500.000 pojavníc, v oklepaju)

4.2.1. Primerjava med iskanjem s kodami MSD in z obrazili

V zgornjih tabelah navedeni podatki nam lahko odgovorijo na različna vprašanja. Če se najprej osredotočimo na primerjavo med pridobivanjem podatkov

s kodami MSD in z obrazili, vidimo, da prihaja do konkretnih razlik med obema metodama analize: v vseh primerih je pri analizi s kodami MSD prišlo do izgube podatkov ali, z drugimi besedami, vsa dejansko prisotna deležja niso bila kategorizirana kot taka. Najmanjša razlika se pojavlja pri korpusu JOS 100K, kjer kode MSD niso zajele 11,7 odstotkov najdenih deležij. Pri znanstvenih besedilih je bilo napačno označenih 30 odstotkov deležij, pri strokovnih pravnih besedilih pa celo 73,3 odstotkov. Tabeli 3 in 5 nam kažeta, kako do razlik prihaja tako pri enem in istem obrazilu (pri pravnih besedilih, na primer, kode MSD označujejo tri primere na -oč, medtem ko jih je prisotnih pet. Pravilno je na primer kategorizirano deležje v primeru (9), napačno pa v primeru (10).⁶

(9) ... šest mesecev in ne daljši od dveh let, računajoč od dne pravnomočnosti sodbe.

(10) ... kadar v upravnih stvareh, neposredno uporabljajoč predpise, odločajo o pravicah ...

Razlogi za napačno označevanje ležijo gotovo v skladijskem okolju deležij, ki so navadno zaradi bližine besede, kategorizirane kot samostalník, zmotno označena kot deležniki (seveda do napak prihaja skoraj vedno ravno tam, kjer so oblike deležnika in deležja identične). Sklepamo torej, da ima uporabljeni sistem za označevanje določene systemske težave z razdvoumljanjem med tema dvema glagolskima oblikama, ki se lahko pojavljata v na videz podobnih skladijskih strukturah.

Zaključimo lahko, da nam avtomatska analiza deležij s pomočjo kod MSD sicer lahko nudi podatke o tem, ali so deležja prisotna v določenem korpusu besedil ali ne, vendar za natančnejše kvantitativne analize podatki žal niso dovolj zanesljivi.

4.2.2. Primerjava pojavljanja deležja v obravnavanih besedilnih tipih

V nadaljevanju se vrnimo na osnovno vprašanje, ki ga obravnavamo v pričujočem prispevku, pojavljanje deležja v izbranih besedilnih tipih. Pri tem se bomo oprli na podatke v Tabelah 4 in 5, ki zajemata praktično celotno število deležij, prisotnih v analiziranih korpusih, in bomo opazovali normirane podatke na 500.000 pojavníc.

Največ deležij je zabeleženih v korpusu splošne slovenščine JOS 100K (115 primerov na 500.000 pojavníc); čeprav gre podatke jemati z dozo rezerve zaradi majhnega obsega korpusa, vsekakor nudijo osnovno idejo o pogostnosti deležja v splošnem (pisnem) jeziku. V ostalih dveh besedilnih tipih je deležij bistveno manj (pol manj v znanstvenih besedilih in celo 60 odstotkov manj v strokovnih pravnih besedilih).

Zanimiva je tudi velika razlika med obema analiziranimi tipoma znanstvenih besedil: veliko bolj pogosto se deležje namreč pojavlja v humanističnih in družboslovnih znanstvenih prispevkih (42,4 primerov na 500.000 pojavníc) kot v naravoslovnih (15,3 primerov na 500.000 pojavníc). Očitno gre pri zadnjem tipu besedil skorajda za izogibanje tej obliki: skoraj vsi obravnavani primeri imajo obrazilo -oč in večinoma gre za primere, v katerih so uporabljene ustaljene fraze, ko vidimo v

⁶ Oba primera sta vzeta iz korpusa strokovnih pravnih besedil.

primerih (11) in (12), ne pa za deležja v njihovi klasični glagolski funkciji v polstavkih, kot v primeru (13).

- (11) V preteklih desetih letih je bil tako *rekoč* v vseh državah v regiji ...
- (12) Za Slovensko populacijo imamo, *zahvaljujoč* že omenjenim raziskavam ...
- (13) *Zavedajoč* se problema, mnogi iščejo ustrezne ...

V družboslovnih in humanističnih besedilih je raznolikost uporabljenih deležij večja; čeprav tudi pri teh izrazito prevladujejo deležja na *-oč*. Najdemo primere z vsemi obrazili razen *-de* in *-eje*. In če so med primeri z obrazilom *-oč* v naravoslovnih besedilih samo štirje primeri od 15 taki, kjer deležje ni ustaljena fraza, temveč živa oblika, je takih deležij v humanističnih in družboslovnih besedilih več kot polovica (23 primerov od 43 deležij na *-oč*). Torej ne gre samo za razliko v pogostnosti, marveč tudi v raznovrstnosti deležij.

V strokovnih pravnih besedilih kljub temu, da se na splošno deležja precej redko pojavljajo (46,7 primerov na 500.000 pojavnic), ne najdemo ustaljenih fraz, podobnih zgoraj navedenim, temveč izključno različna deležja v 'pravih' polstavkih.

Tudi v korpusu splošnih besedil JOS 100K je opazna določena raznolikost, tako glede prisotnih obrazil kot tudi glede posameznih deležij, ki se pojavljajo.

Če podatke pogledamo še transverzalno, opazimo, da se med obrazili skoraj povsod na prvem mestu pojavlja *-oč* (razen pri strokovnih pravnih besedilih, kjer je na drugem mestu, za *-če*), ostala obrazila pa se (spet z izjemo pravnih besedil in tokrat tudi splošne slovenščine) pojavljajo neprimerno redkeje, če sploh.

Mogoče je, da se priponi *-oč* in *-če* na splošno najpogosteje pojavljata tudi zaradi podobnosti oblike z deležniki, ki se vsekakor zdijo veliko bolj prisotni in uporabni/uporabljeni od deležij:⁷ prisotnost in sprejemljivost enake formalne oblike bi lahko pogojevala večjo sprejemljivost druge, vsaj v določenih besedilnih tipih.

Zbrani podatki torej nakazujejo, da se deležje precej različno uporablja v različnih besedilnih tipih in da je pravzaprav veliko pogostejše v splošnem pisnem jeziku kot v znanstvenem ali strokovnem pravnem jeziku, v nasprotju z nekaterimi teoretičnimi predpostavkami, izraženimi v uvodnih delih prispevka. Poleg tega se izkazuje tudi razlika med dvema analiziranima tipoma znanstvenih besedil, saj je v našem humanistično-družboslovnem podkorpusu skoraj trikrat več deležij kot v naravoslovnem. Besedila s teh dveh področij se gotovo zelo razlikujejo po slogu in načinu argumentiranja, zato ta razlika ni posebno presenetljiva. Pravzaprav se celo odlično ujema s podatki o deležjih v literarnih besedilih, ki jih navajamo v nadaljevanju.

⁷ Dejansko o tem, kolikor je znano, ne obstaja nobena študija, gre preprosto za kombinacijo površinskega opazovanja korpusov in osebnega intuitivnega mnenja. Sicer pa je primerjava z deležniki omenjena tako zaradi površinske podobnosti oblik kot tudi zaradi dejstva, da gre v obeh primerih za neosebne glagolske oblike, čeprav je njihova pogostnost v jeziku seveda medsebojno neodvisna.

4.3. Primerjava pojavljanja deležja v analiziranih neliterarnih korpusih in v literarnih besedilih

Za konec smo dobljene rezultate primerjali s podobno raziskavo, izvedeno na korpusu izvirnih in prevedenih literarnih besedil Spook (Mikolič Južnič, v tisku). Kot omenjeno zgoraj, gre za raziskavo, ki se posveča tudi vplivom izvirnega jezika (v tem primeru italijanščine) na jezik literarnih prevodov, vendar se na tem mestu omejujemo le na pojavljanje deležja v slovenskih besedilih kot takih.

Izkaže se, da je število deležij v literarnih besedilih, tako v izvirnih kot v prevedenih, na splošno veliko večje kot v zgoraj omenjenih splošnih, znanstvenih ali strokovnih besedilih. Podrobni podatki o številu pojavnic v uporabljenih korpusih in o najdenih deležjih so podani v Tabeli 6.

	Izvirna literarna besedila	Prevedena literarna besedila
<i>Št. pojavnic</i>	1.454.275	478.591
<i>-oč</i>	337 (115,9)	35 (36,6)
<i>-eč</i>	137 (47,1)	34 (35,5)
<i>-aje</i>	70 (24,1)	14 (14,6)
<i>-ši</i>	4 (1,4)	0
<i>-če</i>	401 (137,9)	87 (90,9)
<i>-de</i>	345 (118,6)	47 (49,1)
<i>-te</i>	52 (17,9)	7 (7,3)
<i>-eje</i>	4 (1,4)	6 (6,3)
Skupno	1.350 (464,2)	230 (240,3)

Tabela 6: Prisotnost deležja v izvirnih in prevedenih literarnih besedilih (povzeto po Mikolič Južnič, v tisku; v oklepajih so rezultati, normalizirani na 500.000 pojavnic)

Če vzamemo povprečje pojavljanja deležja v vseh obravnavanih literarnih besedilih, se številka giblje okoli 352,25 primerov na 500.000 pojavnic, kar je več kot trikrat toliko kot v splošnih besedilih korpusa JOS 100K, kjer je bila pogostnost deležja v naši raziskavi največja. Pomenljivo je dejstvo, da se deležje pogosteje (skoraj dvakrat pogosteje) pojavlja v izvirnih literarnih besedilih kot v prevedenih delih, in to kljub dejstvu, da je italijanščina bogata z neosebnimi glagolskimi strukturami in bi lahko s tega vidika pričakovali kako interferenco v prid povečanega števila deležij (kot npr. opaža Schlamberger Brezar 2005 za prevode političnih besedil iz francoščine), vendar podatki tega ne potrjujejo.⁸

Razlog za tako izrazito različno pojavnost morda lahko iščemo v odnosu do deležja, ki ga imajo govorniki slovenščine. O tem, koliko je deležje prisotno v spontanem govoru oz. drugih, s pravili manj obremenjenih besedilnih tipih, ni podatkov, vendar smo govorniki večinoma vsi zrasli z idejami, ki jih širijo obstoječe slovnice (npr. Toporišič 1991), o tem, da je deležje zastarelo, nerodno, malo uporabno, tako da ni nič kaj

⁸ Smotno bi bilo preveriti, ali se podobno dogaja tudi v literarnih prevodih iz drugih jezikov, vendar naj bo to predmet prihodnjega raziskovanja.

presenetljivo, da se mu pri tvorjenju besedil marsikdo zavestno ali podzavestno izogiba. Le pri kreativnem pisanju – literaturi – si tvorci besedil očitno suvereno upajo posegati po tem slovničnem orodju slovenščine. Zdi se, da bolj kot je besedilo oddaljeno od te kreativnosti, bolj se deležje izgublja: v slogu humanistično-družboslovnih znanstvenih besedil, na primer, pogosto lahko zasledimo splošne ambicije po leporečju in posledično poseganje tudi po nevsakdanjih jezikovnih oblikah (v našem primeru deležjih), redkeje pa so take težnje prisotne v naravoslovnih ali strokovnih pravnih besedilih. Po drugi strani se prevajalci pri prevajanju literarnih besedil morajo soočiti s številnimi težavami, ne nazadnje z izogibanjem negativnemu transferu in s tem povezani morebitni posledični hiperkorektnosti, zaradi katere je število deležij dejansko bistveno manjše kot v izvorni slovenski literaturi. Tako se literarna dela, prevedena iz italijanščine, dejansko po številu prisotnih deležij postavljajo nekako vmes med splošni in literarni jezik ter na ta način deloma sledijo tradicionalnemu slovničarskemu videnju deležja, deloma se prilagajajo živemu vplivu slovenske literature, deloma pa verjetno tudi drugim literaturam, kjer so podobne slovnične strukture pogostejše kot v našem jeziku.

5. Sklep

Deležje se je izkazalo kot zanimiva tema za korpusno raziskovanje, kljub temu da se samodejno označevanje besedne vrste v tem primeru ni najbolje obneslo, saj je bilo iskanje po obrazilih kot preprostih zaporedjih črk precej bolj uspešno.

Raziskava je ponudila zanimiv vpogled v pojavljanje deležja v različnih besedilnih tipih slovenskega jezika. Tudi če gre precejšen del pojavitev deležja pripisati stalnim besednim zvezam, pri katerih ni prisoten pridih starinskosti, se namreč izkaže, da se ta glagolska oblika v jeziku ne pojavlja enakomerno, marveč se zdi vezana na tiste žanre, kjer avtorji čutijo večjo kreativno svobodo in jezikovno samozavest, to je v izvirnih literarnih besedilih. Besedila, v katerih se tega elementa kreativnosti ne pričakuje, kot so na primer znanstvena besedila s področja naravoslovja ali strokovna pravna besedila, vsebujejo izrazito manj deležij (tudi 30-krat manj) od literarnih besedil.

Veliko vprašanj glede deležja ostaja še nerešenih in torej odprtih za nadaljnje raziskovanje: besedilnih tipov, v katerih bi lahko iskali prisotnost deležja, ostaja še mnogo, na prvem mestu pa so zagotovo različne tipologije govorenega jezika; poleg tega je nujno tudi preverjanje, kako se deležje pojavlja v različnih tipih literarnih besedil in ali prihaja do razlik pri prevodih iz različnih tujih jezikov.

Kar zadeva označevanje slovenskih besedil s kodami MSD pa bi bilo prav glede razdvoumljanja med deležji in deležniki potrebno vložiti še precej truda, preden bi

korpusi omogočali dovolj zanesljivo samodejno iskanje tovrstnih elementov.

6. Literatura

- Arhar, Š. (2007). *Kaj početi z referenčnim korpusom Fidaplus*. Ljubljana: Filozofska fakulteta. http://www.fidaplus.net/Files/Kaj%20poceti%20s%20korpusom%20FidaPLUS_navadna%20dvostranska.pdf, dostop 31.07.2012.
- Erjavec, T., D. Fišer, S. Krek, in N. Ledinek (2010). The JOS Linguistically Tagged Corpus of Slovene. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Malta. <http://www.lrec-conf.org/lrec2010/>, dostop 31.07.2012. 1806–1809.
- Fidaplus. Referenčni korpus slovenskega jezika. <http://www.fidaplus.net/>, dostop 31.07.2012.
- Jesenšek, M. (1998). *Deležniki in deležja na -č in -ši. Razširjenost oblik v slovenskem knjižnem jeziku 19. stoletja*. Maribor: Slavistično društvo Maribor.
- Korpus JOS 100K. <http://nl.ijs.si/jos/>, dostop 31.07.2012.
- Krek, S. in A. Kilgarriff (2006). Slovene Word Sketches. V T. Erjavec in J. Žganec Gros (ur.), *Zbornik 5. slovenske in 1. mednarodne konference Jezikovne tehnologije 2006*. Ljubljana: Institut Jožef Štefan. 62-67
- Mezeg, A. (2011). *Korpusno podprta analiza francoskih polstavkov in njihovih prevedkov v slovenščino*. Doktorska disertacija, Univerza v Ljubljani.
- Mikolič Južnič, T. (2011). Vpliv besedilnih tipov na pojavljanje nominalizacije v slovenščini: korpusna raziskava. V S. Kranjc (ur.), *Meddisciplinarnost v slovenistiki*. Ljubljana: Znanstvena založba Filozofske fakultete. 321-327.
- Mikolič Južnič, T. (v tisku). Neosebne glagolske oblike v prevodni in izvorni slovenščini: primer deležja. V Š. Vintar (ur.), *Slovenski prevodi skozi korpusno prizmo*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Schlamberger Brezar, M. (2005). Politična besedila kot tip besedil in postopki prevajanja stalnih formul. V N. Kocijančič Pokorn, E. Prunch in A. Riccardi (ur.), *Beyond Equivalence – Jenseits der Äquivalenz – Oltre l'equivalenza – Onkraj ekvivalence*. Gradec: Institut für Theoretische und Angewandte Translationswissenschaft. 121-135.
- Sketch Engine. <http://www.sketchengine.co.uk/>, dostop 31.07.2012.
- Toporišič, J. (1991). *Slovenska slovnica*. Maribor: Obzorja.
- Vintar, Š. (2009). Slovenski prevodoslovniki korpus. V M. Stabej (ur.), *Infrastruktura slovenščine in slovenistike*. Ljubljana: Znanstvena založba Filozofske fakultete. 385-391
- Vintar, Š. (ur.) (v tisku). *Slovenski prevodi skozi korpusno prizmo*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Žagar, F. (2001). *Slovenska slovnica in vadnica*. Maribor: Obzorja.