Univerza v Ljubljani
Filozofska fakulteta

Peter Holozan

Računalniško postavljanje vejic v slovenščini

Doktorska disertacija

Mentorica: prof. dr. Andreja Žele
Somentorica: prof. dr. Špela Vintar

Študijski program: Humanistika in družboslovje
Področje: Slovenistika

Ljubljana, 2016

Zahvala

Zahvalil bi se rad vsem, ki so kakor koli pripomogli k izvedbi te naloge, še posebej pa:

- mentorici prof. dr. Andreji Žele in somentorici prof. dr. Špeli Vintar za potrpežljivost, nasvete in ideje;
- podjetju Amebis na čelu z direktorjem Mirom Romihom, mojim delovnim mentorjem, za ves čas, ki sta mi ga dala na voljo, in ker sta mi omogočila v delu uporabo tehnologij, razvitih v podjetju Amebis (Besana, Ases);
- lektorici Aleksandri Kocmut za mnenja pri nekaterih vejicah;
- lektorici Marti Kocjan-Barle za mnenja pri nekaterih vejicah;
- dr. Tomu Korošcu za mnenja pri nekaterih vejicah;
- dr. Tomažu Erjavcu za pomoč pri uporabi repozitorija CLARIN.SI;
- dr. Mojci Kompara za pomoč pri prevajanju izvlečka v angleščino;
- Branku Gradišniku za lektoriranje in mnenja pri nekaterih vejicah in
- svoji Ireni, za vse ure, ki sem jih preživel za računalnikom namesto z njo, in ker me je spodbujala, da sem začeto tudi dokončal.

Doktorski študij je delno sofinancirala Evropska unija, in sicer iz Evropskega socialnega sklada. Sofinanciranje se izvaja v okviru Operativnega programa razvoja človeških virov za obdobje 2007-2013, 1. razvojne prioritete Spodbujanje podjetništva in prilagodljivosti; prednostne usmeritve 1. 3: Štipendijske sheme.

Ireni za potrpežljivost

Izvleček
Računalniško postavljanje vejic v slovenščini

Pričujoče delo raziskuje možnosti uporabe računalnikov za postavljanje in popravljanje vejic v slovenščini. Zbrana je obsežna zbirka primerov rabe vejice, v kateri so označene manjkajoče in odvečne vejice (138.626 pravilno postavljenih vejic, 17.768 manjkajočih in 4608 odvečnih); zbirka je prosto dostopna pod licenco Creative commons. Primeri so uporabljeni za postavljanje vejic s strojnim učenjem, najboljši rezultat je dosežen z metodo ADTree z uporabo lem, oblikoskladenjskih oznak in rezultatov skladenjskega razčlenjevalnika, označevalnik in razčlenjevalnik pa morata biti naučena na učnem korpusu z odstranjenimi vejicami (priklic pri postavljanju vseh vejic 66,10 % pri natančnosti 89,20 %). Za popravljanje programov, ki uporabljajo pravila za postavljanje vejic, sta bili razviti dve metodi: 1) iskanje okolic vejic in 2) določanje vrst vejic, ki navede razloge za postavitev vejice. S tema metodama so izboljšani rezultati programov LanguageTool (priklic manjkajočih vejic s 40,15 % na 44,62 %) in Besana (s 57,45 % na 63,47 %).

Ključne besede: postavljanje vejic, popravljanje napačnih vejic, slovnični pregledovalnik, skladenjski razčlenjevalnik, veznik

Abstract
Automatic Comma Placing in Slovenian

The work explores the possibilities of computer-assisted comma placing and correction in Slovenian. In the extensive collection of instances of comma usage, all missing and superfluous commas are marked (in total 138,626 correctly placed commas, 17,768 missing commas and 4608 superfluous commas). The collection is freely available under the licence Creative commons. The instances are used for machine assisted comma placing, the best result is achieved using ADTree with lemmas, morphosyntactic descriptors and the results of syntax analyser, the tagger and analyser are trained on corpora with removed commas (recall for comma placing is 66.10% at precision 89.20%). For improving the rule-based programs for comma error correction, two methods were developed. The first method searches the words around the commas, and the second determines the type of comma and explains the reason for placing them. Thanks to these methods, the results of LanguageTool (recall of finding missing commas from 40.15% to 44.62%) and Besana program (from 57.45% to 63.47%) were improved.

Keywords: comma placing, comma error correction, syntax checker, dependency parser. conjunction

Kazalo

1 Uvod ... 1

1.1 Vejica v zgodovini ... 2
1.1.1 Ime ločila ... 2
1.1.2 Stičnost ... 2
1.1.3 Pravila v starih slovnicah ... 3
1.2 Pravila za rabo vejice ... 4

2 Priprava zbirke primerov rabe vejice ... 6

2.1 Struktura zbirke ... 6
2.2 Viri primerov ... 7
2.2.1 Korpus Šolar ... 7
2.2.2 Korpus KUST ... 18
2.2.3 Korpus Lektor ... 19
2.2.4 Wikipedija ... 20
2.3 Zbirka primerov napačne in pravilne rabe vejice ... 20

3 Dosedanje delo na področju postavljanja vejic ... 22

3.1 Dosedanji programi za slovenščino ... 22
3.1.1 Mspell ... 22
3.1.2 Besana ... 22
3.1.3 LanguageTool ... 24
3.1.4 Strojno učenje ... 25
3.2 Delo za druge jezike ... 25
3.2.1 Postavljanje vejic z uporabo pravil ... 25
3.2.2 Strojno učenje ... 26

4 Evalvacija dosedanjih rezultatov ... 27

4.1 Ocenjevanje rezultatov ... 27
4.2 Metoda evalvacije ... 27
4.2.1 Besana ... 28
4.2.2 LanguageTool ... 28
4.3 Iskanje napak pri postavljanju vejic ... 30
4.4 Postavljanje vseh vejic ... 35
4.5 Analiza težav pri iskanju napak pri postavljanju vejic ... 37
4.5.1 Okoliške besede ... 37
4.5.2 Program za določanje vrst vejic ... 39

5 Izboljševanje metod ... 49

5.1 Preizkus strojnega učenja ... 49
5.1.1 Zasnova poskusa ... 49
5.1.2 Preizkušanje ... 55
5.1.3 Iskanje napak pri postavljanju vejic ... 65
5.2 LanguageTool ... 66
5.2.1 Metode za ugotavljanje manjkajočih vejic ... 66
5.2.2 Metode ugotavljanja odvečnih vejic ... 71
5.3 Besana ... 72
5.3.1 Uporaba analize povedi ... 73

5.3.2 Uporaba okolice napak ... 85
5.3.3 Iskanje odvečnih vejic ... 92

6 Implementacija metod in rezultati ... 94

6.1 Spremljanje rezultatov strojnega učenja ... 94
6.1.1 Primerjava z drugimi rezultati za slovenščino ... 94
6.1.2 Primerjava z rezultati za druge jezike ... 95
6.1.3 Primerjava rezultatov pri popravljanju vejic ... 96
6.2 Spremljanje rezultatov za LanguageTool po izboljšavah ... 97
6.3 Spremljanje rezultatov Besane po izboljšavah ... 99
6.4 Primerjava rezultatov po izboljšavah ... 101
6.4.1 Iskanje manjkajočih vejic ... 101
6.4.2 Iskanje odvečnih vejic ... 103
6.4.3 Postavljanje vseh vejic ... 104

7 Sklep ... 107

7.1 Možnosti nadaljnjega dela ... 107
7.1.1 Zbirka primerov rabe vejice ... 108
7.1.2 Strojno učenje ... 108
7.1.3 LanguageTool ... 109
7.1.4 Besana ... 109

8 Viri in literatura ... 111

9 Priloge ... 115

9.1 Pravila za rabo vejice v SP 2001 ... 115
9.2 Vejica v starejših slovnicah in pravopisih ... 119
9.2.1 Slovenska slovnica za srednje šole (Josip Šuman, 1884) ... 119
9.2.2 Slovenski pravopis (Fran Levec, 1899) ... 119
9.2.3 Slovenska slovnica za srednje šole (Anton Breznik, 1921) ... 122
9.2.4 Slovenski pravopis (Anton Breznik, Fran Ramovš, 1935) ... 124
9.3 Vejica v srednješolskih učbenikih ... 125
9.3.1 Slovenščina 3: Z besedo do besede ... 125
9.3.2 Slovenščina 4: Z besedo do besede ... 126
9.3.3 Na pragu besedila 3 ... 126
9.3.4 Govorica jezika 2 in 3 ... 127
9.3.5 Besede 1 ... 129
9.3.6 Slovenska jezikovna vadnica ... 131
9.3.7 Slovenski knjižni jezik 1 ... 133
9.3.8 Slovenski jezik ... 134
9.4 Spremembe pravil za LanguageTool ... 137
9.4.1 Popravljena pravila ... 137
9.4.2 Dodana pravila ... 139

10 Kazala ... 150

10.1 Kazalo tabel ... 150
10.2 Kazalo slik ... 151

1 Uvod

Vejica je ločilo, ki kljub jasnim pravopisnim pravilom dela težave velikemu številu ljudi, celo bodočim učiteljem na razredni stopnji (Šek Mertük 2011) (še toliko bolj pa imajo težave potem učenci (Žagar 1987: 139–140)). Po drugi strani pa so že iz zgodovine znani primeri o pomembnosti oz. pomenski odločilnosti pravilno postavljenih vejic (»Streljati ne pomilostiti.«), seveda pa ne manjka tudi modernejših primerov, kjer lahko ena napačno postavljena vejica v pogodbi pomeni več kot dva milijona dolarjev dodatnih stroškov za podjetje (Robertson 2006).

Pravila o rabi vejice so obsežno našteta v Slovenskem pravopisu 2001 (Toporišič 2001: 35–41), dodatno pa so razložena še v številnih jezikovnih priročnikih (npr. (Kocjan-Barle 1992: 74–91), (Cedilnik 1995: 44–51), (Verovnik 2005: 17–26) in (Žagar 1991: 159–160)), diplomskih nalogah (npr. (Žibert 2006)) in seveda učbenikih (npr. (Vogel et al. 2010: 181), (Toporišič 1975: 88–96)). Zavedati pa se je treba, da pravila, kot na primer, da mora biti vejica med nadrednim in odvisnim stavkom, računalniku nič ne pomagajo, če ne ve, kako naj najprej ugotovi, kaj sploh je nadredni ali odvisni stavek v povedi. In določanje zgradbe povedi (Toporišič 1984: 497–524) je za računalnike še zelo težak problem, ki se pa dodatno zaplete, ker računalnik pri analizi pričakuje slovnično pravilen stavek, ki pa ga zaradi manjkajoče vejice ne more biti. Že samo določanje povedka v stavku je za računalnik lahko zelo zapleten problem, saj ne more npr. pri besedi »moram« zlahka odločiti, ali gre za samostalnik ali glagol (med petimi najpogostejšimi besedami v slovenščini sta tako dvoumni »je«, ki je lahko glagol »biti«, glagol »jesti« ali osebni zaimek, in »da«, ki je lahko veznik, členek ali glagol »dati«). Dodatno je pri tem treba paziti, da ne pademo v začarani krog, ko bi za določitev vejice potrebovali pravilno stavčno analizo, za pravilno analizo pa bi potrebovali točno postavljeno vejico.

Računalniški program, ki bi znal dovolj dobro postavljati vejice, ne bi bil uporaben le za avtorje besedil oz. lektorje (kot so uporabni npr. črkovalniki), ampak na primer tudi pri razpoznavi govora (Huang, Zweig 2002), saj iz samega govora velikokrat ni mogoče slišati, kje bi morale biti vejice, in jih je torej treba dodajati programsko. Pravilno postavljene vejice boljšajo tudi označevanje besedil z oblikoskladenjskimi oznakami (Hillard et al. 2006), kar lahko izboljša oblikoskladenjsko označevanje korpusov, v katerih so napačno postavljene vejice.

V nekaterih drugih jezikih obstajajo programi za preverjanje slovnice (slovnični pregledovalniki), ki med drugim iščejo tudi napake pri vejicah. Večinoma gre za komercialne izdelke in o njihovem delovanju ni veliko objavljenega. So pa pravila za vsak jezik drugačna, zato jih je treba za slovenščino napisati posebej.

Za slovenščino že obstajata odprtokodni slovnični pregledovalnik LanguageTool in slovnični pregledovalnik Besana, ki ga pri podjetju Amebis razvijam že od leta 1988. Oba že opozarjata na nekatere tipe manjkajočih in odvečnih (slednje le Besana) vejic. Pričujoča naloga predpostavlja še intenzivnejše delo na tem področju, posledica pa bo izboljšano delovanje slovničnega pregledovalnika pri manjkajočih in odvečnih vejicah. Rezultat dela utegne torej izboljšati tudi uporabo vejice pri vsakodnevni pisni komunikaciji uporabnikov slovničnega pregledovalnika.

Tovrstne izboljšave so danes nujne zaradi potreb vzdrževanja in izboljševanja pismenosti, še posebej, ker se pisci pri pisanju vedno bolj zanašajo na računalniške programe, ki tako postajajo najvplivnejši normativni priročnik (Dobrovoljc in Jakop 2011: 27-28).

1.1 Vejica v zgodovini

1.1.1 Ime ločila

Za vejico so bila uporabljena zelo različna imena, v jezikovnih priročnikih jo je prvi imenoval Marko Pohlin leta 1783, in sicer dolgè potéſaj. Valentin Vodnik je leta 1811 uporabil izraz rés, Anton Martin Slomšek pa je v različnih izdajah učbenika Blaže in Nežica v nedeljski šoli uporabil naslednje izraze: prenehlej, sareſza, vejza (1842); prenehlej, zaresca (vejca) (1848); prenehlej, zaresca, vejca, klinček (1857). Jožef Muršec je leta 1847 uporabil izraza véjica in protec, Fran Malavašič pa 1849 rez. Anton Janežič je v različnih izdajah Slovenske slovnice uporabil naslednja imena: klinčič, protec (1854), klinček (1863) in vejica (1900). Fran Levec je v Slovenskem pravopisu leta 1899 uporabil ime vejica in od takrat se ime ni več spreminjalo. (povzeto po Gojčič (2009))

1.1.2 Stičnost

V slovenskih tiskanih besedilih 16. stoletja so vejico večinoma pisali levostično, včasih pa tudi nestično ali kar stično (Weiss 2002: 254). Sčasoma se je uveljavilo le levostično pisanje.

1.1.3 Pravila v starih slovnicah

V starih slovnicah vejici niso namenjali veliko pozornosti, nekoliko več je o njeni uporabi pisal šele Dajnko. Vendar iz tega ne moremo sklepati, da pravil ni bilo, kajti v starih besedilih vejice so, in so tudi dovolj dosledno postavljene.

1.1.3.1 Zimske urice (Adam Bohorič, 1584)

Bohorič je vejico v svoji slovnici le omenil, ni pa napisal pravil za njihovo rabo (Gojčič 2009: 8).

Seveda pa to ne pomeni, da pravila niso obstajala, dalo bi se jih izpeljati iz slovenskih primerov v njegovi knjigi. Tako npr. primer *Retki ſo dobri, kumaj jih je v'zhiſli tulikajn, kuliko je vrat, v'mejſti Thebe, oli kuli je ſtrug bogatiga Nila.* pokaže pisanje vejice v ločnem priredju (ali pa morda celo vrivek).

1.1.3.2 Krajnska gramatika (Marko Pohlin, 1768 in 1783)

Pohlin je v prvi izdaji zapisal, da »vejica ločuje pomen in razumevanje enega stavka od drugega. *Sé she nise nagledal, ke videsh, da so uſſe zirqve is kositarjam pokrite?*« Dodal je opozorilo, da je pravilna stava tega ločila zelo pomembna: »Če te vejice niso dobro postavljene, delajo zmedo kot tista prerokba: *Bosh shl, naſaj pershl nabosh umerl na vojski*, kjer se ne ve, ali *nabosh* spada k *naſaj pershl* ali k *umerl*.« (po Gojčič (2009: 20))

V drugi izdaji je zapisal drugačno pravilo o rabi: »Vejica deli besede, ki sicer ne sodijo neposredno skupaj, temveč k celotnemu stavku. *Menem, de se ſhe ſadoſti nagledal, ke videsh, de so usi zirqveni ſgvoniki is kositarjam pokryti.*« Opozorilo o pravilni stavi je prepisal in dodal enak zgled. (po (Gojčič 2009: 20))

1.1.3.3 Pismenost ali Gramatika za Perve Shole (Valentin Vodnik, 1811)

1) Rés (,) To prepono dévamo med imena, perloge in glagole, kadar jih vezh sapored pride, de jih narasen réshemo, kakor: mir, pokoj, pasenje, pokorſhina ſo v' ſholi potrebne. Lubesen je krotka, poterpeshliva, dobrotliva. Sedim, glédam, piſhem, kar vuzhenik na tablo poſtavla.

Rés rudi raspénja isréke, poſtavim: vſak vé, de tize po nebu letajo, ribe pa po vodi plavajo.

1.1.3.4 Lehrbuch der Windischen Sprache (Peter Dajnko, 1824)

Dajnko je postavil štiri pravila o rabi vejice (po (Gojčič 2009: 20–21)):

a) Pred vsemi oziralnimi zaimki: *Kniga, kero sem bral, je lepa. Vse, kaj se sveti, je ne zlato. Ti, ki živiš, se veseli.*

b) Pred in za vrinjenimi besedami in stavki. *To, priatel, maš mojo roko, ino, kaj ti rečem, mi veri.*

c) Med več zaporednimi besedami ene vrste, ko niso povezane niti z ino, niti z ali: *Sunce, mesec, zvezde, nebo, zemlo, vse je stvoril Bog; njemi bodmo hvalni, podložni, pokorni; njega častmo, slavmo, molmo, lybmo!*

Posamezne besede, povezane z ino, ali, ne potrebujejo vejice: *Pobožno ino pravično življenje se dopadne Bogi ino vsakemu človeki.*

Ko pa je cel stavek, ali več členov stavka povezano z vezniki ino, ali, vejice vmes ne smemo izpustiti: *Božji sin sedi na desnici svojega očeta, ino nas vyči, da bode drygoč prišel, na sodni den, nas plačat, ali kaštigat, potem, kak živimo.*

d) V vseh dvodelnih stavkih, katerih členi so kratki, t. j. niso znatno dolgi: *Tega ne rečem rad, alipa resnice zatajiti ne smem.*

1.1.3.5 Blaže in Nežica v nedeljski šoli (Anton Martin Slomšek, 1842, 1848 in 1857)

O vejici ni povedano veliko, o postavljanju slovenskih vejic pravi le:

V pismi se za vsakim povedkom prenehlej (,) naredi /.../.

1.2 Pravila za rabo vejice

Pravila za rabo vejice so obširno našteta v SP 2001 (v celoti so navedena v prilogi 9.1 na strani 115).

Zanimiva dopolnitev teh pravil je predlagana v (Korošec 2003), kjer sta uporabljena koncepta desnosmerne in levosmerne vejice, pri čemer so desnosmerne vejice pred odvisniki, polstavčnimi prilastki, pristavki, vrinjenimi stavki ipd., levosmerne vejice pa so za zadnjo besedo stavčnih prilastkov, polstavčnih prilastkov, odvisnikov in delov priredij ter dostavki oz. za zadnjo besedo vrinjenih stavkov.

Pravila za postavljanje vejic so tudi del (srednješolskih) učbenikov, primeri so navedeni v prilogah od 9.3.1 do 9.3.8. Zanimivo je, da je opis pravil v novejših učbenikih

večinoma krajši od opisov v starejših učbenikih (tudi zato, ker novejši učbeniki bolj spodbujajo samostojno iskanje pravil iz zgledov namesto suhoparnega naštevanja pravil).

Pravila za vejico so redno tudi v različnih jezikovnih priročnikih, ki poskušajo pravila razložiti na bolj poljuden način.

2 Priprava zbirke primerov rabe vejice

Zbirka primerov rabe (in sicer tako pravilne kot napačne) vejice je zelo pomembna, ker nam omogoča, da lahko računalniško (brez ročnega dela) preverjamo, kako uspešne so metode za postavljanje vejic. Dodatno se da ta zbirka primerov uporabiti tudi kot vzorec za učenje postavljanja vejic s statističnimi metodami (če popravimo vse označene napake).

Za slovenščino taka zbirka še ni izgotovljena, delno se da v ta namen uporabiti korpus Šolar (podrobneje opisan v točki 2.2.1), ki sicer ima označene učiteljske popravke, vendar se je pri preverjanju pokazalo, da marsikatera napaka ni označena (bodisi ker jo je učitelj spregledal bodisi ker je bilo pred tem že toliko napak, da je obupal nad nadaljnjim popravljanjem), v nekaterih primerih pa so bili popravki vejic tudi neustrezni za naš namen, ker je učitelj popravil okoliško besedilo in zdaj vejica ni na takem mestu, kot je bila v originalnem besedilu.

Drugi problem je, kako dobiti primere besedil, ki bi imeli res natančno postavljene vejice. V korpusu Fida so nekatera besedila sicer imela oznako o lektoriranosti, vendar ta oznaka ni bila popolnoma zanesljiva, po drugi strani pa tudi lektorji lahko kaj spregledajo, še posebej v naglici, ki jo zahteva lektoriranje časopisnih člankov. Naslednja težava pri tem je še konsistentnost, kajti nekatera pravila za postavljanje vejic dopuščajo, da vejice so ali pa jih ni, predvsem za statistično učenje pa je ugodno, da je postavljanje vejic čim bolj konsistentno.

Zato je v okviru te disertacije pripravljena nova zbirka primerov, ki bo na voljo pod licenco Creative Commons (CC, Ustvarjalna gmajna), kar omogoča, da bodo lahko zbirko uporabljali tudi drugi, ob čemer se bodo dali tudi primerjati rezultati različnih metod računalniškega postavljanja vejic.

2.1 Struktura zbirke

Zbirka je izvedena kot seznam vrstic s tremi stolpci.

V prvem stolpcu je oznaka dela (ki ločuje podatke glede na vir).

V drugem stolpcu je številčni podatek o poddelu. Primeri iz korpusa Kust (točka 2.2.2) so razdeljeni glede na prvi jezik učenca (nemščina, angleščina, španščina, italijanščina, srbščina/hrvaščina/bosanščina), primeri iz korpusa Šolar (točka 2.2.1) pa glede na razred oz. letnik in vrsto šole (6. do 9. razred osnovne šole, 1. do 3. in 5. letnik poklicne šole, 1. do 4. letni srednje strokovne šole, 1. do 4. letnik gimnazije ter maturitetni tečaj). Možno bi bilo tudi združiti podatke iz korpusa Šolar v le štiri kategorije (osnovna šola, poklicna šola, strokovna

šola in gimnazija), vendar me je zanimalo tudi, koliko se rezultati znotraj teh skupin ujemajo po letnikih, ne pa le povprečki. Slabost te odločitve so manj pregledne tabele (seveda pa je možno podatke potem po potrebi združevati).

V tretjem stolpcu so primeri stavkov. Manjkajoče vejice so označene z znakom ¤, odvečne pa nadomeščene z znakom ÷.

Baza je izvedena kot preglednica v programu Excel, da pa jo programi potem laže uporabljajo, se naredi izvoz v besedilno obliko, v kateri so vrstice ločene s kodo za novo vrstico, stolpci pa s kodo za tabulator.

2.2 Viri primerov

Izbrani so le taki viri, ki so dostopni pod licenco CC, da se potem tudi zbirka primerov napačne (in pravilne) rabe vejice objavi pod isto licenco. Taka prostodostopna zbirka omogoča, da bodo lahko drugi raziskovalci preizkušali svoje metode postavljanja vejic na istih primerih, s čimer bo rezultate mogoče laže primerjati.

2.2.1 Korpus Šolar

Korpus šolskih pisnih izdelkov Šolar je korpus besedil, ki so jih učenci slovenskih osnovnih in srednjih šol samostojno tvorili pri pouku. Zajeta so besedila, kjer je slovenščina materni jezik avtorjev, ki niso bila napisana posebej za projekt, ampak so del šolske produkcije, jezikovni popravki pa so realni, torej, takšni, kakršne so naredili učitelji (Rozman et al. 2010).

V korpusu so besedila učencev od 6. do 9. razreda osnovnih šol in od 1. do 5. letnika srednjih šol ter maturitetnega tečaja.

V korpus so vključeni tudi popravki napak, in sicer tako popravki, ki so jih naredili učitelji, kot popravki, ki so jih vnesli sestavljavci korpusa (vendar je bilo to narejeno le na delu korpusa).

V prvem poskusu, ki je bil uporabljen v (Holozan 2013), so bile v zbirki primerov le povedi, v katerih je bila vsaj ena napaka pri rabi vejic (bodisi manjkajoče bodisi odvečne vejice). Vendar se je pokazalo, da je za izračun natančnosti popravljanja vejic potreben vzorec, v katerem so ob primerih napačne rabe tudi pravilni zgledi. To je potrebno zato, ker se lahko program za postavljanje vejic zmoti in dodaja odvečne vejice tudi v povedih, kjer prej ni bilo napak. Vendar je očitno, da izbira besedila zelo vpliva na rezultat – dobro lektorirano

besedilo brez manjkajočih vejic bo imelo slabšo natančnost kot besedilo, v katerem manjka veliko vejic.

Ker v Šolarju niso v celoti označeni popravki, so izbrana tista besedila, v katerih je bil vsaj en učiteljski popravek. Pokazalo pa se je, da je kljub temu veliko napak neoznačenih: v besedilih z veliko napakami se je očitno dogajalo, da so učitelji še pred koncem besedila obupali in sploh niso več popravljali napak. Dodatna težava so primeri, kjer so učitelji, namesto da bi popravili vejico, v popravku raje spremenili poved, da vejica ni bila več potrebna. Nemalo primerov pa je bilo tudi takih, ko so učitelji kakšno napako kratko malo spregledali ali pa so jo celo dodali.

2.2.1.1 Ročno preverjanje primerov

Zato je bilo treba primere iz korpusa Šolar ročno pregledati in dopolniti popravke vejic. Gre pa za skoraj 50.000 povedi oz. 3.222.000 znakov brez presledkov oz. 2148 lektorskih strani. Možnost je bila, da bi to naredili zunanji popravljavci, pri čemer bi lahko kak kos pregledala dva popravljavca, s čimer bi se dalo potem določiti, koliko sta si skladna. Vseeno sem se odločil, da bom primere pregledal sam in bo s tem označevanje poenoteno (ne nazadnje pa tako delo poglobi tudi poznavanje pravil o postavljanju vejic), pri nekaterih primerih sem se o postavitvi vejice posvetoval tudi z mentorico in drugimi, navedenimi v zahvali.

2.2.1.1.1 Preizkušanje zanesljivosti označevanja

Za večjo zanesljivost bi bilo sicer dobro ta postopek ponoviti vsaj dvakrat in potem primerjati rezultate, vendar bi to zahtevalo veliko preveč dela, tako da sem se zadovoljil s tem, da sem ponovno pregledal le manjši vzorec povedi (500 povedi oz. približno 1 % primerov).

Vsi popravki so zbrisani (tudi učiteljski) in na novo ročno dodani, razlik je bilo 23 (4,6 %), in sicer naslednje (v prvi vrstici je rezultat prvega označevanja, v drugi pa drugega (preizkusnega)). Boljše[1] različice so odebeljene, manjkajoče vejice so označene z znakom ¤, odvečne pa z znakom ÷:

Dostikrat pa se je tudi spreminjal na podlagi doživetega, kot npr. pri čarovnici Kirki.
Dostikrat pa se je tudi spreminjal na podlagi doživetega÷ kot npr. pri čarovnici Kirki.[2]

[1] Po moji presoji oz. v nekaterih primerih sem za mnenje vprašal tudi druge in se potem glede na to odločil.
[2] Čeprav bi v tem primeru vseeno lahko rekli, da je bila prvotna izbira boljša, ker »kot« tukaj stoji namesto »tako«. Ta primer nam kaže, da se je v nekaterih primerih pravzaprav težko odločiti, kako je z vejico.

Ampak Odisej se je znašel tako, da je svojim mornarjem v ušesa vlil vosek¤ sebe pa privezal na jambor¤ tako÷ da je lahko slišal petje siren brez strahu, da bi pobegnil k njim, ostali mornarji z zadelanimi ušesi pa so mirno odpluli mimo.
Ampak Odisej se je znašel tako, da je svojim mornarjem v ušesa vlil vosek¤ sebe pa privezal na jambor tako, da je lahko slišal petje siren brez strahu, da bi pobegnil k njim, ostali mornarji z zadelanimi ušesi pa so mirno odpluli mimo.

Dandanes na človekovo srečo vplivajo razni materialni viri¤ kot so denar, hiša÷ … in družba, npr. družina¤ ljubezen, prijatelji.
Dandanes na človekovo srečo vplivajo razni materialni viri¤ kot so denar, hiša÷ ... in družba, npr. družina ljubezen, prijatelji.

To knjigo bi zaključil drugače: William Shekspere Romeo in Julija
To knjigo bi zaključil drugače: William Shekspere¤ Romeo in Julija

Župnik izdelal načrt, da spije uspavalo in da bojo vsi mislili¤ da je umrla, Romeu, pa bi med tem povedali za načrt.
Župnik izdelal načrt, da spije uspavalo in da bojo vsi mislili¤ da je umrla, Romeu÷ pa bi med tem povedali za načrt.

Saj menim, da bi zgodba lahko imela srečen konec in tu je moja različica.
Saj menim, da bi zgodba lahko imela srečen konec¤ in tu je moja različica.

In res obe družini sta se odzvali besedilu.
In res¤ obe družini sta se odzvali besedilu.

Dogodek¤ za katerega sem se opredelil¤ in katerega bi zaključil drugače¤ je ta, ko sta Romeo in Julija umrla zaradi sporov med družinama, ki nista bila povezana z njimi.
Dogodek¤ za katerega sem se opredelil in katerega bi zaključil drugače¤ je ta, ko sta Romeo in Julija umrla zaradi sporov med družinama, ki nista bila povezana z njimi.

To večer, ko je Julija hotela spit napoj, se je Romeo počasi vtihotapil v hišo prek njenega balkona¤ ji preprečil¤ da to naredi in sta počasi odšla iz hiše, brez da bi jih kdo opazil.
To večer, ko je Julija hotela spit napoj, se je Romeo počasi vtihotapil v hišo prek njenega balkona ji preprečil¤ da to naredi¤ in sta počasi odšla iz hiše, brez da bi jih kdo opazil.

Namesto skrite poroke¤ bi priredil zabavo ter na njo povabil starše obeh družin.
Namesto skrite poroke bi priredil zabavo ter na njo povabil starše obeh družin.

Spoznali bi, da ljubezen prepreči vse nesporazume in postali bi velika družina.
Spoznali bi, da ljubezen prepreči vse nesporazume¤ in postali bi velika družina.

Za tak konec zgodbe, sem se odločil zato, ker nimam rad, da bi se stvari skrivale.
Za tak konec zgodbe÷ sem se odločil zato, ker nimam rad, da bi se stvari skrivale.

Zgodo Romeo in Julija je napisal eden največjih dramatikov William Shakespeare, knjiga je tragedija, ker ima tragičen konec, nastala pa je v letu 1595¤ v obdobju renesanse.
Zgodo Romeo in Julija je napisal eden največjih dramatikov William Shakespeare, knjiga je tragedija, ker ima tragičen konec, nastala pa je v letu 1595 v obdobju renesanse.

A na poti do ljubezni sta srečala na oviro, njeni družini sta se sovražili in to tako dolgo¤ da so že pozabili¤ zakaj.
A na poti do ljubezni sta srečala na oviro, njeni družini sta se sovražili¤ in to tako dolgo¤ da so že pozabili¤ zakaj.

Namesto da nakonci oba umreta¤ bi se mogla Julija prej zbuditi in da bi imela pravo in veliko poroko¤ bi se družini pobotali tudi brez njune smrti.
Namesto da nakonci oba umreta¤ bi se mogla Julija prej zbuditi¤ in da bi imela pravo in veliko poroko¤ bi se družini pobotali tudi brez njune smrti.

Takoj, ko je videl Julijo¤ se je zaljubil v njo.
Takoj÷ ko je videl Julijo¤ se je zaljubil v njo.

Ko je videl Julijo mrtvo, je spil strup, ki ga je kupil na poti¤ in umrl.
Ko je videl Julijo mrtvo, je spil strup, ki ga je kupil na poti in umrl.

Po poroki Julija prosi Lorenza za uspavalni napoj, ki ga spije¤ in pade v trd neprebudljiv spanec, Lorenzo pa Romeu pošlje pismo, v katerem piše¤ da Julija ni mrtva.
Po poroki Julija prosi Lorenza za uspavalni napoj, ki ga spije¤ in pade v trd neprebudljiv spanec, Lorenzo pa Romeu pošlje pismo, v katerem piše da Julija ni mrtva.

Družini bi morali prej dojeti, da¤ kar delata¤ prinaša samo zlo.
Družini bi morali prej dojeti, da kar delata¤ prinaša samo zlo.

Capuletovi pridejo¤ da bi se še poslovili od Julije, ampak so našli tudi mrtvega Romea.
Capuletovi pridejo da bi se še poslovili od Julije, ampak so našli tudi mrtvega Romea.

moja tragedija, ni praprav ni tragedija¤ je veselje in sreča.
moja tragedija÷ ni¤ praprav ni tragedija¤ je veselje in sreča.

Domišljal si je, da bi tuti on rad pomagal ljudem ter se boril proti zljikovcem in tudi to je naredil.
Domišljal si je, da bi tuti on rad pomagal ljudem ter se boril proti zljikovcem¤ in tudi to je naredil.

Po letu dni bivanja je Juliji bilo vedno bolj žal, da je odšla in se je spraševala, če jo Romeo še sploh ljubi, saj ga redko vidi in še sama ne ve¤ če še isto čuti do njega.
Po letu dni bivanja je Juliji bilo vedno bolj žal, da je odšla¤ in se je spraševala, če jo Romeo še sploh ljubi, saj ga redko vidi in še sama ne ve¤ če še isto čuti do njega.

Preverjanje je pokazalo, da je 7 primerov bilo boljših v prvem označevanju, 16 pa v drugem. Glede na ta rezultat je torej po prvem označevanju bilo okoli 3 % napak, kar pri 50.000 primerih pomeni okoli 1500 napak, ki so verjetno ostale tudi po ročnem preverjanju primerov iz korpusa Šolar.

Dve od napak sta taki, kjer se je način postavljanja vejic določil šele med označevanjem (2.2.1.1.2.6 in »takoj ko«), tukaj so bili podobni primeri poiskani in popravljeni z urejevalnikom besedil. Bolj zapletene težave, ki najbrž tudi niso bile

konsistentno popravljane, pa je mogoče najti, ko je postavljanje vejic dovolj zanesljivo oz. pri preverjanju rezultatov postavljanja vejic. Zato je bilo smiselno nadaljevati delo tudi z nepopolno zbirko primerov, potem pa na koncu narediti novo verzijo zbirke primerov s popravljenimi napakami, ki so se med delom našle, in po potrebi, če je napak toliko, da bi lahko vplivale na rezultate, ponoviti preizkušanje metod.

2.2.1.1.2 Dileme, ki so se pokazale pri preverjanju

Ob ročnem preverjanju se je pokazalo nekaj zanimivih primerov, pri katerih se je bilo treba odločiti, kako je najboljše postaviti vejice, pri nekaterih pa bi bilo treba tudi premisliti o morebitni dopolnitvi pravil postavljanja vejice, vendar to že preseda temo pričujočega dela..

Mogoče je postaviti celo hipotezo, da je v korpusu Šolar veliko zanimivih primerov za postavljanje vejic ravno zaradi tega, ker pisci slabo poznajo pravila. Običajno se pisci izogibamo takim stavčnim konstrukcijam, pri katerih nismo prepričani, kako postavimo vejice – stavek pač predelamo, da se ognemo težavam. In tudi pri marsikaterem od primerov iz Šolarja je tako, da bi ga sicer raje spremenili tako, da vejica ne bi bila več problematična[3].

2.2.1.1.2.1 »tako, da« proti »tako da«

V SP 2001 je »tako da« naveden kot podredna vezniška zveza, vendar pravilo tudi pravi, da v nekaterih primerih to zvezo lahko razdelimo. Tu se seveda zastavlja vprašanje, kdaj pravzaprav se to zgodi. Primeri kažejo, da do tega pride, kadar se glagol možno veže s prislovnim določilom načina:

Bog je torej Kajna kaznoval÷ tako¤ da je blodil po Zemlji vse dni svojega življenja.
Negativno pa lahko vpliva tako¤ da se otroci iz novejših filmov naučijo kletvic, vidijo veliko nasilja, lahko postanejo zasvojeni in zaradi tega imajo manj socialnega življenja, ker cel dan sedijo za televizijo in se ne družijo s prijatelji.
Jaz bi zgodbo zaključil tako¤ da bi družini že ob spoznanju Romea in Julije pripravili vse za njuno poroko ter vidno prej zgladili vse spore.
Moje mnenje je tako¤ da se bom poročil in imel dva do tri otroke¤ in če bomo imeli možnost živeti v podeželju¤ bomo z veseljem.
Frank ni bil sam kriv za svojo usodo, saj vojna je skušala razdvojiti brata tako¤ da sta se v vojni bojevala eden proti drugemu.
Če bi jaz bil Medeja¤ bi Jazonu pomagal tako¤ da oče o tem nebi nič vedel ali sumil.

Primeri, kjer »tako da« obravnavamo kot vezniško zvezo, pa so naslednji:

[3] Za računalniško postavljanje vejic je problematično tudi veliko število zatipkanih besed, pri katerih računalnik težko ugane, kaj je pravzaprav bilo mišljeno. Pri praktični uporabi ima računalnik lažje delo, ker pisec s pomočjo črkovalnika naprej popravi tipkarske napake, po čemer lahko slovnični pregledovalnik bolje popravi vejice.

A seveda Matiček ni neumen in zna razmišljat s svojo glavo, tako da se med sodno razpravo skoraj ubrani.
Mami in Clairi je prevajal razne pogovore, tako da sta se tudi oni dve lahko pogovarjali s slovenci.
Menil je¤ da vsak človek enkrat umre, tako da je vseeno¤ kdaj in kako se to zgodi.
Simon se zato ni uspel dolgo upirati tem lutkam sistema, tako da je kmalu sprejel novo identiteto in se imel za Poljaka.
Nekateri učenci znajo to profesorjevo usmiljenost tudi s pridom izkoristiti, tako da jih lepo pogledajo in že se jih profesor usmili.
Polikarpa so neprestano tlačile nočne more, tako da se je moral preseliti v klet, kjer ga ponoči ni nihče slišal ječati.

2.2.1.1.2.2 »prosim« kot členek

Besedo »prosim« bi bilo mogoče obravnavati tudi kot členek oz. v členkovni vlogi in potem ne pisati vejic (podobno kot pri *recimo* in *denimo*). Vendar Pravopis trenutno tega še ne predvideva, zato so v primerih vejice označene.

Toda preden to narediš, se¤ prosim¤ posvetuj z njimi.

2.2.1.1.2.3 »takoj, ko« proti »takoj ko«

Spet je »takoj ko« vezniška zveza, ampak v nekaterih primerih, ko je »takoj« bolj poudarjen, je vseeno smiselno to pisati z vmesno vejico (pri odločanju si lahko pomagamo tako, da besedo »takoj« poskusimo nadomestiti z besedo »nemudoma«, če je to mogoče, je vejica za »takoj«).

Antigona pri svojih odločitvah ni popuščala, jaz pa tudi ne, zato sem velel, naj jo obesijo, takoj ko bo priložnost.
Takoj ko sva z očetom prišla domov, me je prepisal na drugo šolo.
Takoj÷ ko sta mi dovolila¤ sem si oblekla prečudovito, čisto novo rumeno obleko in čevlje.
Micka je bila videti sproščena in vesela, a takoj ko je prišel njen sin Marko v kuhinjo¤ se je začela jeziti.
Takoj÷ ko sem se bolje počutila, sem se vrnila domov.
Takoj÷ ko je odprla oči¤ je Romeo skočil k njej ter jo poljubil.
Takoj÷ ko vstopis¤ vstaneta Sonja in Jaka ter mi začneta ploskata.

Da ne bi trpel, bi jo lahko tudi ubil takoj, ko jo je zagledal.
Družbene razmere pa nanjo vplivajo že takoj, ko postane spremljevalka nemškega vojaka.
Ob branju Shakespearjevih del sem ugotovil, da se je s težavami potrebno spopasti takoj, ko jih ugotovimo, kajti÷ potem se bodo razvili le še večji problemi.
Drama je tragedija in govori o nesrečni ljubezni med Romeom in Julijo, ki se je začela takoj, ko je Romeo zagledal Julijo na plesu.
Zmenila sta se, da bosta začela plesati takoj, ko bosta stopila na trg.

2.2.1.1.2.4 »potem, ko« proti »potem ko«

Spet je »potem ko« vezniška zveza, ki pa jo lahko pišemo z vmesno vejico, kadar potem zahteva, da poudarimo »potem«.

Potem÷ ko se je Hamlet pogovoril z duhom, se je najprej obotavljal in želel priti resnici do dna.
Potem÷ ko se je uprl gospodarju, ga ne omenja več kaj dosti niti ne nakaže, da je ravnal prav.
Potem ko opravim vse šolske obveznosti, se odpravim do prijatelja.
Potem ko so slišal, da sta srečna, da imata otroka¤ sta bila zelo presenečana in bili veseli.
Potem ko je videl oz. zagledal Jeklenolaso¤ se mu je povrnila tista moč, odročnost do življenja.
Potem ko smo jih spustili¤ pa je ena izmed rib padla v stričev čoln.
Potem ko so že odšli, pa se je od nikjer prikazal Piksna in naju kar spustil, da pobegneva.
Tudi potem÷ ko se krsti¤ to ne stori zaradi verovanja, vendar zaradi ljubezni do Bogomile.
Bogomila je pa ženska, ki je močno predana bogu, saj se odločiti pokristjaniti¤ potem÷ ko izve¤ da se je Črtomir vrnil živ iz boja.

Se pa tukaj zastavi vprašanje, ali bi bilo vsaj v nekaterih zgornjih primerih potem »potem« bolje nadomestiti s »po tem« (oziroma še bolje »po tistem«).

Zato Polikarp takrat, takoj potem, ko je zagrešil umor, ne pokaže niti malo obžalovanja.
Pozneje¤ ko mu je Sanče Pansa pomagal in mu pravil, da to ni bilo pametno¤ je Don Kihot hitro našel izgovor, da jih je bilo preveč in da¤ če bi bil samo eden¤ ga bi lahko premagal.
Njen Laert misli, da je kriv Hamlet, zato se maščuje z dvobojem¤ v katerem Hamlet zmaga, vendar umre potem, ko razkrinka Klavdija ljudstvu in ga umori.
Kmalu potem, ko se je Julija zbudila¤ jo je Romeo prijel za roko in na konju sta odjezdila v neznano.
Ta strah je občutil še potem¤ ko je bil doma in mu po glavi so mu rojile podobe, ki jih je videl.

2.2.1.1.2.5 »med tem, ko« proti »medtem ko«

Če je pisec napisal »med tem ko« namesto »medtem ko«, je potrebna vejica.

Na plesu sedim za mizo med tem¤ ko se ostali zabavajo.
Ker očeta ni bilo doma, da bi med tem¤ ko smo mi popravljali motor¤ hodil v kontrolo in nam kaj svetoval¤ smo delali ”prosto po Prešernu.”

2.2.1.1.2.6 Oziralni zaimki za vezniki

Pravilo v SP 2001 sicer omenja le, da med zaporednima podrednima vezniškima besedama ni vejice (§ 327), zato se postavi vprašanje, kaj narediti v primerih, ko oziralni zaimek sledi vezniku. Zdi se nam smiselno, da oziralne zaimke obravnavamo enako kot veznike oz. so lahko vezniške besede in torej vmes ne postavljamo vejic.

Družini bi morali prej dojeti, da kar delata¤ prinaša samo zlo.
Tako tudi mi÷ želimo ”prepovedan sadež”, kajti kar je prepovedano¤ je še bolj mikavno.

2.2.1.1.2.7 Naštevanje prislovov in prislovnih odvisnikov

Ti primeri so bili umetno narejeni, se je pa podoben primer našel tudi v zbirki primerov.

Jutri ali ko bom take volje, bom prišel.
Ko bom take volje ali jutri bom prišel.
Prišel bom jutri ali ko bom take volje.
Prišel bom, ko bom take volje ali jutri.

2.2.1.1.2.8 Trije zaporedni vezniki

V besedilu so se pojavljali primeri, kjer so bili trije zaporedni vezniki, in sicer vedno tako, da sta bila prva »in da«, temu pa je sledil še tretji veznik. Po razmisleku so bili primeri označeni tako, da je vejice le pred začetnim »in«, in sicer zato, ker je možno v vseh primerih »da« na drugem mestu izpustiti, s čimer bi potem dobili kombinacijo dveh veznikov in vejico pred prvim, ni pa smiselno, da izpust vpliva na postavitev vejic.

Takrat sem spoznal, da niso vsi taki¤ kot se zdijo¤ in da četudi se na prvi pogled zdijo prijatelji, se lahko v njem skriva nekdo, ki me prezira.
Bogomila mu je razložila, da se je pokristjanila zaradi njega, ker krščanska vera uči o življenju po smrti¤ in÷ da če se nebi več videla¤ bi bila le v posmrtnem življenju lahko skupaj.
Ljudje pravijo, da v peklu živijo hudiči¤ in da če prideš v pekel¤ te tam ne čaka srečno življenje.
Pozneje¤ ko mu je Sanče Pansa pomagal in mu pravil, da to ni bilo pametno¤ je Don Kihot hitro našel izgovor, da jih je bilo preveč¤ in da če bi bil samo eden¤ ga bi lahko premagal.
Njun odnos je nečloveški, Volodja se grdo obnaša do njega, na nek način mu hoče dokazati, da je on glavni¤ in da če ga ne bo ubogal¤ se ne bo izteklo dobro.

2.2.1.1.2.9 Tipkarske napake, ki spremenijo pomen

Pri teh primerih je bila dilema, ali postaviti vejice tako, kot je bilo besedilo verjetno mišljeno, ali tako, kot je v resnici napisano. Ker bi tako popravljanje besedila preseglo trenutne zmožnosti računalnika, je bila odločitev, da se besedilo upošteva tako, kot je napisano.

Opravičil se je učiteljici in odšel¤ da svoje klopi.
Sestra mi jih ni hotela, zato sem jih samin¤ uzel sem kladivo in klešče.
Mati je sklenila roke ubogemu¤ Tončka pa je čevljar prijel za glavo.
Pesem kot celota pa sporoča¤ da je človek¤ postane močnejši le skozi bolečino.
Stekla — oziroma z rolerji÷ sem prirolala¤ da naše hiše ob obali in z rolerji na nogah stekla po mamo.
VSAK VIKEND MESECU MAJU VSI MLADI¤ MOŽNOST TABORJENJA V PARKU.
Vsak vojak naj si vzame svoje orožje¤ bori¤ bo svojih močeh.
Zvečer v televizijski oddaji so poročevali¤ a oblaku smoga¤ ki je zajel pokrajino.
Moje mnenje o družini je, da klub temu÷ da so ločeni drug od drugega, da ostanejo skupaj in so v stiku.
Baron je primer zvestega moža, saj kljub temu, da je bil poročen, dvoril Nežki.
tam so se imeli lepo, otroci so hodili v šolo, sin je delal, mama (nevesta÷ pa je bla doma pospravljalja¤ kuhala….
Vse je je na prodaj in mnogo ljudi misli, da se tudi ljubezen÷ da kupiti.
Vem¤ da je bil to moj naj srečnejši danko¤ sem te spoznala in ko sva postala par¤ v mojem življenju si zdaj ti in ti boš vstal tudi v mojem srcu¤ saj te ljubim zelo in sem srečna s teboj.
Ljubezen na prvi pogled÷ kod nekateri veste¤ naj ne bi obstajala dolgo, saj je še od takrat naprej veliko življenja pred teboj.
Za njuno ločitev je bil povod to, da se nista znala organizirati skupno življenje, da klub temu÷ da sta bila oba zaposlena.
Jaz tega mami še nisem naredila, ker se mi zdi¤ zdelo nesramno narediti kaj takega.

2.2.1.1.2.10 »ne pa da«

Odločitev je bila, da pri zvezi »ne pa da« ni smiselno postavljati vejice, ampak je to boljše obravnavati kot vezniško zvezo (tudi SP 2001 vsebuje primer »*Delat bi šel, ne pa da postopaš*«).

Če bi bil v današnjem času¤ ga bi z dokazi poslal v zapahe, ne pa÷ da bi ga umoril.
In od takrat naprej tudi sam težave rešujem takoj, ne pa÷ da me spremljajo v življenju.
Ne pa÷ da se otroku iz bogate družine že ob rojstvu odpre pot do razkošja; revnemu pa pot še do večje revščine.
Pričakoval bi, da bi on sam poskušal kaj narediti, ne pa÷ da pokoro dobesedno preloži na Izidorja.
Ne pa÷ da se kregata in nimata skupnih ali podobnih razmišljan.
Ta stavek si predstavljam, kot da se Antigona zaveda¤ da smo bili na svet ustvarjeni zato, da ljubimo, ne pa da sovražimo.

2.2.1.1.2.11 »kljub temu da« proti »kljub temu, da«

Zvezo »kljub temu da« tipično obravnavamo kot vezniško zvezo s pomenom »čeprav«, zato vejico pišemo spredaj. Vendar je pri drugem primeru morda vseeno smiselno pisati vejico pred »da«, ker je »temu« soodnosnica odvisniku oz. odvisnikom, ki se nadaljujejo z »da«.

Besedilo tudi nekako Ani, saj je bil zelo prizadet, ker se je odločila dati otroka v rejo¤ kljub temu÷ da se oni ni strinjal.
Matera govori svojemu otroku, da kljub temu¤ da je nezakonska mama, da ga ima rada÷ ter da mu bo pomagala vse življenje.

2.2.1.1.2.12 Dvoumne povedi

Pri teh primerih je vejica odvisna od pomena, ki ga želimo izraziti. Označeni so bili tako, kot je najverjetnejši pomen, je pa malo verjetno, da bi jih računalniku uspelo pravilno postavljati.

Osebno poznam tak primer¤ ko se je moški ubil, ker ga je ženska zapustila in ostal je sam.
Sklenem, da se ne udeležim nobenega plesa več, in živim življenje, kot je bilo prej.
Črtomir se odloči, da se preda in zbeži.
Najbolj mi je bila všeč črtica Mačka v bloku, ker se mi je mačka smilila¤ in iz te črtice si se lahko naučil nauk.
Mislil je, da so v zavodu vsi zmešani¤ skupaj z osebjem.
Ker imam naslednji dan kakšen predmet¤ kjer bom vprašana ali pišemo, si grem pogledat, naučit in ponovit snov.
Tak prizor÷ ni bilo lahko gledati.

2.2.1.1.2.13 »tudi če«

Zvezo »tudi če« je smiselno obravnavati kot vezniško zvezo s pomenom »četudi« in zato ne pisati vejice vmes.

Dobro dejanje je bilo, da ni naredil samomor, ampak je podredil življenje ljubezni¤ in tudi÷ če se ni mogel poročiti z Bogomilo, dal krstiti in čakal¤ da se združita v posmrtnem življenju.

Vem, da tudi če bi bila jaz Črtomir¤ bi oklevala pri pokristjanjevanju sebe, čeprav bi zelo ljubila Bogomilo.
Dobro dejanje je bilo, da ni naredil samomor, ampak je podredil življenje ljubezni¤ in tudi÷ če se ni mogel poročiti z Bogomilo, dal krstiti in čakal¤ da se združita v posmrtnem življenju.

2.2.1.1.2.14 Prilastek ali povedkovo določilo

V tem primeru je vprašanje, ali zadnji del obravnavati kot desni prilastek (kar bi zahtevalo vejico) ali kot povedkovo določilo, pri čemer vejice ni. Prevladala je druga možnost.

Tatjana je punca÷ približno istih let kot Bubi.

2.2.1.1.2.15 Členek pred odvisnikom

Kadar je členek pred odvisnikom na začetku povedi, vmes tipično ni vejice, kadar pa se to zgodi sredi povedi, pride vejica vmes. Nekateri primeri pa kažejo, da morda to pravilo ni tako absolutno.

Takrat se je Antigona že obesila¤ ravno÷ ko je dokončala svoj cilj (pokop Polinejka).

Najhuje je, da se iz napak÷ šele¤ ko se zgodijo, kaj naučimo.

2.2.1.1.2.16 Deležje z odvisnikom

Pravilo v Slovenskem pravopisu pravi, da vejice ni pri golem deležju (§ 334), sicer pa je (§ 333). V teh primerih je po eni strani deležje sicer res golo, po drugi strani pa se nanj navezuje odvisnik.

Pogumen in prebrisan Odisej mu je daval vrče vina÷ misleč, da jih bo Polifem v zameno izpustil.

Simon se je nenadoma znašel v zavodu, v katerem je pristal nevedoč¤ zakaj.

Baron baronici÷ misleč¤ da govori z Nežko, razlaga¤ kakšna mora biti ženska.

Pri tem se ni opiral na mnenja meščanov, rekoč¤ da je Antigona nedolžna, ampak je vztrajal pri svojem.

2.2.1.1.2.17 Deležijski polstavek na začetku priredja

Ta primer vsebuje pravzaprav celo dve dilemi. Prva je, ali sploh obravnavati »sodeč po izrazih na obrazih« kot polstavek ali reči le, da gre za prislovno določilo. Po mojem mnenju gre za polstavek, zgrajen okoli negolega deležja, je pa v samem stavku v vlogi prislovnega določila.

Druga dilema pa potem je, kje je treba postaviti vejico na začetku. Logično nadaljevanje pravila o zaporednih veznikih nas pripelje do tega, da je vejica pred sprednjim veznikom, v tem primeru torej pred »in«.

Večina učencev je naročila prav to¤ in sodeč po izrazih na obrazih¤ je bilo dobro.

2.2.1.1.2.18 Nenavaden oz. nenaraven besedni red

Nenavaden besedni red kdaj povzroči, da se pomen povedi tako spremeni, da to vpliva na postavitve vejic.

Lik žene v antičnih bajkah÷ mislim¤ da je zelo pomemben, saj se marsikatera zgodba ne bi odvijala÷ tako¤ kot se je brez ženskega lika.
Mislim, da zato, da bi pri Simonu vzbudil nek občutek, da je zavod nekaj dobrega÷ ter¤ kjer se bo lahko pozdravil.
Vlasta Korošec ima kratke lase, rjave oči÷ srednje rasti.
Zdaj treniram troskok, daljino÷ pa tudi kratke proge občasno.

2.2.1.1.2.19 Desni prilastek na levi

V tem primeru bi sicer morda lahko rekli, da gre za pristavčni del.

V mladosti mu je Atlantida pomenila neko zanimivo, polno upanja, državo, kjer je vse drugače kot v Rusiji.

2.2.1.1.2.20 Odvečna podpičja

V korpusu je bilo nekaj primerov, kjer je bilo za začetek pristavkov uporabljeno podpičje namesto vejice, v teh primerih bi pravzaprav morali zamenjati podpičje z vejico, ker pa te vrste napake v vzorcih nimamo predvidene, so bili taki primeri izločeni.

Kreon; takratni vladar¤ je dal ukaz nad truplo enega izmed bratov.

2.2.1.1.2.21 Pika v navedku

Vprašanje je, ali tudi navedke obravnavati kot premi govor in na koncu spreminjati pike v vejice ali ne.

„V marsikaterem očesu so se zalesketale solze.”¤ je navedeno v odlomku.

Spoznal je tudi, da je imela Antigona prav, da je rekla: „Ne da sovražim, da ljubim¤ sem na svetu.”, saj nima pomena sovražiti in biti zaradi tega potrt ali jezen.

Ker pa učenje odlašamo, si rečemo: „Ah, pa sai se lahko učim tudi jutri.”¤ imamo pri testih in ustnem spraševanju težave, kajti nemogoče se je naučiti toliko snovi v enem popoldnevu.

2.2.1.1.2.22 Klicaj v navedku

Ali za klicajem v navedku oz. premem govoru lahko nadaljujemo z vejico, zdi se smiselno.

Če komentiram še Descartesovo misel: „Mislim, torej sem!”, menim, da je z njo hotel poudariti pomen razuma.

Naj zaključim še z mislijo slavnega razsvetljenskega misleca Descartesa, ki pravi: „Mislim, torej sem!”, ki predstavlja bistvo razsvetljenske miselnosti.

Tudi ko je tekla za vozom¤ so se ji vsi posmehovali „No, teci, teci, teci zdaj!", namesto da bi se ustavili in ji pomagali, saj si je po mojem mnenju zaslužila iti na romanje÷ prav tako÷ kot vsi drugi.

Materin nagovor otroku je: "Kaj pa je tebe treba bilo¤ dete ljubó, dete lepo!", kar pomeni, da si ga ni "želela" oz. da je čisto nepričakovano.

Ko je vzkliknila: „Počakajte, ljudje božji!", mi da misliti, da ji je bilo zelo pomembno vkrcati se na voz.

2.2.1.1.2.23 Manjka vprašaj v premem govoru

V tem primeru sicer v resnici manjka vprašaj, ampak če že ni vprašaja, mora biti pa vsaj vejica, zato je bilo označeno, da manjka vejica (ker je izhodišče, da se ne spreminja nič razen vejic).

„Kaj je narobe¤" vprašam zaskrbljeno.

2.2.1.1.2.24 »skoraj da«

Odločitev je bila, da je »skoraj da« členkovna zveza v pomenu »skorajda«, zato vmes ni vejice.

V današnjem času pa je tehnologija že zelo razvita in otroci skoraj÷ da ne poznajo navadnih igrač¤ ampak samo še računalniške igrice in internet.

2.2.1.1.2.25 »in to«

Za razliko od »in sicer«, kjer ni dvoumnosti, se je tukaj večkrat težko odločiti, ali gre za vezniško zvezo ali pa je »in« veznik in »to« osebek.

Problem v nas ženskah je, da smo preveč naivne¤ in to nas lahko pošteno izuči.

2.2.1.1.2.26 ne glede na

Odločitev je bila, da pred »ne glede na« ni vejice (vsaj ne desnosmerne, lahko pa je levosmerna zaradi morebitnega prilastkovega odvisnika pred tem).

In ker ima Bog vse ljudi enako rad÷ ne glede na grehe÷ in ker se je sin kesal za svoje dejanje, mu je oče odpustil.
Do vseh želi biti enakopraven, zato sledi zakonom÷ ne glede na to, kdo je storil napako÷ in koga mora zaradi tega ubiti.
Antigona se odloči, da pokoplje brata÷ ne glede na prepoved trenutnega kralja Kreona.
Kajti s svojo smrtjo se bi že sprijaznila, bolelo pa bi jo, če bi brata÷ ne glede na to¤ kakšen je bil, pustila nepokopanega.

2.2.2 Korpus KUST

Korpus KUST (korpus usvajanja slovenščine kot tujega jezika) je zbirka besedil, ki so jih napisali govorci drugih jezikov, ki se učijo slovenščine. Tak korpus je bil predlagan v (Stritar 2006), besedila so bila zbrana v okviru projekta ESS Uspešno vključevanje otrok,

učencev in dijakov migrantov v vzgojo in izobraževanje, ki ga je izvajal Center za slovenščino kot drugi/tuji jezik Filozofske fakultete Univerze v Ljubljani (Rozman et al. 2010). Besedila so napisana na roko in pretipkana v okviru projekta, dodani so podatki o tipu dokumenta, vrsti in stopnji tečaja, na katerem je besedilo nastalo, starosti avtorja besedila, državi izvora, kraju bivanja in njegovem prvem jeziku. V besedilih so prekriti podatki, ki razkrivajo identiteto avtorja (če je v besedilu omenjeno npr. ime Janez, je to nadomeščeno z XImeX, in to ne glede na sklon). V korpusu je skupaj 32.117 besed v 306 besedilih (Rozman et al. 2010). Glede na prvi jezik avtorja besedila so deleži (glede na število besed) naslednji: španščina 31,7 %, italijanščina 29,9 % (+ 1,8 % dvojezično s slovenščino), angleščina 11,2 %, srbščina 8,4 %, nemščina 4,3 % (+ 1,5 % dvojezično s slovenščino), bosanščina 2,7 %, hrvaščina 0,8 % (sledi makedonščina s 3,7 %, drugi jeziki imajo največ 1,8 %) (Rozman et al. 2010). Starost avtorjev je bila med 13 in 21 let, večina jih je imela med 16 in 19 let (Rozman et al. 2010).

Primeri iz korpusa KUST so bili vključeni zaradi domneve, da je tukaj več napak kot v besedilih rojenih govorcev in da bi lahko slovnični pregledovalnik po potrebi prilagodili, da bi bil še bolj uporaben za tujce, ki se učijo slovensko.

Primeri za ta del so bili pripravljeni, preden sem se odločil, da bodo vsebovali tudi zglede pravilnih povedi, zato so v tem delu le povedi, kjer je bila kakšna manjkajoča ali odvečna vejica. Vejice se bile popravljene ročno, zbranih je bilo 388 povedi.

2.2.3 Korpus Lektor

Korpus lektorskih popravkov Lektor je nastal v okviru doktorske naloge (Popič 2014) in vsebuje približno milijon besed. V njem so besedila, ki so jih napisali pisci, ki se profesionalno ukvarjajo s pisanjem, dobra polovica je prevodov. Dostopen je na naslovu www.korpus-lektor.net. Sama baza je v formatu XML, v besedilu so označeni vsi popravki, ki so jih naredili lektorji.

Podobno kot pri korpusu Šolar je korpus pretvorjen v besedilni format z označenimi napakami pri vejicah, ni pa tukaj narejeno ročno preverjanje, rezultati so taki, kot so jih naredili lektorji, tako da so v tej zbirki primerov vejice postavljene bolj realno in ne nujno tako poenoteno.

Ročno so bile izločene le povedi, ki so bile v celoti v tujem jeziku (v enem od besedil je npr. bilo večje število francoskih primerov), ker so prvi poskusi pokazali, da to zelo moti analizator, ki potem v takih povedih postavlja popolnoma nesmiselne vejice. Vendar to še ni

izvedeno v celoti, potreba po dodatnem čiščenju je odvisna od rezultatov evalvacije postavljanja vejic.

2.2.4 Wikipedija

Zadnja skupina virov so članki iz Wikipedije, pri čemer je bilo treba paziti, da se izberejo članki, ki niso preveč lektorirani (npr. s pomočjo kategorije »Članki, ki so potrebni čiščenja«). Seveda je bilo treba potem ročno označiti vse napačne vejice, kar je enako kot v drugih delih izvedeno z znakoma ¤ in ÷.

Wikipedija je prosta enciklopedija na spletu, ki jo lahko urejajo vsi. Slovenska Wikipedija vsebuje 143.350 člankov (stanje 16. decembra 2014). Wikipedija je bila izbrana zato, ker lahko predpostavimo, da pisci poskušajo pisati čim bolj pravilno (to npr. ne velja pri forumih), po drugi strani pa niso nujno preveč slovnično podkovani. Dobra lastnost Wikipedije je še licenca Creative Commons.

Za izdelavo zbirke primerov je bilo izbranih 9 člankov, in sicer 15. 7. 2013. Ta del je sicer relativno majhen v primerjavi s Šolarjem in Lektorjem, namen je bil dobiti testne primere, ki so iz druge domene kot drugi deli.

2.3 Zbirka primerov napačne in pravilne rabe vejice

Končna verzija zbirke primerov (imenovana Vejica 1,0) vsebuje 113.308 povedi, pri čemer je označenih 17.768 (11,36 %) manjkajočih vejic, 4.608 (3,22 %) vejic pa je označenih za odvečne. Zbirko sestavljajo štirje deli (Šolar, Kust, Lektor in Wikipedija), pri čemer so velikosti posameznih delov oz. poddelov naslednje:

Vejica 1,0	del	poddel	povedi	vejic	manjkajočih	odvečnih	delež manjkajočih	delež odvečnih
KUST	**4**	**0**	**388**	**218**	**420**	**104**	**65,83 %**	**32,30 %**
KUST de	4	14	8	2	11	0	84,62 %	0,00 %
KUST en	4	15	98	29	66	53	69,47 %	64,63 %
KUST es	4	16	110	100	130	32	56,52 %	24,24 %
KUST it	4	17	61	49	74	8	60,16 %	14,04 %
KUST sh	4	18	111	38	139	11	78,53 %	22,45 %
Šolar	**5**	**0**	**49444**	**49233**	**16180**	**3759**	**24,74 %**	**7,09 %**
Šolar OŠ 6	5	23	678	433	237	28	35,37 %	6,07 %
Šolar OŠ 7	5	24	2461	1288	722	116	35,92 %	8,26 %
Šolar OŠ 8	5	25	4003	2509	1427	215	36,26 %	7,89 %
Šolar OŠ 9	5	26	3398	2532	696	172	21,56 %	6,36 %
Šolar PŠ 1	5	27	1137	1166	598	71	33,90 %	5,74 %
Šolar PŠ 2	5	28	619	626	325	50	34,17 %	7,40 %
Šolar PŠ 3	5	29	966	819	519	81	38,79 %	9,00 %
Šolar PŠ 5	5	30	472	436	291	41	40,03 %	8,60 %

Šolar SŠ 1	5	31	4431	3575	1863	344	34,26 %	8,78 %
Šolar SŠ 2	5	32	3533	3818	1389	302	26,68 %	7,33 %
Šolar SŠ 3	5	33	3703	3216	1500	270	31,81 %	7,75 %
Šolar SŠ 4	5	34	2940	3178	1225	242	27,82 %	7,08 %
Šolar G 1	5	19	7241	8490	2117	758	19,96 %	8,20 %
Šolar G 2	5	20	3134	3930	805	209	17,00 %	5,05 %
Šolar G 3	5	21	3613	4399	824	296	15,78 %	6,30 %
Šolar G 4	5	22	6893	8632	1530	551	15,06 %	6,00 %
Šolar MT	5	35	222	186	112	13	37,58 %	6,53 %
Lektor	**7**	**0**	**62606**	**88242**	**1050**	**678**	**1,18 %**	**0,76 %**
Wikipedija	**6**	**0**	**870**	**933**	**118**	**67**	**11,23 %**	**6,70 %**
skupaj			113308	138626	17768	4608	11,36 %	3,22 %

Tabela 1: Sestava zbirke primerov Vejica 1,0

Tabela 1 prikazuje sestavo zbirke primerov Vejica 1,0. Iz rezultatov je očitno, da je število napak pri vejicah po delih zelo različno, največ napak je v delu KUST, najmanj pa v delu Lektor. Dela KUST in Šolar sta dodatno razdeljena na poddele (KUST glede na materni jezik, Šolar pa glede na razred/letnik). Velikosti delov so zelo različne, največja sta Lektor in Šolar, Wikipedija in KUST pa veliko manjša.

Delež manjkajočih vejic je izračunan tako, da delimo število manjkajočih vejic z vsoto števila napisanih vejic in števila manjkajočih vejic (niso bile upoštevane odvečne vejice). Delež odvečnih vejic je bil izračunan kot kvocient števila odvečnih vejic z vsoto števila napisanih vejic in števila odvečnih vejic (niso bile upoštevane manjkajoče vejice).

Zbrana zbirka primerov napačne in pravilne rabe vejice v slovenščini je objavljena v repozitoriju slovenske raziskovalne infrastrukture CLARIN.SI pod imenom Vejica 1.0 s spletno povezavo http://hdl.handle.net/11356/1055 pod licenco CC BY-NC-SA 4.0.

3 Dosedanje delo na področju postavljanja vejic

Tako za slovenščino kot tudi za druge jezike je bilo narejenih že precej programov oz. raziskav za postavljanje vejic v besedilo oziroma za njihovo popravljanje.

3.1 Dosedanji programi za slovenščino

Za slovenščino sta v tem trenutku aktivna dva programa za popravljanje vejic v besedilih, Besana in LanguageTool, Mspell pa se ne razvija več in ga tudi ni mogoče več preizkusiti.

Programa za popravljanje vejic v slovenščini je preizkusil že Holozan (2012), pri čemer so bili uporabljeni primeri iz korpusa Šolar (točka 2.2.1).

3.1.1 Mspell

Mspell je črkovalnik za slovenščino, ki ga je napisal Miha Mazzini leta 1990, opozarjal pa je tudi na nekatere slovnične in slogovne napake, med drugim tudi na manjkajoče vejice (pred vezniki, ki tipično zahtevajo vejico)[4]. Napisan je bil za operacijski sistem DOS.

3.1.2 Besana

Besana (kar je okrajšava za BESedna ANAliza) je slovnični pregledovalnik, ki ga razvija podjetje Amebis. Namenjen je iskanju napak v besedilih, in sicer predvsem slovničnih, to je takih, ki jih črkovalnik ne more odkriti.

Odkriva npr. neujemanje med pridevniki in samostalniki v sklonu, spolu in številu, napačne sklone za predlogi, napačne variante predlogov s/z oz. k/h, napačne predloge pri krajevnih imenih, zanikanje s tožilnikom, napačno tvorbo trpnika, nekatere tipične neknjižne uporabe, napačno dvojino, napačne velike/male začetnice, presledke pri ločilih ipd. Pomemben del pa je tudi opozarjanje na napačne vejice, in sicer tako na manjkajoče kot odvečne.

Besana lahko deluje kot samostojen program (Besana Mini) ali pa je vključena kot preverjanje slovnice v Microsoft Word ali LibreOffice.

Dodatek pri Besani je še pregibnik, tj. program za pregibanje (spreganje, sklanjanje) besed. Poleg oblik besed pri krajevnih imenih pokaže še dodatne podatke (tipične predloge,

[4] Za informacije o programu Mspell se zahvaljujem avtorju Mihi Mazziniju.

pridevnik, imena prebivalcev). Pregibnik je dostopen tudi na spletni strani http://besana.amebis.si/pregibanje/.

Besana odkriva morebitne napake na dva načina: osnovni način je stavčni analizator, ki ima vgrajene tudi tipične napake (opis nekaterih tipičnih napak, ki jih analizator spregleduje, in težav, ki lahko zaradi tega nastanejo, je v Holozan (2006)) oz. so tipične napake skupaj s podatki o vrstnih oznakah vgrajene že v leksikalno podatkovno zbirko ASES (Arhar in Holozan 2009). Kadar pa analizatorju analiza ne uspe, uporabi Besana pomožna pravila, ki so vgrajena neposredno v kodo programa in jih uporabniki ne morejo prilagajati (lahko pa pri vseh vrstah napak nastavijo, ali želijo, da jih Besana opozarja nanje).

3.1.2.1 Kratka zgodovina Besane

Prva zabeležena verzija programa Besana je nastala 9. septembra 1988, in sicer kot kratek program, napisan v jeziku Basic. Ta je za vneseno besedo napisal, v katerem sklonu je. V slovarju so bile štiri besede: zveza, hiša, reka in voda. Nastala je kot srednješolska raziskovalna naloga (avtorji Peter Holozan, Marko Šimunovič in Iztok Grilc, mentorja Dušan Sitar in Mojca Poznanovič), katere prvotni cilj je bil narediti strojni prevajalnik iz slovenščine v angleščino, ker pa se je pokazalo, da je že prvi korak do strojnega prevajalnika, pri katerem se besedilo analizira, zajeten zalogaj, se je na koncu raziskovalna naloga omejila na problem besedne analize slovenskih besedil.

Že med razvojem pa se je pokazalo, da metode, ki so uporabljene za izločanje odvečnih možnosti (oblikoskladenjskih oznak) pri besedah, občasno dosežejo, da pri besedi ne ostane več kot možna niti ena možnost. To se zgodi, če se npr. pridevnik in samostalnik, ki stojita zapored, ne ujemata v sklonu, spolu in številu. In to se je največkrat dogajalo v primerih, ko je bila v besedilu napaka. Podobno se je z dodajanjem vedno večjega števila besed v slovar izkazalo, da se vedno pogosteje dogaja, da je še neznana beseda zatipkana, tako pravzaprav delujejo črkovalniki.

Po nekaj letih dela je Besana dosegla stopnjo, ko je bila dovolj uporabna, da je lahko postala prodajni izdelek, in sicer predvsem v vlogi iskanja slovničnih in tipkarskih napak, čeprav je v ozadju še vedno ostala tudi sama besedna analiza. Program je bil narejen za sistem DOS in ga je podjetje Amebis kar nekaj let uspešno prodajalo in seveda tudi počasi dopolnjevalo.

Po nekaj letih premora je bila na novo narejena besedna analiza, in sicer najprej za strojni prevajalnik Presis. Temu pa je (leta 2005) sledila še nova verzija Besane za okolje Windows, ki je sicer uporabljala nekaj starih metod delovanja, velika večina pa je bila

narejena na novo, in sicer predvsem na podlagi stavčne analize s pomočjo glagolskih predlog[5].

3.1.2.2 Dosedanje metode odkrivanja napačnih vejic

Besana je uporabljala dve glavni metodi za iskanje manjkajočih vejic. Prva je preverjala vejice pred tipičnimi vezniki, ki zahtevajo vejico, pri čemer je pazila na morebitne druge veznike ali členke spredaj, pa tudi na to, da je npr. pri »da« na to opozarjala le, kadar je bilo iz povedi jasno, da na tem mestu ne more biti »da« kot oblika glagola »dati«.

Druga metoda pa je temeljila na opozarjanju, da najbrž manjka vejica, v primerih, ko sta v stavku dve besedi, ki sta glagol (in ne moreta biti nič drugega), vmes pa ni nobenega veznika ali ločila. Težava pri tej metodi je, da program ne more povedati točnega mesta, kje vejica manjka, ampak lahko le opozori na to, da manjka kje v stavku.

3.1.3 LanguageTool

LanguageTool je odprtokodni program za preverjanje sloga in slovnice. Podpira angleščino, francoščino, nemščino, poljščino, nizozemščino, romunščino in še množico drugih jezikov, med katerimi je tudi slovenščina. Odkriva napake, ki jih črkovalniki ne morejo. Program deluje s pravili, ki so narejena za vsak jezik posebej. Osnovna pravila so zapisana v formatu XML, podpira pa tudi kompleksna pravila, napisana v jeziku Java.

Je prosto dostopen pod licenco LGPL. Glavni vzdrževalec za slovenska pravila je Martin Srebotnjak, 30. 6. 2012 je bilo za slovenščino 85 pravil (od teh 41 za manjkajoče vejice). Do verzije 2.7 (stanje 12. decembra 2014) se pravila za vejico niso spremenila, zadnja sprememba se je pojavila v verziji 1.6 (21. decembra 2011).

LanguageTool lahko deluje kot samostojni program ali pa je vključen v LibreOffice oz. OpenOffice.

Izboljšava pravil za LanguageTool je bila narejena v Piškur (2015). V delu so bila preizkušena na korpusu Lektor, opis v evalvaciji pravil pa kaže, da so nekatera nova pravila naredila več težav kot koristi, zato ta nova pravila niso bila dodatno preizkušena.

[5] Glagolske predloge so vzorci, ki opisujejo vezljivost glagola v stavku in so del podatkovne zbirke ASES. Podroben opis je v Holozan (2011).

3.1.4 Strojno učenje

Ob začetku pisanja pričujočega dela za slovenščino še ni bilo preizkušeno strojno učenje (statistične metode) za postavljanje vejic (strojno učenje pa je bilo uporabljeno za druge jezike, kar je opisano v točki 3.2.2).

Vmesni rezultati uporabe strojnega učenja za postavljanje vejic v slovenščini so objavljeni v Holozan (2013).

Med našim preizkušanjem strojnega učenja in zaključkom pričujočega dela strojno učenje pri postavljanju vejic obravnava še Krajnc (2015), pri čemer uporablja isto zbirko primerov iz korpusa Šolar. Zbirka primerov iz Holozan (2013) je označena kot Šolar1, zbirka primerov, ki je bila pripravljena v okviru te naloge, pa kot Šolar2. Dodatno je Krajnčeva preizkusila klasifikator RandomForest[6] (naključni gozdovi) in metodo podpornih vektorjev (SVM[7]). Preizkušala je tudi različne nabore atributov, pri čemer je zanimiva njena ideja, da uporabi kot atribute rezultate pravil, napisanih za LanguageTool.

Zgornja raziskava je dopolnjena še v Krajnc in Robnik-Šikonja (2015).

3.2 Delo za druge jezike

Veliko dela na področju postavljanja oz. popravljanja vejic je bilo narejenega tudi za druge jezike, vendar ti rezultati niso nujno neposredno uporabni oz. primerljivi, zaradi različnih pravil za postavljanje vejic v različnih jezikih.

3.2.1 Postavljanje vejic z uporabo pravil

Ni veliko objavljenega o postavljanju vejic z uporabo pravil za druge jezike. Pravila uporabljajo komercialni programi za preverjanje slovnice (npr. Grammarly[8], Grammatik[9]), ki pa ne objavljajo podrobnosti o načinu delovanja.

Zanimivo pa je delo, ki sta ga opravila Jakubíček in Horák (2010) za češčino, še posebej zato, ker so pravila za postavljanje vejic v češčini do neke mere podobna slovenskim pravilom[10]. Uporabila sta dopolnjen razčlenjevalnik za določitev strukture povedi, iz katere sta potem določila mesta, na katerih morajo biti vejice. Podobno je delal tudi Kovář (2014), ki

[6] https://en.wikipedia.org/wiki/Random_forest
[7] https://sl.wikipedia.org/wiki/Metoda_podpornih_vektorjev
[8] https://www.grammarly.com
[9] https://en.wikipedia.org/wiki/Grammatik
[10] https://en.wikipedia.org/wiki/Czech_orthography#Punctuation

je uporabil drug razčlenjevalnik (SET) in dosegel zelo dobre rezultate (natančnost 95 % pri priklicu 50 %).

3.2.2 Strojno učenje

V drugih jezikih so strojno učenje že večkrat uporabili za učenje postavljanja vejic, večinoma pa so raziskovali problem, ko je treba v besedilo postaviti vse vejice (oz. nekateri celo vsa ločila), kar je pomembno predvsem pri sistemih za razpoznavo govora (Huang & Zweig 2002).

Beeferman at al. (1998) so preizkušali postavljanje vejic v angleščini s pomočjo skritega markovskega modela[11] in z uporabo Viterbijevega algoritma[12].

V danščini je postavljanje vejic preizkušal Hardt (2001), in sicer z uporabo Brillovega označevalnika[13], vendar se je omejil le na ugotavljanje odvečnih vejic, pri čemer so bile odvečne vejice dodane naključno.

Zhang et al. (2002) so preizkušali strojno učenje za vejice v angleščini in nemščini, in sicer z odločitvenimi drevesi z uporabo skladenjskih podatkov.

Shieber in Tao (2003) sta preizkušala postavljanje vejic za angleščino; pomembna je njuna ugotovitev, da se statistični označevalnik splača učiti njegovega dela na učnem korpusu brez vejic.

Alegria et al. (2006) so preizkušali strojno učenje v baskovščini. Uporabili so program WEKA[14] in preizkušali različne metode strojnega učenja.

Israel et al. (2012) so se ob problemu postavljanja vseh vejic v angleškem besedilu lotili tudi problema popravljanja napačnih (manjkajočih in odvečnih) vejic v njem.

Hara et al. (2013) so se ukvarjali s postavljanjem vejic v kitajščini, uporabili so kombiniran pristop, kjer v prvem koraku postavljajo vejice s pomočjo strojnega učenja, v drugem koraku pa uporabijo filter, ki temelji na pravilih.

---

[11] https://sl.wikipedia.org/wiki/Skriti_model_Markova
[12] https://en.wikipedia.org/wiki/Viterbi_algorithm
[13] https://en.wikipedia.org/wiki/Brill_tagger
[14] http://www.cs.waikato.ac.nz/ml/weka/

4 Evalvacija dosedanjih rezultatov

4.1 Ocenjevanje rezultatov

Za ocenjevanje rezultatov sta enako kot v Holozan (2013) uporabljeni metriki natančnost (delež pravilno postavljenih vejic) in priklic (delež odkritih manjkajočih vejic) ter metrika F1, ki je harmonična sredina natančnosti in priklica in se izračuna kot 2 * natančnost * priklic / (natančnost + priklic)[15].

Natančnost je lahko v veliki meri odvisna od preizkusnih primerov. Če je med primeri zelo malo napak, je natančnost slabša, ker se poveča verjetnost napačnih popravkov, če v primerih ni nobene napake, je natančnost enaka 0, takoj ko se najde vsaj en napačen popravek (Helfrich, Music 2000).

Vprašanje je, čemu dati prednost pri skupni oceni, natančnosti ali priklicu. Vsekakor ne le eni metriki, kajti hitro se da pokazati, da to ni smiselno (program, ki bi vsem besedam pripisal vejico, bi tako imel priklic 100 % pri zelo slabi natančnosti, po drugi strani pa bi program, ki bi postavljal le najbolj zanesljive vejice, imel natančnost 100 % pri zelo slabem priklicu. V ta namen sicer imam metriko F1, vendar sta tam natančnost in priklic enakovredno zastopana, kar pa ni nujno najboljša rešitev. Zanimivo vprašanje je celo, ali bi bilo želeno natančnost smiselno prilagajati potencialnemu uporabniku. Pri uporabniku, ki o postavljanju vejic ne ve veliko, mora biti natančnost toliko večja, saj ga drugače program lahko hitro zavede. Po drugi strani pa je uporabniku, ki zna sam postavljati vejice, program pa uporablja bolj za to, da česa ne spregleda, mogoče ponuditi manjšo natančnost, da potem sam izbere tiste popravke, ki so pravilni (če za to dobi ustrezno višji priklic). Za to bi lahko uporabili metriki F2 (ta poudari priklic) in F0,5 (ta poudari natančnost).

4.2 Metoda evalvacije

Metoda evalvacije je bila zasnovana v (Holozan 2012), izkazalo pa se je, da je nekaj zmot, kar se tiče odvečnih vejic. Te napake so zdaj odpravljene. Podobno se je pokazalo nekaj težav tudi pri evalvaciji rezultatov pri LanguageTool.

[15] https://en.wikipedia.org/wiki/F1_score

Za evalvacijo je narejen program Vejicar[16], ki pregleda zbirko primerov in za vsak primer naredi primerjavo med pravilno postavljenimi vejicami in vejicami, kot jih postavita Besana in LanguageTool. Rezultat se na koncu zapiše v datoteki v tekstovnem formatu v obliki tabele, kjer so podatki med seboj ločeni s tabulatorji. S tem je omogočeno preprosto kopiranje v preglednico, narejeno v Excelu, v katerem se izračunajo dodatni podatki (deleži) in oblikujejo grafi. V Excelu narejene tabele in grafi se na koncu skopirajo v Wordov dokument.

4.2.1 Besana

Podpora za Besano je vgrajena v program Vejicar. V povedi se ustrezno postavijo vejice glede na to, kaj želimo poiskati (če ugotavljamo napake pri vejicah, zbrišemo oznake za manjkajoče vejice in spremenimo oznake za odvečne vejice v vejice; če preverjamo postavljanje vseh vejic, zbrišemo vse vejice ter oznake za manjkajoče in odvečne vejice). Besana kot rezultat vrne seznam potencialnih napak, v seznamu se poiščejo tiste potencialne napake, ki zadevajo vejice, potem pa se na mesta, ki so navedena pri potencialnih napakah, postavijo oznake za manjkajoče oz. odvečne vejice.

Pri Besani so dodatno še oznake za primere, ko opozori, da nekje manjka vejica, ne ve pa točnega mesta. Ta mesta se označijo z znakom ×, se pa pri teh znakih potem evalvacija zaplete, ker lahko ena manjkajoča vejica povzroči tudi več opozoril, da neznano kje manjka vejica. Ker so opozorila, ki točno povedo, kje manjka vejica, veliko bolj uporabna od tistih, da vejica nekje manjka, so v rezultatih prikazani še posebej (označeni z »brez nekje«).

4.2.2 LanguageTool

Pokazalo se je, da je evalvacijo za LanguageTool najbolj preprosto narediti tako, da se rezultati, ki jih vrne LanguageTool, vnaprej pripravijo kot besedilne datoteke in se potem dodajo kot dodatni stolpci v datoteko z zbirko primerov, ki jo bere program Vejicar. Ta potem prebere te dodatne stolpce in uporabi pravega, ki je zapisan enako kot rezultati pri Besani, tako da je nadaljnji postopek evalvacije enak kot pri Besani.

Stolpec s primeri se najprej izvozi v besedilno datoteko. Ta besedilna datoteka se potem pretvori v tri ločene besedilne datoteke: v prvi se zbrišejo oznake za manjkajoče vejice, oznake za odvečne vejice pa zamenjajo z vejicami; v drugi se oznake za manjkajoče vejice zamenjajo z vejicami, oznake za manjkajoče vejice pa se zbrišejo (kot rezultat dobimo

[16] Pravilno ime bi sicer bilo Vejičar, vendar imajo računalniki še vedno kdaj težave s črkami, ki niso del angleške abecede, zato potem kar Vejicar.

primere brez napak); v tretji se zbrišejo vse vejice in vse oznake za manjkajoče in odvečne vejice (dobimo primere popolnoma brez vejic za preizkušanje postavljanja vseh vejic).

V naslednjem koraku se vse tri besedilne datoteke označijo z LanguageTool (uporabljena je bila verzija 2.7). To se doseže s tem, da se pravila za slovenščino v LanguageTool predelajo tako, da pri popravkih namesto vejice dodajajo znak za manjkajočo vejico (¤). Hkrati sem popravil tudi pravilo za vejico pred »temveč tudi«, ki je razen dodane vejice zamenjalo tudi »temveč« z »ampak« (kar je najbrž napaka, do katere je prišlo, ker je bilo pravilo za vejico pred »temveč tudi« narejeno tako, da se je skopiralo pravilo za manjkajočo vejico pred »ampak tudi«). Da drugi popravki (npr. brisanje podvojenih besed, popravljanje besed, ki jih ne prepozna črkovalnik) ne bi motili delovanja, je bilo treba LanguageTool nastaviti tako, da se onemogočijo vsa druga pravila razen pravil za napake pri vejicah (pri tem je treba paziti, da se ne uporabi pravilo za manjkajoče vejice pred »to je«, ki je privzeto kot izključeno in je zelo nenatančno, saj na primer v povedi »In to je zares žalostno.« predlaga vejico: »In¤ to je zares žalostno.«). Ukazna vrstica, uporabljena za označevanje, je naslednja:

java -jar languagetool-commandline.jar -l sl -eo -e TODA_BREZ_VEJICE,SAJ_BREZ_VEJICE,KAJTI_BREZ_VEJICE,ZATO_BREZ_VEJICE,ZATOREJ_BREZ_VEJICE,TOREJ_BREZ_VEJICE,TJ_BREZ_VEJICE,AMPAK_TUDI_BREZ_VEJICE,TEMVEČ_TUDI_BREZ_VEJICE,AMPAK_BREZ_VEJICE,KI_BREZ_VEJICE,PREDLOG_KATERI_BREZ_VEJICE,PREDLOG_KAR_BREZ_VEJICE,KAR_BREZ_VEJICE,PREDLOG_KAJ_BREZ_VEJICE,KAJ_BREZ_VEJICE,KATERI_BREZ_VEJICE,KAKRŠEN_BREZ_VEJICE,KAKŠEN_BREZ_VEJICE,ZARADI_BREZ_VEJICE,KO_BREZ_VEJICE,KADAR_BREZ_VEJICE,KAMOR_BREZ_VEJICE,KOLIKOR_BREZ_VEJICE,V_KOLIKOR_BREZ_VEJICE,KAKOR_BREZ_VEJICE,KAKO_BREZ_VEJICE,KER_BREZ_VEJICE,DA_BREZ_VEJICE,KOT_BREZ_VEJICE,ČE_BREZ_VEJICE,NE_DA_BI_BREZ_VEJICE,PREDEN_BREZ_VEJICE,KAJNE_BREZ_VEJICE,KJE_BREZ_VEJICE,KJER_BREZ_VEJICE,KDAJ_BREZ_VEJICE,KAM_BREZ_VEJICE,DOKLER_BREZ_VEJICE,VSE_DOKLER_BREZ_VEJICE -a primeri.txt > primeriozn.txt

Ker je postopek izveden kot ukazna datoteka in ker so v imenih pravil tudi črke č, š in ž, je bilo ukazno datoteko treba pretvoriti iz kodne tabele CP 1250 (ki je uporabljena v sistemu Windows) v kodno tabelo CP 852 (ki je bila uporabljena v sistemu DOS, še vedno pa je uporabljena v ukaznem oknu).

Ko so bile v pravilih dodane vejice zamenjane z znakom ¤, se je pokazalo, da LanguageTool v tem primeru pred znake ¤ doda še presledek, zato so bili v naslednjem koraku vsi presledki pred ¤ zbrisani. Na koncu so bile besedilne datoteke z rezultati v obliki dodatnih stolpcev ročno skopirane v preglednico z zbirko primerov v Excelu in potem je bila cela zbirka primerov skupaj z dodatnimi stolpci spet izvožena v besedilno datoteko, ki jo uporablja program Vejicar.

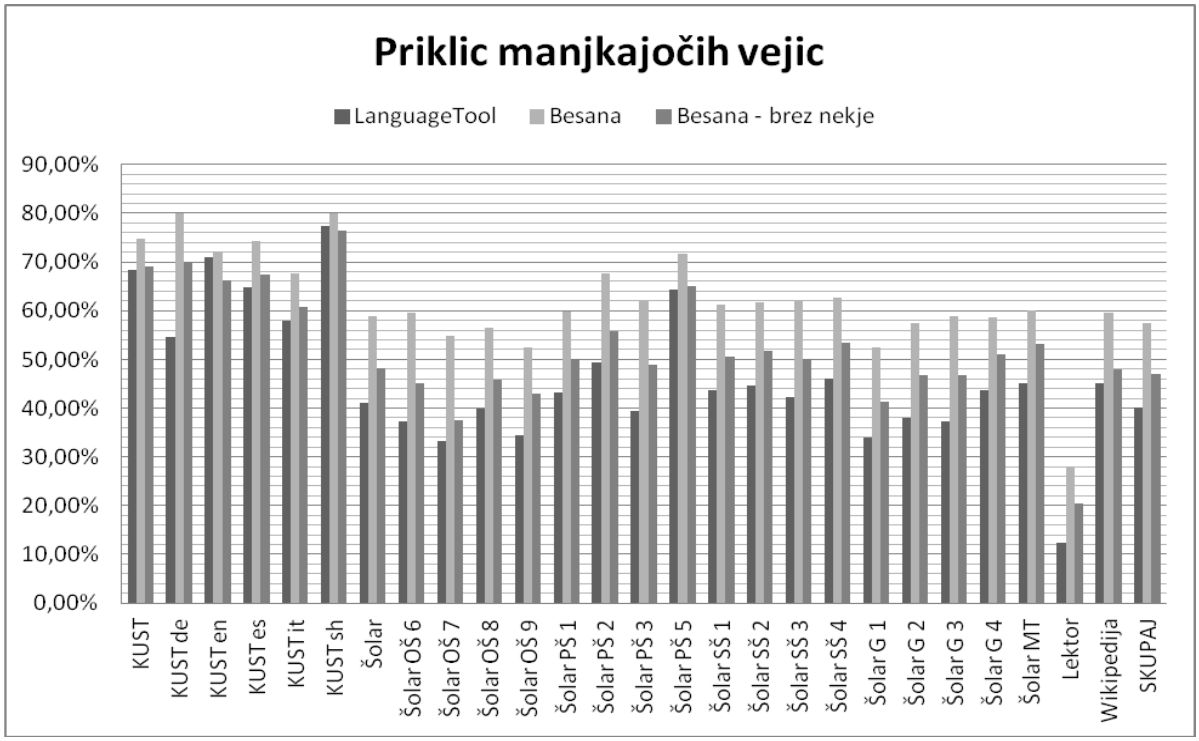
4.3 Iskanje napak pri postavljanju vejic

Primerjava med LanguageTool in Besano je bila objavljena v (Holozan 2012), vendar je bila pri tem uporabljena starejša zbirka primerov iz korpusov KUST in Šolar. Zato je bila primerjava ponovno izvedena z rabo nove zbirke primerov (v točki 6.2 je odkrit manjši hrošč pri evalvaciji rezultatov za LanguageTool, a brez bistvenih sprememb rezultatov).

Tabela 2 prikazuje rezultate evalvacije za LanguageTool po posameznih delih in poddelih. LanguageTool ne odkriva odvečnih vejic, zato so stolpci, kar se tega tiče, prazni. Rezultati so zelo različni po delih, izrazito najslabši rezultat pa velja za del iz korpusa Lektor.

Tabela 3 prikazuje rezultate evalvacij za Besano po posameznih delih in poddelih. Rezultati so zelo odvisni od tega, za kateri del gre, opazni pa so slabši rezultati pri delu iz korpusa Lektor.

V naslednjem koraku smo primerjali rezultate programov LanguageTool in Besana.



Slika 1: Priklic manjkajočih vejic – dosedanji rezultati

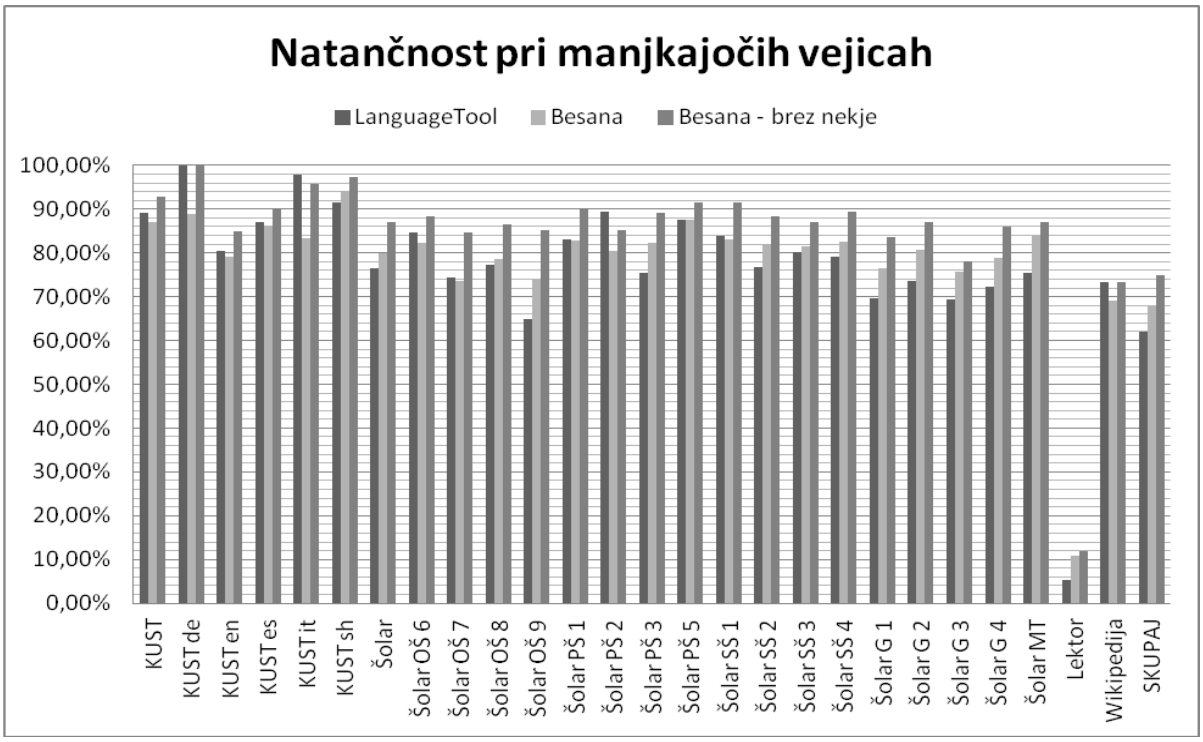
Tabela 2: Rezultati evalvacije za LanguageTool - dosedanji rezultati

LT	zahtevana najdena	zahtevana nekje najdena	zahtevana nenajdena	odvečna prečrtana	odvečna neprečrtana	ni in najdena	ni in nekje najdena	dobra in prečrtana	priklic manjkajočih	priklic odvečnih	natančnost manjkajočih	natančnost odvečnih	F1 manjkajočih	F1 odvečnih
KUST	295	0	136	0	104	36	0	0	68,45%	0,00%	89,12%	/	77,43%	/
KUST de	6	0	5	0	0	0	0	0	54,55%	/	100,00%	/	70,59%	/
KUST en	49	0	20	0	53	12	0	0	71,01%	0,00%	80,33%	/	75,38%	/
KUST es	87	0	47	0	32	13	0	0	64,93%	0,00%	87,00%	/	74,36%	/
KUST it	44	0	32	0	8	1	0	0	57,89%	0,00%	97,78%	/	72,73%	/
KUST sh	109	0	32	0	11	10	0	0	77,30%	0,00%	91,60%	/	83,85%	/
Šolar	6716	0	9665	0	3759	2067	0	0	41,00%	0,00%	76,47%	/	53,38%	/
Šolar OŠ 6	89	0	149	0	28	16	0	0	37,39%	0,00%	84,76%	/	51,90%	/
Šolar OŠ 7	240	0	484	0	116	83	0	0	33,15%	0,00%	74,30%	/	45,85%	/
Šolar OŠ 8	575	0	863	0	215	168	0	0	39,99%	0,00%	77,39%	/	52,73%	/
Šolar OŠ 9	241	0	460	0	172	130	0	0	34,38%	0,00%	64,96%	/	44,96%	/
Šolar PŠ 1	262	0	345	0	71	53	0	0	43,16%	0,00%	83,17%	/	56,83%	/
Šolar PŠ 2	169	0	173	0	50	20	0	0	49,42%	0,00%	89,42%	/	63,65%	/
Šolar PŠ 3	208	0	320	0	81	68	0	0	39,39%	0,00%	75,36%	/	51,74%	/
Šolar PŠ 5	190	0	105	0	41	27	0	0	64,41%	0,00%	87,56%	/	74,22%	/
Šolar SŠ 1	823	0	1064	0	344	159	0	0	43,61%	0,00%	83,81%	/	57,37%	/
Šolar SŠ 2	628	0	777	0	302	190	0	0	44,70%	0,00%	76,77%	/	56,50%	/
Šolar SŠ 3	642	0	877	0	270	158	0	0	42,26%	0,00%	80,25%	/	55,37%	/
Šolar SŠ 4	572	0	669	0	242	152	0	0	46,09%	0,00%	79,01%	/	58,22%	/
Šolar G 1	728	0	1418	0	758	317	0	0	33,92%	0,00%	69,67%	/	45,63%	/
Šolar G 2	309	0	506	0	209	111	0	0	37,91%	0,00%	73,57%	/	50,04%	/
Šolar G 3	311	0	522	0	296	137	0	0	37,33%	0,00%	69,42%	/	48,56%	/
Šolar G 4	677	0	870	0	551	261	0	0	43,76%	0,00%	72,17%	/	54,49%	/
Šolar MT	52	0	63	0	13	17	0	0	45,22%	0,00%	75,36%	/	56,52%	/
Lektor	129	0	915	0	678	2273	0	0	12,36%	0,00%	5,37%	/	7,49%	/
Wikipedija	55	0	67	0	67	20	0	0	45,08%	0,00%	73,33%	/	55,84%	/
SKUPAJ	7195	0	10783	0	4608	4396	0	0	40,02%	0,00%	62,07%	/	48,67%	/

Tabela 3: Rezultati evalvacije za Besano – dosedanji rezultati

Besana	zahtevana najdena	zahtevana nekje najdena	zahtevana nenajdena	odvečna prečrtana	odvečna neprečrtana	ni in najdena	ni in nekje najdena	dobra in prečrtana	priklic manjkajočih	priklic odvečnih	natančnost manjkajočih	natančnost odvečnih	F1 manjkajočih	F1 odvečnih
KUST	293	24	107	14	90	23	24	2	74,76%	13,46%	87,09%	87,50%	80,46%	23,33%
KUST de	7	1	2	0	0	0	1	0	80,00%	/	88,89%	/	84,21%	/
KUST en	45	4	19	9	44	8	5	0	72,06%	16,98%	79,03%	100,00%	75,38%	29,03%
KUST es	89	9	34	2	30	10	6	2	74,24%	6,25%	85,96%	50,00%	79,67%	11,11%
KUST it	45	5	24	3	5	2	8	0	67,57%	37,50%	83,33%	100,00%	74,63%	54,55%
KUST sh	107	5	28	0	11	3	4	0	80,00%	0,00%	94,12%	/	86,49%	/
Šolar	7801	1730	6655	707	3052	1172	1226	183	58,88%	18,81%	79,90%	79,44%	67,80%	30,42%
Šolar OŠ 6	105	34	94	6	22	14	16	3	59,66%	21,43%	82,25%	66,67%	69,15%	32,43%
Šolar OŠ 7	269	125	324	16	100	49	92	4	54,87%	13,79%	73,64%	80,00%	62,89%	23,53%
Šolar OŠ 8	648	148	615	43	172	102	114	11	56,41%	20,00%	78,66%	79,63%	65,70%	31,97%
Šolar OŠ 9	301	66	334	31	141	52	78	5	52,35%	18,02%	73,84%	86,11%	61,27%	29,81%
Šolar PŠ 1	299	58	239	7	64	33	41	5	59,90%	9,86%	82,83%	58,33%	69,52%	16,87%
Šolar PŠ 2	185	39	107	5	45	32	22	0	67,67%	10,00%	80,58%	100,00%	73,56%	18,18%
Šolar PŠ 3	246	65	191	5	76	30	37	1	61,95%	6,17%	82,28%	83,33%	70,68%	11,49%
Šolar PŠ 5	191	20	83	6	35	18	12	1	71,77%	14,63%	87,55%	85,71%	78,88%	25,00%
Šolar SŠ 1	943	200	722	56	288	86	147	6	61,29%	16,28%	83,07%	90,32%	70,53%	27,59%
Šolar SŠ 2	718	139	533	43	259	94	94	12	61,65%	14,24%	82,01%	78,18%	70,39%	24,09%
Šolar SŠ 3	747	178	567	43	227	110	101	9	62,00%	15,93%	81,43%	82,69%	70,40%	26,71%
Šolar SŠ 4	656	113	460	43	199	78	84	7	62,57%	17,77%	82,60%	86,00%	71,20%	29,45%
Šolar G 1	884	236	1019	141	617	172	170	26	52,36%	18,60%	76,61%	84,43%	62,20%	30,49%
Šolar G 2	380	86	346	44	165	57	54	18	57,39%	21,05%	80,76%	70,97%	67,10%	32,47%
Šolar G 3	387	99	341	52	244	109	47	11	58,77%	17,57%	75,70%	82,54%	66,17%	28,97%
Šolar G 4	782	116	635	165	386	127	113	64	58,58%	29,95%	78,91%	72,05%	67,24%	42,31%
Šolar MT	60	8	45	1	12	9	4	0	60,18%	7,69%	83,95%	100,00%	70,10%	14,29%
Lektor	213	79	752	119	559	1585	786	516	27,97%	17,55%	10,97%	18,74%	15,75%	18,13%
Wikipedija	58	14	49	16	51	21	11	7	59,50%	23,88%	69,23%	69,57%	64,00%	35,56%
SKUPAJ	8365	1847	7563	856	3752	2801	2047	708	57,45%	18,58%	67,81%	54,73%	62,20%	27,74%

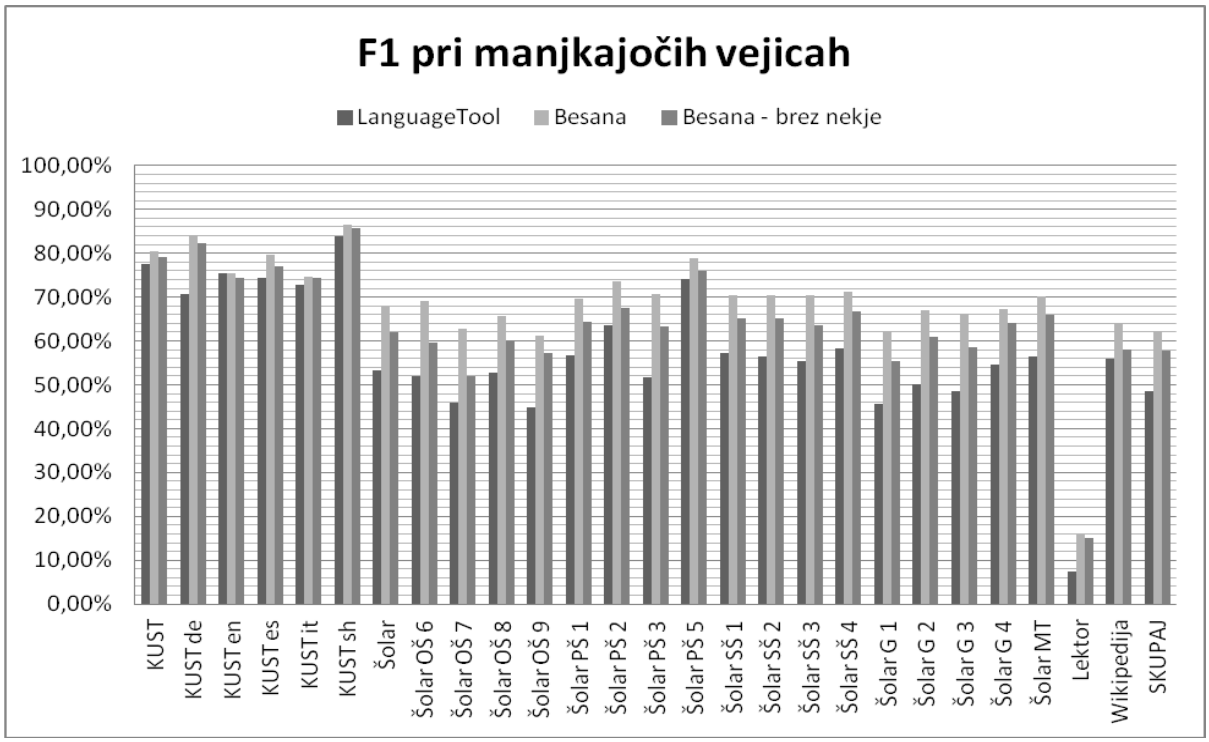
Slika 1 prikazuje dosežen priklic iskanja manjkajočih vejic v programih LanguageTool in Besana, pri čemer sta za Besano dva rezultata, od katerih prvi upošteva vejice, glede katerih Besana le opozori, da manjkajo, ne navede pa točnega mesta, drugi pa upošteva le vejice, glede katerih Besana sporoči točno mesto (to je v rezultatih označeno kot »brez nekje«). Rezultati kažejo, da je priklic zelo odvisen od tega, kakšen je del (manj poddel) zbirke primerov: najboljši rezultati so za del iz korpusa KUST, najslabši pa je rezultat za del iz korpusa Lektor, in to zelo izrazito (saj izkazuje manj kot pol natančnosti, ki velja glede drugih delov). Razmerje med LanguageTool in Besano pa je v vseh delih podobno, Besana brez vejic, ki niso na točnih mestih, je okoli 6 odstotnih točk uspešnejša kot LanguageTool, če pa upoštevamo še opozorila za vejice brez točnega mesta, je priklic boljši za okoli 15 odstotnih točk. Relativno je največja razlika pri delu iz korpusa Lektor, kjer je Besana več kot dvakrat uspešnejša od LanguageToola.



Slika 2: Natančnost pri manjkajočih vejicah – dosedanji rezultati

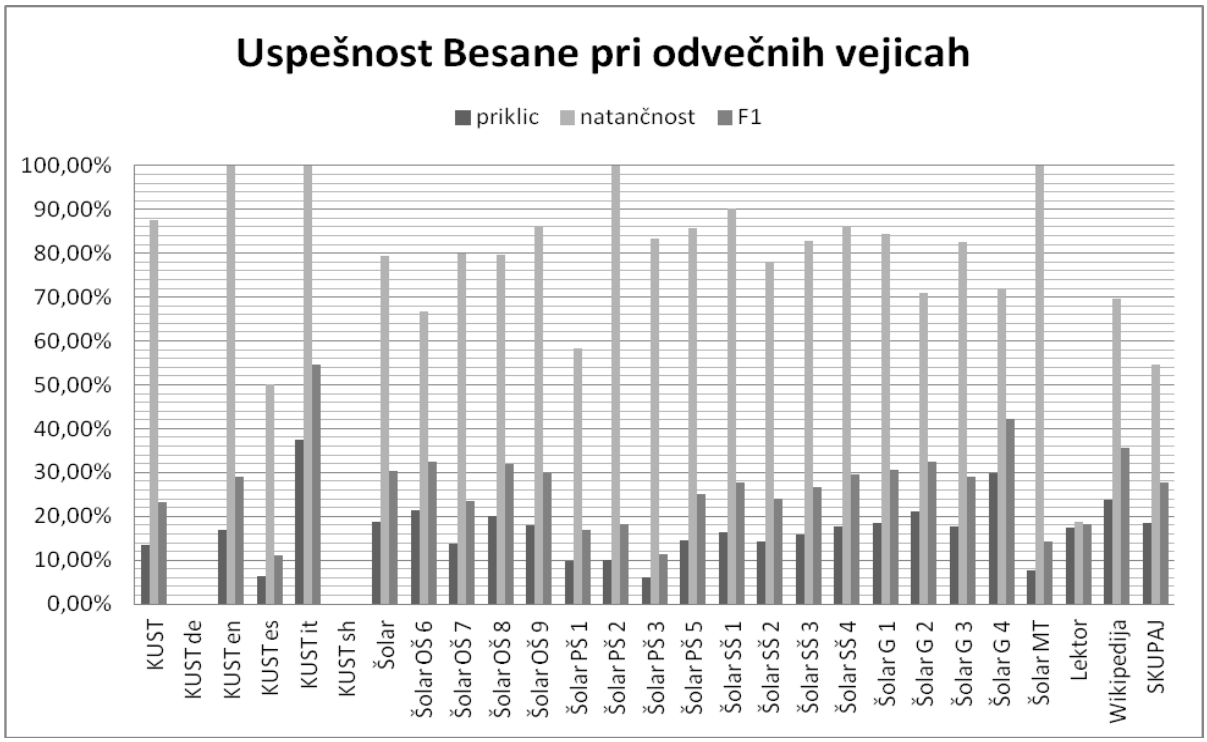
Slika 2 prikazuje doseženo natančnost pri iskanju manjkajočih vejic. Tudi tokrat je rezultat izrazito odvisen od dela pri korpusu Lektor, kjer je rezultat veliko slabši kot v drugih delih. To morda pojasnjuje Tabela 1 na strani 21, ki kaže, da je delež napačnih vejic v delu iz korpusa Lektor veliko manjši kot v drugih delih, zato je naloga popravljanja vejic v tem delu pač veliko težja. Razlika med LanguageTool in Besano je pri natančnosti manjša (v nekaterih

primerih je LanguageTool tudi boljši), zanimivo pa je, da je Besana bol natančna, če ne upoštevamo popravkov brez točno določenega mesta vejice.



Slika 3: F1 pri manjkajočih vejicah – dosedanji rezultati

Slika 3 prikazuje rezultate za metriko F1 pri iskanju manjkajočih vejic. Rezultati so podobni rezultatom za priklic, najboljša je Besana (skupaj z vejicami brez točno določenega mesta).

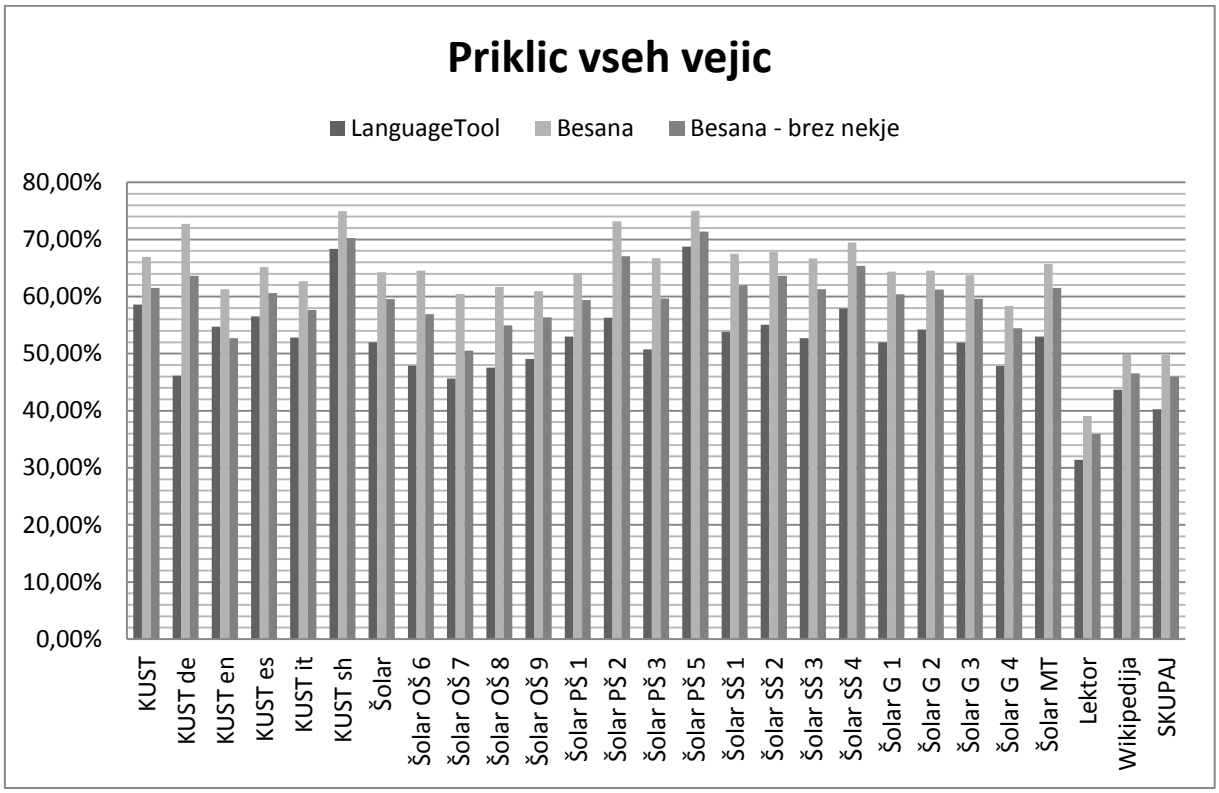


Slika 4: Uspešnost Besane pri odvečnih vejicah – dosedanji rezultati

Slika 4 prikazuje uspešnost Besane pri odkrivanju odvečnih vejic (LanguageTool odvečnih vejic ne išče), na isti sliki so prikazani priklic, natančnost in F1. Priklic je precej manjši kot pri manjkajočih vejicah, je pa v večini primerov visoka natančnost, zelo slab pa je spet rezultat pri delu iz korpusa Lektor (kjer je del razlage spet to, da je delež odvečnih vejic v tem delu zelo majhen). V poddelih iz korpusa KUST je vidno veliko nihanje rezultatov, kar je možno pripisati majhnemu številu primerov v teh poddelih.

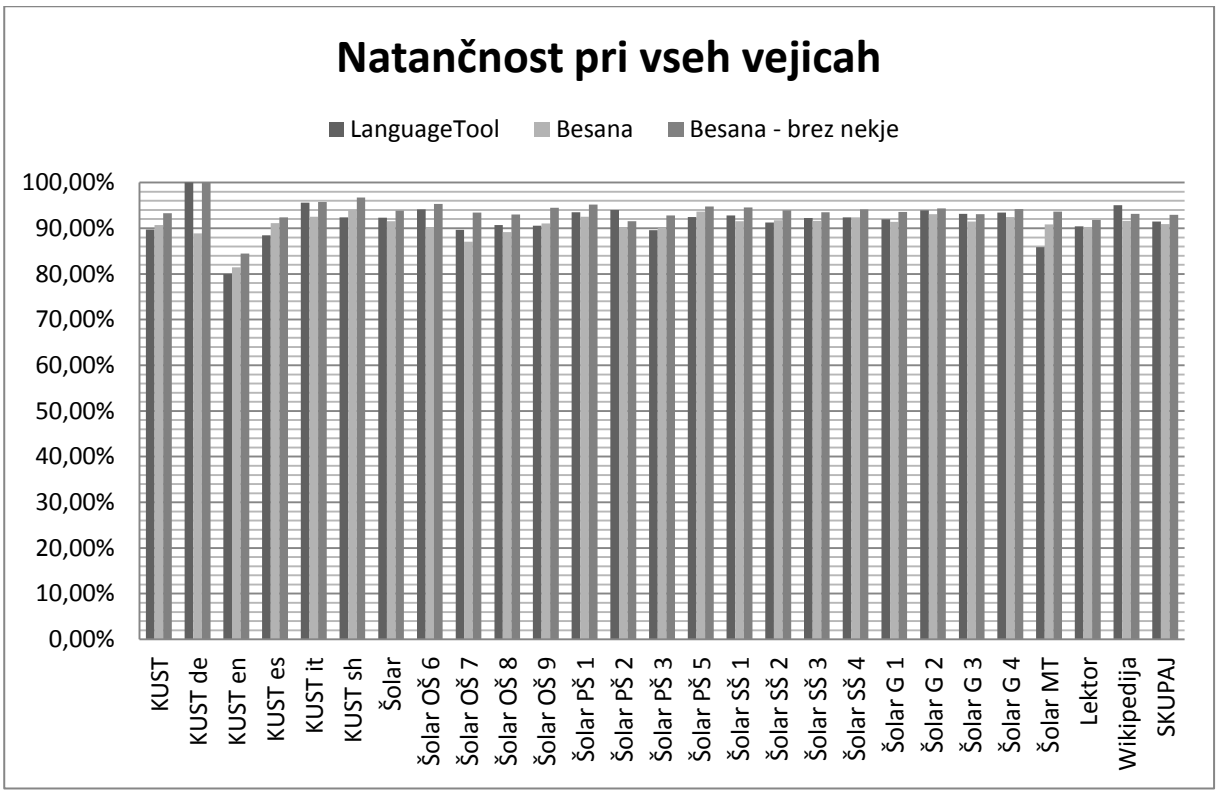
4.4 Postavljanje vseh vejic

V nekaterih primerih želimo v besedilo postaviti vse vejice. To je recimo koristno pri razpoznavi govora, kjer iz samega govora ni mogoče določiti mest, kjer bi morale stati vejice (Huang in Zweig 2002). Zato je bilo zanimivo preizkusiti Besano in LanguageTool tudi pri tem problemu, čeprav programa v bistvu nista namenjena reševanju tega problema (ne eden ne drugi npr. ne opozarjata na manjkajoče vejice v naštevanju).



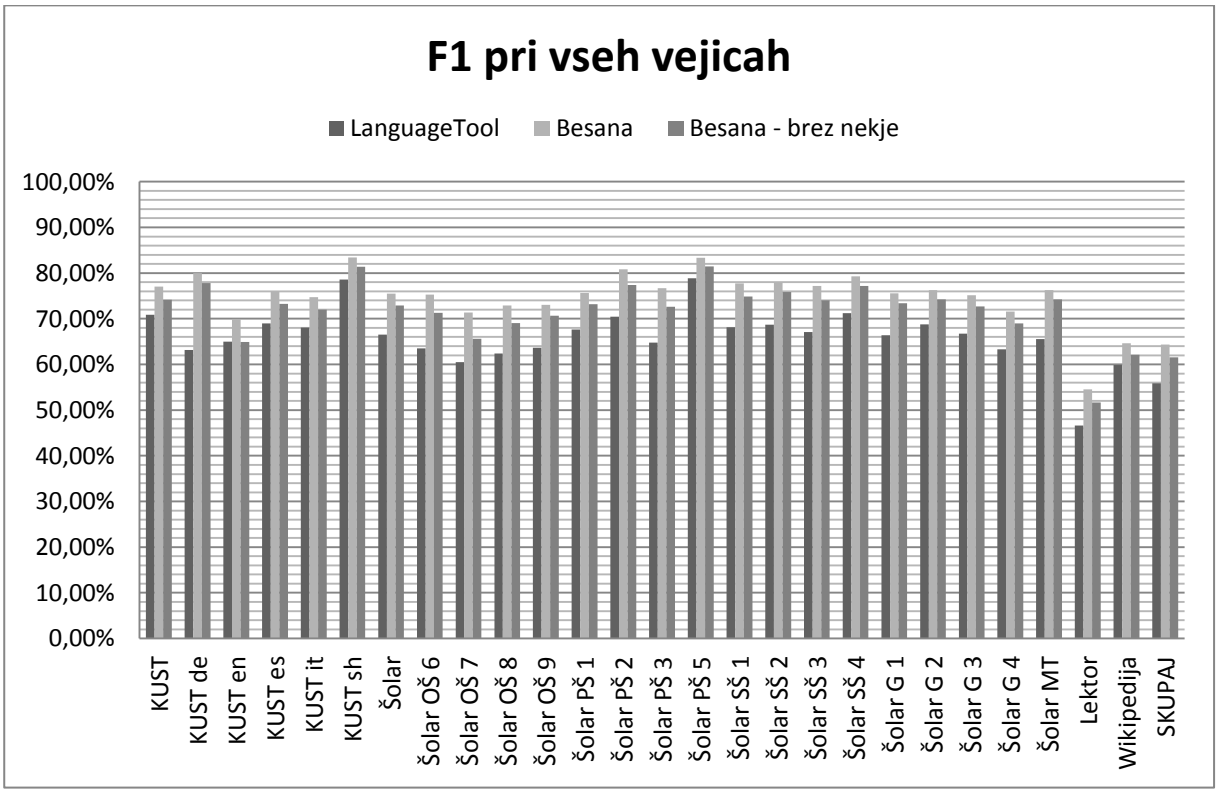
Slika 5: Priklic vseh vejic – dosedanji rezultati

Slika 5 prikazuje priklice LanguageTool in Besane pri postavljanju vseh vejic. V povprečju so rezultati podobni kot pri iskanju manjkajočih vejic, je pa rezultat boljši pri delu iz korpusa Lektor.



Slika 6: Natančnost pri iskanju vseh vejic – dosedanji rezultati

Slika 6 prikazuje natančnost pri iskanju vseh vejic, rezultati so zelo dobri in tudi zelo konsistentni po delih in poddelih (razen poddelov iz korpusa KUST, ki pa so zelo majhni).



Slika 7: F1 pri iskanju vseh vejic – dosedanji rezultati

Slika 7 prikazuje še rezultat F1 za problem iskanja vseh vejic. Rezultati so podobni kot pri iskanju le manjkajočih vejic, manj pa izstopa rezultat za del iz korpusa Lektor. Vendar pa moramo vseeno upoštevati, da je pri iskanju vseh vejic v istem besedilu kljub vseeno številčno več manjkajočih vejic, ker je pri iskanju le manjkajočih vejic teh že na začetku le manjši del vseh vejic, saj je večino vejic napisal že pisec besedila (v primerih iz korpusa Lektor manjka le 1,18 % vejic, tako da tudi pri najdenih le 20 % manjkajočih vejic je na koncu več kot 99 % postavljenih). Iz tega sledi, da je pri postavljanju na koncu npr. v povprečju postavljena v besedilo le polovica potrebnih vejic, kar ni rezultat, s katerim bi lahko bili zadovoljni.

4.5 Analiza težav pri iskanju napak pri postavljanju vejic

Ker se je pri evalvaciji iskanja napak pri postavljanju vejic pokazalo, da imata tako LanguageTool kot Besana v mnogo primerih še težave pri popravljanju vejic, se je smiselno vprašati, ali gre te napake morda razvrstiti v kategorije. Preprosta metoda je razvrščanje glede na okoliške besede, bolj napredna pa to, da vsem vejicam v besedilu določimo razlog, zakaj so na tistem mestu, in potem razvrstimo napake glede na te razloge.

4.5.1 Okoliške besede

Ker pravila v LanguageTool delujejo predvsem na nivoju besed in sosednjih besed, je bilo zanimivo preizkusiti, v katerih primerih ima LanguageTool (in potem tudi Besana, ki delno prav tako uporablja taka pravila) največ težav. V ta namen je program Vejicar dopolnjen tako, da naredi dodatno izhodno datoteko, kjer za vsako mesto, kjer potencialno lahko pride vejica, napiše vrstico, kjer napiše besedo za tem mestom, stanje na tistem mestu in okolico mesta (dve besedi spredaj in dve zadaj).

Tabela 4: Možna stanja pri analizi napak

stanje	pomen stanja
1	v zbirki manjka vejica, program je to ugotovil
2	v zbirki manjka vejica, program tega ni ugotovil
3	v zbirki je odvečna vejica, program je to ugotovil
4	v zbirki je odvečna vejica, program tega ni ugotovil
5	v zbirki je vejica, program jo je pustil

6	v zbirki je vejica, program meni, da je odvečna
7	v zbirki je presledek, program meni, da je tako v redu
8	v zbirki je presledek, program meni, da na tem mestu manjka vejica

Tabela 4 prikazuje možna stanja, soda stanja so napake, liha pa so na mestih, kjer program pravilno označi stvari.

Primer rezultata v izhodni datoteki je naslednji:

*da 7 Ne ***da sovražim, da*
*sovražim 7 Ne da ***sovražim, da ljubim¤*
*da 5 da sovražim,*** da ljubim¤*
*ljubim 7 sovražim, da ***ljubim¤ sem na*
*sem 2 da ljubim¤*** sem na*
*na 7 ljubim¤ sem ***na svetu.*
*svetu 7 sem na ***svetu.*

Gre za poved »Ne da sovražim, da ljubim¤ sem na svetu.«, kjer v zbirki primerov pred »sem« manjka vejica, česar LanguageTool ne zmore ugotoviti.

V naslednjem koraku je bil napisan postopek, ki izhodno datoteko z okolicami uredi po abecedi (da pridejo skupaj vse pojavitve iste besede). Potem program za vsako besedo prešteje, kolikokrat se pojavi kakšno stanje, in rezultat izpiše v naslednjo datoteko (izpiše le besede, ki se pojavijo vsaj petkrat, saj redkejše besede samo po nepotrebnem večajo preglednico, ker za tako redke besede ni smiselno pisati pravil), kjer je vsaka beseda v svoji vrstici, potem pa sledijo števila posameznih stanj, ločena s tabulatorjem. Ta datoteka se preprosto skopira v preglednico v Excel, ki potem omogoča urejanje po posameznih stolpcih, s čimer lahko dobimo besede, pred katerimi je največ napak.

Vendar pa na vejice vplivajo tudi sosednje besede, po eni strani naslednje besede zaradi stalnih zvez (»in sicer«, »medtem ko«) oz. zaradi zaporednih veznikov (»in če«, kjer lahko vejica zaradi »če« stoji pred »in«), po drugi strani pa spredajšnje besede (npr. drugi vezniki ali pa členki). Zato se je postavilo vprašanje, kako omogočiti, da bi bilo možno gledati te podatke in se glede nanje odločati o novih pravilih. Podobnim iskanjem so namenjeni konkordančniki, zato sem se odločil, da uporabim katerega od njih. Velik del konkordančnikov je zdaj namenjen rabi na spletu, kar je bilo za namen naloge nerodno; treba je bilo torej poiskati konkordančnik, ki deluje kot samostojen program in omogoča vsa potrebna iskanja. Najprej sem preizkusil brezplačni program AntConc[17], vendar se je

[17] http://www.laurenceanthony.net/software/antconc/

pokazalo, da ima z večjimi količinami podatkov težave. Potem sem preizkusil WordSmith[18], in ta je deloval odlično. Pri tem programu je sicer težava, da je plačljiv, ker pa brezplačna predstavitvena verzija prikazuje prvih 25 rezultatov, je bilo to za potrebe te analize povsem dovolj (še posebej ker že v Excelu poiščemo kandidate, ki jih želimo preverjati).

Analiza napak iz okoliških besed je bila uporabljena za izboljšanje programa LanguageTool v točki 5.2.

4.5.2 Program za določanje vrst vejic

Analiza okoliških besed lahko navede le del razlogov za vejice, saj vejice niso vezane na besede, ampak na strukturo povedi. Ročno pregledovanje primerov, kot je npr. opravljeno v (Šek Mertük 2001), je zamudno delo (ker nas zanima tudi relativno število težav, je treba označiti tudi pravilne primere), kar še posebej velja, če želimo za primerjavo označiti tudi vse pravilno postavljene vejice. Zato je bilo smiselno poskusiti, ali se da narediti program, ki bi vejicam določil vzrok, zakaj so na določenem mestu, da bi se dalo potem videti, na katerih mestih je največ manjkajočih (oz. odvečnih vejic).

Delovna verzija programa je uporabljena v (Holozan 2015), kjer so uporabljeni tudi nekateri opisi delovanja iz te točke.

4.5.2.1 Analizator povedi

Uporabljen je bil analizator podjetja Amebis, ki prevede poved v Amebisov vmesni jezik, iz katerega je mogoče razbrati strukturo povedi oz stavkov. Analizator sicer še ne zmore analizirati vseh povedi, zato je bilo treba ugotoviti tudi, kateri primeri mu delajo težave in katere bi veljalo popraviti – v ta namen program naredi seznam povedi, ki jih analizator ni uspel analizirati. V prvem poskusu je bilo takih povedi 49.039 (od skupaj 113.308, torej 43 %), zato so bile potem pregledane, da bi se ugotovilo, ali bi bilo mogoče analizator kaj dopolniti, da bi jih zmogel analizirati več.

Nekaj primerov tistega, česar analizator ni uspel analizirati:

Vanjo se je zateklo približno 900 ljudi, ki so jo branili polnih deset mesecev.
Lahko povzroči apoptozo (celično smrt) rakavih celic.
Linhart je zraven vključil rokokojske like¤ in sicer Jerico in Tončka.
Linhart jo je skrajšal, preoblikoval ter jo vstavil v slovenski jezik.
Linhart je komedijo Matiček se ženi napisal z nekim razlogom.
In kdor to zmore¤ je pravo drevo v hudourniku, ki upogiba veje.
Usoda je popolnoma odvisna od njih samih.
Važno je priti na ”grič”.

[18] http://www.lexically.net/wordsmith/

(razpravljalni esej)
Anton je nekdanji španski borec, ki umre absurdne smrti.
Mora se razosebiti in ubijati.
Ni pomembno¤ koga.
Zadnji junak romana Senčni ples pa je Silvan Kandor.
Če se Berk ne bi odločil ubijati in za vsako ceno preživeti, bi ga vojna pokopala.
Težko se mu je odločiti.
Sama pa svoje krivde ne prizna niti ne zanika.
V njemu se odvija notranji boj, ki ga žre.
Vsi ljudje kdaj pa kdaj potujemo.
Nežki je pomagal tudi Matiček, baronici pa študent, ki ga je finančno podpirala, a sta le sledila navodilom.

Vmes so se pojavile tudi povedi, pri katerih so notri take napake, ki analizatorju onemogočajo pravilno analizo:

V zgodbi na Cankar pokaže, da revež vedno ostane revež.
Skoraj polovica otrok s cerebralno paralizo je nedononedonošenčkov.
Ljudje so se nerazumno in hoteli Jermana dobiti v roke.
Tu je prihajalo za tekmovalnost med dvem oblastem v državi.
Vsaka kitici se začne z nedoločnikom.
Namesto poroko je bila tragedija.
Ko sta se spozna, sta se zaljubila.
Vem¤ da je Miha dovolj močna oseba in bi se lahkouprel očetu in izrazil želje in cilje.
Osebo je vedno potrebno spoznati, predenj je prepozno.
Bogovi ga niso marali, ker je vse naučil ljudje.
Noč je bil dolga in nemirna.
Odlomek govori o vojni med krščansko in poganska vero.
Čedalje več je neporočenih ljudeh.
Seboj pa bi še vzela Bronjinega sina Inga.
Našo mrtvi kolegi so ga častno prestali.

V teh primerih lahko poskusimo narediti analizator bolj odporen na napake, ni pa to vedno možno.

Poseben problem je s stvarmi, ki jih pravzaprav ni mogoče smiselno analizirati:

*mis < *mens 'isto'*
**meh1s(e)nko-) > stcslovan.*
mesęc?'mesec'
hrv. mj?sec 'isto'
šć (npr. plašćeniceju, stišćenima, nišćim, otpušću), gr. g' > čak.
[6] Tu habites le quartier?

Pri teh bi bilo smiselno razmišljati o tem, da bi se jih pobrisalo iz zbirke.

Pokazali pa so se tudi primeri, kjer gre za napake v zbirki, ki jih je smiselno popraviti za naslednjo verzijo:

Oče mu v čast, pripravi veselico.
Ampak poznam in tudi moja dva bratranca¤ imata podobno zgodbo.
Barva značilna za dečke pa je modra.
Prepovedana nikoli mogoča in nasploh razveljavljena.

Postaven, močen,, bogat.

V nekaterih primerih gre tudi za napačen razrez na povedi, in tudi to bi veljalo popraviti v zbirki:

Ima ga. In jaz jo razumem.
Za zgodovino glotodidaktike je zanimivo njegovo delo Discours physique de la parole (1668, slov.
Razprava o fiziki govora), v katerem je pojasnil svoje poglede na naravo govora.

Stavčni analizator je bil glede na te problematične primere dopolnjen. Po eni strani so bile dodane manjkajoče besede (predvsem lastna imena, manjkali pa so prej tudi prislovi, ki so enaki oblikam pridevnikov, npr. »začudeno«, »mesarsko« in »zvito«), dopolnjene oz. dodane so bile glagolske predloge (npr. pri »umreti« je bila dodana možnost »umreti {PR2}«, kar pomeni, da se glagol »umreti« lahko veže z rodilnikom, s čimer pokrijemo primere tipa »umrl je absurdne smrti«). Dopolnjeno je bilo tudi avtomatizirano iskanje tipkarskih napak za primere, ko sta zlepljena predlog in samostalnik, ki sledi (»zaraditega«, »stem«). Pri skladnji je bila dodana podpora za povedkov prilastek (»Tudi Grki niso od nekdaj telovadili goli.«), osebke, ki so namesto v običajni tretji osebi v prvi osebi (»Vsi ljudje kdaj pa kdaj potujemo.«), deležniška vezljivost (»To je bila kratka slovnica italijanščine s približno štiridesetimi dialogi, zapisanimi v dveh stolpcih, v italijanščini in angleščini.«), podpora za narekovaje (»Važno je priti na ”grič”.«) in podpora za oklepaje.

Popravki analizatorja povedi so od 219 izbranih neanaliziranih povedi poskrbeli za uspešno razrešitev 126 povedi; ostalo jih je še 93. V celotni zbirki se je število neanaliziranih povedi zmanjšalo na 40.816, torej na 36 %, izboljšanje je zneslo 7 odstotnih točk oz. dodatno je bilo zanaliziranih okoli 17 % povedi, ki prej niso bile.

Analizator povedi je možno še dopolniti, predvsem pri izpustitvah in napakah (ker druge napake lahko zmotijo analizator, ki potem zaradi tega ne popravi pravilno vejic). Vendar to presega obseg tega dela, zato se bomo zadovoljili s trenutnim stanjem in delali z analizami, ki jih imamo na voljo (v Amebisu sicer analizator povedi ves čas po malem dopolnjujemo), dodatno pa uporabiti še metode, ki ne potrebujejo analize povedi.

4.5.2.2 Dopolnjevanje programa

Najprej smo naredili osnovno verzijo programa, ki je poleg označevanja vrst vejic dodatno naredila še seznam povedi, ki niso bile uspešno analizirane (ker je bilo uporabljeno za dopolnjevanje analizatorja povedi), in seznam neoznačenih vejic, v katerem so posebej označene tiste, ki so bile spregledane v sicer uspešno analiziranih povedih. Ti primeri kažejo,

da moramo program ustrezno dopolniti bodisi s tem, da ustrezno upoštevamo analizo, bodisi s tem, da dodamo še potrebne nove oznake za vrste vejic.

Po popravkih program še enkrat uporabimo na zbirki primerov in ga tako iterativno izboljšujemo. Popolnega označevanja sicer ne moremo doseči (predvsem nas omejuje analizator povedi); naš cilj je, da razvrstimo vsaj tri četrtine vejic.

4.5.2.3 Oznake vrst vejic

Oznake so sestavljene tako, da je na začetku velika črka, ki označuje osnovno vrsto vejice. Temu sledi števka, ki označuje podvrsto, tej pa bi lahko sledila še mala črka, ki bi podrobneje določila vrsto vejice (vendar to za zdaj še ni bilo uporabljeno). Klasifikacija je bila narejena po razdelitvi v (Šek Mertük 2011), ki izhaja iz pravil v Slovenskem pravopisu 2001.

Tabela 5: Oznake vrst vejic

oznaka	vrsta	primer
A	nestavčni enakovredni deli	
A1	naštevalna	Večja slovenska mesta so Ljubljana$A1$, Maribor$A1$, Celje.
A1a	*naštevalna vezalna*	
A1b	*naštevalna ločna*	
A2	protivni	Fant je malo počasen$A2$, a zanesljiv.
A3	*vzročni*	
A4	*posledični*	
A5	*pojasnjevalni*	
A6	*ločni*	
B	pristavčni deli	
B1	samostalniška fraza	Pri Prešernu$B1$, največjem slovenskem pesniku$B1'$, so se učili vsi kasnejši pesniški rodovi.
B1a	*ni ime v pristavku*	
B1b	*je ime v pristavku*	
B2	*nestavčno/stavčni*	
C	stavčno enakovredni deli	
C1	vezalno priredje	Pleše, kot da je ponorela$E3'$$C1$, in poje na ves

		glas.
C2	*stopnjevalno*	
C3	ločno	Umreti pogumno, za svoja načela$C3$, ali živeti notranje mrtvi.
C4	protivno	Vse ima$C4$, vendar ni zadovoljen.
C5	vzročno	Odhiteli so pospravljat seno$C5$, kajti pripravljalo se je k nevihti.
C6	posledično/sklepalno	Cesta je bila mokra$C6$, zato so avtomobili vozili počasi.
C7	pojasnjevalno	Veliki kazalec ure v stolpu je kazal na dvanajst$C7$, in sicer je bila ura poldne.
C8	vezalno brez veznika	Pisal je pesmi$C8$, bral povesti$C8$, užival je v lepi domači besedi.
D	prilastkov odvisnik	
D1	ki	Prebral sem knjigo$D1$, ki je bila napisana lani.
D2	kateri	Knjiga$D2$, katera je bila napisana lani$D2'$, je bila zelo zanimiva.
D3	predlog + kateri	Koruza pokovka skače iz ponve$D3$, v kateri jo grejemo.
D4	deležniški –n/-t	Prebral sem knjigo$D4$, napisano lani.
D5	deležniški -č	Človek$D5$, sloneč ob zidu$D5'$, si je prižgal cigareto.
D6	pridevniški	Pije napitek$D6$, poln vitaminov.
E	odvisniki	
E1	krajevni	Z družino smo se vračali domov oz. v apartma$E1$, kjer smo bivali.
E2	časovni	Prišel sem le$E2$, ko je deževalo.
E3	načinovni	Vsak je dobil$E3$, kakor si je zaslužil.
E4	vzročni	Gore ljubimo$E4$, ker so lepe.
E5	*namerni*	
E6	*primerjalni*	
E7	osebkov/predmetni/povedkov	Obljubila je$E7$, da bo delo opravila v dogovorjenem času.

		Njegova želja je$E7$ da bi postal dober obrtnik. Novica$E7$ da je zmagalo naše moštvo, nas je zelo razveselila. Zdi se$E7$ da bo deževalo.
E8	pogojni	Največja sreča za človeka je$E8$, če je zdrav.
E9	dopustni	Prav zato je bil večkrat sumljiv za oblast$E9$, čeprav ji ni škodil.
F	polstavki	
F1	deležje	V Nemčijo je odšel$F1$, hrepeneč po bogastvu.
F2	*nedoločnik*	
F3	*samostalnik*	
G	izpostavki/dostavki	
G1	*izpostavek*	
G2	*dostavek*	
G3	zvalnik	Miha$G3$, pridi sem!
G4	medmet/členek	Da$G4'$$C8$, vse je čisto.
G5	kajne	Jutri prideš$G5$, kajne?
H	govor	
H1	premi govor	»Prišla bom jutri$H1'$,« mu je rekla.
H2	odvisni govor	Bila je pogumna in je točno vedela$H2$, kaj hoče.
H3	posamezna vprašalnica	Ni pomembno$H3$, koga.
H4	odvisni govor z naj ...	Rekel je$H4$, naj pridem v šolo.

Tabela 5 vsebuje oznake vrst, uporabljena pa so potem še dodatna imena, tako npr. E? označuje vse odvisnike od E1 do E9, če nas ne zanimajo rezultati za vsak odvisnik posebej. Ležeče so napisane oznake, ki še niso podprte.

Opuščaj (') na koncu označuje, da gre za zaključek elementa; E?' tako zaključuje konec odvisnika. Če sta na istem mestu dva razloga, se oznaki sestavita. D1'E? tako pomeni, da gre za konec prilastkovega odvisnika s »ki«, ki mu takoj sledi drug odvisnik (npr. »Miha, ki je prišel domov$D1'$$E?$, ko je začelo deževati«). Z znakom + so označene situacije, ko kaki oznaki neposredno desno sledi druga oznaka (npr. »Miha bo prišel$C1$+$E?$, in če ne bo deževalo, pokosil travo.«). V označenem besedilu so oznake obdane z znakoma $, da se laže ločijo od besedila.

4.5.2.4 Način dela programa

Program za določanje vrst vejic uporablja rezultat analizatorja povedi, ki je zapisan v Amebisovem vmesnem jeziku (Holozan 2011), iz katerega je mogoče razbrati tudi strukturo povedi.

Program naredi tabelo razlag za vsako besedo in ločilo in jo napolni s praznimi nizi. Potem gre po analizah stavkov, ki so del povedi, in dodaja nize za začetek oz. konec posamezne vrste.

Za odvisnike se pogleda glava analize stavka. Če se npr. začne z »(-POV:(-STAj«ali »(-POV:(=STAj«, se doda na začetek stavka oznaka »$E8$«, za besedo, ki sledi stavku (in vsem morebitnim odvisnikom, ki so del stavka), pa oznako »$E8'$«.

Za priredja se pogleda, ali obstaja pred njim element »(-VEZp« in vejica »(-LOCzv«. Vezniški element vsebuje tudi podatke o vrsti priredja, konec pa se pri priredjih ne označuje.

Podobni postopki so sprogramirani tudi za druge vrste vejic, ki se označujejo.

Na koncu se oznake izpišejo le v primeru, če so na vejici, čeprav so lahko pripisane tudi drugim besedam in ločilom. Pogledajo se še morebitne oznake sosednjih besed in dodajo s + oz. -.

4.5.2.5 Evalvacija programa

Program sicer razvrsti 78 % vejic, vendar je vprašanje, koliko so te razvrstitve napačne (ker npr. lahko že analizator napačno razume poved). Zato smo ročno preverili 100 naključno izbranih primerov razvrščenih vejic glede tega, ali so vejicam pripisane prave oznake.

Preverjanje je pokazalo, da je 83 oznak ustreznih, 17 pa neustreznih. Neustrezni primeri so na primer naslednji:

Za mestni tloris velja, da je zelo trdoživ element morfološke zgradbe mesta$E7'$$D1$, ki se ohranja skozi različna časovna obdobja in razvojne faze mesta. – Tukaj se odvisnik E7 v resnici še nadaljuje do konca prilastkovega odvisnika.

In tukaj pridemo do vprašanja$C3$, ali naj se Antigona drži državnih zakonov in si s tem ne ogrozi svojega življenja ali naj svojega brata pokoplje in mu s tem omogoči, da po njihovemu verovanju pride njegova duša na cilj. – Tukaj se pokaže, da manjka možnost, da je vprašanje desni prilastek samostalnika »vprašanje«, zato se tukaj napačno analizira kot ločno priredje, kar bi potem pomenilo tudi, da bi bila ta vejica odveč. Sorodni so primeri, kjer so vprašanja v vlogi prilastkovega odvisnika pri kazalnih zaimkih:

O tem, kaj bi na njihovem mestu in v njihovem položaju storila jaz$C8$, je zelo težko govoriti.

dolge generacije na Trški gori svoj vinograd, vinograd z zelo lepo zidanico in lahko rečem, torej, kar sem jaz dožvela, tega$A1$-$E7$, kar je …

Sama se prilagodim življenju, a sledim svojim ciljem in prepričanju, med tem ko Antigona sledi le svojim načelom in naredi le tisto$E2'$$E2$, kar misli, da je po njenem najbolje.

Kar nekaj težav je tudi pri detektiranju pristavkov, ki se zamenjujejo z naštevanjem (razmisliti bo treba o možnosti, da se pri odločanju upoštevajo členki in pridevniki, ki kažejo na večjo verjetnost pristavkov, kakršni so: *recimo, denimo, na primer, npr., med drugim, tako imenovan*), taki primeri so:

V besedilu govori o tem, kako se je nad mestom pojavil velik črni oblak$H2'$$C8$, tako imenovani smog, ljudje pa se želijo rešiti.

Zaradi številnih vojaških in trgovskih stikov se je latinščina razširila po vseh delih cesarstva_ imperija$C8$, na celotno območje Sredozemlja, govorili so jo v večjem delu antične Hispanije, v Galiji, današnji Italiji, Severni Afriki, na Balkanu, v delih Azije, obvladovala je tretjino sveta, ki ga je poznal antični človek.

"bzw. beziehungsweise ? oziroma'; oz. f. in ff. folgende?sledeči' (f. pomeni, da se misel nadaljuje na eni strani, ff. pa, da se nadaljuje na več straneh; v slovenščini$A1$, npr. v ESSJ, ustrezno uporabljamo s. in ss.)Hg. ali Hrsg."

Zlasti višje vodstvo se mora zavzeti za izobraževanje in spodbujanje zaposlenih za podporo okolju prijaznih praks$A1$, med drugim ločevanja in zbiranja odpadkov.

V marsikateri afriški državi so v zadnjih letih namenjali več proračunskih sredstev za vojsko kot za zdravstvo, kar je predvsem problematično v državah z repa lestvice indeksa človekovega razvoja$E3$, kot so DR Kongo, Čad, Srednjeafriška republika, Sierra Leone in podobne.

V nekaterih primerih ni bil pravilno najden prilastkov odvisnik s *kateri*, kadar je rabljen pridevniško:

Nemci nam nimajo ničesar odpustiti, saj neposredna krivda za povzročitev 2. svetovne vojne in njen zverinski potek izključno pripada nemškemu imperializmu in fašizmu$C8$, katerih naslednik je Nemška federalna republika.«

Aramejščina, eden od semitskih jezikov$C8$, katerega zgodovina je trajala celih tri tisoč let, je bil jezik verskih obredov, jezik administracije velikih imperijev, v njem so bili napisani Talmud in tudi več knjig Svetega pisma.

Pokazalo se je, da je treba analizator povedi še nekoliko dopolniti, kritični so predvsem prilastkovi odvisniki, kadar se v tej vlogi pojavljajo drugi odvisniki in vprašanja. Naslednja pomembna dopolnitev je pridevniška raba zaimka *kateri*, še zadnja pa določanje pristavkov s pomočjo tipičnih členkov. Odločil sem se, da analizator povedi dopolnim s temi tremi dodatki in da se potem še enkrat označijo vrste vejic in ponovi evalvacija označevanja. Dopolnitve analizatorja povedi in samega programa za določanje vrste vejic so rešile 11 težav, 6 pa jih je ostalo.

Pri ponovnem preizkusu so se pokazale še nekatere težave, ki jih je veljalo popraviti pred nadaljevanjem. Najprej primer, ko se je pokazalo, da analizator prevečkrat pričakuje izpust (elipso) veznika, v tem primeru »ko« pred »svoje matere …«, stavek torej razume kot bi se glasil »Ko je prispel do mesta in do hiralnice, ko svoje matere ni hotel še poslednjič videti, le odšel je do ravnatelja, da bi uredil vse potrebne formalnosti.«:

Ko je prispel do mesta in do hiralnice, svoje matere ni hotel še poslednjič videti$E2'$, le odšel je do ravnatelja, da bi uredil vse potrebne formalnosti.

Pokazalo se je, da analizator daje previsoko prioriteto možnosti, da sta na isti samostalnik vezana dva prilastkova zaimka.

Na tablo nalepite prazno polo papirja in določite nekoga od udeležencev, da na tablo zapisuje asociacije$E7'$, ki jih skupina poveže z vodo.

Največ težav pa je delalo ločevanje pristavkov od naštevanja, pokazalo se je, da je treba pri tej analizi biti pozoren na vejico pred »in« in pojasnjevanje osebnih imen.

Klavdij je tisti, ki ubije kralja, Hamletovega očeta$C1$, in se poroči z njegovo mamo.

Ko je truplo sina videla Evridika$E2'$$C8$, Kreonova žena, ga je preklela, odšla v svojo spalnico in se zabodla.

Ali so v kampanjah$A1$, trženju, prisotni sprožilci, ki ciljajo neracionalne motive?

Ponoven preizkus, kjer je bilo spet izbranih 100 označenih vejic, je pokazal, da je bilo tokrat napačno označenih 18 vejic (rezultat je slabši od začetnega, ker je precej odvisen od sreče pri naključni izbiri primerov, ta vpliv bi lahko zmanjšali tako, da bi delali na večjem številu izbranih primerov (npr. 1000)). Vsi ti popravki analizatorja so izboljšali tudi uspešnost le-tega, in sicer se je število neanaliziranih povedi zmanjšalo na 39.484 (s 40.816).

Napake pri označevanju so bile raznolike, kažejo se težave, ki jih analizatorju naredijo izpusti, npr.:

Da je najbolje, če si na terenu pišeš slovenska imena$A1$, doma pa latinska imena.

Prvi je napisan epsko$A1$, drugi in tretji pa lirsko, saj sta ta dva dela v obliki pesmi.

Po drugi strani se pojavljajo primeri, ko analizator napačno predvidi izpust veznika:

Ko sem šla s staršoma v nakupovalni center$E2'$$E2$, sem srečala Sonjo, prvo jo sploh nisem spoznala, potem mi je postajala vsebol znana.

Zanimiv primer napake je še:

Vegan, ki zavrača vse$D1'$, kar povzroča mučenje živali.

Do te napake je prišlo zato, ker je analizator dal prednost stavku pred samostojnim osebkom, zato je kot osnovni stavek razumel »Vegan kar povzroča mučenje živali.«

Kaže pa se, da ni kakšnih napak, ki bi se pojavljale zelo množično, ampak gre za veliko število različnih malenkosti, s katerimi ima analizator težave, zato ni mogoče hitro doseči bistvenega izboljšanja. Potemtakem je po pameti uporabiti program za razvrščanje takšen, kot je bil narejen, nekatere od bolj motečih napak pa potem popraviti pri samem popravljanju vejic.

5 Izboljševanje metod

Poglavje 4 je pokazalo, kako uspešni so dosedanji programi za postavljanje vejic v slovenščini. Ker pa menimo, da je mogoče te rezultate izboljšati, bomo v tem poglavju poiskali načine, kako to doseči.

5.1 Preizkus strojnega učenja

Na področju jezikovnih tehnologij se vedno bolj uveljavljajo statistične metode oz. metode strojnega učenja iz podatkov. Velika prednost teh metod je, da jih je lažje prenašati med jeziki, če le imamo na voljo podatke.

Ta raziskava je objavljena v obliki članka (Holozan 2013), vendar mu ni identična. Večji del tega poglavja je resda enak članku, je pa tam uporabljena starejša verzija primerov iz korpusa Šolar, ki je vsebovala veliko napak (vsebovala pa je le povedi, v katerih so bile označene kakšne napake pri vejicah), zato sem poizkus ponovil na delu zbirke primerov rabe vejice iz korpusa Šolar, primerjalno pa še na delu zbirke primerov iz korpusa Lektor.

5.1.1 Zasnova poskusa

Namen poskusa je bil preizkusiti metode strojnega učenja v slovenščini, in sicer najprej za problem postavljanja vseh vejic (ki mu je bila doslej namenjena večina tujih raziskav in ki je pomemben pri razpoznavi govora), potem pa še za problem popravljanja napačnih vejic (kar je uporabno v slovničnih pregledovalnikih, ki tako pomagajo piscem besedil postavljati vejice).

Osnova ideja poskusa postavljanja vseh vejic je povzeta po Alegria et al. (2006) in je takale: uporabi se korpus s pravilno postavljenimi vejicami, s tem da se oblikoskladenjsko označi, lematizira in skladenjsko razčleni (pri čemer je treba upoštevati, da pri praktični uporabi nimamo vejic vnaprej, zato je to treba preizkusiti označevanje tudi brez vejic, na kar sta opozorila že Shieber in Tao (2003), medtem ko Alegria et al. (2006) tega niso posebej preizkušali). Vsaka beseda se z določenim okoliškim oknom pretvori v seznam atributov, doda pa se atribut, ki pojasnjuje, ali ji sledi vejica (ta atribut je potem razred pri klasifikacijskem problemu). Tako zapisane besede se uvozijo v program za strojno učenje, tam pa se izvedejo eksperimenti.

Enako kot pri Alegria et al. (2006) je bil uporabljen program WEKA, ki ima vgrajenih veliko klasifikatorjev. Preizkusili smo večje število klasifikatorjev, potem pa izbrali nekaj najboljših (pri čemer smo upoštevali še to, da so čim bolj različni), in ti so potem rabili v nadaljnjih preizkusih, kjer so bili preizkušeni različni atributi, velikost okna, vpliv označevanja in parametri klasifikatorja.

Za preizkušanje je uporabljeno 10-kratno prečno preverjanje, pri čemer se primeri razdelijo na 10 delov, devet delov se uporabi za učenje, preostali del pa za preizkušanje, kar se 10-krat ponovi z različnim delom za preizkušanje, in se potem izračunata povprečni priklic in natančnost.

Dobljene rezultate bo mogoče primerjati z rezultatoma za Besano in LanguageTool iz točke 0.

Drugi poskus sestavljata prenos ugotovitev iz prvega poskusa v popravljanje napačnih vejic ter primerjava s programoma Besana in LanguageTool iz točke 4.4. Preizkušanje v tem poskusu je namreč bolj zapleteno, zato je najboljšo kombinacijo za strojno učenje lažje poiskati pri problemu iskanja vseh vejic in jo potem uporabiti še pri popravljanju napačnih vejic.

5.1.1.1 Zbirka primerov

V raziskavi je bil uporabljen del zbirke primerov iz korpusa Šolar, dodatno pa je bil narejen poskus na delu iz korpusa Lektor. Dela KUST in Wikipedija nista bila uporabljena, ker sta majhna in bi se program tam težko naučil pravil. Možno bi bilo sicer uporabiti hkrati skupaj vse dele, vendar sem želel doseči čim večjo enotnost pri podatkih, ker to povečuje možnost uspešnega učenja. Del Wikipedija bi bilo mogoče uporabiti za preizkus, kjer bi se npr. program naučil postavljanja vejic na delu Šolar, potem pa poskusil postaviti vejice v delu Wikipedija, s čimer bi bilo možno preizkusiti delovanje postopka ob spremembi domene (je pa v obeh teh delih vejice popravljal isti označevalec, tako da je glede tega že poenotena).

Dela sta bila shranjena v besedilni format (pri tem so bili odstranjeni podatki o delu in poddelu) v formatu UTF8, kar je primeren vhod za označevalnik in lematizator.

Za problem postavljanja vseh vejic (in tudi za učenje pri popravljanju vejic) je bil korpus predelan tako, da so bile vse vejice popravljene (znaki ¤ zamenjani z vejicami, znaki ÷ pa pobrisani), s čimer smo dobili korpus s pravilno postavljenimi vejicami.

Druga možnost za postavljanje vseh vejic bi bila uporaba dela katerega od obstoječih korpusov (npr. Gigafide[19]), vendar se tu postavi vprašanje, kako natančno so lektorirana v korpuse vključena besedila. So se pa za to rešitev odločili pri večini tujih raziskav (med drugim tudi Alegria et al. (2006), ki so med drugim uporabili časopisna besedila).

5.1.1.1.1 Označevanje

Tako Hardt (2001) kot tudi Alegria et al. (2006) so eksperimentirali z označenimi korpusi, saj lahko pravilne oblikoskladenjske oznake in poznavanje strukture povedi pomagajo pri postavljanju vejic.

Zato smo tudi za slovenščino uporabili označevanje, in sicer smo si pomagali z oblikoslovnim označevalnikom in lematizatorjem Obeliksom[20] ter s skladenjskim razčlenjevalnikom[21], razvitima v okviru projekta Sporazumevanje v slovenskem jeziku[22]. Pri skladenjskem označevanju se je pojavila težava zaradi tega, ker je programu zmanjkalo pomnilnika, rešitev za to je bila, da se v okolju Java, v katerem deluje Obeliks, omejitev porabe pomnilnika poveča na 16 GB.

Pri poskusih za baskovščino in danščino ni posebej specificirano, ali so označevali korpus s pravilno postavljenimi vejicami ali z napačno postavljenimi (oz. brez vejic); zdi se, da so uporabili varianto s pravilno postavljenimi vejicami. Ker pa pravilnost vejic lahko vpliva na natančnost označevalnika (Hillard et al. 2006) in ker pri praktični uporabi (npr. popravljanju napačnih vejic v besedilu, pa tudi pri postavljanju vseh vejic) ni mogoče vnaprej imeti pravilno postavljenih vejic, smo preizkusili obe varianti označevanja, da bi se videlo, kako to vpliva na delovanje.

5.1.1.2 Ocenjevanje rezultatov

Kot je opisano v točki 4.1, so za ocenjevanje rezultatov uporabljene metrike natančnost, priklic in F1. Problem postavljanja vejic predstavimo z razredom, ki pove, ali zadevni besedi sledi vejica. V korpusu je 23.291 mest, kjer vejica mora biti, kar znese 11,1 % vseh besed, večinski razred pa je, da besedi ne sledi vejica, kakor je v 88,9 % primerov.

Program WEKA je rezultate izračunal za primer, ko ni vejice, kot tudi za primere, ko vejica je. Ker je cilj postaviti vejice v besedilo, je zanimiv predvsem rezultat pri primerih, ko

---

[19] http://www.gigafida.net
[20] http://www.slovenscina.eu/tehnologije/oznacevalnik
[21] http://www.slovenscina.eu/tehnologije/razclenjevalnik
[22] http://www.slovenscina.eu

vejica je, saj nam to pove, koliko manjkajočih vejic bi metoda odkrila. Natančnost je pomembnejša od priklica, ker pri slovničnem pregledovalniku nočemo preveč lažnih opozoril, seveda pa tudi priklic ne sme biti premajhen (manj kot 50 %, bi rekel), da je metoda uporabna, zato je pomemben tudi rezultat za F1, ki ga tudi izračunava program WEKA.

Rezultati so izračunani na besede, ker je beseda (z okoliškim oknom) element pri strojnem učenju.

Referenčna vrednost uspešnosti je rezultat, ki ga dosežeta programa, ki postavljata vejice s pomočjo pravil, kar je izračunano v točki 4.4 za problem postavljanja vseh vejic in v točki 4.3 za problem iskanja napak pri vejicah.

5.1.1.3 Priprava podatkov

Program WEKA potrebuje podatke v formatu ARFF, v katerem sledi glavi, ki vsebuje opis atributov, podatkovni del, v katerem vsaka vrstica predstavlja en primer. Rezultat označevanja besedil je v formatu XML-TEI[23], zato smo za pretvorbo napisali program v Perlu. Ta program za vsako besedo določi atribute, potem pa pri izvozu v ARFF ob sami besedi izpiše še atribute za prejšnje in naslednje besede glede na nastavitev okna (privzeta vrednost je -5+5, torej pet besed spredaj in pet besed zadaj, s čimer so začeli tudi Alegria et al. (2006)). Vejice niso besede, ampak le atribut je-vejica na besedi neposredno pred vejico. Ta atribut je potem uporabljen kot razred pri strojnem učenju.

Program za izvoz v ARFF izvozi vse atribute (razen podatka o obstoju vejice) kot nize, s čimer pa večina klasifikatorjev ne zna delati, zato jih je treba najprej spremeniti v nominalne atribute, kjer je pri definiciji atributa našteta zaloga možnih vrednosti. V ta namen je bil v programu WEKA uporabljen filter StringToNominal.

5.1.1.3.1 Atributi

Osnovni atributi za vsako besedo so oblika (sama beseda, taka, kot je napisana, npr. "mize"), lema (osnovna oblika besede, npr. "miza") in oblikoskladenjska oznaka (ali MSD (ang. morpho-syntactic descriptor), npr. "Sozer") po oblikoskladenjskih specifikacijah JOS[24], ki pove besedno vrsto, podatke o sklonu, spolu, številu ipd. Ker ločila nimajo oblikoskladenjskih oznak, jim je bila pripisana oznaka "Y", da se lahko obravnavajo enako kot besede. Neobstoječim besedam znotraj okna so bili vsi atributi nastavljeni na "*", vsak stavek je enota zase in okno ne sega na sosednje stavke.

---

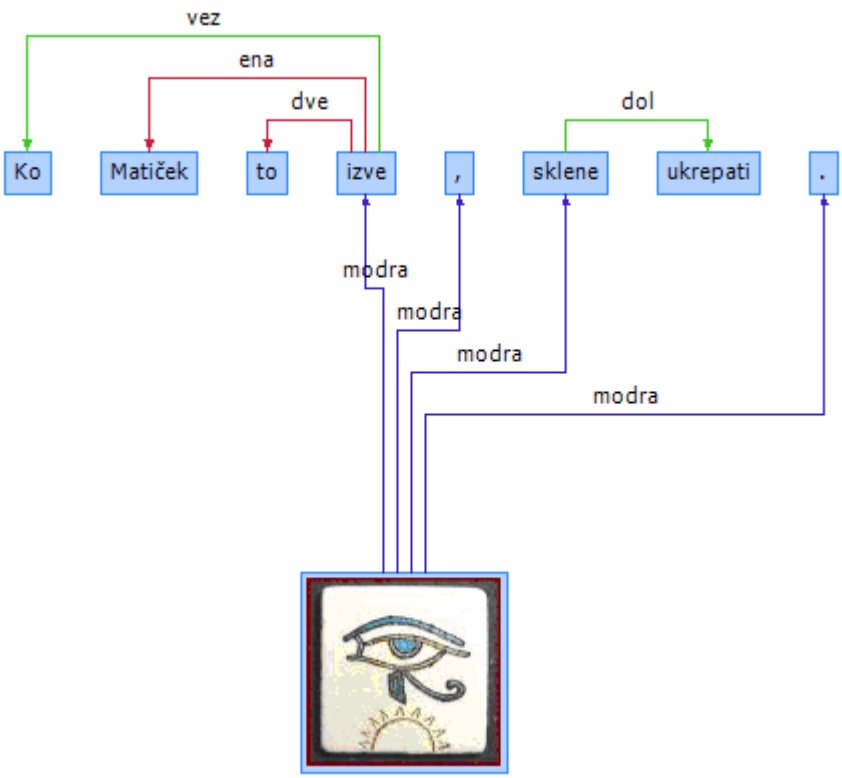
[23] http://www.tei-c.org/Guidelines/P5/
[24] http://nl.ijs.si/jos/msd/html-sl/index.html

Atributi so našteti tako, da so najprej atributi za zadevno besedo (položaj 0), temu sledijo atributi za predhodne besede (od -1 do -5) in tem atributi za naslednje besede (od +1 do +5).

Po celotnem MSD je bil narejen še poskus z delnim MSD, kjer so atributi ločeno prvi znak MSD, drugi znak MSD in pri samostalnikih še sklon.

Delni MSD2 je bil poskus, kako čim bolje prenesti informacije iz MSD in se ogniti uporabi celotnega MSD (zaradi predpostavke, da veliko število različnih MSD lahko ovira učenje). Vsak MSD bil razdeljen v dva atributa, prvi je kot prvo črko vseboval besedno vrsto, druga črka pa je bila vrsta pri posamezni besedni vrsti (pri samostalnikih, pridevnikih, glagolih, zaimkih, številnikih in veznikih). Drugi atribut je vseboval sklon pri samostalnikih, pridevnikih, zaimkih, predlogih in številnikih, sicer pa "**".

Naslednji poskus je bil uporaba podatkov skladenjskega razčlenjevalnika, kjer pa je rezultate težje pretvoriti v atribute kot pri oblikoslovnem označevalniku in lematizatorju, saj so rezultat skladenjskega razčlenjevalnika povezave, ki gradijo drevo.



Slika 8: Rezultat skladenjskega razčlenjevalnika

Slika 8 kaže rezultat skladenjske razčlembe za poved "Ko Matiček to izve, sklene ukrepati.". Za postavljanje vejic so pomembne predvsem povezave[25] "vez", ki kaže na veznike, "modra", ki kaže na del povedi, in rdeče povezave "ena", "dve", "tri" in "štiri", ki kažejo na osebke, predmete in prislovna določila, pri čemer nas pri modrih in rdečih povezavah zanima začetek bloka, zato moramo upoštevati še vse naslednje povezave, da pridemo do začetka tega bloka.

Rezultat razčlenjevalnika (skupaj z rezultatom lematizatorja in oblikoskladenjskega analizatorja) je zapisan v formatu XML, in sicer na naslednji način (izpuščene so značke "<S />", ki označujejo presledke):

```
<s xml:id="0.0">
  <w lemma="ko" msd="Vd" xml:id="0.0.1">Ko</w>
  <w lemma="Matiček" msd="Slmei" xml:id="0.0.2">Matiček</w>
  <w lemma="ta" msd="Zk-set" xml:id="0.0.3">to</w>
  <w lemma="izvedeti" msd="Ggdste" xml:id="0.0.4">izve</w>
  <c xml:id="0.0.5">,</c>
  <w lemma="skleniti" msd="Ggdste" xml:id="0.0.6">sklene</w>
  <w lemma="ukrepati" msd="Ggnn" xml:id="0.0.7">ukrepati</w>
  <c xml:id="0.0.8">.</c>
  <links>
    <link afun="vez" dep="0.0.1" from="0.0.4" />
    <link afun="ena" dep="0.0.2" from="0.0.4" />
    <link afun="dve" dep="0.0.3" from="0.0.4" />
    <link afun="modra" dep="0.0.4" from="0.0.0" />
    <link afun="modra" dep="0.0.5" from="0.0.0" />
    <link afun="modra" dep="0.0.6" from="0.0.0" />
    <link afun="dol" dep="0.0.7" from="0.0.6" />
    <link afun="modra" dep="0.0.8" from="0.0.0" />
  </links>
</s>
```

Slika 9: Zapis označevanja in skladenjske razčlembe v formatu XML

Slika 9 je primer, zapisan v formatu XML, ki je rezultat označevanja in skladenjskega razčlenjevalnika. Značke "<s" so povedi, značke "<w" besede, značke "<c" ločila in značke "<link" skladenjske povezave.

Za skladenjske atribute so bili izbrani (vrednost je 1, če je trditev resnična, oz. 0, če ni):

- je vez: beseda, na katero kaže povezava "vez";
- začetek modrega bloka: prva beseda v bloku, na katerega kaže povezava "modra";
- začetek rdečega bloka: prva beseda v bloku, na katerega kaže rdeča povezava.

[25] Vsi tipi povezav so opisani na http://www.slovenscina.eu/tehnologije/razclenjevalnik

```
'Ko','ko','Vd','1','0','1','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','*','
*','*','*','*','*','Matiček','Matiček','Slmei','0','1','0','to','ta','Zk-set','0','1','0','izve','izvedeti','Ggd
ste','0','0','0','sklene','skleniti','Ggdste','0','0','0','ukrepati','ukrepati','Ggnn','0','0','0',ni-vejice
'Matiček','Matiček','Slmei','0','1','0','Ko','ko','Vd','1','0','1','*','*','*','*','*','*','*','*','*','*','*','*','*'
,'*','*','*','*','*','*','*','*','*','*','*','to','ta','Zk-set','0','1','0','izve','izvedeti','Ggdste','0','0','0','sklen
e','skleniti','Ggdste','0','0','0','ukrepati','ukrepati','Ggnn','0','0','0','.','.','Y','1','0','0',ni-vejice
'to','ta','Zk-set','0','1','0','Matiček','Matiček','Slmei','0','1','0','Ko','ko','Vd','1','0','1','*','*','*','*','
*','*','*','*','*','*','*','*','*','*','*','*','*','*','izve','izvedeti','Ggdste','0','0','0','sklene','skleniti','Ggd
ste','0','0','0','ukrepati','ukrepati','Ggnn','0','0','0','.','.','Y','1','0','0','*','*','*','*','*','*',ni-vejice
'izve','izvedeti','Ggdste','0','0','0','to','ta','Zk-set','0','1','0','Matiček','Matiček','Slmei','0','1','0','
Ko','ko','Vd','1','0','1','*','*','*','*','*','*','*','*','*','*','*','*','sklene','skleniti','Ggdste','0','0','0','ukr
epati','ukrepati','Ggnn','0','0','0','.','.','Y','1','0','0','*','*','*','*','*','*','*','*','*','*','*','*',je-vejica
```

Slika 10: Zapis začetka zgornjega primera v formatu ARFF z oknom -5+5

Slika 10 kaže, kako je začetek zgornjega primera zapisan v formatu ARFF, ki ga zna brati programski paket WEKA.

5.1.2 Preizkušanje

Za problem, ko je treba postaviti vse vejice besedilu, je bilo narejeno večje število preizkusov, da bi se našla najboljša kombinacija klasifikatorja, atributov, velikosti okna, načina označevanja in parametrov klasifikatorja.

Zaradi velikega števila možnih kombinacij ni bilo mogoče preizkusiti vseh možnih kombinacij, ampak se je po posameznih delnih preizkusih ožil izbor (na podlagi natančnosti in delno tudi F1 na mestih, kjer so vejice), katere kombinacije je najbolj smiselno preizkušati naprej.

5.1.2.1 Izbiranje klasifikatorja in vpliv velikosti korpusa

Preizkušeno je bilo večje število klasifikatorjev, ki jih podpira program WEKA, vsi so bili uporabljeni s privzetimi parametri.

Tabela 6: Vpliv klasifikatorja in velikosti; Šolar, celoten MSD, brez skladenjskih atributov, okno -5+5

		ni vejice			je vejica		
	klasifikator	natančnost	priklic	F1	natančnost	priklic	F1
100 %	**ADTree**	0,955	0,989	0,971	0,820	0,525	0,640
	HyperPipes	0,913	0,992	0,951	0,307	0,035	0,062
	AdaBoostM1	0,944	0,991	0,967	0,822	0,409	0,546
	RBFNetwork	0,966	0,973	0,970	0,707	0,652	0,678
	NaiveBayes	0,976	0,951	0,964	0,607	0,765	0,677
	VFI	0,943	0,934	0,939	0,389	0,423	0,405
	DecisionStump	0,944	0,991	0,967	0,822	0,409	0,546
	MultiBoostAB	0,944	0,991	0,967	0,822	0,409	0,546

	ClassificationViaClustering	0,926	0,601	0,729	0,113	0,514	0,185
	DecisionTable	0,959	0,996	0,997	**0,993**	0,568	**0,706**
	BayesNet	0,981	0,931	0,955	0,536	**0,815**	0,647
25 %	NaiveBayes	0,956	0,993	0,974	0,882	0,529	0,661
	BayesNet	0,969	0,978	0,974	0,754	**0,683**	**0,717**
	DecisionTable	0,960	0,992	0,976	0,881	0,573	0,694
	ADTree	0,955	0,989	0,972	0,819	0,526	0,641
	RBFNetwork	0,946	0,997	0,971	**0,926**	0,415	0,573
	HyperPipes	0,923	0,993	0,957	0,699	0,156	0,254
	NaiveBayesUpdateable	0,956	0,993	0,974	0,882	0,529	0,661
	DTNB	0,965	0,978	0,972	0,742	0,641	0,688
	DecisionStump	0,945	0,991	0,968	0,823	0,411	0,548
	VFI	0,942	0,950	0,946	0,446	0,407	0,426
	ClassificationViaClustering	0,914	0,609	0,731	0,094	0,414	0,153
	CVParameterSelection	0,911	1,000	0,953	0,000	0,000	0,000
	LogitBoost	0,948	0,996	0,972	0,921	0,448	0,602
	MultiBoostAB	0,945	0,991	0,968	0,823	0,411	0,548
	AdaBoostM1	0,945	0,991	0,968	0,823	0,411	0,548
5 %	J48	0,910	1,000	0,953	0,000	0,000	0,000
	ADTree	0,949	0,992	0,970	0,844	0,458	0,594
	RBFNetwork	0,950	0,995	0,972	0,900	0,470	0,618
	NaiveBayes	0,944	0,999	0,970	**0,969**	0,397	0,564
	DecisionTable	0,957	0,992	0,974	0,876	**0,551**	**0,676**

Preizkušeno je bilo še več klasifikatorjev, kjer pa izračunavanje bodisi ni uspelo (SMO, LibSVM, HNB, MultilayerPerceptron, Bagging, FT, Prism, J48) bodisi je trajalo predolgo (LWL, KStar, Id3, NBTree, BFTree, LADTree, SimpleCart, REPTree). Je pa seveda možno, da bi se dala katera od teh metod usposobiti z ustreznimi parametri klasifikatorja, ustrezno zmanjšanim oknom, manjšim korpusom ali več potrpljenja (potrebnega za to, da se na rezultat čaka nekaj dni).

Če želimo iskati manjkajoče vejice, nas zanima predvsem natančnost pri možnosti, ko vejica je, vendar seveda tudi priklic ne sme biti preslab, tako da iščemo tudi dober F1.

Kot uspešni klasifikatorji so se pokazali DecisionTable, NaiveBayes, BayesNet, AdaBoostM1 (ta je dosegel enake rezultate kot DecisionStump in MultiBoostM1, zato druga dva nista bila uporabljena za naprej) in RBFNetwork. Slaba stran klasifikatorja DecisionTable pa je, da je preizkušanje počasno do neuporabnosti (preizkus na celotnih podatkih je trajal tri dni), zato je bil pri nadaljnjem preizkušanju namesto njega uporabljen ADTree (ki se je pokazal kot uspešen v Holozan (2013); klasifikatorji za nadaljnje preizkušanje so bili namerno izbrani tako, da pripadajo različnim skupinam klasifikatorjem in med seboj niso

preveč podobni). Morda velja DecisionTable še enkrat preizkusiti na koncu, ko bodo izbrani drugi parametri (atributi, označevanje).

Klasifikatorje, ki niso bili uspešni pri celotnem korpusu, smo preizkusili še na zmanjšanem korpusu, da vidimo, ali bi bili morda uspešni tam (nekatere klasifikatorje pa smo ponovno uporabili za primerjavo, koliko vpliva velikost korpusa).

Klasifikator J48, uporabljen v Alegria et al. (2006), je bil uspešen le v 5 % primerov (vendar je tudi tu uporabil le večinski razred in povsod odgovoril, da vejice ni), pri 25 % in polnem korpusu preizkus ni bil uspešen. Klasifikator SMO pa sploh ni bil uspešen (ker ni imel dovolj pomnilnika) niti pri 5 %. Ta rezultat je presenetljiv, Alegria et al. (2006) so uporabljali korpus s 130.000 besedami za preizkuse (100.000 besed za učenje in 30.000 za preizkušanje) in malo manjše okno (-5+2), kar pomeni, da 5 % korpusa v našem poskusu ne bi smelo biti problem. Zato bi bilo smiselno reč v prihodnosti še enkrat preizkusiti z ustrezno nastavitvijo parametrov klasifikatorjev.

Manjšanje korpusa je slabšalo rezultate, zato je bilo smiselno nadaljevati s celotno zbirko primerov.

5.1.2.2 Atributi

Vprašanje je, kateri podatki so pomembni v tem smislu, da naj se dodajo kot atributi. Osnovna podatka sta sama beseda in lema besede, narejen pa je bil poskus, kako uporabiti oblikoskladenjske oznake (MSD).

Tabela 7: Vpliv atributov; Šolar, okno -5+5

		ni vejice			je vejica		
	klasifikator	natančnost	priklic	F1	natančnost	priklic	F1
celoten MSD	NaiveBayes	0,976	0,951	0,964	0,607	0,765	0,677
	RBFNetwork	0,966	0,973	0,970	0,707	0,652	0,678
	BayesNet	0,981	0,931	0,955	0,536	0,815	0,647
	ADTree	0,955	0,989	0,971	0,820	0,525	0,640
	AdaBoostM1	0,944	0,991	0,967	0,822	0,409	0,546
delni MSD	NaiveBayes	0,978	0,936	0,957	0,549	0,790	0,648
	RBFNetwork	0,970	0,957	0,963	0,613	0,696	0,652
	BayesNet	0,982	0,915	0,947	0,490	0,831	0,616
	ADTree	0,955	0,990	0,972	0,836	0,531	0,650
	AdaBoostM1	0,944	0,991	0,967	0,822	0,409	0,546
brez oblik	NaiveBayes	0,977	0,949	0,963	0,598	0,773	0,674
	RBFNetwork	0,968	0,971	0,969	0,695	0,675	0,685
	BayesNet	0,979	0,937	0,958	0,557	0,800	0,656
	ADTree	0,955	0,989	0,971	0,820	0,525	0,640
	AdaBoostM1	0,944	0,991	0,967	0,822	0,409	0,546
delni	NaiveBayes	0,978	0,936	0,957	0,550	0,789	0,648

MSD2	RBFNetwork	0,969	0,957	0,963	0,614	0,690	0,650
	BayesNet	0,982	0,913	0,946	0,485	0,833	0,613
	ADTree	0,948	0,997	0,972	**0,942**	0,448	0,607
	AdaBoostM1	0,944	0,991	0,967	0,822	0,409	0,546
MSD + delni MSD2	NaiveBayes	0,981	0,926	0,953	0,521	0,814	0,635
	RBFNetwork	0,963	0,961	0,962	0,613	0,623	0,618
	BayesNet	0,984	0,906	0,943	0,469	0,846	0,604
	ADTree	0,948	0,997	0,972	**0,942**	0,448	0,607
	AdaBoostM1	0,944	0,991	0,967	0,822	0,409	0,546
MSD + skladnja	NaiveBayes	0,982	0,915	0,947	0,490	0,832	0,617
	RBFNetwork	0,973	0,950	0,961	0,592	0,732	0,655
	BayesNet	0,985	0,891	0,936	0,439	0,865	0,583
	ADTree	0,960	0,988	0,974	0,828	0,577	0,680
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
delni MSD + skladnja	NaiveBayes	0,983	0,900	0,940	0,455	0,844	0,591
	RBFNetwork	0,962	0,952	0,957	0,561	0,621	0,590
	BayesNet	0,986	0,876	0,928	0,409	0,872	0,557
	ADTree	0,959	0,990	0,975	0,853	0,573	0,685
	AdaBoostM1	0,960	0,963	0,961	0,610	0,590	0,600
delni MSD2 + skladnja	NaiveBayes	0,983	0,896	0,938	0,445	0,846	0,583
	RBFNetwork	0,967	0,948	0,957	0,558	0,667	0,608
	BayesNet	0,986	0,869	0,924	0,397	0,874	0,546
	ADTree	0,967	0,984	0,976	0,802	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
MSD + delni MSD2 + skladnja	NaiveBayes	0,984	0,894	0,937	0,443	0,855	0,583
	RBFNetwork	0,952	0,964	0,958	0,582	0,506	0,542
	BayesNet	0,986	0,872	0,926	0,403	**0,877**	0,552
	ADTree	0,967	0,984	0,976	0,802	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631

Delni MSD (ločeno prvi znak MSD, drugi znak MSD in pri samostalnikih še sklon) je malce izboljšal rezultate pri klasifikatorjih ADTree in AdaBoostM1, poslabšal pa pri NaiveBayes in RBFNetwork.

Zanimiv rezultat je prinesla ukinitev atributov z oblikami (torej so ostale le leme), kjer je bil pri ADTree in AdaBoostM1 rezultat popolnoma enak, pri NaiveBayes in RBFNetwork pa se je poslabšal.

Delni MSD2 je bil poskus, kako čim bolje prenesti informacije iz MSD in se izogniti uporabi celotnega MSD (zaradi predpostavke, da veliko število različnih MSD lahko ovira učenje). Vendar je tudi ta poskus samo poslabšal rezultate (je sicer izboljšal natančnost pri ADTree, vendar za ceno velikega poslabšanja priklica) (rezultat je poslabšal celo delni MSD 2 in dodan celotni MSD), tako da je očitno najbolj smiselno uporabiti kar celotni MSD.

Atributi s podatki o skladnji v kombinaciji z delnim MSD 2 so sicer poslabšali rezultat pri klasifikatorjih NaiveBayes in RBFNetwork, vendar so ga popravili pri ADTree in AdaBoostM1, in to toliko, da je F1 pri ADTree postal najboljši, zato je bila za nadaljnje poskuse izbrana ta kombinacija (oz. kombinacija, ki vsebuje še celotni MSD, ker je pri BayesNet tam najboljši priklic, F1 pri ADTree pa je enak kot v primeru, če ni celotnega MSD).

5.1.2.3 Velikost okna

Preizkusili smo vplivanje velikosti okna, tj. števila besed pred in za besedo, glede katere ugotavljamo, ali ji sledi vejica.

Tabela 8: Vpliv okna; Šolar, MSD + delni MSD 2 + skladnja

		ni vejice			je vejica		
okno	klasifikator	natančnost	priklic	F1	natančnost	priklic	F1
-5+5	NaiveBayes	0,984	0,894	0,937	0,443	0,855	0,583
	RBFNetwork	0,952	0,964	0,958	0,582	0,506	0,542
	BayesNet	0,986	0,872	0,926	0,403	0,877	0,552
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-4+5	NaiveBayes	0,985	0,891	0,936	0,438	0,861	0,581
	RBFNetwork	0,952	0,966	0,959	0,600	0,510	0,551
	BayesNet	0,987	0,871	0,925	0,401	0,880	0,551
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-3+5	NaiveBayes	0,985	0,888	0,934	0,433	0,863	0,577
	RBFNetwork	0,963	0,965	0,964	0,637	0,622	0,629
	BayesNet	0,987	0,870	0,924	0,400	0,881	0,550
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-2+5	NaiveBayes	0,985	0,884	0,932	0,424	0,866	0,569
	RBFNetwork	0,964	0,966	0,965	0,650	0,635	0,642
	BayesNet	0,987	0,867	0,923	0,395	0,882	0,545
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-1+5	NaiveBayes	0,985	0,873	0,926	0,402	0,867	0,549
	RBFNetwork	0,970	0,966	0,968	0,668	0,696	0,682
	BayesNet	0,987	0,857	0,917	0,378	**0,882**	0,529
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,956	0,958	0,570	0,594	0,581
-0+5	NaiveBayes	0,985	0,855	0,915	0,370	0,864	0,519
	RBFNetwork	0,969	0,966	0,967	0,666	0,684	0,675
	BayesNet	0,986	0,841	0,908	0,352	0,876	0,502
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-5+4	NaiveBayes	0,984	0,913	0,947	0,489	0,845	0,620

	RBFNetwork	0,971	0,948	0,959	0,575	0,710	0,635
	BayesNet	0,986	0,895	0,939	0,451	0,872	0,594
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-5+3	NaiveBayes	0,982	0,927	0,954	0,527	0,825	0,643
	RBFNetwork	0,975	0,954	0,964	0,617	0,747	0,676
	BayesNet	0,985	0,912	0,947	0,490	0,855	0,623
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-5+2	NaiveBayes	0,980	0,934	0,956	0,546	0,802	0,650
	RBFNetwork	0,972	0,960	0,966	0,639	0,717	0,676
	BayesNet	0,983	0,922	0,951	0,513	0,837	0,636
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-5+1	NaiveBayes	0,976	0,926	0,950	0,506	0,773	0,611
	RBFNetwork	0,966	0,961	0,964	0,626	0,658	0,641
	BayesNet	0,980	0,910	0,944	0,471	0,815	0,597
	ADTree	0,967	0,984	0,975	0,801	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-5+0	NaiveBayes	0,958	0,811	0,879	0,250	0,639	0,360
	RBFNetwork	0,910	1,000	0,953	0,000	0,000	0,000
	BayesNet	0,964	0,782	0,863	0,241	0,704	0,359
	ADTree	0,911	0,999	0,953	0,455	0,011	0,021
	AdaBoostM1	0,910	1,000	0,953	0,000	0,000	0,000
-1+2	NaiveBayes	0,980	0,932	0,955	0,538	0,804	0,644
	RBFNetwork	0,966	0,966	0,966	0,656	0,656	0,656
	BayesNet	0,981	0,925	0,952	0,520	0,821	0,637
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-0+1	NaiveBayes	0,975	0,935	0,954	0,534	0,754	0,625
	RBFNetwork	0,967	0,964	0,965	0,645	0,663	0,654
	BayesNet	0,975	0,935	0,954	0,534	0,754	0,625
	ADTree	0,967	0,984	0,975	0,801	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-6+6	NaiveBayes	0,984	0,877	0,927	0,405	0,853	0,550
	RBFNetwork	0,928	0,990	0,958	0,686	0,217	0,329
	BayesNet	0,986	0,849	0,912	0,363	0,876	0,514
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
-7+7	NaiveBayes	0,983	0,862	0,918	0,376	0,846	0,521
	RBFNetwork	0,930	0,989	0,959	0,699	0,250	0,368
	BayesNet	0,985	0,828	0,899	0,332	0,869	0,480
	ADTree	0,967	0,984	0,976	**0,802**	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631

Tabela 3 kaže, da klasifikator ADTree uporablja le trenutno besedo in še eno naprej. Vendar (razen na hitrost) večanje okna ne vpliva negativno na rezultat, zato je pri nadaljnjih

preizkusih uporabljeno kar okno -5+5, tudi zaradi domneve, da začne klasifikator pri spreminjanju parametrov klasifikatorja ADTree (torej večanju drevesa) upoštevati tudi besede zunaj okna -0+1, ki se je pokazalo kot zadostno tukaj (drevo, ki je rezultat poskusa s parametrom –B 70, res vsebuje tudi položaje +4, +3, +2, -1 in -3, torej bi bilo tam optimalno drevo -3+4, kar potrjuje to domnevo).

Pri drugih klasifikatorjih je spreminjanje okna veliko bolj vplivalo na rezultat. Pri NaiveBayes se je rezultat izboljšal, ko se je okno naslednjih besed manjšalo, najboljši rezultat je bil pri -5+2. RBFNetwork je dosegel najboljši rezultat pri -1+5, BayesNet pri -5+2, na AdaBoostM1 pa spreminjanje okna ni vplivalo (dokler ni postalo premajhno). Na prvi pogled ni videti smiselno, da manjšanje količine podatke izboljšuje rezultate, vendar se to zgodi zaradi tega, ker nekatere metode med množico atributov težko izberejo, kateri so res pomembni.

Rezultati kažejo, da je potrebno okno zelo majhno, vendar je treba upoštevati, da so dobljeni atributi (še posebej skladenjski) že pridobljeni z večjim oknom, kar pomeni, da ne moremo posplošiti, da sta pri ADTree pomembni le dve zaporedni besedi pri določanju vejice.

V vseh primerih (razen pri ekstremnem -5+0, kjer je bil malenkost boljši NaiveBayes), je bil najboljši klasifikator ADTree, zato smo v nadaljevanju preizkušanja uporabljali le njega.

5.1.2.4 Vpliv označevanja

Rezultati postavljanja vejic so zelo uspešni, vendar vsebujejo problematično predpostavko: pri oblikoslovnem označevanju in skladenjski razčlembi je uporabljeno besedilo, ki je vsebovalo pravilno postavljene vejice. To pa seveda ni realna situacija, saj v primeru, da hočemo v neko besedilo postaviti vejice, teh vnaprej pač ne vemo.

Zato je imel naslednji poskus nalogo ugotoviti, kaj se zgodi, če oblikoslovni označevalnik in skladenjski razčlenjevalnik nimata vejic v vhodnem besedilu. Iz korpusa smo izbrisali vse vejice in ga ponovno označili in pretvorili v format ARFF. Ker pa je bil seveda povsod zapisan podatek, da ni vejice, je bilo treba iz datoteke ARFF za korpus z vejicami prenesti stolpec s podatki za vejico v datoteko ARFF korpusa brez vejic. Pri tem postopku je potrebna previdnost: nujno je treba preveriti, da se število besed ujema in besede pokrivajo. Nekateri tipi napak v izvornem korpusu namreč naredijo težave pri brisanju vejic, tak primer je npr. manjkajoč presledek za vejico, kjer brisanje vejice potem zlepi besedi in povzroči, da je v korpusu brez vejic ena beseda manj. Težava je tudi, da tokenizator (rezalnik besedila na posamezne besede) včasih spreminja vezavo pike na predhodno besedo različno (na primer

pri arabskem zapisu vrstilnih števnikov), če je blizu vejica. Te primere je bilo treba v označenem XML potem ročno popraviti, da so se besede ujemale.

Tabela 9: Vpliv označevanja; Šolar, MSD + delni MSD2 + skladnja, okno -5+5

		ni vejice			je vejica		
	klasifikator	natančnost	priklic	F1	natančnost	priklic	F1
označevanje z vejicami	NaiveBayes	0,984	0,894	0,937	0,443	0,855	0,583
	RBFNetwork	0,952	0,964	0,958	0,582	0,506	0,542
	BayesNet	0,986	0,872	0,926	0,403	**0,877**	0,552
	ADTree	0,967	0,984	0,976	0,802	0,663	**0,726**
	AdaBoostM1	0,960	0,973	0,966	0,681	0,587	0,631
označevanje brez vejic	NaiveBayes	0,983	0,887	0,933	0,426	0,849	0,567
	RBFNetwork	0,931	0,986	0,958	0,645	0,263	0,374
	BayesNet	0,986	0,864	0,921	0,388	0,873	0,537
	ADTree	0,956	0,988	0,972	0,814	0,542	0,615
	AdaBoostM1	0,944	0,991	0,967	0,820	0,405	0,542
označevanje brez vejic, označevalnik naučen brez vejic	NaiveBayes	0,984	0,888	0,933	0,427	0,850	0,568
	RBFNetwork	0,940	0,976	0,957	0,597	0,366	0,454
	BayesNet	0,986	0,864	0,921	0,388	0,874	0,537
	ADTree	0,959	0,987	0,973	0,809	0,569	0,668
	AdaBoostM1	0,957	0,969	0,963	0,644	0,563	0,600
	DecisionTable	0,966	0,994	0,980	**0,910**	0,645	**0,755**

Tabela 9 pove, da so se rezultati ugotavljanja vejic v primeru, ko besedilo pri označevanju ni imelo vejic, poslabšali (čeprav ne zelo izrazito, največja razlika je bila pri klasifikatorju AdaBoostM1), kar se sklada tudi s splošnimi ugotovitvami Hillarda et al. (2006), da pravilno postavljene vejice izboljšujejo oblikoslovno označevanje besedil.

Preizkusili smo še, ali lahko označevanje (in s tem posledično določanje vejic) izboljšamo s tem, da lematizator, oblikoslovni označevalnik in skladenjski razčlenjevalnik učimo iz učnega korpusa brez vejic (to sta uporabila že Shieber in Tao (2003)). V ta namen so bile v učnem korpusu SSJ500k izbrisane vse vejice (in povezave na vejice pri skladenjski razčlenitvi) in na novo naučeni modeli za lematizator, oblikoslovni označevalnik in skladenjski razčlenjevalnik (ta postopek predvsem za oblikoslovni označevalnik porabi veliko procesorskega časa (dobrih 20 ur), vendar ga je treba narediti le enkrat). Rezultati so se izboljšali, niso pa dosegli primera, ko je bilo besedilo označeno z vejicami, kar kaže na to, da so vejice pomembne za razdvoumljanje. Vseeno pa se je pokazalo, da je v primeru, ko je treba v besedilu dodati vse vejice, smiselno naučiti označevalnike z učnim korpusom brez vejic.

Tukaj smo dodatno preizkusili še klasifikator DecisionTable, ki je bil pri izbiranju klasifikatorjev zelo uspešen, vendar ga zaradi dolgega trajanja preizkusa nismo izbrali za nadaljnje preizkušanje.

5.1.2.5 Parametri klasifikatorja

Klasifikator DecisionTable je sicer dosegel najboljši rezultat, vendar je en poskus trajal tri dni. Zato smo vplivanje parametrov preverjali raje pri sicer drugouvrščenem klasifikatorju ADTree (izvirno ime v prevodu pomeni »alternirajoče odločitveno drevo«), ki je bil občutno hitrejši.

Tabela 10: Vpliv parametrov na ADTree; Šolar, MSD + delni MSD2 + skladnja, okno -5+5, označevanje brez vejic, naučeno brez vejic

	ni vejice			je vejica		
parametri	natančnost	priklic	F1	natančnost	priklic	F1
-B 10 -E -3	0,959	0,987	0,973	0,809	0,569	0,668
-B 8 -E -3	0,957	0,987	0,972	0,804	0,553	0,656
-B 6 -E -3	0,957	0,986	0,971	0,796	0,554	0,653
-B 4 -E -3	0,957	0,984	0,971	0,778	0,554	0,647
-B 12 -E -3	0,959	0,986	0,972	0,803	0,573	0,669
-B 15 -E -3	0,961	0,986	0,973	0,803	0,592	0,682
-B 20 -E -3	0,960	0,988	0,974	0,831	0,584	0,686
-B 30 -E -3	0,963	0,990	0,976	0,863	0,612	0,716
-B 50 -E -3	0,965	0,993	0,979	0,894	0,638	0,744
-B 70 -E -3	0,967	0,992	0,980	0,892	**0,661**	**0,759**
-B 10 -E -2	0,957	0,988	0,972	0,819	0,545	0,655
-B 30 -E -2	0,961	0,992	0,976	0,879	0,589	0,706
-B 50 -E -2	0,962	0,993	0,977	0,898	0,603	0,722
-B 70 -E -2	0,963	0,994	0,978	**0,906**	0,616	0,733
-B 10 -E -1	0,957	0,988	0,972	0,819	0,545	0,655
-B 50 -E -1	0,962	0,993	0,977	0,898	0,603	0,722
-B 70 -E -1	0,963	0,994	0,978	**0,906**	0,616	0,733

Tabela 10 prikazuje spreminjanje rezultatov glede na spreminjanje parametrov. Parameter -B pove število ponovitev dodajanj vozlišč pri gradnji drevesa in tako povečuje drevo, ki je rezultat učenja, hkrati pa podaljšuje čas, potreben za izračun.

Parametri -3, -2 in -1 povedo, na kakšen način išče klasifikator nova potencialna vozlišča. Pri parametru 3 preveri vse možnosti, pri -2 in -1 pa preiskovanje omeji, kar pospeši iskanje, rezultat pa ni nujno optimalen (najboljše možno odločitveno drevo za dano število vozlišč).

Zadnji rezultat (s 101 listom v odločitvenem drevesu) je presegel tudi rezultat s privzetimi parametri pri klasifikatorju DecisionTable, se pravi, da je najboljši sploh. V prihodnosti bi veljalo preizkusiti različne parametre tudi pri drugih klasifikatorjih, da bi se tako našla optimalna kombinacija

Dodatna prednost klasifikatorja ADTree je, da izpiše odločitveno drevo, ki bi se dalo relativno preprosto interpretirati (Freund in Mason 1999) in tako uporabiti v drugih programih.

```
: -1.159
| (1)je_vez1 = 1: 1.266
| | (2)lem1 = in: -1.503
| | (2)lem1 != in: 0.519
| (1)je_vez1 != 1: -0.401
| | (3)msda0 = So: 0.57
| | (3)msda0 != So: -0.234
| | | (7)msda0 = Gp-: -1.017
| | | (7)msda0 != Gp-: 0.083
| | (4)msda2 = *: -1.787
| | (4)msda2 != *: 0.093
| | | (5)zac_modrega1 = 0: -0.163
| | | (5)zac_modrega1 != 0: 0.661
| | (10)msd1 = Vd: 1.718
| | (10)msd1 != Vd: -0.03
| (6)msda0 = D: -2.143
| (6)msda0 != D: 0.041
| | (8)msd0 = Vd: -1.958
| | (8)msd0 != Vd: 0.025
| | | (9)lem0 = in: -2.658
| | | (9)lem0 != in: 0.025
| | | | (11)zac_modrega0 = 1: -0.424
| | | | (11)zac_modrega0 = 0: 0.082
| | | | | (12)msda2 = Gp-: 0.575
| | | | | (12)msda2 != Gp-: -0.063
Legend: -ve = ni-vejice, +ve = je-vejica
Tree size (total number of nodes): 37
Leaves (number of predictor nodes): 25
```

Slika 11: Odločitveno drevo za ADTree -B 12 -E -3

Slika 11 prikazuje primer odločitvenega drevesa pri -B 12 (s 25 listi). Na verjetnost, da gre za vejico, najbolj vpliva podatek iz skladenjskega razčlenjevalnika, da na naslednjo besedo kaže povezava "vez". Če je tako, v naslednjem koraku preveri, ali je naslednja beseda »in«, v tem primeru se verjetnost zelo zmanjša.

Zanimiv je vpliv msda2 z vrednostjo * (kar pomeni, da te besede ni), kar z drugimi besedami pomeni, da vejica tik pred koncem stavka ni posebno verjetna. Zanimiv je tudi pogoj msda0 D, predlog torej močno zmanjša verjetnost, da neposredno za njim stoji vejica.

5.1.3 Iskanje napak pri postavljanju vejic

Dosedanji rezultati strojnega učenja povedo, kako dobro postavijo programi vejice v besedilo, v katerem ni na začetku nobenih vejic, kar je npr. uporabno pri razpoznavi govora, ki ne zaznava vejic. Vprašanje pa je, kako dobro se programi obnesejo pri popravljanju pravih napak, saj te niso naključno razporejene, ampak nekateri tipi vejic delajo piscem več težav kot drugi.

V ta namen lahko uporabimo zbirko primerov (spet smo se omejili na primere iz korpusa Šolar). Vendar pa je primerov napačnih vejic veliko manj kot vseh primerov vejic, pa še štiri možna stanja so (ob *je vejica* in *ni vejic*e še *ni manjkajoče vejice* in *je odvečna vejica*) in je zato vprašanje, ali bi bilo 16.180 primerov manjkajoče vejice in 3759 primerov odvečne vejice dovolj za uspešno učenje (vseh vejic je sicer 65.413).

Zato smo izbrali drugačen postopek: in sicer se program WEKA uporabi tako, da je prvih (izključi se privzeto naključno izbiranje) 80 % primerov učni korpus, zadnjih 20 % pa se uporabi kot testni korpus, pri čemer se rezultat preizkušanja izpiše za vsak primer posebej. Ker so v korpusu Šolar primeri sicer razporejeni po razredih in letnikih oz. vrstah šol, ne bi bilo primerno, če bi vsi preizkusni primeri prišli iz istega letnika oz. z istovrstnih šol – Holozan (2012) je pokazal, da so rezultati popravljanja vejic različni glede na letnik oz. šolo – in zato smo najprej izvedli postopek, ki je delno premešal primere, s tem da je najprej izločil vsako peto poved, te izločene povedi pa so bile potem dodane na koncu.

Rezultat preizkušanja (stolpec, ki pove, katero stanje vejice je izbral klasifikator) je bil potem poravnan s podatki o vejicah iz korpusa (pri čemer je bilo treba paziti, da se je poravnalo z zadnjimi primeri in ne s prvimi, težave pa pri poravnavanju povzroča še to, da se razrez na besede lahko zaradi dodanih oziroma odvzetih vejic spremeni), oboje je bilo sestavljeno v eno tabelo, potem pa prešteto, kolikokrat pojavlja katera kombinacija.

```
  3465-je-vejica      1:ni-vejic
  6190-je-vejica      2:je-vejic
  1624-manjka-vejica        1:ni-vejic
  1550-manjka-vejica        2:je-vejic
130973-ni-vejice    1:ni-vejic
  1250-ni-vejice      2:je-vejic
   554-prevec-vejica        1:ni-vejic
   179-prevec-vejica        2:je-vejic
```

Slika 12: Rezultat primerjave rezultatov preizkušanja s podatki iz korpusa

Slika 12 prikazuje tak (grob) rezultat za primer, ko je bil korpus označen brez vejic (torej smo zbrisali tudi vejice, ki so bile napisane) z označevalnikom, naučenim brez vejic: spredaj je število primerov, drugi stolpec prikazuje stanje v zbirki primerov, tretji stolpec pa je rezultat preizkušanja klasifikatorja, torej je npr. v 1550 primerih, ko je vejica manjkala, klasifikator menil, da bi tam morala biti vejica, v 1624 primerih pa, da vejice tam ni, po drugi strani pa je v 1250 primerih postavil vejico, kjer je ne bi smelo biti, natančnost (kakšen delež dodanih vejic je pravilen) je tako 1550 / (1550 + 1250) oz. 55,4 %.

Tak postopek smo ponavljali glede na različne načine označevanja, nismo pa izvedli 10-kratnega prečnega preverjanja, ker bi bil ta postopek precej zapleten (in bi ga bilo treba prej bolj avtomatizirati, medtem ko so bili tokrat nekateri koraki izvedeni ročno za vsak primer posebej). Samo 10 % primerov pri preizkušanju pa bi bilo morda tudi premalo, da bi lahko potem dovolj zanesljivo dobili rezultat pri primerjavi z napakami v korpusu, zato smo izbrali razdelitev 80:20. Preizkušanje smo, da ne bi predolgo trajalo, izvedli le s klasifikatorjem ADTree s parametri (-B 30 -E -3), rezultati in primerjava z LanguageToolom in Besana so v točki 6.1.3.

5.2 LanguageTool

Po uporabi strojnega učenja smo se lotili izboljševanja LanguageToola. Pravila za program LanguageTool smo izboljšali s pomočjo rezultatov analize napak pri iskanju napak pri vejicah, izvedene pod točko 4.5.1. Narejeni popravki pravil so v prilogi 9.4.

5.2.1 Metode za ugotavljanje manjkajočih vejic

Popravke lahko razdelimo v dve skupini. V prvi je treba ugotoviti, katere manjkajoče vejice je program LanguageTool spregledal in ali lahko to rešimo z dodatnimi pravili, v drugi skupini pa je bilo ugotoviti, v katerih primerih program najde manjkajoče vejice na mestih, kjer v resnici niso potrebne, in ali z dopolnitvijo pravil lahko zmanjšamo število takšnih primerov.

5.2.1.1 Spregledane manjkajoče vejice

S pomočjo postopka, opisanega v točki 4.5.1, smo dobili seznam najpogostejših besed, pred katerimi je manjkajoča vejica, ki je LanguageTool ni postavil.

Tabela 11: Najpogostejše besede za manjkajočimi vejicami, ki jih LanguageTool ne odkrije

beseda	1	2	3	4	5	6	7	8
in	0	1305	0	456	1768	0	44937	0
je	0	903	0	139	2143	0	57212	0
se	0	501	0	85	1346	0	28994	3
da	1689	442	0	474	18940	0	2027	488
kot	33	427	0	419	1568	0	5805	4
naj	0	293	0	5	406	0	1613	0
so	0	267	0	54	713	0	16071	1
pa	0	259	0	144	1260	0	13404	1
a	0	240	0	3	1508	0	239	0
vendar	0	236	0	0	1499	0	180	0
ali	0	232	0	150	486	0	4483	0
sem	0	229	0	25	330	0	5187	0
v	129	186	0	68	1984	0	31387	9
bi	0	161	0	47	366	0	9336	1
zakaj	0	158	0	1	209	0	75	0
saj	577	134	0	2	3720	0	8	34
ne	4	129	0	74	2096	0	8905	1
sta	0	106	0	17	143	0	3150	0
ker	276	95	0	33	2337	0	139	97
ni	0	93	0	12	286	0	6497	0
ki	393	87	0	3	16009	0	81	8
ko	373	83	0	172	2725	0	676	187
ga	0	82	0	25	226	0	5329	0
tako	0	79	0	48	685	0	4348	1
si	0	76	0	13	173	0	3984	2
kako	450	74	0	1	1087	0	204	73
s	53	72	0	27	735	0	6443	3
če	217	70	0	68	1605	0	554	166
na	34	67	0	33	986	0	17539	12
lahko	0	66	0	10	371	0	5685	0
ter	0	65	0	327	138	0	2151	0
kaj	344	63	0	5	802	0	356	554
to	0	61	0	9	489	0	6105	0
jo	0	61	0	4	111	0	3169	0
kar	377	59	0	30	2017	0	414	518
smo	0	58	0	9	130	0	2596	0
čeprav	0	56	0	1	501	0	118	0
”	1	56	0	0	49	0	16	0
tudi	0	55	0	24	279	0	9545	0
mu	0	55	0	3	129	0	3105	0
bo	0	51	0	0	96	0	3055	0
ampak	233	47	0	1	947	0	1	26
kdo	0	46	0	1	98	0	251	0

še	0	45	0	14	299	0	4739	0
samo	0	42	0	4	86	0	1666	0
zato	146	41	0	20	1301	0	628	354
jih	0	41	0	4	137	0	3809	0
potem	0	40	0	2	276	0	494	0
mi	0	38	0	8	56	0	2376	0
za	46	36	0	30	555	0	13065	19
npr	0	34	0	1	190	0	309	0
ima	0	33	0	4	83	0	1668	0
koliko	0	33	0	0	79	0	32	0
z	0	32	0	21	261	0	9472	0
me	0	29	0	5	60	0	1151	0
temveč	1	27	0	4	693	0	0	0
ji	0	27	0	1	42	0	1244	0
po	4	26	0	28	487	0	4849	2
nato	0	26	0	0	233	0	276	0
le	0	25	0	8	202	0	2658	0
niso	0	25	0	2	61	0	1323	0
kljub	0	23	0	25	87	0	305	0
kakšna	0	23	0	1	34	0	26	0
kakor	14	21	0	25	135	0	138	4
katera	0	21	0	0	56	0	37	0
on	0	21	0	0	100	0	789	0
od	2	20	0	8	251	0	4448	1
zelo	0	20	0	8	65	0	3749	0
medtem	0	20	0	0	285	0	28	0
kjer	127	19	0	0	1173	0	15	9

Tabela 11 prikazuje besede in števce po posameznih stanjih pred to besedo (Tabela 4 opisuje pomen posameznih stanj), besede so urejene po števcu za stanje 2, ki pove, kolikokrat je bila pred to besedo manjkajoča vejica, ki je program LanguageTool ni odkril.

Seveda pa samo dejstvo, da je pred neko besedo pogosto manjkajoča vejica, še ne pove, da bi bilo pred to besedo treba vedno dodati vejico. Najpogostejša taka beseda je »in«, kjer manjka vejica v 1305 primerih, ampak po drugi strani je »in« v 44.937 primerih pravilno brez vejice, kar pomeni, da bi postavljanje vejice v vseh primerih zelo poslabšalo natančnost, zato je v tem primeru treba pravilo bolj natančno določiti.

Kot dodatne veznike, ki zahteva vejice, bi tako lahko uvedli »vendar«, »zakaj«, »čeprav«, »koliko«, »temveč« in »medtem«, morda tudi »a«. Dodatno pa bi lahko okrepili pravilo za vejico pred »saj« in »ampak«, saj podatki kažejo, da je v teh primerih vejica tako rekoč vedno na mestu, tako da je trenutno pravilo očitno preveč omejujoče.

Vseeno pa se splača tudi pri drugih besedah preveriti, ali bi bilo mogoče uporabiti zvezo dveh besed za iskanje manjkajočih vejic. Če pogledamo, katere besede se pojavljajo za »in« za stanjem 2 (pred katerim manjka vejica in LanguageTool tega ne ugotovi), najdemo zvezi »in to« in »in sicer«, temu sledijo zveze z vezniki »in ko«, »in če«, »in da«. Vendar od tega lahko uporabimo le »in sicer«, pred katerim edinim je vejica zanesljivo.

Zanimiva je tudi beseda »kot«, tukaj pogosto manjkajo vejice pred »kot je«, »kot so«, »kot sem«, »kot sta«, »kot smo«, kar se na prvi pogled sicer zdi uporabno glede na to, da je pred veznikom »kot« vejica, če sledi osebna glagolska oblika, vendar je težava v tem, da je »kot« lahko tudi samostalnik, pri čemer pa potem ni vejice v navedenih kombinacijah (enako pa je ni tudi za stopnjevanimi pridevniki oz. prislovi: »bolj ko je«). Dalo bi se sicer reči, da je »kot« redkeje samostalnik kot veznik, vendar iskanje pokaže, da je npr. zveza »kot je« za stanjem 2 v 180 primerih, za stanjem 7 pa v 66, kar pomeni natančnost 73 % in kar ni bistveno slabše od dosedanje natančnosti za del iz korpusa Šolar (76 %), tako da je vredno razmisleka, ali bi bilo res treba dodati to pravilo. Vendar je natančnost LanguageToola že zdaj slabša od natančnosti Besane, zato je bolj po pameti izboljševati natančnost (kar smo naredili pod točko 5.2.1.2), ne pa jo še slabšati na račun izboljševanja priklica.

Zveza, ki jo je smiselno dodati, je še »medtem ko«, pri čemer je tudi nekaj primerov z vmesno vejico »medtem, ko«, kjer je treba še vmesno vejico označiti kot odvečno.

Pri besedici »v« je treba dodati zvezo »v katerem«, saj s tem rešimo 84 primerov (od 186).

Pri nekaterih besedah je smiselno pogledati tudi, kaj stoji spredaj. Kadar je spredaj neodkrita manjkajoča vejica, so pred »da« najpogosteje »tako«, »mi«, »je«, »pomeni«, »to«, »zato«, »mislim«, »zdi«; pri nekaterih se resda ne moremo prepričati, da je vejica vedno, pri »je«, »pomeni«, »mislim« in »zdi« (ter še »rekel«) pa smo o tem lahko že kar prepričani.

Podobno je zanimivo tudi, kaj stoji pred besedo »naj«. Besede, kjer lahko sklepamo na manjkajočo vejico, so »rekel«, »prosi«, »ukazal« in »prosila«, pri »rekel«, »rekla« in »naročil« pa ne moremo vedeti, da ne gre za samostalnik (čeprav je najbrž redek, tako da je to vseeno vredno razmisleka). Tudi pri »naroči« in »ukaže«, ne moremo biti prepričani, da je vmes vejica zaradi primerov tipa »Naroči naj pivo.«, pri katerih to ne velja Vprašanje je, kaj narediti z »je«, pri katerem se v večini primerov izkaže, da gre za glagol »biti«, in je potemtakem vejica pred »naj«, lahko pa gre za glagol »jesti«, in v tem primeru vejice ni.

5.2.1.2 Odvečne manjkajoče vejice

Poiščemo seznam besed, pri katerih LanguageTool največkrat postavi vejico, ki v resnici ni potrebna (stanje 8); s tem poskušamo poboljšati natančnost.

Tabela 12: Najpogostejše besede, pred katere LanguageTool postavi odvečno vejico

beseda	1	2	3	4	5	6	7	8
kaj	344	63	0	5	802	0	356	554
kar	377	59	0	30	2017	0	414	518
da	1689	442	0	474	18940	0	2027	488
zato	146	41	0	20	1301	0	628	354
ko	373	83	0	172	2725	0	676	187
če	217	70	0	68	1605	0	554	166
kdaj	33	2	0	1	51	0	46	125
ker	276	95	0	33	2337	0	139	97
kakšno	17	4	0	1	35	0	65	76
kakšen	32	6	0	0	56	0	52	74
kako	450	74	0	1	1087	0	204	73
kakšne	20	2	0	0	53	0	31	52
katerih	1	1	0	1	117	0	509	52
katerega	14	3	0	0	123	0	287	45
saj	577	134	0	2	3720	0	8	34
kje	42	11	0	1	86	0	31	27
ampak	233	47	0	1	947	0	1	26
katere	27	7	0	0	150	0	284	25
kam	23	7	0	0	30	0	7	22
kakšnih	0	0	0	0	0	0	15	21
za	46	36	0	30	555	0	13065	19
kakšnega	2	0	0	0	0	0	7	19
dokler	38	7	0	1	137	0	25	17
kateri	27	4	0	2	91	0	450	14
kadar	18	5	0	0	165	0	23	13

Tabela 12 prikazuje dobljeni rezultat. Spet moramo s pomočjo konkordančnika preveriti besedo za besedo.

Začnemo z besedo »kaj«, s katero je največ težav. Pogledamo besede, ki v teh primerih nastopajo spredaj, to so »ve«, »al«, »je«, »ni«, »se«, »a«, »če«, »lahko«, »pa«, »le«, »bo«, »si«, »da«, »bi«, »ampak«. Pri »ve«, »al« (pogovorno za »ali«), »a«, »če«, »da«, »ampak« smo lahko precej prepričani, da vmesne vejice v nobenem primeru ni, in to reši kar precejšnje število primerov (pri »ve« npr. 31).

Pri besedi »kar« je daleč najpogostejša beseda spredaj »je« (82 primerov), vendar v tem primeru ne moremo zanesljivo trditi, da vejice nikoli ni (»Povedal je, kar je želel.«), zato

tega žal ne moremo uporabiti, tudi »pa« in »se« in druge besede spredaj niso uporabne (beseda »kar« je zelo zahtevna za postavljanje vejic zaradi različnih vlog: poleg oziralnega zaimka, kjer tipično stoji vejica, je še členek, kjer vejice ni).

Pri »da« so spredaj največkrat »tko«, »se«, »pa«, »kljub temu« in še »da«, »brez«, »namreč«. Razen »se« in »pa« jih lahko dodamo v pravilo, vprašanje je morda »tko« (tako).

Pri »zato« so spredaj najpogostejše »je«, »prav«, »se«, »so«, »bil« in »tudi«, najbrž je smiselno vse dodati kot izjeme, čeprav seveda s tem izgubimo tudi kakšno pravilno najdeno manjkajočo vejico.

Pred »ko« najpogosteje najdemo »da«, »bolj«, »šele«, »s tem«, »že«, »saj«, »takoj«, »ker«. Vse razen »šele« in »že« lahko dodamo kot izjeme, kjer ni vejice, morda je vprašanje delno le »takoj«.

Kot zelo problematična se kaže beseda »kdaj«, saj je dodana vejica v večini primerov napačna. Od besed, ki se pojavljajo spredaj, lahko uporabimo le »še«, a to pokrije le 15 primerov, tako da bi bilo razmisliti, da pravilo za postavljanje vejic pred besedo »kdaj« privzeto izključimo (pravilo je našlo 33 vejic in pri tem postavilo 125 odvečnih).

5.2.2 Metode ugotavljanja odvečnih vejic

Program LanguageTool do zdaj ni imel pravil za iskanje odvečnih vejic v slovenščini. Zato smo opravili analizo, da bi se ugotovilo, ali je možno sestaviti kakšno dovolj zanesljivo pravilo. V ta namen poiščemo besede za stanjem 4.

Tabela 13: Besede, pred katerimi je največ odvečnih vejic, ki jih LanguageTool ne odkrije

beseda	1	2	3	4	5	6	7	8
da	1689	442	0	474	18940	0	2027	488
in	0	1305	0	456	1768	0	44937	0
kot	33	427	0	419	1568	0	5805	4
ter	0	65	0	327	138	0	2151	0
ko	373	83	0	172	2725	0	676	187
ali	0	232	0	150	486	0	4483	0
pa	0	259	0	144	1260	0	13404	1
je	0	903	0	139	2143	0	57212	0
…	0	2	0	93	34	0	1001	0
se	0	501	0	85	1346	0	28994	3
zaradi	12	8	0	85	193	0	1812	1
ne	4	129	0	74	2096	0	8905	1
če	217	70	0	68	1605	0	554	166
v	129	186	0	68	1984	0	31387	9
so	0	267	0	54	713	0	16071	1
tako	0	79	0	48	685	0	4348	1

bi	0	161	0	47	366	0	9336	1
oziroma	0	10	0	44	17	0	712	0
do	3	14	0	41	134	0	4393	3
itd	0	0	0	40	9	0	265	0
ker	276	95	0	33	2337	0	139	97
na	34	67	0	33	986	0	17539	12
kar	377	59	0	30	2017	0	414	518
za	46	36	0	30	555	0	13065	19
oz	0	7	0	30	14	0	1038	0
po	4	26	0	28	487	0	4849	2
s	53	72	0	27	735	0	6443	3
kakor	14	21	0	25	135	0	138	4
sem	0	229	0	25	330	0	5187	0
ga	0	82	0	25	226	0	5329	0
kljub	0	23	0	25	87	0	305	0

Tabela 13 prikazuje, pred katerimi besedami je največ odvečnih vejic. Če primerjamo še s stolpcem za stanje 5, ki pove, kolikokrat je pred to besedo vejica, ki ni odvečna, odkrijemo, da sta kandidata za preverjanje »...«, »itd.«, morda tudi »oziroma« in »oz.« (vendar je pri teh dveh vseeno večja možnost (ni pa to nemogoče niti pri prvih dveh), da je spredaj levosmerna vejica zaradi odvisnika, tako da bi bilo tvegano avtomatsko prepovedati vejico).

Če analiziramo, kaj se največkrat pojavi pred »da«, kadar je pred njim odvečna vejica, dobimo naslednje besede, pri katerih lahko precej zanesljivo sklepamo, da bi morala biti vejica na začetku sklopa »in da«, »kljub temu da«, »ter da«, »če da«, »oziroma da«, »brez da«; ne moremo pa biti dovolj prepričani pri drugih: »kot da«, »tako da«, »samo da«, »kakor da«. Podobno pri »ko« lahko prestavljamo vejico na začetek sklopa »medtem ko«, »in ko«, »s tem ko«, vprašanje pa je s »potem ko« in »takoj ko«. Pri »ker« lahko uporabimo »in ker«, ne moremo pa »zato ker«, pri »če« uporabimo »in če«.

5.3 Besana

Za LanguageToolom je prišla na vrsto še Besana. Popravljanje vejic v Besani se je sicer izboljšalo že zaradi popravkov analizatorja povedi, izvedenih za program za določanje vrst vejic (točka 4.5.2.1). Izboljšanje sicer ni bilo zelo izrazito, F1 pri manjkajočih vejicah se je popravil iz 65,03 % na 65,66 %.

Izboljšav Besane smo se lotili na dva načina: z uporabo analize povedi, pri čemer bodisi uporabimo že obstoječe analize in s pravilnim interpretiranjem opozorimo na napačne

vejice bodisi dopolnimo stavčni analizator tako, da zna tudi pri napačno postavljenih vejicah ugotoviti pravilno analizo povedi, iz česar potem lahko spet opozarjamo na napake pri vejicah.

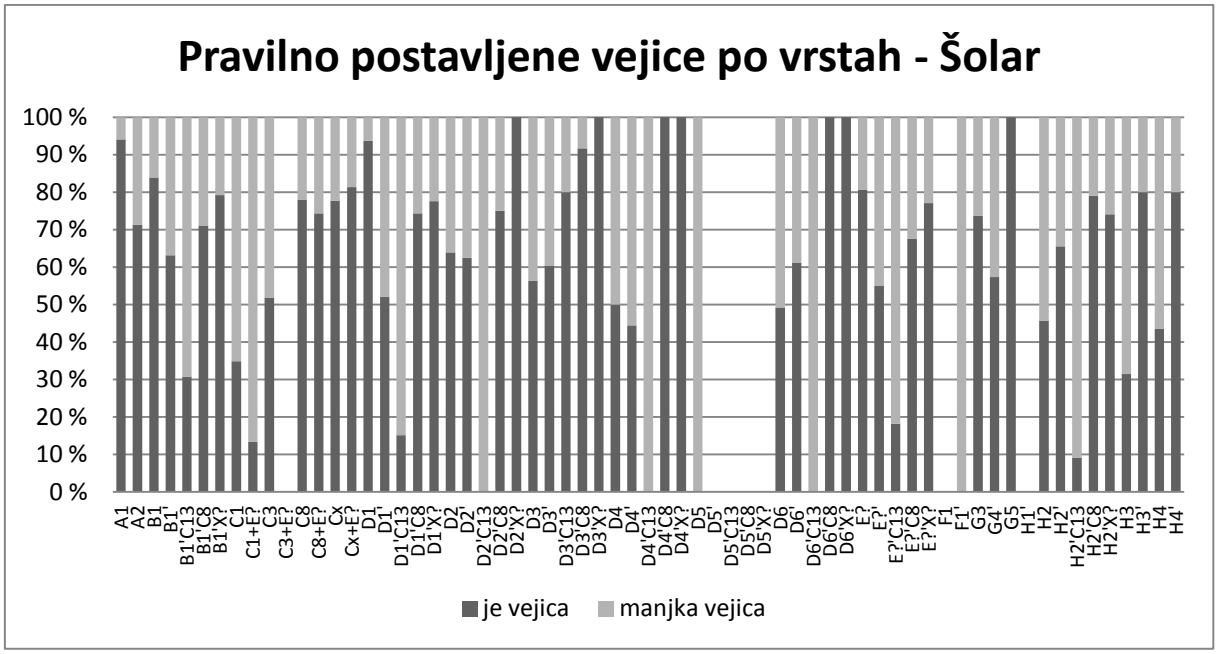
Drugi način pa je pomožna metoda, ki deluje v primerih, ko analizatorju ne uspe uspešno analizirati povedi; v teh primerih uporabimo besede, ki so v okolici napak, podobno kot je to narejeno za LanguageTool pod točko 5.2.

5.3.1 Uporaba analize povedi

Premislek je pokazal, da je bolj smiselno najprej izboljšati postavljanje vejic z uporabo analize povedi in šele potem uporabiti okolico napak (točka 5.3.2), kar deluje tudi na povedih, ki jih analizatorju ne uspe analizirati.

5.3.1.1 Analiza rezultatov za korpusa Šolar in Lektor

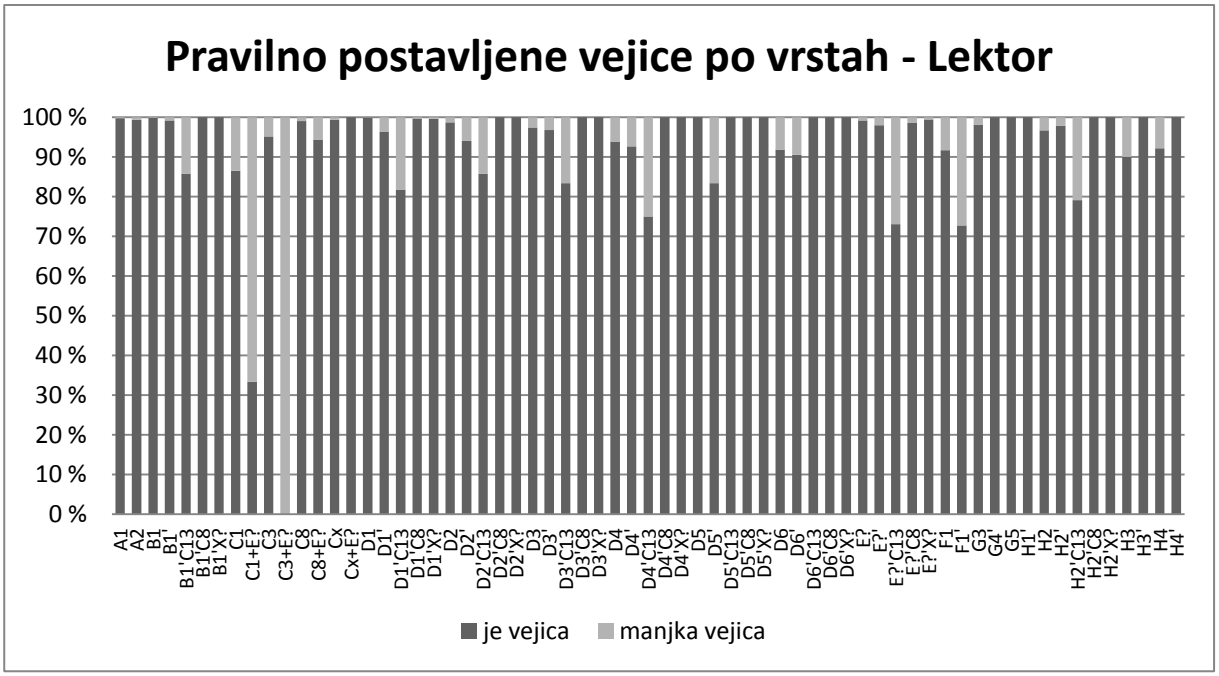
V ta namen je uporabljen program za določanje vrst vejic, in sicer tako, da se za vsako vrsto vejic ugotovi, koliko težav imajo z njo pisci in kako uspešno jo Besana popravlja. Na podlagi teh rezultatov se da ugotoviti, pri katerih analizah bi morala Besana predlagati vejico oziroma pri katerih je treba popraviti analizator, da bo uspešno deloval tudi brez vejic.



Slika 13: Deleži pravilno postavljenih vejic po vrstah v korpusu Šolar

Slika 13 prikazuje delež vejic, ki so jih učenci pravilno postavili v korpusu Šolar. Vidne so velike razlike glede na posamezne vrste (razlaga oznak vrst je v točki 4.5.2.3 na strani 42), učenci imajo zelo malo težav z vejicami pri naštevanju (A1), v večini primerov tudi

dobro postavijo vejice takrat, ko se pokrijeta levosmerna in desnosmerna vejica (kot ko npr. zaključku enega odvisnika takoj sledi začetek naslednjega); taki primeri so D1'X?, E?'X? ipd. Še vedno dobro so postavljene vejice na začetkih odvisnikov (D2, E?), pri čemer sta precej slabši vrsti D4 in D6, torej deležniški in pridevniški prilastkovi odvisniki, pa tudi pri D3 je rezultat slab. Več težav je pri zaključkih odvisnikov (npr. E?'), izrazito slabi pa so rezultati v primerih, ko zaključku odvisnika sledi vezalno ali ločno priredje, ki običajno spredaj nima vejice (D1'C13, E?'C13), in pa vrsta C1+E? (tukaj gre npr. za: »Miha bo prišel$C1$+$E?$, in če ne bo deževalo, pokosil travo.«)



Slika 14: Deleži pravilno postavljenih vejicah po vrstah v korpusu Lektor

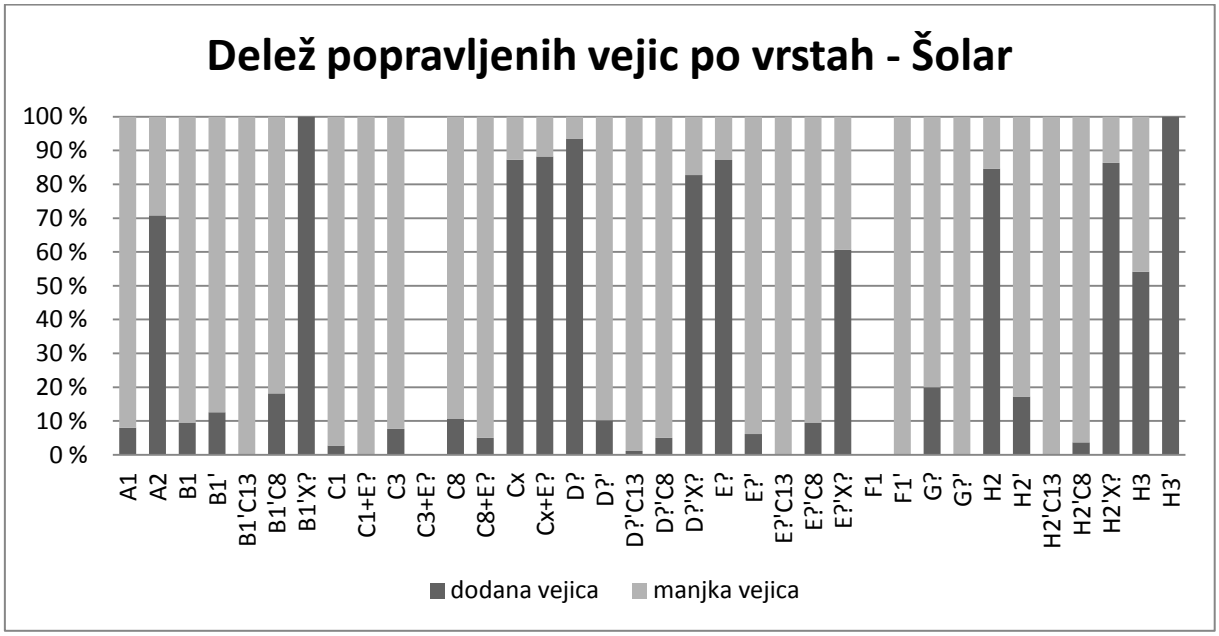
Slika 14 prikazuje deleže še za korpus Lektor. Rezultati so tu veliko boljši kot pri korpusu Šolar, izrazito slab je rezultat le pri C1+E? in C3+E? (vendar je pri slednjem v korpusu le en primer, tako da težko govorimo o statistiki). Rezultati pa so slabši tudi v primerih, ko zaključku odvisnikov sledi vezalno ali ločno priredje, tipično gre za veznik »in«. Precej manjkajočih vejic je še pri tipu F1', torej pri zaključku deležijskih polstavkov.

Primerjava med korpusoma Šolar in Lektor pokaže še, da učenci v korpusu Šolar zelo malo uporabljajo deležijske polstavke (F1) in prilastkove odvisnike z deležnikom na -č (D5).

Ker je tabela zaradi množice vrst precej nepregledna, sem se odločil za naslednje korake: podatke za različne prilastkove odvisnike v skupno oznako D?, G3, G4 in G5 združiti v G?; H4 pridružiti H2; rezultate za H1 pa izločiti, ker zaradi priprave primerov iz korpusa Šolar ni primerov premega govora.

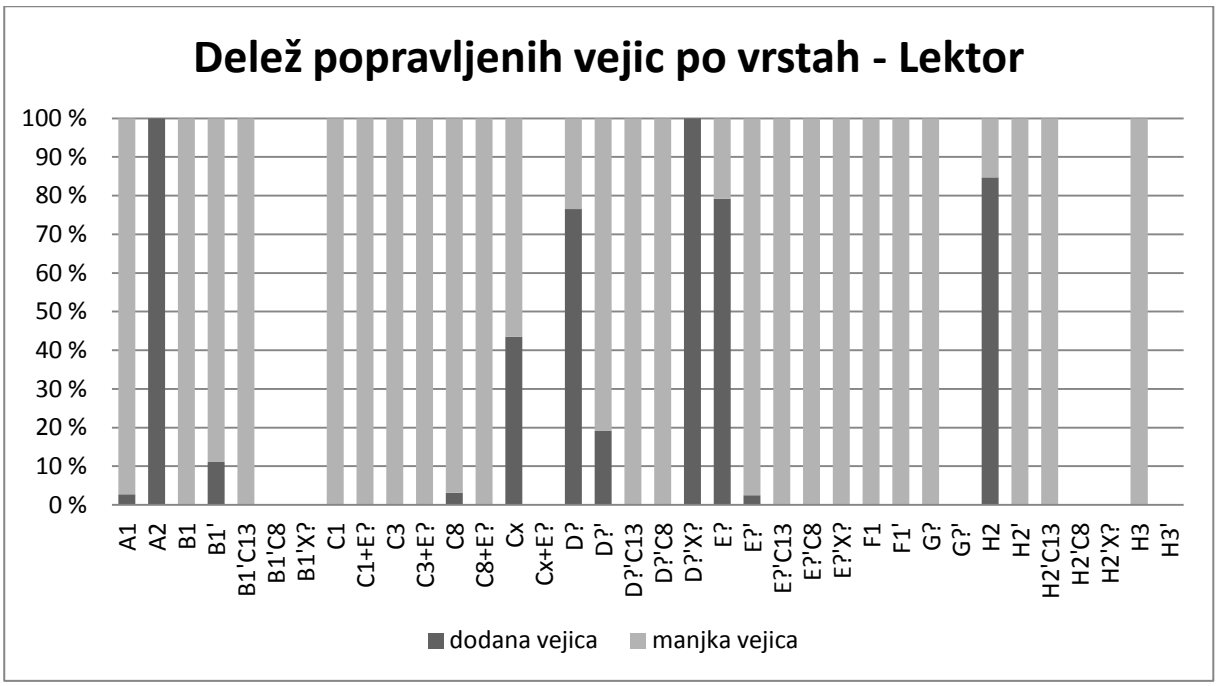
Skupaj je v rezultatih označevanja vrst vejic v korpusu Šolar 56.469 mest, kjer je potrebna vejica, ta manjka v 13.820 primerih, kar je 24,5 %. V korpusu Lektor je 59.682 mest, vejica manjka v 751 primerih, torej v 1,3 % primerov, manjkajočih vejic je v korpusu Lektor torej bistveno manj kot v korpusu Šolar.

Besana pri razvrščenih vejicah v korpusu Šolar doda 7006 vejic, torej najde 50,7 % manjkajočih vejic. V korpusu Lektor doda 230 vejic oziroma 30,6 %.



Slika 15: Delež popravljenih vejic po vrstah v korpusu Šolar

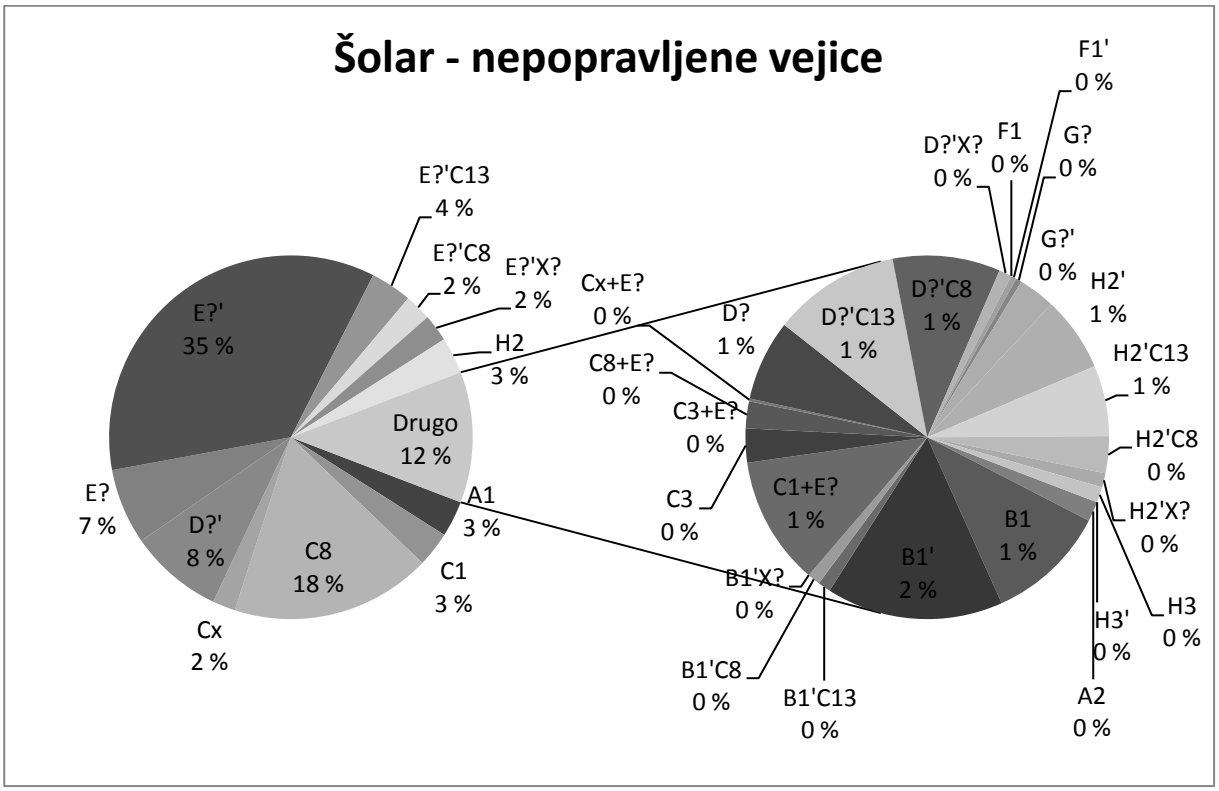
Slika 15 prikazuje, kako uspešna je Besana pri posameznih vrstah v korpusu Šolar. Vidi se, da je zelo uspešna pri začetkih odvisnikov (oziroma zaključkih odvisnikov, pri katerih takoj sledi začetek naslednjega odvisnika), pri protivnih priredjih (A2) in pri vprašalnicah v vlogi odvisnega govora (H3 in H3'). Veliko manj uspešna pa je pri levosmernih vejicah na koncu odvisnikov; pravzaprav je uspešnost popravljanja po vrstah do neke mere podobna uspešnosti samih učencev.



Slika 16: Delež popravljenih vejic po vrstah v korpusu Lektor

Slika 16 prikazuje uspešnost Besane še v korpusu Lektor. Spet dobro najde začetke odvisnikov in protivna priredja, drugih vejic pa praktično ne zmore popraviti.

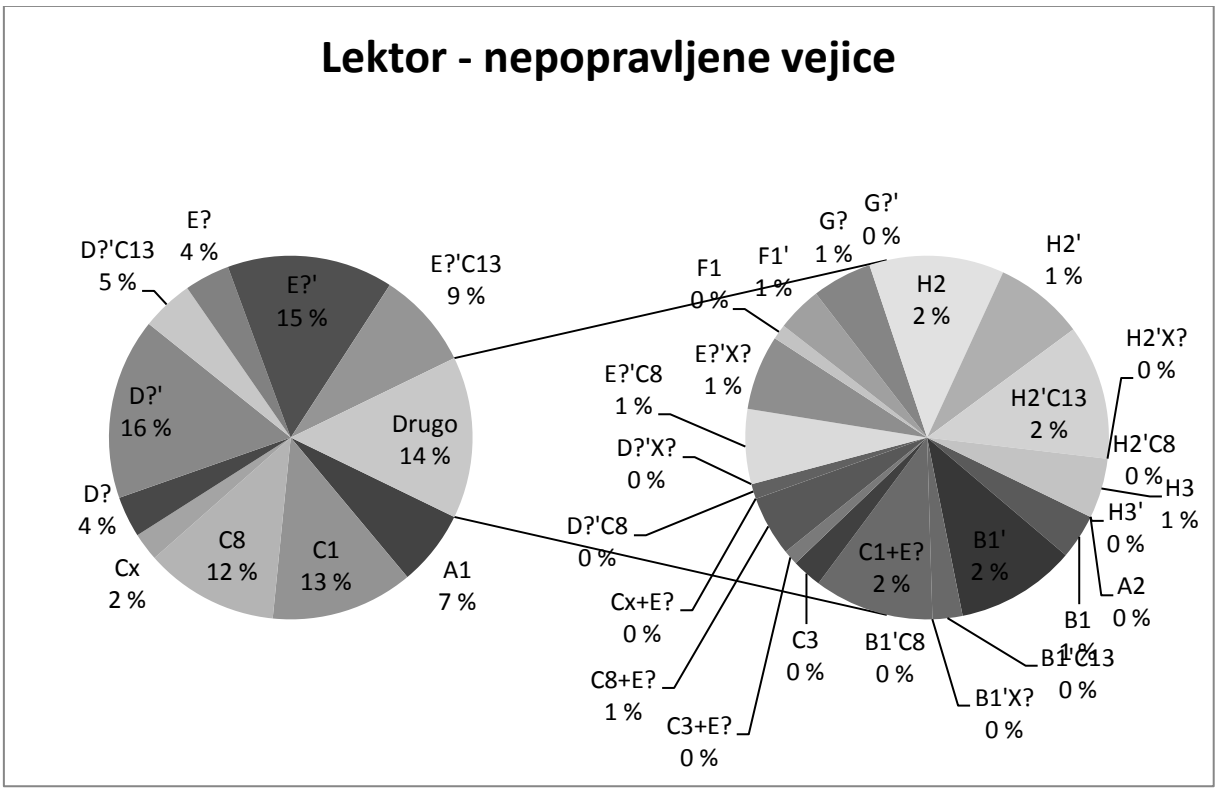
Po drugi strani pa so nekatere vrste oznak veliko manj pogoste kot druge, splača se popraviti predvsem delovanje pri tistih vrstah, ki nastopajo najbolj pogosto.



Slika 17: Deleži vrst nepopravljenih vejic v korpusu Šolar

Slika 17 nam prikazuje, katere vrste manjkajočih vejic ostanejo nepopravljene, ko izločimo primere, ki jih Besana popravi. Predvidljivo je najvišji delež pri zaključkih odvisnikov (E?' in D?'), skupaj 43 % (lahko pa bi dodali še primere, ko je na tem mestu še začetek nečesa, npr. E?'C8, E?'C13 ipd.). Vendar pa je 7 % nepopravljenih vejic tudi na začetku odvisnikov (E?), in to je treba raziskati, da se ugotovi vzrok, da v teh primerih Besana ni uspešna. Preseneča tudi visok delež pri C8 (priredje brez veznika), ki znaša 18 %. Treba je preveriti posamezne primere, da se ugotovi, ali gre tu res za take primere manjkajočih vejic ali pa ima morda program za določanje vrst težave pri razporejanju v C8.

Slika 18 prikazuje rezultate še za korpus Lektor. Tudi tu je velik delež zaključkov odvisnikov (D?' in E?'), vendar je ta delež v primerjavi s Šolarjem manjši, 31 %. Je pa podvojen delež pri E?'C13 (9 %), torej pri zaključkih odvisnikov, ki jim sledi vezalno priredje, tipično z veznikom »in«. Delež pri C8 je tudi tu visok (12 %), kar je dodaten argument, da je treba te primere preveriti. Visok je tudi delež neodkritih manjkajočih vejic pri vezalnem priredju (C1) – v teh primerih običajno ni vejice, kar kaže, da je program za razvrščanje vejic tem primerom verjetno napačno določil vrsto, kar je treba preveriti. Podobno je smiselno preveriti še A1 (naštevanje).



Slika 18: Deleži vrst nepopravljenih vejic v korpusu Lektor

5.3.1.2 Analiza problematičnih vrst

Ker vemo, katere vrste nepopravljenih vejic nas zanimajo, lahko analiziramo vsako vrsto posebej. Na seznam sta dodani še vrsti Cx in C1+E?, slednja sicer absolutno ni tako pogosta, vendar relativno dela največ težav piscem v korpusu Lektor in tudi v korpusu Šolar.

5.3.1.2.1 D?' in E?'

Nekaj primerov za to vrsto (mesta, kjer manjka ta vejica, so označena z znakom §):

5 : $D1'$: 2 : 11 : Po eni strani bralce, ki se jim je Matiček priljubil§ skrbi, kako se bo vse izšlo zanj ,in če ne bo imel sreče ter se bo moral res poročiti s Smrekarico.

5 : $D1'$: 2 : 12 : Še zadnji poizkus v njegovem življenju, ki mu je kaj pomenil§ se je izjalovil.

5 : $D1'$: 2 : 16 : Človek se bori za golo preživetje, a se od vseh travm, ki jih doživi§ ne nauči nič.

5 : $E4'$: 2 : 5 : Ker ga je punca prevarala§ je prosil Meursaulta, da ji napiše pismo in jo povabi k njemu na kavo.

Teh primerov analizator zaradi manjkajoče vejice ne uspe analizirati, saj poskuša vse analizirati kot celoto.

5 : $E3'$: 2 : 6 : Te ljubezni, kot pravi Romeo§ ne more premagati kamniti zid.

Tukaj se pokaže, da včasih analizator izsili analizo, ki pa ni pravilna, v tem primeru privzame, da je vejica pred »kot« odveč in da je »pravi« pridevnik.

5 : $D2'$: 2 : 9 : Znotraj skupine ljudi, na katere je čustveno navezan§ spoznava osnovna življenjska znanja.

5 : $D2'$: 2 : 9 : Priimek prve družine, iz katere je prihajal Romeo§ je bil Monteg.

Pri oznaki D2' je nekaj primerov, ki bi morali imeti oznako D3', do tega pride po vsem videzu zato, ker določevalnik vrst ne upošteva pravilno predlogov, ki so del glagolskih predlog (vezani na glagol).

5.3.1.2.2 D? in E?

Najdeni primeri kažejo, da so težava manjkajoče vejice za narekovaji in zaklepaji:

5 : $D1$: 2 : 16 : Vojna je huda stvar, hudo zlo, ki se zgodi zaradi nekaterih „norcev”§ ki imajo od tega korist.

5 : $D1$: 2 : 11 : Tonček je študent (to v današnjem času pomeni gimnazijec)§ ki so mu všeč vse ženske.

5 : $D1$: 2 : 6 : V romanu „Visoška kronika„§ ki ga je napisal Ivan Tavčar, je zgodba, ki govori o Izidorjevi družini.

5 : $E1$: 2 : 8 : Vsi poznamo zgodbo „Cvetje v jeseni”§ kjer sta močno zaljubljena Janez in Metka.

Veliko je težav pri dvobesednih veznikih:

5 : $E2$: 2 : 11 : Slabo dejanje pa je, da je edini odšel iz boja§ medtem ko so drugi pomrli, in namesto da bi bil junak, je izpadel strahopetec, in da je zapustil svojo vero.

5 : $E7$: 2 : 10 : Ker ni upoštevala kraljevih ukazov, je tudi zelo tvegala§ samo da bi pokopala brata.

Nasploh so očitne težave v primerih, ko so pred vezniki (potencialni) členki:

5 : $E3$: 2 : 6 : Starejši sin je seveda ravnal tako§ kot se mu je tisti trenutek zdelo potrebno, vendar je spregledal neko pomembno stvar, da je njegov oče pravičen, in če bi ga on prosil za kozliča, bi mu ga oče dal.

5 : $E4$: 2 : 4 : Zaključek bi spremenil zato§ ker bi rad izvedel, kako bi se nadaljevalo.

5 : $E7$: 2 : 10 : Njegov lastni stric mu je povzročil te težave le zato§ da je on lahko postal kralj.

5 : $E7$: 2 : 3 : Mislil je torej§ da lahko počne, kar hoče.

V nekaterih primerih zgreši tudi analizator in določi napačno vrsto vejice:

7 : $E8$: 2 : 6 : Zberite to vrsto papirja ločeno in§ če je mogoče, določite skupni tiskalnik za uporabo tega papirja (tako da enostransko natisnjen papir ne bo pomešan s »čistim«).

V nekaterih primerih pa na prvi pogled ni jasno, zakaj ima analizator težave, domnevati je možno, da v prvem primeru dela težave to, da je »če« tudi pogovorna oblika glagola »hoteti«, v drugem pa dvoumnost pri besedi »kar«, ki je lahko tudi členek:

5 : $E8$: 2 : 5 : V nekaterih državah po svetu§ če kdo koga ubije, gre ta na električni stol.

5 : $E7$: 2 : 5 : Verjetno zato, ker vse§ kar je napisano in tudi znano, je tudi plačano.

5.3.1.2.3 E?'C13 in D?'C13

Glede na to, da je postavljanje levosmernih vejic na koncu odvisnikov že samo po sebi težek problem, kar kaže točka 5.3.1.2.1, se ne gre čuditi temu, da je v primerih, ko temu sledi priredni veznik, pred katerim običajno ni vejice, težava še večja:

5 : $E1'$$C1$: 2 : 12 : Osrednja prireditve potekajo na mestih, kjer se posamezna športna panoga izvaja§ in v olimpijski vasi, kjer stanujejo športniki.

5 : $E1'$$C1$: 2 : 12 : Pred pogrebom je Lorenzo odšel v mrtvašnico, kjer je imel prijatelje§ in je odpeljal Julijo v samostan, kjer jo pričaka Romeo.

5 : $E2'$$C1$: 2 : 9 : Lahko je govoriti, ko nisi v takih okoliščinah§ in modrovati o svoji lastni moči, a resnica je velikokrat drugačna.

5 : $E2'$$C1$: 2 : 9 : Zdaj je že nekaj let, odkar je odšel§ in ni mu žal, da je odšel.

5 : $E2'$$C1$: 2 : 11 : Ljudje so se zaljubili, ko je bil še čas renesanje§ in ljudje se zaljubamo sedaj.

5 : $E4'$$C1$: 2 : 7 : Dano besedilo je epika, ker pripoveduje§ in je roman, ker je daljše besedilo.

5 : $D1'$$C1$: 2 : 9 : Vedno več je ljudi, ki plavajo s tokom§ in vedno manj je tistih, ki izrazijo svoje mnenje, svoj prav.

5 : $D1'$$C1$: 2 : 12 : Kralj Ojdip je nekako izvedel za grozoto, ki jo je storil§ in se je zato dal oslepiti in je šel umret.

Popolnega postavljanja vejic te vrste praktično ni mogoče narediti z računalnikom, in sicer že zato ne, ker je v nekaterih primerih tako, da je poved možna z vejico ali brez nje, spremeni pa se pomen (kot v primeru: »General ukaže vojaku, naj pospravi dvorišče in odide.«). Vendar naloga ni tako nemogoča, če pogledamo strukturo celotne povedi. V zadnjih dveh primerih je tako možno sklepati, da bi vejica morala biti, ker se drugače v naštevanju prilastkovih odvisnikov pri izpuščenem prilastkovem zaimku »ki« spremeni osebek, kar je malo verjetno. Pri drugem primeru pa ne moremo storiti kaj dosti, ker je poved »*Pred pogrebom je Lorenzo odšel v mrtvašnico, kjer je imel prijatelje in kjer je odpeljal Julijo v samostan, kjer jo pričaka Romeo.*« slovnično popolnoma smiselna (če odmislimo, da bi moral namesto drugega »kjer« stati »od koder«).

5.3.1.2.4 C1

Primeri kažejo, da so vejice pri tej vrsti primerov večinoma posledica napačnega določanja vrste, pri katerih manjka še prvi del, ki je tipično zaključek odvisnika:

5 : $C1$: 2 : 6 : Postane mu vseeno, kaj bo§ in ravno, ko si hoče vzeti življenje, zasliši materin glas.

5 : $C1$: 2 : 12 : Simon predstavlja nas, ki smo vsi notri in kričimo in brcamo§ in nič ne pomaga.

5 : $C1$: 2 : 13 : Občutek, da lahko narediš karkoli, greš, kamor te je volja§ in brez strahu izražaš svoje mnenje.

V nekaterih primerih je spregledana tudi vrsta C1+E?:

5 : $C1$: 2 : 11 : Ko jo gregor vidi, ga prevzame „brezumna slast"§ in ker jo mora ubiti, „zadnja groza".

5 : $C1$: 2 : 6 : Postane mu vseeno, kaj bo§ in ravno, ko si hoče vzeti življenje, zasliši materin glas.

5 : $C1$: 2 : 8 : Vsi so bili ujeti v totalitarni komunistični režim§ in kdor je bil sumljiv, je bil odstranjen.

Kot potencialna dopolnitev za postavljanje vejic pa so primeri, kjer gre res za vezalno priredje in je pred njim vzročno priredje:

5 : $C1$: 2 : 24 : Bubi vojne in okupacije ne jemlje tako resno, kot bi jo moral, saj zna govoriti nemško, malo italijansko in seveda slovensko§ ter je prepričan, da se lahko, v primeru nesporazuma, zmeni, kar se mora.

5 : $C1$: 2 : 11 : Odrezali so mu nogo, saj jo je imel Drohojewski tudi§ in sam Simon je začel verjeti, da je res Drohojewski in začel govoriti v poljskem jeziku.

Lahko pa gre tudi za protivno (ali pa posledično) priredje z veznikom »in«, česar analizator še ne zna ločiti:

7 : $C1$: 2 : 13 : Če pa že morajo pisati teste znanja, naj bodo naloge objektivnega tipa§ in ne esejistične.

5.3.1.2.5 C8

Pri prvem primeru te vrste se pokaže, da je treba analizator dopolniti s tem, da je v primeru, da glagolska predloga pričakuje nedoločnik, namesto tega lahko tudi kazalni zaimek »to«:

5 : $C8$: 2 : 4 : In kdor to zmore§ je pravo drevo v hudourniku, ki upogiba veje.

Nasploh večina primerov kaže na težave analizatorja, ki ne identificira odvisnikov spredaj (pri »pozanimati se« npr. manjka ustrezna glagolska predloga):

7 : $C8$: 2 : 5 : Kako nemočno je Društvo narodov§ se je pokazalo tudi 1935, ko so Italijani napadli Etiopijo.

5 : $C8$: 2 : 12 : Težko se vživim v njihove situacije, ampak kar so oni storili§ je predvsem razumljivo.

5 : $C8$: 2 : 5 : Šele ko sem se pozanimal§ sem ugotovil, kaj se je dogajalo pred začetkom.

Nekaj pa je res napak te vrste, vendar gre za relativno redko napako, zato se ne bomo posebej ukvarjali z njo:

5 : $C8$: 2 : 5 : Lojzki je to bilo nagnusno§ ni želela imeti družino, ki se dobro razume samo pred drugimi gosti.

5.3.1.2.6 A1

Pri naštevanju se pokaže, da imajo pisci težave pri vejicah predvsem pri pridevnikih:

5 : $A1$: 2 : 8 : Njegov značaj je opisan kot grob, nasilen§ pohlepen.

7 : $A1$: 2 : 2 : V sodobnem§ vsestransko povezanem svetu je Podsaharska Afrika v marsikaterem pogledu še vedno odrinjena na rob svetovnega dogajanja.

7 : $A1$: 2 : 9 : "se s pomočjo novih socialnih veščin naučijo novih§ prilagojenih načinov samoocenjevanja in samopotrditve;"

Vendar je postavljanje vejic v teh primerih za računalnik zelo težko, potrebna bi bila klasifikacija pridevnikov, da bi program lahko ugotovil, kdaj so zaporedni pridevniki dovolj sorodni, da zahtevajo vmes vejico. Te klasifikacije v bazi Ases ni in tako je mogoče le predlagati, da se Ases sčasoma dopolni s tem, kar bo omogočilo, da Besana opozarja na vejice tudi v teh primerih.

Primeri kažejo tudi na to, da ima analizator še vedno težave pri razločevanju med naštevanjem in pristavki v primerih, ko je pristavek na koncu stavka in mu zato ne sledi vejica:

5 : $A1$: 2 : 3 : Ima dva sinova§ starejšega Izidorja ter mlajšega Jurija.

7 : $A1$: 2 : 18 : "Drugo skupino pa sestavljajo stavki, ki imajo nezadostno slovnično strukturo; oseba izpusti en del stavka§ in to pogosto bistveni del."

Pokažejo pa se tudi napačne analize povedi, ki mejo med stavkoma zamešajo za naštevanje:

5 : $A1$: 2 : 18 : Komedija se tako konča s tem, ko se baron in baronica pobotata, in z dvema porokama§ večno ljubezen si končno obljubita Matiček in Nežka in Žužek in Smrekarica

5 : $A1$: 2 : 15 : Kajti ljubezen v obeh deli, v Romeu in Juliji in prav tako v Hamletu§ v renesančnem obdobju ljubezen premaga vse uvire.

5 : $A1$: 2 : 6 : Primarna socializacija je prva stopnja socializacije§ Poteka v zgodnjem otroštvu znotraj družine.

5.3.1.2.7 Cx

Primeri kažejo, da je največ težav z vejicami pred veznikom »saj«:

5 : $C5$: 2 : 9 : Sam sem ode zgodbi doživljal kot nauk za življenje§ saj imata obe zgodbi nek nauk, ki je povedan na zelo preprost način, tako da ga lahko vsak razume.

5 : $C5$: 2 : 6 : Sicer pa je oče zelo usmiljen§ saj ljubi svojega sina in mu zato odpusti vse njegove napake, ki jih je naredi, ker ga je zavedla strast oz. želja po užitku.

5 : $C5$: 2 : 18 : V tistem času so se ljudje tepli za preživetje, bali so se lakote, hudih razmer …§ saj je potekala tridesetletna vojna zoper katoliške in protestantske cerkve.

5 : $C5$: 2 : 8 : Moja spoznanja o ljubezni so manjša od Antigoninih§ saj kar nekaj ljudi sovražim ali mi niso posebej pri srcu.

5 : $C5$: 2 : 21 : Imam zastavljene visoke življenske cilje (priti na OI 2014 v Sočiju, priti med tri v svetovnem pokalu …)§ saj vem, česa sem zmožna.

Verjetno težave povzroča dvoumnost besede »saj«, ki je lahko tudi oblika samostalnika »saje«, vprašanje pa je, zakaj analizatorju teh primerov ne uspe analizirati. Pri drugem primeru (»Sicer pa je oče ...«) se npr. izkaže, da dela težave napaka pri besedi »naredi« (namesto »naredil«); če se ta napaka popravi, Besana takoj opozori na manjkajočo vejico pred »saj«, kar je pravzaprav velik uspeh, saj je stavek »Sicer pa je oče zelo usmiljen saj ljubi svojega sina.« slovnično pravilen tudi brez vejice, če predpostavimo, da sta »saj« in »ljubi« samostalnika (prvi v imenovalniku, drugi v dajalniku ednine).

5.3.1.2.8 C1+E?

Pokazalo se je, da analizatorju ne uspe pravilno analizirati strukture povedi tega tipa, in to niti v primerih, ko so vejice na pravem mestu:

5 : $C1$+$E2$: 2 : 4 : Sam je nezvest mož§ in ko ugotovi, da se njegova žena srečava z neznanim gospodom, je še ljubosumen.

5 : $C1$+$E2$: 2 : 3 : Pripravil je igro§ in ko se je Klavdij zavedel, kaj prikazuje, je zapustil prizorišče, tako Hamlet ni imel nobenega suma več o tem, kako je umrl njegov oče.

Zato je treba dopolniti analizator, da bo veznik »in« pravilno prisodil glavnemu stavku in ne vrinjenemu deležniku.

5.3.1.3 Predlog izboljšav analizatorja in Besane

Na podlagi analize problematičnih primerov iz prejšnje točke je treba v analizator povedi dodati naslednje:

- Analizator na mestih, kjer bi lahko tipično bila vejica (za začetek so to osebne glagolske oblike in naslonske oblike osebnih zaimkov), predvidi možnost, da tam manjka vejica. Analize s to manjkajočo vejico morajo imeti manjšo prioriteto, prednost naj imajo analize, kjer vejica ne manjka.
- Opozarjanja na manjkajoče vejice na začetku odvisnikov ne smejo motiti narekovaji in zaklepaji.
- Na vejice na začetku odvisnikov je treba opozoriti tudi, kadar je spredaj členek.

- Vsaj za nekatere primere je treba dodati preverjanje smiselnosti nadaljevanja pri »in« brez vejice za koncem odvisnika.
- Pri vezalnem priredju za pojasnjevalnim priredjem obstaja možnost, da manjka vejica.
- Analizator mora pri elementu NED dovoliti tudi kazalni zaimek »to«.
- Povečati je treba verjetnost, da je beseda »saj« veznik, in preveriti, zakaj v primerih iz točke 5.3.1.2.7 analizatorju ne uspeva ustrezno razdvoumljanje.
- Vgraditi je treba podporo za povedi tipa C1+E?, veznik »in« ne sme biti del odvisnika v teh primerih, ampak mora postati del naslednjega stavka. Podporo je treba vgraditi tako za primere, ko vejice ni, kot tudi za primere, ko vejica je, in to tako za pravilno možnost, ko je vejica pred »in«, kot tudi za napačno možnost, ko je vejica za »in«.

Dolgoročno je treba bazo Ases dopolniti s podatki, ki bodo omogočali dovolj podrobno klasifikacijo vejic, da bo mogoče opozarjati na vejice med pridevniki.

Dopolniti je treba še program Vejicar, da bo pravilno označeval vrsto D3 namesto D2 v primerih, ko je oznaka oziralnosti različna od »a« (različna je v primeru, ko je predlog del glagolske predloge).

5.3.1.4 Rezultati izboljšav analizatorja in Besane

Kot rezultat izboljšav se je na vseh testnih primerih število najdenih manjkajočih vejic povečalo z 9062 na 10825 (za 1763), torej za 19 %. Ob tem se je sicer število opozoril, da nekje manjka vejica, zmanjšalo s 1961 na 1116 (za 845), vendar so, prvič, pomembnejša točna opozorila, in drugič, napredek je velik, tudi če upoštevamo to zmanjšanje.

Kot bolj problematično pa se kaže, da se je poslabšalo iskanje odvečnih vejic (s 1050 na 883). To poslabšanje je treba nujno preveriti in ustrezno dopolniti Besano (prvo preverjanje kaže, da je težava pri vejicah pred »kot«, eden od primerov, kjer je prej Besana pravilno označila odvečno vejico, zdaj je pa ne več, je: »Raje bi ga nosil, kot poslušal.«). Poslabšala se je tudi natančnost pri iskanju manjkajočih vejic, saj je občutno več lažnih opozoril (iz 2770 na 3383), in tudi to je treba še preveriti.

Na splošno pa je (še posebej, če se bodo dali omiliti stranski učinki popravkov) napredek zadovoljiv. Kot zelo uspešno se je pokazalo odkrivanje levosmernih vejic na koncu odvisnikov, pri vejicah pri vmesnem odvisniku za prirednim veznikom (C1+E?) pa se je pokazalo, da bo treba za uspešno detektiranje še dopolniti analizator povedi, čeprav pri

posameznih primerih že deluje (npr. »Pripravil je igro§ in ko se je Klavdij zavedel, kaj prikazuje, je zapustil prizorišče, tako Hamlet ni imel nobenega suma več o tem, kako je umrl njegov oče.«).

Smiselno je, da se ta način izboljševanja Besane še iterativno ponavlja, vendar je treba pred tem še izboljšati analizator povedi.

5.3.1.5 Analiza napak pri izboljšavah analizatorja in Besane

Najbolj problematično je bilo veliko poslabšanje iskanja odvečnih vejic. Pokazalo se je, da je vzrok pravzaprav banalen, pri dodajanju nove vrste odvečnih vejic v Vejicar je bil v pogoju pomotoma napisan operator && (in) namesto || (ali), kar je povzročilo, da se niso pravilno označile vejice, označene kot odvečne pred besedo »kot«.

Pri lažnih opozorilih se je pokazalo, da je bil po eni strani problematičen veznik »torej«, in sicer zaradi tega, ker se prekriva s pogostim členkom »torej«. Zato je bila v analizator dodana omejitev, da vejica pred »torej« manjka le v primeru, če se da njegova okolica potem uspešno analizirati.

Dodane so bile še omejitve za iskanje levosmernih vejic na koncu odvisnikov, na primer to, da vejica ni med »si se«, »sem se«, »sem si« ipd. Zaradi tega sicer lahko izgubimo nekaj najdenih manjkajočih vejic, vendar se je pokazalo, da je šlo v večini takih primerov za napačno interpretacijo povedi, kar je potem povzročilo, da je Besana vrivala odvečne vejice (to se je dogajalo v primerih, ko Besani ni uspelo analizirati celega stavka, ker npr. ni imela ustrezne glagolske predloge, z delnima stavkoma pa ni imela težav).

Po teh popravkih je rezultat sledeč: število najdenih manjkajočih vejic se je sicer zmanjšalo na 10771 (z 10825), vendar se je hkrati močno zmanjšalo število napačno najdenih vejic (s 3383 na 2983), kar pomeni, da se je izboljšala natančnost, ki je tudi zelo pomembna. Število najdenih odvečnih vejic je zraslo na 1149 (z 883), pred začetkom popravkov je bil ta rezultat 1050. Izboljšanje je torej tudi tu precejšnje, čeprav smo se ukvarjali predvsem z manjkajočimi vejicami, in je posledica tega, da zdaj pri dveh vrstah napak vejico prestavimo (pri sestavljenem vezniku »kljub temu da« izpred »da« pred »kljub«; pri vmesnih odvisnikih za vezalnim priredjem pa izza »in« pred »in«).

5.3.2 Uporaba okolice napak

Ker analizator povedi še ne zmore uspešno analizirati vseh povedi, bi bilo treba poskusiti uporabiti tudi dopolnila pravila na podlagi sosednjih besed, podobno kot je to narejeno za LanguageTool v točki 5.2.1.

5.3.2.1 Spregledane manjkajoče vejice

Spet s pomočjo postopka, opisanega v točki 4.5.1, dobimo seznam najpogostejših besed, pred katerimi je manjkajoča vejica, ki je Besana ne najde.

Tabela 14: Najpogostejše besede za manjkajočimi vejicami, ki jih Besana ne odkrije

beseda	1	2	3	4	5	6	7	8
in	188	1115	46	410	1624	144	44793	144
pa	15	244	65	79	1165	95	13301	104
saj	**507**	**205**	**0**	**2**	**3720**	**0**	**43**	**0**
se	306	196	44	41	1337	9	28938	62
je	712	191	70	69	2123	20	57086	127
kot	272	188	199	220	1499	69	5635	175
ali	93	138	12	138	482	4	4369	114
kar	**316**	**120**	**0**	**30**	**2001**	**16**	**897**	**77**
ne	42	91	10	64	2079	17	8866	42
da	2050	82	75	399	18930	10	2332	232
ga	0	82	9	16	226	0	5326	3
naj	214	79	3	2	401	5	1585	28
v	239	77	3	65	1978	6	31380	18
lahko	0	66	3	7	370	1	5685	1
ni	30	63	3	9	284	2	6493	4
jo	0	61	1	3	111	0	3168	1
to	0	61	1	8	487	3	6105	0
ter	5	60	60	267	124	14	2144	7
kaj	350	59	0	5	801	1	872	81
tako	20	59	1	47	682	3	4306	43
vendar	**177**	**59**	**0**	**0**	**1498**	**1**	**166**	**14**
so	209	58	32	22	708	5	16052	22
zato	130	57	0	20	1299	2	996	11
”	0	57	0	0	49	0	17	0
sem	173	56	16	9	323	7	5175	12
mu	0	55	3	0	129	0	3104	1
tudi	0	55	8	16	279	0	9532	13
na	47	54	0	33	985	1	17540	11
ko	412	44	5	167	2724	1	770	108
še	3	42	0	14	296	3	4733	7
jih	0	41	3	1	137	0	3808	1
samo	1	41	0	4	85	1	1664	2
mi	0	38	2	6	55	1	2376	0
si	40	36	4	9	173	0	3978	9
npr	0	34	0	1	190	0	297	12
za	51	31	1	29	555	0	13075	11
a	211	29	0	3	1508	0	45	194
me	0	29	0	5	60	0	1152	0

z	4	28	0	21	261	0	9471	1
ji	0	27	0	1	42	0	1243	1
kako	498	26	0	1	1086	1	249	32
sta	80	26	6	11	143	0	3144	6
bi	136	25	27	20	362	4	9319	20
le	0	25	1	7	202	0	2657	1
ima	10	23	3	1	82	1	1668	1
potem	17	23	0	2	276	0	488	6
ker	349	22	4	29	2338	0	194	48
po	8	22	0	28	486	1	4846	5
s	103	22	0	27	735	0	6434	12
smo	36	22	3	6	128	2	2597	1
on	0	21	0	0	100	0	790	0
nato	6	20	0	0	233	0	272	4
zelo	0	20	1	7	65	0	3749	0
vse	0	18	0	3	143	0	2715	0
jaz	0	17	2	9	224	0	820	0
ona	0	17	0	0	68	0	403	0
»	0	17	0	0	22	0	165	0
bo	35	16	0	0	95	1	3054	1
iz	13	16	0	6	175	0	4078	3
te	0	16	0	4	46	0	1393	0
jim	0	15	0	0	48	0	1204	1
kam	**15**	**15**	**0**	**0**	**30**	**0**	**31**	**0**
ti	0	15	0	7	68	0	870	2
spada	0	14	0	2	12	0	246	0
do	4	13	2	39	133	1	4393	3
nam	0	13	2	2	46	0	1050	0
ta	0	13	2	3	147	0	1613	0
mislim	0	12	1	1	213	1	270	0
od	10	12	0	8	251	0	4447	2
predvsem	0	12	0	2	190	0	762	0

Tabela 14 prikazuje besede in števce po posameznih stanjih pred to besedo (Tabela 4 opisuje pomen posameznih stanj), besede so urejene po števcu za stanje 2, ki pove, kolikokrat je pred navedeno besedo manjkala vejica, ki je Besana ni odkrila.

Zanimajo nas predvsem primeri, kjer je vrednost v stolpcu 2 veliko večja od vsote stolpcev 7 in 8 (kolikokrat manjka vejica v primerjavi s tem, da vejice pred navedeno besedo ni, ker je ni treba).

Daleč največ je neugotovljenih manjkajočih vejic pred »in«, in sicer 1115; Besana pred »in« pravilno postavi 188 vejic, napačno pa 144, kar pomeni, da je natančnost teh opozoril že zdaj slaba.

Kot kandidatke za dodatno postavljanje vejic se kažejo predvsem besede »saj«, »kar«, »vendar« in »kam«, a edino pri »saj« se izkaže, da je število primerov, pri katerih vejice ne sme biti, manjše od števila nenajdenih manjkajočih vejic, tako da je »saj« pravzaprav edini pravi kandidat. Imamo pa še besedi »ga« in »jo«, ki sta kandidatki za levosmerne vejice na zaključku odvisnikov.

Zanimiv je rezultat pri »ki«. Če pogledamo primere, pri katerih Besana postavi odvečno vejico, dobimo tele primere (to je vseh 6 primerov, vejica pred »ki« pa sicer ni bila postavljena v 3 primerih, v katerih bi morala biti):

Tudi sem dobil novi prijatelji§ ki smo dobri in smešni.

Prav tako v dvoboju zgresen vbod nima hitrega povratka oz. velike moznosti za nadalne kombinacije, vbod§ ki zadane, je sicer navadno smrten, vendar nima hitrega povratka v blok (npr. v boju z več nasprotniki).

Lenizacija je znana v beneški italijanščini in furlanščini, izguba labialnega dela labiovelara je pogostejša v furlanščini kot beneščini, palatalizacija iz tega nastalega§ ki> čipa je samo furlanska, prim. ben. it. che : furl.

Na to sem jaz dodal, da so argumenti, s katerimi se utemeljujejo nekatere genetske naddružine ali file, ki se§ ki se navajajo tudi v standardnih priročnikih, kakršni so penutijski jeziki (da ne govorimo o Greenbergovi »amerindijščini«) šibkejši od tistih, na katerih je Illič-Svityč osnoval nostratsko jezikovno naddružino.

Zvezdica na levi strani stično pred besedo ali delom besede§ ki je tiskan ležeče, pomeni, da je to, kar ji sledi, rekonstruirano: sln. krava < psl.

, vsakič, ko je blo treba rezat, ko je blo treba vezat, ko je blo treba kopat, ko je blo treba potem škropit, z galico, in trgat, takrat je ona šla iz Novga mesta, to je bla ura hoda, vsak dan, v Trško goro, potem, kadar so bla dela§ ki so trajala več dni, je ostala v Trški gori, kadar pa ne, kadar je pa bil samo en dan, se je potem vračala domov.

Razen tretjega primera gre povsod drugod za napake v zbirki primerov.

5.3.2.2 Odvečne manjkajoče vejice

Poiščemo seznam besed, pri katerih Besana največkrat postavi vejico, ki v resnici ni potrebna (stanje 8); s tem skušamo poboljšati natančnost.

Tabela 15: Najpogostejše besede, pred katere Besana postavi odvečno vejico

beseda	1	2	3	4	5	6	7	8
da	2050	82	75	399	18930	10	2332	232
a	211	29	0	3	1508	0	45	194

kot	272	188	199	220	1499	69	5635	175
če	277	12	1	67	1603	2	575	165
in	188	1115	46	410	1624	144	44793	144
je	712	191	70	69	2123	20	57086	127
ali	93	138	12	138	482	4	4369	114
ko	412	44	5	167	2724	1	770	108
pa	15	244	65	79	1165	95	13301	104
kaj	350	59	0	5	801	1	872	81
kar	316	120	0	30	2001	16	897	77
se	306	196	44	41	1337	9	28938	62
ker	349	22	4	29	2338	0	194	48
tako	20	59	1	47	682	3	4306	43
ne	42	91	10	64	2079	17	8866	42
kako	498	26	0	1	1086	1	249	32
kdaj	29	6	0	1	51	0	157	29
naj	214	79	3	2	401	5	1585	28
A	0	1	0	0	129	0	82	27
ampak	279	2	0	1	947	0	1	26
A=	0	0	0	1	13	0	3	23
dokler	45	0	0	1	137	0	19	23
so	209	58	32	22	708	5	16052	22
kdo	37	9	0	1	97	1	229	22
bi	136	25	27	20	362	4	9319	20

Tabela 15 prikazuje dobljeni rezultat. Zelo velik je delež preveč postavljenih vejic pred »a«. Pregled primerov pokaže, da so težava citati iz angleščine, npr. »World as ***a chess game« in francoščine »gens nés ***a Lausanne sont«. Za rešitev tega bi bilo treba analizator dopolniti tako, da bi odkrival tujejezične citate, zasilna rešitev pa je, da se pogledajo sosednje besede, na primer dve spredaj in dve zadaj, in če so med njimi vsaj 3 besede, ki jih v slovenščini ni (oziroma jih analizator ne prepozna), lahko rečemo, da tu ni smiselno zahtevati vejice.

Katastrofalno slab je rezultat pri besedi »pa«, Besana le 15-krat pravilno opozori na manjkajočo vejico, hkrati pa so tu 104 opozorila preveč, in glede na ta rezultat je najbolj smiselno, da Besana sploh ne poskuša postavljati vejic pred »pa« (pri »pa« so težava še izpusti, ki jih analizator še ne prepozna, zato potem pri njih lahko najde zelo nenavadne interpretacije povedi).

Zanimivo je, da je veliko postavljenih odvečnih vejic pri veznikih, ki sicer zelo zanesljivo zahtevajo vejice (da, če, ko, ker), zato je koristno te besede preverjati s pomočjo konkordančnika.

Pri »da« se pokaže, da je največ preveč predlaganih vejic v primerih, ko so pred »da« besede »pa«, »se«, »s tem«, »res«, »da« in »lahko«, pojavljajo se pa tudi značilni vezniki, kot so »ampak«, »če« ipd. Pri »ko« je izrazito število primerov z »bolj«, pojavljajo pa se tudi s »s tem«, dodatno pa še z »a«, »da« in »tudi«. Pri »ker« imamo »zato«, »pa«, »a«.

Ročni pregled teh primerov je pokazal tudi na nekaj primerov napak v zbirki primerov (mesta, kjer bi morala biti oznaka, da manjka vejica, so označena z znakom §):

Jaz mislim§ da Mladinska poletna šola bo dobra za mene zato¤ ker jaz hočem naučit bol slovenščino.

Mislim§ da slovenska šola je dobra zato¤ ker učil vsem slovensko in pomaga da kultura živi.

Imam dva brata in dve sestre, Jaz pričakujem§ da jaz bomo govorila še slovenščina z moji starši÷ in moji staristarši.

Pričujem§ da bom največ poznal slovenijo¤ domovino¤ kjer so bili rojeni moji stari starši in hvala njem ima to možnost.

Smrtni so le§ če zadanejo kritično točko (vrat, zile).

Vsi ljudje imamo svoje sanje, vendar kaj§ ko nam usoda vedno znova pokaže, da ni vse tako lepo.

Zelo slaba natančnost je tudi pri besedi »kdaj«. Vendar je tu premalo primerov za dobro statistiko, trije primeri so za »se« in »še«. Je pa tudi nekaj primerov, kjer bi pravzaprav sam postavil vejico, npr. »Ne vem kdaj, …« (če ni v smislu »neznano kdaj«), zanimiva pa je še fraza »kdaj pa kdaj«. Vendar se izkaže, da je ta fraza že vnesena v Ases, težava pa je v primeru:

Vsi se kdaj pa kdaj sprašujemo kam bomo odšli ko bomo umrli in nas več nebo

Preizkus pokaže, da to napačno obvestilo povzroči to, da manjka vejica pred »kam«, zaradi česar analizator povedi ne zna pravilno določiti analize povedi. Ker je beseda »kam« zanimiva kandidatka tudi pri spregledanih manjkajočih vejicah, jo je smiselno dodati v analizator na seznam mest, kjer je pogosta meja med stavkoma v povedi. Popravek povzroči, da Besana v tem primeru popravi vejici.

Smo pa tudi našli dva primera neoznačenih napak v zbirki primerov:

Tudi mi je bilo všeč§ kdaj smo pogledali nastop dijakov iz XDržavaX in XDržavaX¤ ker so bili zelo veseli in so imali barvaste obleke¤ katere so mi bile všeč¤ ker so drugače kot naše.

Smrt je jemal kot nekaj normalnega¤ in če je človeku namenjena¤ ni važno§ kdaj in kako ga doleti.

Preverili smo tudi primere, ko je taka vejica postavljena pred »in«. Pokazalo se je, da gre v veliko primerih za napake v zbirki primerov, kjer niso bile označene manjkajoče vejice pred »in«:

Doma se učimo od prvi razred§ in dokler ne pridemo sem¤ ne vemo¤ kaj se učimo.

Že prej je napisal Županovo Micko, vendar se je pri drugi bolj osredotočil na tedanjo družbo§ in da bi bilo vse še malo bolj zapleteno, je ustvaril štiri ljubezenske kroge.

S tem je Linhart prikazal¤ kako se plemiči niso zanimali za nič drugega kot le, kako bi bilo njim najbolj lepo§ in čeprav je bilo od njih odvisno toliko življenj, se tega sploh niso zavedali.

Po njeni smrti se je poskusil ponovno postaviti na noge§ in ko se mu je že nasmehnila kapljica sreče, je kruta usoda odigrala svojo vlogo in mu pobrala še tisto malo upanja, kar ga je še ostalo.

Vendar pa gre še vedno za eno življenje§ in to osebe, ki ti nekaj pomeni.

"fenilketonurija; to je dedna motnja, pri kateri manjka eden od encimov za presnovo aminokisline fenilalanina v krvi§ in ker ni encima za to pretvorbo, se fenilalanin kopiči v krvi in toksično deluje na možgane oziroma povzroča duševno motnjo;"

Glasbeni narodopisci so se praviloma osredotočali na ljudsko glasbo podeželjskega prebivalstva v lastnem nacionalnem okolju, primerjalni muzikologi pa na glasbe zunaj Evrope§ in sicer t.i. primitivnih ljudstev in t.i. visokih kultur Orienta.

"Le da našo pacienti potem ne smejo ravnati tako kot Kant – ta je nekoč moral odpustiti svojega tatinskega slugo; vendar pa tega ni mogel preboleti§ in da bi se k temu prisilil, je na steno v sobi obesil tablo z napisom: »Svojega slugo moram pozabiti«."

»Kaj niste poznali dr. J.?« me je vprašal povsem nepričakovano§ in ko sem mu prikimal, je nadaljeval: »Z njim sem delil celico v slavnem moskovskem zaporu L., tik pred mojo izpustitvijo.

Po njeni smrti se je poskusil ponovno postaviti na noge§ in ko se mu je že nasmehnila kapljica sreče, je kruta usoda odigrala svojo vlogo in mu pobrala še tisto malo upanja, kar ga je še ostalo.

Vendar smo našli tudi precej primerov, ko Besana napačno oceni, da gre za mejo prilastkovega odvisnika in ne za naštevanje:

Etnologija, prej tudi etnografija ali narodopisje, veda, ki preučuje ljudsko kulturo§ in način življenja etničnih skupin na vseh stopnjah njihovega razvoja ter na ravni vsakdanjosti še posebej.

Njegovo prepričanje, da je za učiteljevo izpopolnjevanje nujno obsežno§ in liberalno izobraževanje, ki ga morajo dopolnjevati raziskovanje in obdobja poklicnega usposabljanja, so večinoma sprejeli v Evropi in Združenih državah Amerike.

Zanimiv primer, kjer pride do tega zaradi druge napake v povedi (»nastopata« namesto »nastopajo«) pa je:

Trio Alwan, v katerem skupaj nastopata Egipčan, Nemec in Italijan, je predstavil arabsko glasbo.

Zaradi dvojine prideta v prilastkov odvisnik le Egipčan in Nemec, Italijan se pa pridruži triu Alwan.

5.3.3 Iskanje odvečnih vejic

Težava pri izpopolnjevanju Besane za iskanje odvečnih vejic je v tem, da program za določanje vrst vejic tega ne počne z odvečnimi vejicami. Zato smo omejeni le na okolico odvečnih vejic, kar pa ne zadošča, saj so odvečne vejice v precejšnji meri vezane na strukturo povedi oz. stavka (tak primer so vejice za začetnim prislovnim določilom).

Tabela 16: Besede, pred katerimi je največ odvečnih vejic, ki jih Besana ne odkrije

beseda	1	2	3	4	5	6	7	8
in	188	1115	46	410	1624	144	44793	144
da	2050	82	75	399	18930	10	2332	232
ter	5	60	60	267	124	14	2144	7
kot	272	188	199	220	1499	69	5635	175
ko	412	44	5	167	2724	1	770	108
ali	93	138	12	138	482	4	4369	114
zaradi	15	5	0	85	192	1	1812	1
pa	15	244	65	79	1165	95	13301	104
je	712	191	70	69	2123	20	57086	127
če	277	12	1	67	1603	2	575	165
v	239	77	3	65	1978	6	31380	18
ne	42	91	10	64	2079	17	8866	42
tako	20	59	1	47	682	3	4306	43
se	306	196	44	41	1337	9	28938	62
do	4	13	2	39	133	1	4393	3
na	47	54	0	33	985	1	17540	11
kar	316	120	0	30	2001	16	897	77

ker	349	22	4	29	2338	0	194	48
za	51	31	1	29	555	0	13075	11
po	8	22	0	28	486	1	4846	5
s	103	22	0	27	735	0	6434	12
oziroma	1	9	19	25	16	1	705	7
kljub	14	9	1	24	86	1	297	8
oz	0	7	7	23	11	3	1032	6
so	209	58	32	22	708	5	16052	22
z	4	28	0	21	261	0	9471	1
bi	136	25	27	20	362	4	9319	20
zato	130	57	0	20	1299	2	996	11
tudi	0	55	8	16	279	0	9532	13
ga	0	82	9	16	226	0	5326	3
še	3	42	0	14	296	3	4733	7

Tabela 16 prikazuje, pred katerimi besedami je največ odvečnih vejic, ki jih Besana še ne odkriva (stolpec 4).

Največ težav je pri prirednih veznikih, ki običajno ne zahtevajo vejic (»in«, »ter«, »oziroma«, »oz.« in »ali«), kjer pa vejica vseeno zelo pogosto mora biti, zato je ugotavljanje teh odvečnih vejic zelo težka naloga.

Če primerjamo še stolpec 5, ki pove, kolikokrat je pred to besedo vejica, ki tam mora biti, vidimo, da se splača preverjati besedi »zaradi« in »kljub«.

6 Implementacija metod in rezultati

Po vseh izboljšavah v predhodnem poglavju nas seveda zanima, kakšne rezultate smo dosegli pri posameznih možnostih.

6.1 Spremljanje rezultatov strojnega učenja

Rezultati preizkušanja kažejo, da je za najboljši rezultat treba uporabiti označevanje z označevalniki, ki so se učili iz učnega korpusa z odstranjenimi vejicami, uporabiti je treba skladenjsko razčlenjevanje, kot najuporabnejši se je pokazal klasifikator ADTree (alternirajoče odločitveno drevo). Rezultati se izboljšujejo z večanjem drevesa, vendar hkrati narašča čas, potreben za izračun.

6.1.1 Primerjava z drugimi rezultati za slovenščino

Najboljši pridobljeni rezultat je na koncu treba primerjati s prejšnjimi rezultati, najprej z rezultati metod s pravili za slovenščino, potem pa s statističnimi metodami za druge jezike.

Besano in LanguageTool smo tako za primerjavo uporabili za to, da poiščemo vse vejice v besedilu, kar dosežemo tako, da vse vejice označimo kot manjkajoče. Ne Besana ne LanguageTool nista prilagojena taki uporabi, ker sta napisana za iskanje realnih napak v besedilih, zato lahko pričakujemo, da pri tem ne bosta tako uspešna.

Zastavilo se je vprašanje, kako obravnavati rezultate Besane, in sicer so problem opozorila, kjer ugotovi, da nekje manjka vejica, ne ve pa točno, kje. Ti primeri zahtevajo uporabnika, ki bo opozorjen sam znal postaviti vejico na ustrezno mesto, in niso primerni za samodejno postavljanje vejic, npr. pri razpoznavi govora. Zato ima Besana v tabeli dva rezultata, pri prvem so upoštevane le vejice, ki jih Besana točno postavi, pri drugem pa še tiste, kjer le ugotovi, da bi morala vejica biti, a ne ve, kje natanko.

Za primerjavo smo dodali še rezultat, dosežen v Krajnc (2015), in sicer smo uporabili rezultat z najvišjo metriko F1 za vrednost »je vejica« za korpus Šolar2 (ki je isti, kot je bil uporabljen pri drugih metoda v tabeli), nabor atributov je MSD11, klasifikator pa RandomForest. Iz Krajnc (2015) je dodan še rezultat za klasifikator DecisionTable (DT), tudi na naboru atributov MSD11, vendar le pri desetini podatkov.

Tabela 17 kaže, da je strojno učenje doseglo najboljši priklic in F1 (če ne štejemo pri Besani še opozoril brez točnega mesta vejice), najboljša natančnost pa je pri LanguageTool.

Tabela 17: Rezultati za postavljanje vseh vejic na korpusu Šolar

	ni vejice			je vejica		
parametri	natančnost	priklic	F1	natančnost	priklic	F1
ADTree (-B 70 -E -3)	0,967	0,992	0,980	0,892	**0,661**	**0,759**
LanguageTool				**0,972**	0,522	0,668
Besana				0,921	0,622	0,743
Besana + nekje				0,942	0,660	0,776
Krajnc (2015)	0,957	0,995	0,975	0,913	0,542	0,680
Krajnc (2015) – DT	0,959	0,995	0,977	0,920	0,577	0,709

6.1.2 Primerjava z rezultati za druge jezike

Rezultati samodejnega postavljanja vejic so zelo odvisni od jezika, kar so npr. pokazali Zhang et al. (2002), ki so preizkusili isti metodi pri angleščini in nemščini.

Tabela 18: Rezultati strojnega učenja postavljanja vejic za različne jezike

jezik	preizkus	je vejica		
		natančnost	priklic	F1
angleščina	Beeferman et al. (1998), algoritem A	0,756	0,656	0,702
angleščina	Beeferman et al. (1998), algoritem B	0,784	0,624	0,694
angleščina	Zhang et al. (2002), Amalgam	0,744	0,676	0,709
angleščina	Zhang et al. (2002), jezikovno modeliranje	0,782	0,624	0,694
angleščina	Shieber in Tao (2003)	0,797	0,626	0,748
angleščina	Israel et al. (2012)	0,858	0,663	0,748
nemščina	Zhang et al. (2002), Amalgam	0,854	0,875	0,865
nemščina	Zhang et al. (2002), jezikovno modeliranje	0,896	0,746	0,815
baskovščina	Alegria et al. (2006)	0,696	0,486	0,572
slovenščina	Holozan (2013)	0,861	0,641	0,735
slovenščina	Krajnc (2015)	0,913	0,542	0,680
slovenščina	**ta disertacija, ADTree (-B 70 –E -3)**	**0,892**	**0,661**	**0,759**

Tabela 18 kaže, da je natančnost pri slovenščini podobna kot pri nemščini, priklic pa je slabši. Tudi najboljši rezultat za angleščino (Israel et al. 2012) ima podobno natančnost in priklic, kakor je pri slovenskem rezultatu.

Znotraj istega jezika pa lahko na rezultate vpliva tudi izbira korpusa, s katerim učimo in preizkušamo strojno učenje. Tako lahko npr. primerjamo rezultata za enak nabor atributov in enako označeno z vsemi vejicami za ADTree (z istimi parametri –B 10 –E -3).

Tabela 19: Primerjava rezultatov za isti klasifikator in enak nabor atributov, a različen korpus

	ni vejice			je vejica		
korpus	natančnost	priklic	F1	natančnost	priklic	F1
Šolar	0,967	0,984	0,976	0,802	0,663	**0,726**
Lektor	0,943	0,991	0,967	0,780	0,350	**0,483**

Tabela 19 nam tako pokaže, da je vpliv korpusa na rezultate ogromen, tako da je pravzaprav težko primerjati rezultate, ki niso doseženi pri istem korpusu.

6.1.3 Primerjava rezultatov pri popravljanju vejic

Za izračun popravljanja vejic je uporabljen klasifikator ADTree (-B 30 –E -3), preizkus je narejen na primerih iz korpusa Šolar, uporabljeni so različni načini označevanja besedila.

Tabela 20: Rezultat popravljanja vejic s strojnim učenjem

	Iskanje manjkajočih vejic			Iskanje odvečnih vejic		
način	natančnost	priklic	F1	natančnost	priklic	F1
označeno brez vejic	**59,59 %**	47,39 %	52,79 %	15,18 %	**86,77 %**	25,84 %
označeno brez vejic, označevalnik brez vejic	55,36 %	48,83 %	51,89 %	13,78 %	75,58 %	23,32 %
označeno z vejicami v besedilu	55,51 %	45,42 %	49,96 %	14,32 %	71,07 %	23,84 %
označeno z vejicami v besedilu, označevalnik brez vejic	56,77 %	**50,50 %**	**53,45 %**	**17,39 %**	78,27 %	**28,46 %**
LanguageTool	79,90 %	45,88 %	58,25 %	92,15 %	8,11 %	14,91 %
Besana	83,65 %	58,82 %	69,07 %	**82,19 %**	21,60 %	**34,21 %**
Besana + nekje	**89,48 %**	**64,95 %**	**75,27 %**	/	/	/

Tabela 20 prikazuje rezultat popravljanja napak in primerjavo z LanguageToolom in Besano. Najboljši rezultat za iskanje manjkajočih vejic dobimo, če pri označevanju v besedilu pustimo že napisane vejice, uporabimo pa označevalnik, ki je naučen brez vejic. To je drugače kot v Holozan (2013), kjer je rezultat najboljši, kadar v besedilu pred označevanjem zbrišemo vejice. Vzrok za to razliko je morda to, da so bili v Holozan (2013) uporabljeni primeri, kjer je bil delež napačno postavljenih vejic večji, ker ni bilo primerov brez napak pri postavitvi vejic. Rezultati se približujejo rezultatom za LanguageTool, zaostajajo pa za rezultati Besane.

Isti način označevanja se najbolje izkaže tudi pri iskanju odvečnih vejic. V primerjavi z LanguageToolom in Besano ima strojno učenje veliko boljši priklic (78,27 % proti 8,11 oziroma 21,60 %), vendar veliko slabšo natančnost (17,39 % proti 92,15 oziroma 82,19 %). Taka natančnost je tako rekoč neuporabna, zato bi tukaj veljalo preizkusiti še idejo iz Israel et al. (2012), da ne gre upoštevati le dejstvo, da se je klasifikator odločil, da kake vejice ni, ampak tudi njegovo oceno te odločitve, tako da se vejica označi kot odvečna le, če

ta ocena preseže določeno mejo. Pri manjkajočih vejicah to ne bi tako koristilo, ker tam priklic ni tako visok, da bi ga lahko žrtvovali za višjo natančnost.

Rezultati popravljanja napak pri vejicah s strojnim učenjem so slabši kot pri postavljanju vseh vejic, predvsem je slabša natančnost.

Za angleščino so Israel et al. (2012) dosegli natančnost 0,849 pri priklicu 0,200 (F1 0,324), vendar je to rezultat za vse napačne vejice, ni pa posameznih rezultatov za manjkajoče oz. odvečne vejice.

6.2 Spremljanje rezultatov za LanguageTool po izboljšavah

Za LanguageTool so dodana nova pravila v skladu z analizo težav v točki 5.2. Manjši popravki so narejeni v 5 pravilih, dodanih je 26 novih pravil; vsa popravljena in nova pravila so navedena v prilogi 9.4.

Tabela 21 prikazuje rezultate, ki jih je dosegel LanguageTool s popravki, navedenimi v prilogi 9.4. Rezultati kažejo, da so spremembe uspešne, saj se je priklic pri manjkajočih vejicah popravil s 40,15 na 44,62 % in natančnost z 62,26 na 65,26 %; skupaj je šel F1 z 48,82 na 53,00 %. Dodano je bilo odkrivanje odvečnih vejic, kjer je priklic 8,25 % pri natančnosti 81,55 %; F1 je torej 14,98 %. Priklic pri odvečnih vejicah torej ni posebno velik, je pa zato zelo dobra natančnost.

Tabela 21: Rezultati evalvacije za LanguageTool - novi rezultati

LT – novi	zahtevana najdena	zahtevana nekje najdena	zahtevana nenajdena	odvečna prečrtana	odvečna neprečrtana	ni in najdena	ni in nekje najdena	dobra in prečrtana	priklic manjkajočih	priklic odvečnih	natančnost manjkajočih	natančnost odvečnih	F1 manjkajočih	F1 odvečnih
KUST	292	0	138	3	101	35	0	0	67,91%	2,88%	89,30%	100,00%	77,15%	5,61%
KUST de	6	0	5	0	0	0	0	0	54,55%	/	100,00%	/	70,59%	/
KUST en	49	0	20	0	53	13	0	0	71,01%	0,00%	79,03%	/	74,81%	/
KUST es	92	0	42	3	29	12	0	0	68,66%	9,38%	88,46%	100,00%	77,31%	17,14%
KUST it	42	0	34	0	8	1	0	0	55,26%	0,00%	97,67%	/	70,59%	/
KUST sh	103	0	37	0	11	9	0	0	73,57%	0,00%	91,96%	/	81,75%	/
Šolar	7547	0	8903	305	3454	1915	0	26	45,88%	8,11%	79,76%	92,15%	58,25%	14,91%
Šolar OŠ 6	95	0	144	1	27	16	0	0	39,75%	3,57%	85,59%	100,00%	54,29%	6,90%
Šolar OŠ 7	267	0	460	5	111	78	0	0	36,73%	4,31%	77,39%	100,00%	49,81%	8,26%
Šolar OŠ 8	628	0	813	13	202	178	0	0	43,58%	6,05%	77,92%	100,00%	55,90%	11,40%
Šolar OŠ 9	273	0	431	7	165	114	0	0	38,78%	4,07%	70,54%	100,00%	50,05%	7,82%
Šolar PŠ 1	296	0	314	10	61	46	0	0	48,52%	14,08%	86,55%	100,00%	62,18%	24,69%
Šolar PŠ 2	184	0	159	6	44	25	0	0	53,64%	12,00%	88,04%	100,00%	66,67%	21,43%
Šolar PŠ 3	230	0	300	7	74	62	0	0	43,40%	8,64%	78,77%	100,00%	55,96%	15,91%
Šolar PŠ 5	195	0	101	5	36	28	0	0	65,88%	12,20%	87,44%	100,00%	75,14%	21,74%
Šolar SŠ 1	930	0	962	33	311	140	0	2	49,15%	9,59%	86,92%	94,29%	62,80%	17,41%
Šolar SŠ 2	714	0	699	40	262	177	0	2	50,53%	13,25%	80,13%	95,24%	61,98%	23,26%
Šolar SŠ 3	720	0	806	25	245	134	0	0	47,18%	9,26%	84,31%	100,00%	60,50%	16,95%
Šolar SŠ 4	647	0	599	26	216	146	0	2	51,93%	10,74%	81,59%	92,86%	63,46%	19,26%
Šolar G 1	843	0	1318	49	709	296	0	3	39,01%	6,46%	74,01%	94,23%	51,09%	12,10%
Šolar G 2	353	0	467	34	175	110	0	7	43,05%	16,27%	76,24%	82,93%	55,03%	27,20%
Šolar G 3	353	0	482	22	274	121	0	5	42,28%	7,43%	74,47%	81,48%	53,93%	13,62%
Šolar G 4	760	0	792	21	530	229	0	5	48,97%	3,81%	76,85%	80,77%	59,82%	7,28%
Šolar MT	59	0	56	1	12	15	0	0	51,30%	7,69%	79,73%	100,00%	62,43%	14,29%
Lektor	155	0	890	64	614	2320	0	57	14,83%	9,44%	6,26%	52,89%	8,81%	16,02%
Wikipedija	59	0	64	8	59	17	0	3	47,97%	11,94%	77,63%	72,73%	59,30%	20,51%
SKUPAJ	8053	0	9995	380	4228	4287	0	86	44,62%	8,25%	65,26%	81,55%	53,00%	14,98%

Dodatno smo pri tem odkrili težavo z nasveti pri pravilu PREDLOG_KATERI_BREZ_VEJICE, ki išče različne kombinacije predlogov in oblik zaimka »kateri«, v nasvetu pa je bilo vedno zapisano »za kateri« ne glede na to, kateri predlog in oblika sta bila v vhodnem besedilu. To pravilo je popravljeno tako, da zdaj da v nasvet originalni predlog in obliko zaimka kateri. To smo naredili s pomočjo elementa »match« v elementu »suggestion«. Lahko da je ta težava nekoliko pokvarila prvotno preizkušanje LanguageTool v točki 4 in vplivala tudi na nekatere rezultate v točki 5.2 (ker je program pričakoval, da v besedilih po popravkih ne bo drugih sprememb kot pri vejicah), vendar ne bi smela bistveno vplivati na rezultate (skupni priklic pri vseh primerih se je popravil na 40,15 % (s 40,02 %), natančnost z 62,07 % na 62,26 % in F1 z 48,67 % na 48,82 %; podobno je tudi pri iskanju vseh vejic: priklic 40,53 % (40,24 %), natančnost 92,08 % (91,47 %), F1 56,28 % (55,90 %)), zato grafi iz točke 4 niso ponovljeni, so pa v nadaljevanju uporabljeni novi rezultati z namenom primerjalnega spremljanja izboljšave zaradi spremenjenih in dodanih pravil.

6.3 Spremljanje rezultatov Besane po izboljšavah

Tabela 22 vsebuje rezultat, ki jih je dosegla Besana po popravkih, narejenih v točki 5.3. Spremembe so uspešne, priklic pri manjkajočih vejicah se je popravil s 57,45 % na 63,47 %, natančnost s 74,91 % na 79,81 % in F1 s 65,03 % na 69,44 %.

Še večji napredek je bil, če odštejemo možnost, ko Besana opozori na vejico brez točno določenega mesta. Tu se je priklic izboljšal s 47,06 % na 57,50 %, natančnost s 67,81 % na 72,46 %, F1 pa s 55,56 % na 64,12 %.

Pri iskanju odvečnih vejic se je priklic popravil z 18,58 % na 21,55 %, natančnost s 54,73 % na 60,96 % in F1 s 27,74 % na 31,84 %.

Tabela 22: Rezultati evalvacije za Besano - novi rezultati

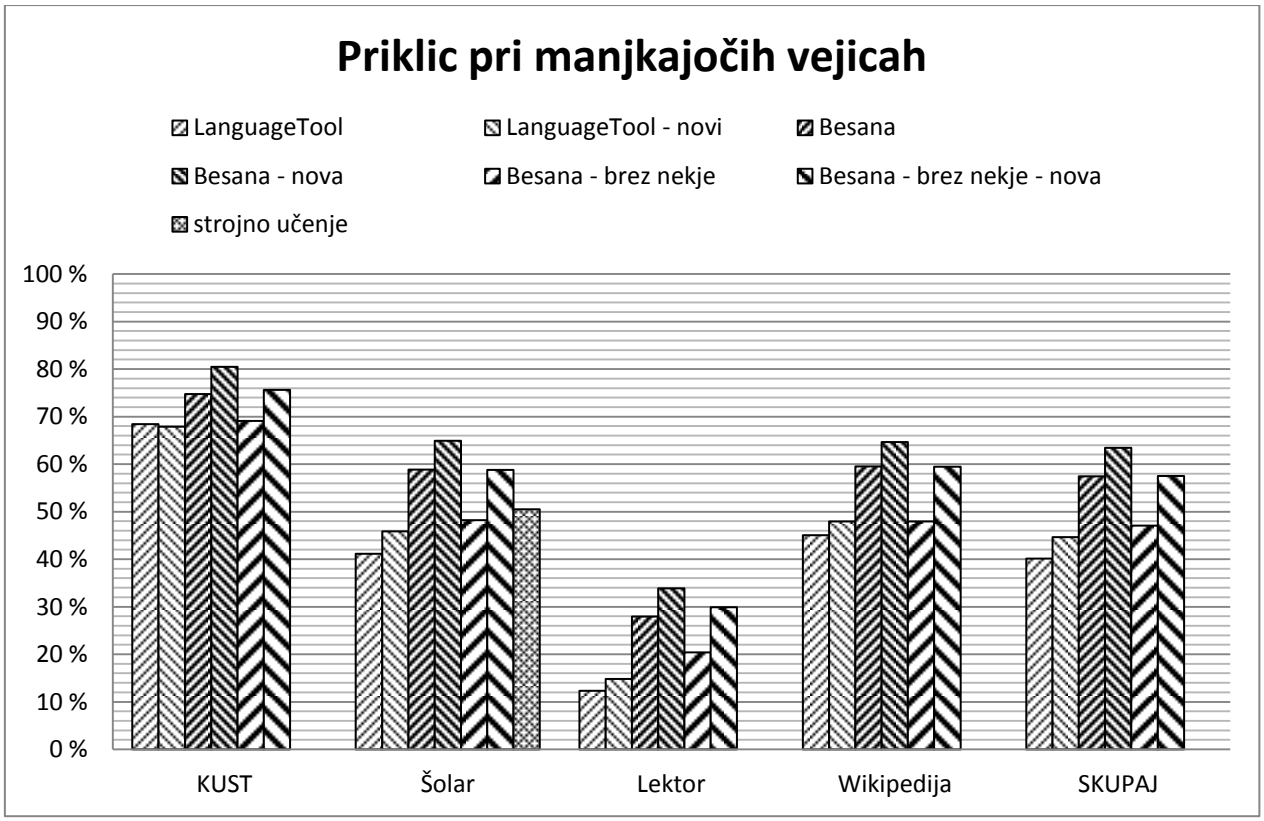
Besana – nova	zahtevana najdena	zahtevana nekje najdena	zahtevana nenajdena	odvečna prečrtana	odvečna neprečrtana	ni in najdena	ni in nekje najdena	dobra in prečrtana	priklic manjkajočih	priklic odvečnih	natančnost manjkajočih	natančnost odvečnih	F1 manjkajočih	F1 odvečnih
KUST	314	20	81	15	89	23	17	0	80,48 %	14,42 %	89,30 %	100,00 %	84,66 %	25,21 %
KUST de	10	0	0	0	0	0	0	0	100,00 %	#DIV/0!	100,00 %	#DIV/0!	100 %	#DIV/0!
KUST en	50	3	12	10	43	7	4	0	81,54 %	18,87 %	82,81 %	100,00 %	82,17 %	31,75 %
KUST es	94	8	27	3	29	10	5	0	79,07 %	9,38 %	87,18 %	100,00 %	82,93 %	17,14 %
KUST it	52	4	17	2	6	1	5	0	76,71 %	25,00 %	90,32 %	100,00 %	82,96 %	40,00 %
KUST sh	108	5	25	0	11	5	3	0	81,88 %	0,00 %	93,39 %	#DIV/0!	87,26 %	#DIV/0!
Šolar	9398	980	5600	812	2947	1105	924	176	64,95 %	21,60 %	83,65 %	82,19 %	73,12 %	34,21 %
Šolar OŠ 6	129	22	80	6	22	13	8	4	65,37 %	21,43 %	87,79 %	60,00 %	74,94 %	31,58 %
Šolar OŠ 7	372	75	264	19	97	47	61	1	62,87 %	16,38 %	80,54 %	95,00 %	70,62 %	27,94 %
Šolar OŠ 8	786	84	532	50	165	89	85	10	62,05 %	23,26 %	83,33 %	83,33 %	71,14 %	36,36 %
Šolar OŠ 9	383	34	274	30	142	56	49	6	60,35 %	17,44 %	79,89 %	83,33 %	68,76 %	28,85 %
Šolar PŠ 1	351	43	193	8	63	37	27	5	67,12 %	11,27 %	86,03 %	61,54 %	75,41 %	19,05 %
Šolar PŠ 2	199	26	92	6	44	20	21	0	70,98 %	12,00 %	84,59 %	100,00 %	77,19 %	21,43 %
Šolar PŠ 3	276	47	171	9	72	28	35	3	65,38 %	11,11 %	83,68 %	75,00 %	73,41 %	19,35 %
Šolar PŠ 5	204	11	74	7	34	14	12	1	74,39 %	17,07 %	89,21 %	87,50 %	81,13 %	28,57 %
Šolar SŠ 1	1123	114	604	65	279	97	109	7	67,19 %	18,90 %	85,72 %	90,28 %	75,33 %	31,25 %
Šolar SŠ 2	847	73	457	54	248	88	70	8	66,81 %	17,88 %	85,34 %	87,10 %	74,95 %	29,67 %
Šolar SŠ 3	896	87	497	52	218	121	78	8	66,42 %	19,26 %	83,16 %	86,67 %	73,85 %	31,52 %
Šolar SŠ 4	755	65	394	46	196	81	76	7	67,55 %	19,01 %	83,93 %	86,79 %	74,85 %	31,19 %
Šolar G 1	1193	125	777	163	595	149	121	32	62,91 %	21,50 %	83,00 %	83,59 %	71,57 %	34,21 %
Šolar G 2	450	55	296	55	154	57	45	17	63,05 %	26,32 %	83,20 %	76,39 %	71,73 %	39,15 %
Šolar G 3	475	53	290	61	235	83	31	15	64,55 %	20,61 %	82,24 %	80,26 %	72,33 %	32,80 %
Šolar G 4	888	64	566	180	371	121	93	52	62,71 %	32,67 %	81,65 %	77,59 %	70,94 %	45,98 %
Šolar MT	71	2	39	1	12	4	3	0	65,18 %	7,69 %	91,25 %	100,00 %	76,04 %	14,29 %
Lektor	312	41	690	148	530	1403	727	453	33,84 %	21,83 %	14,22 %	24,63 %	20,02 %	23,14 %
Wikipedija	69	6	41	18	49	23	12	7	64,66 %	26,87 %	68,18 %	72,00 %	66,37 %	39,13 %
SKUPAJ	10093	1047	6412	993	3615	2554	1680	636	63,47 %	21,55 %	72,46 %	60,96 %	67,67 %	31,84 %

6.4 Primerjava rezultatov po izboljšavah

Primerjamo lahko rezultate za LanguageTool in Besano pred popravki in po njih ter pri iskanju odvečnih vejic in postavljanju vseh vejic še rezultate strojnega učenja.

6.4.1 Iskanje manjkajočih vejic

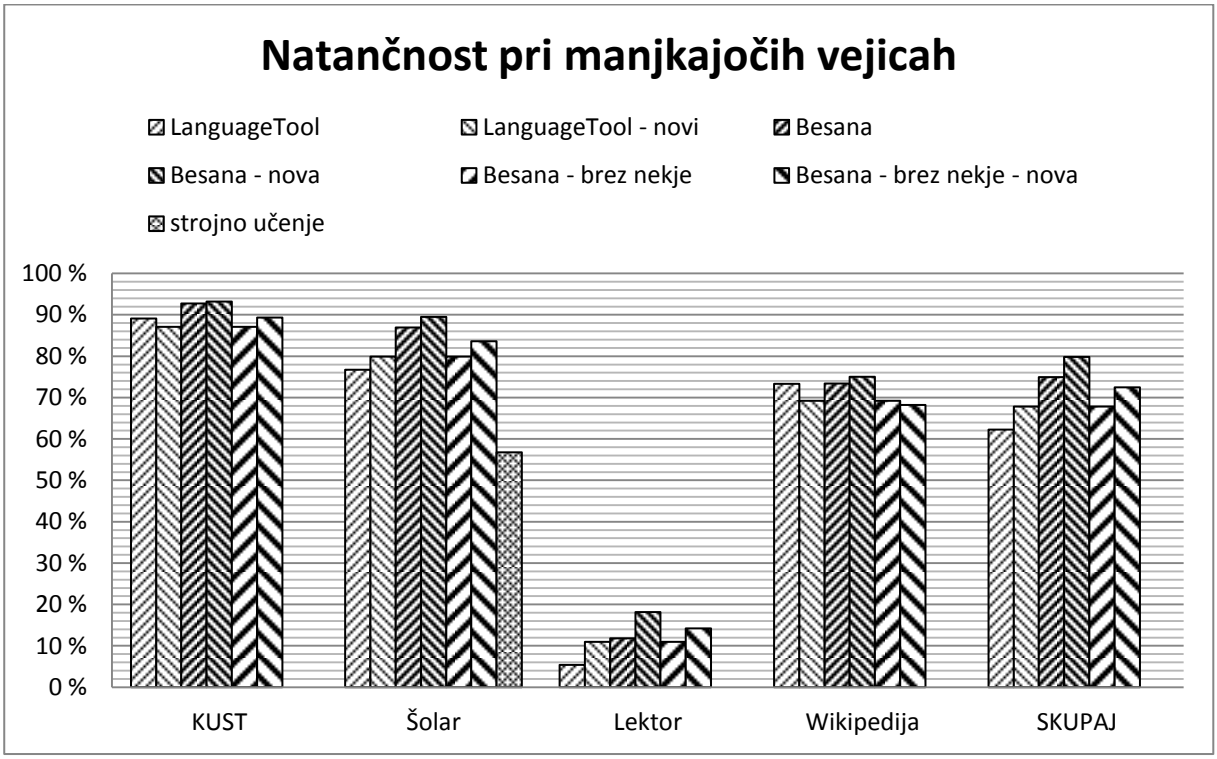
Za problem iskanja manjkajočih vejic so bili doseženi naslednji rezultati:



Slika 19: Končni rezultati priklica manjkajočih vejic

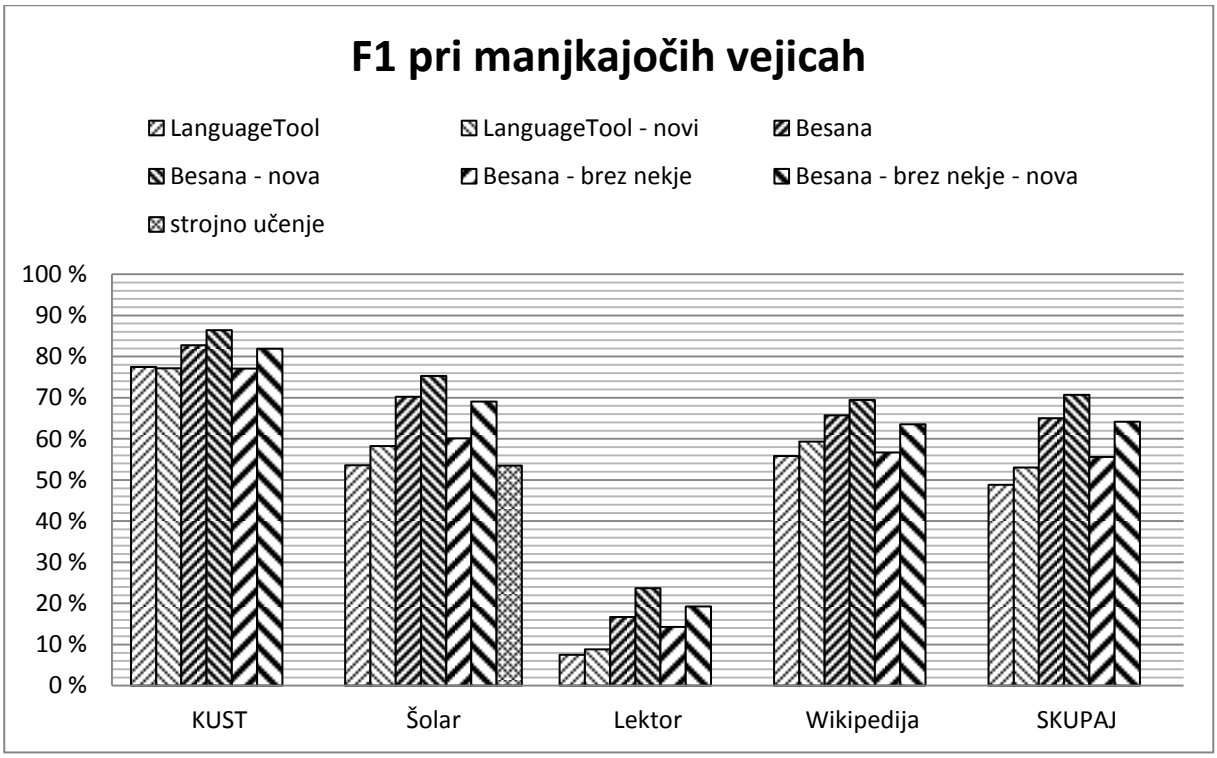
Slika 19 prikazuje rezultate za priklic. Tako LanguageTool kot Besana sta izboljšala rezultate, rezultat s strojnim učenjem je malo nad LanguageToolom, Besana pa ima okoli 14 odstotnih točk prednosti v skupnem rezultatu.

Slika 20 nam pove, da se je popravila natančnost LanguageToola in Besane, razlike so tu manjše, negativno odstopa le strojno učenje.



Slika 20: Končni rezultati za natančnost pri manjkajočih vejicah

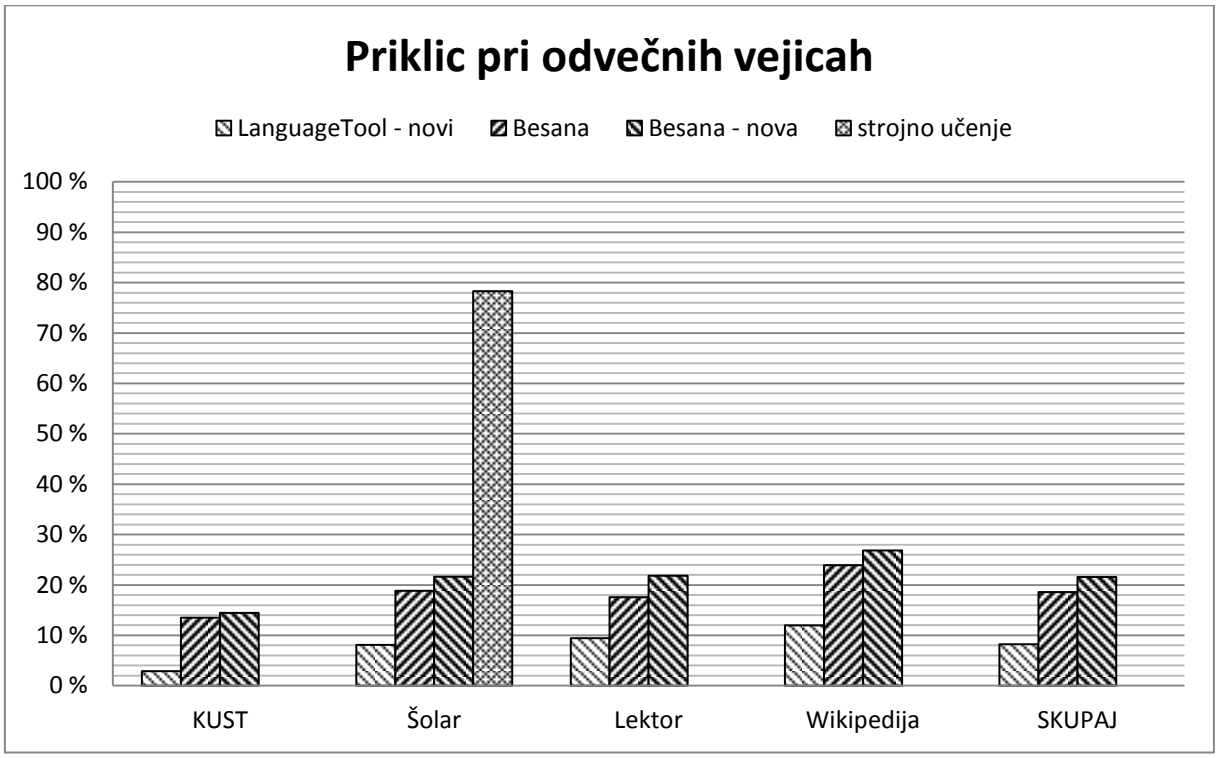
Slika 21 prikazuje še rezultate za F1, tudi tukaj so rezultati izboljšani, Besana ima najboljši rezultat, strojno učenje pa ne zaostaja veliko za LanguageToolom.



Slika 21: Končni rezultati za F1 pri manjkajočih vejicah

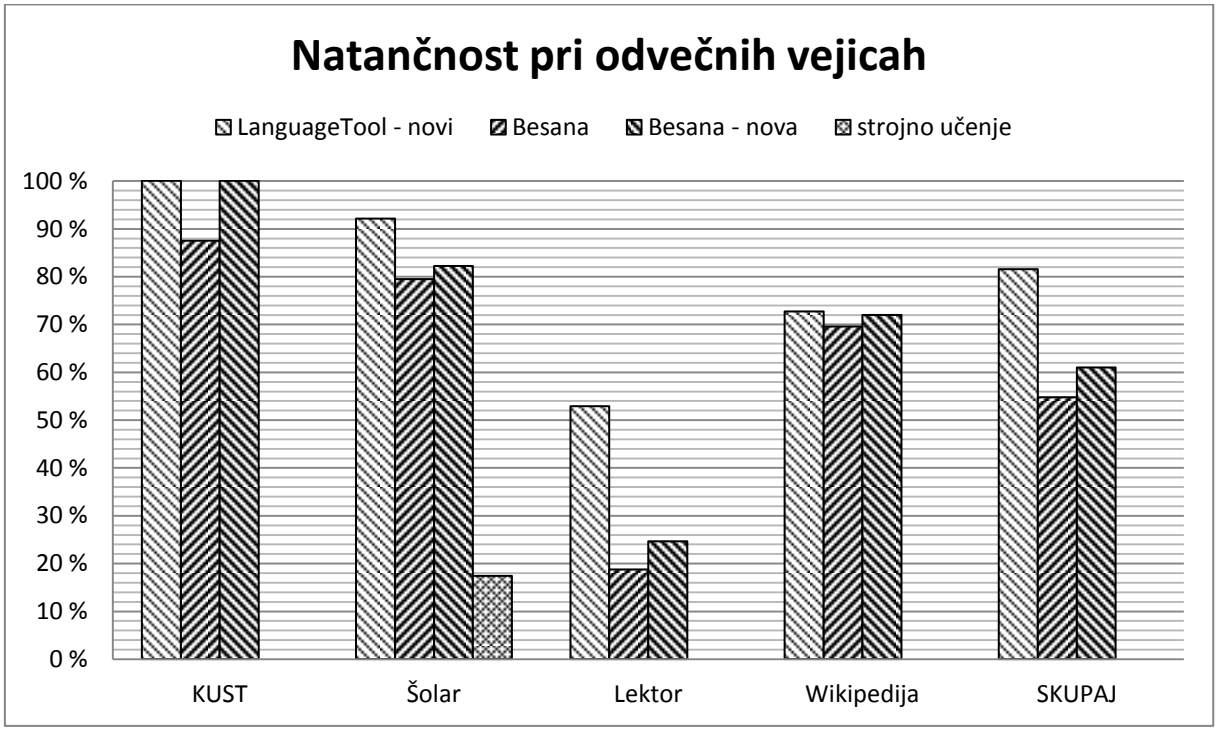
6.4.2 Iskanje odvečnih vejic

Pri problemu iskanja odvečnih vejic so rezultati naslednji:



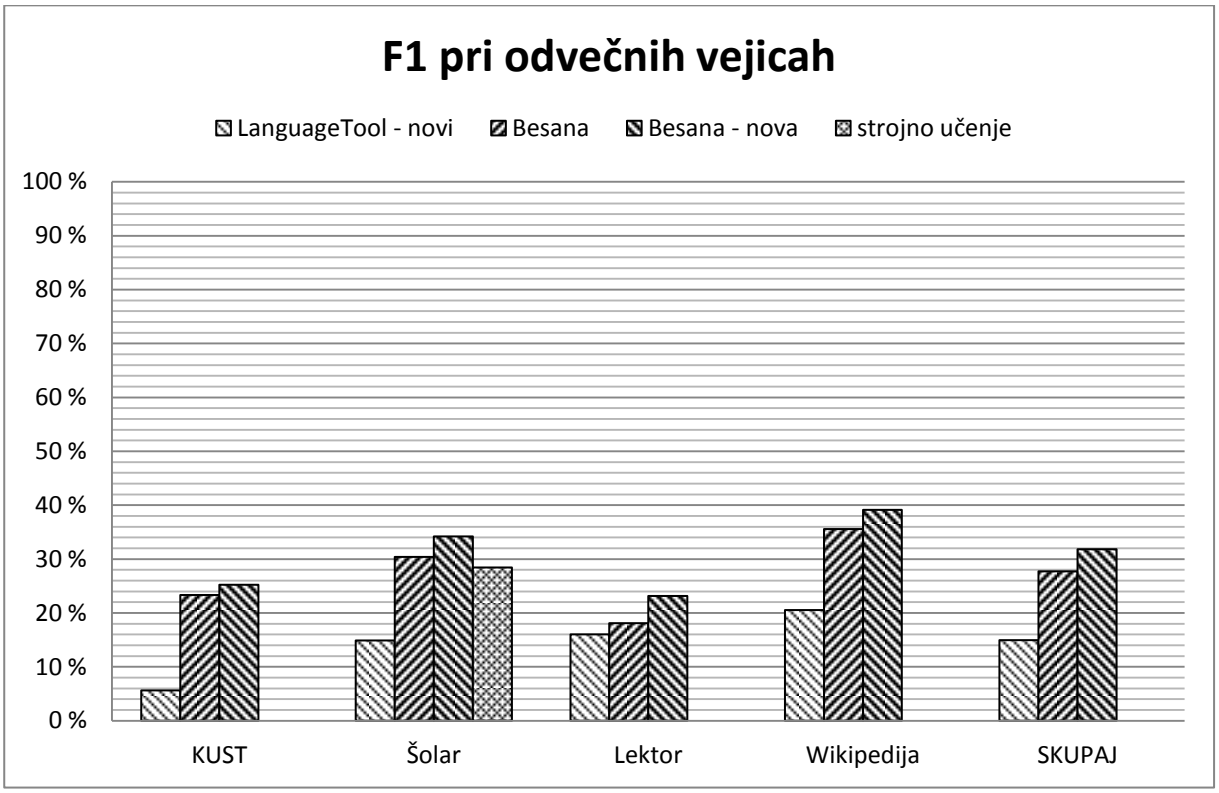
Slika 22: Končni rezultati priklica odvečnih vejic

Slika 22 kaže, da pri priklicu močno prevladuje strojno učenje, Besana je malenkost izboljšala svoj rezultat, LanguageTool zaostaja, vendar pa prej sploh ni iskal teh napak.



Slika 23: Končni rezultati za natančnost pri odvečnih vejicah

Slika 23 prikazuje natančnost pri iskanju odvečnih vejic, LanguageTool ima tu najboljši rezultat, rezultat Besane se je izboljšal, vendar precej zaostaja, in sicer predvsem zaradi izrazito slabega rezultata pri primerih iz korpusa Lektor. Zelo slaba pa je natančnost strojnega učenja.



Slika 24: Končni rezultati za F1 pri odvečnih vejicah

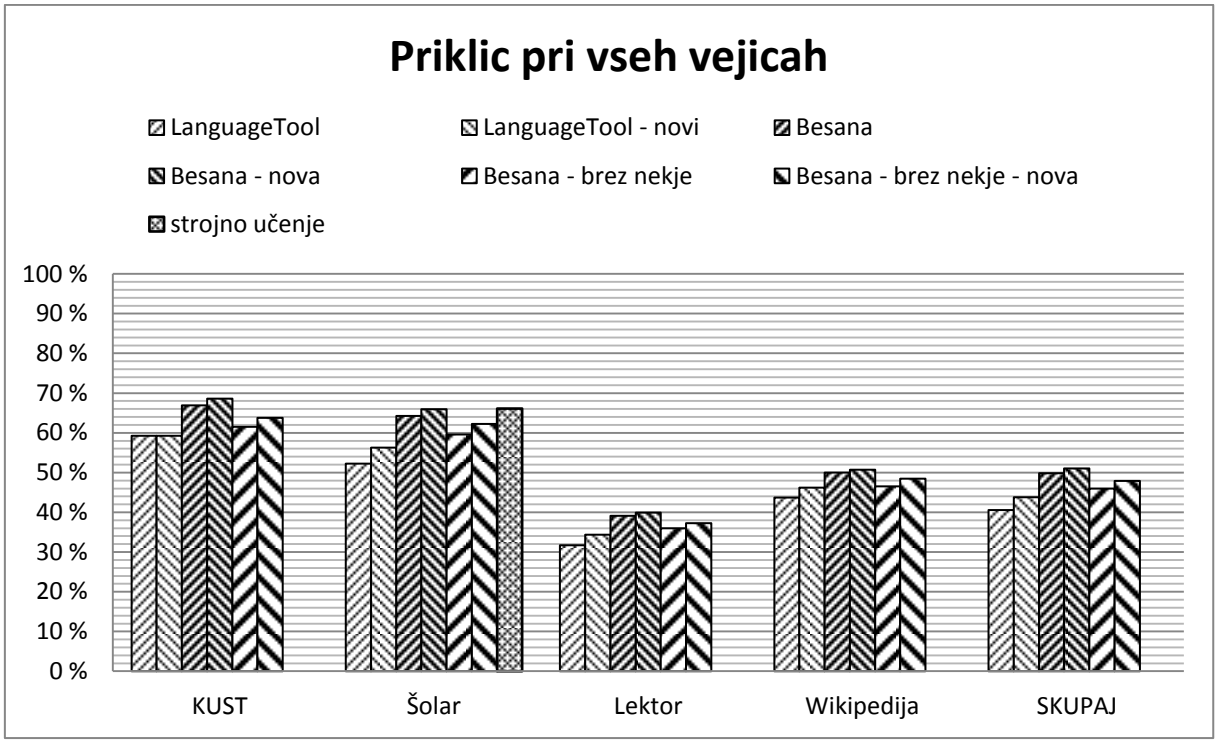
Slika 24 nam pove še F1 pri iskanju odvečnih vejic. Najboljši rezultat, ki se je tudi izboljšal za približno 5 odstotnih točk, ima Besana, strojno učenje je blizu, LanguageTool pa ima pol slabši rezultat, vendar je bil prej njegov rezultat 0, saj sploh ni odkrival odvečnih vejic.

6.4.3 Postavljanje vseh vejic

Za problem postavljanja vseh vejic v besedilo so bili doseženi naslednji rezultati:

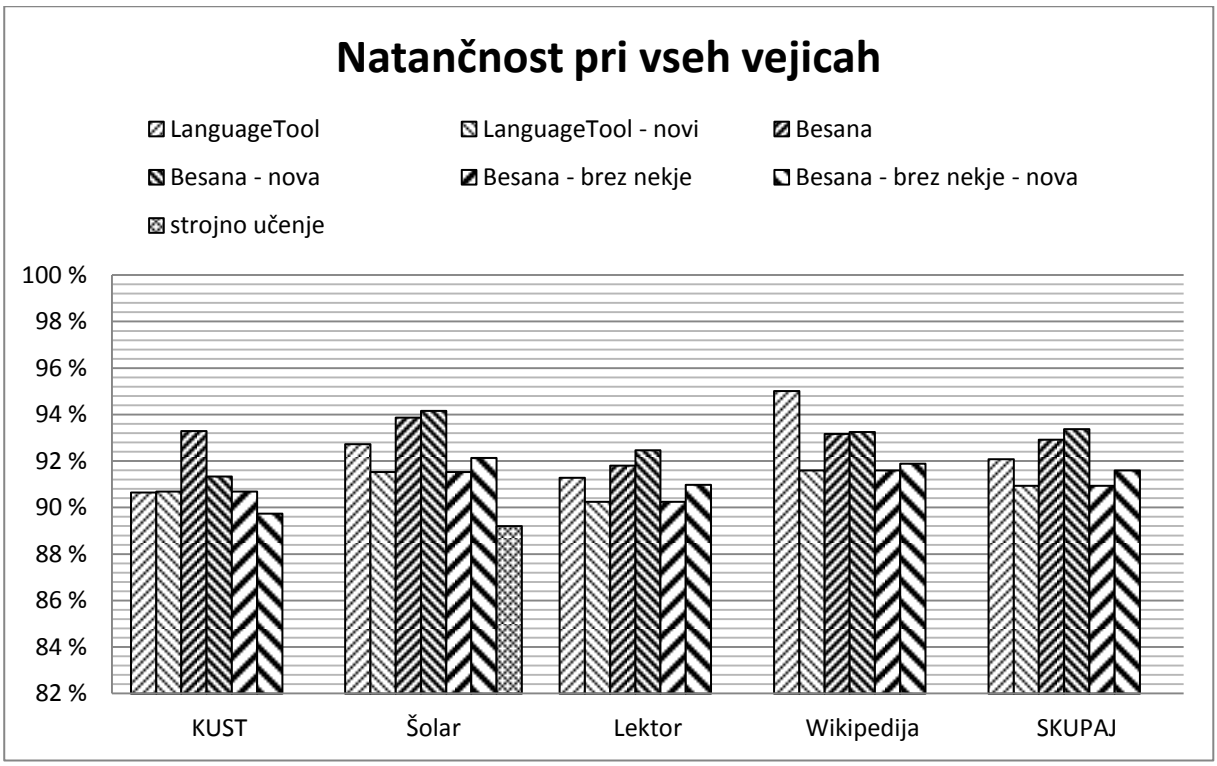
Slika 25 prikazuje rezultate priklica. Rezultati so veliko bolj blizu, kot je bilo to pri popravljanju vejic. Tako Besana kot tudi LanguageTool sta izboljšala rezultat, vendar razlika ni velika. Pri primerih iz korpusa Šolar je pripisan še rezultat, dosežen s strojnim učenjem, ki je dosegel najboljši rezultat, zelo blizu je Besana, vendar je za samodejno postavljanje vseh

vejic uporabna le Besana brez opozoril, da kje manjka vejica (brez nekje), tako da je prava razlika 4 %.

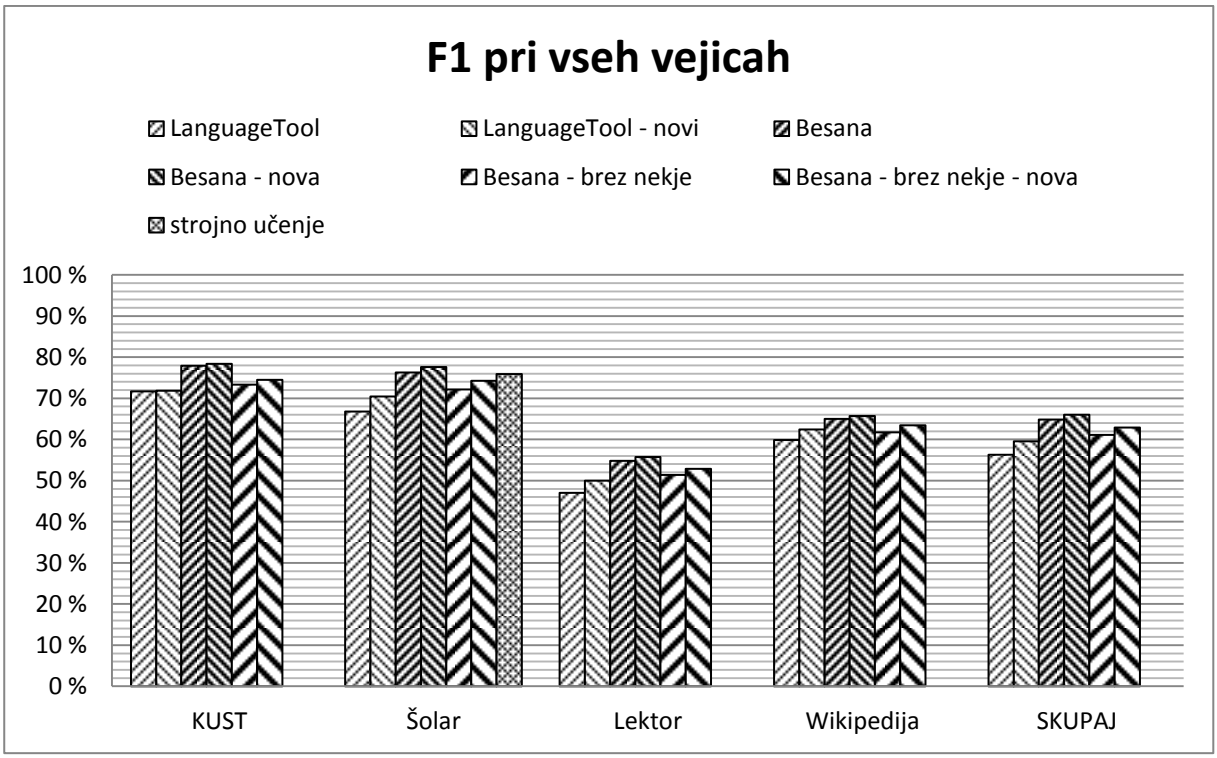


Slika 25: Končni rezultati priklica pri postavljanju vseh vejic

Slika 26 pokaže, da so rezultati pri natančnosti še bolj usklajeni (merilo na levi je drugačno!). Pri Besani je rezultat malenkost boljši, pri LanguageToolu pa malenkost slabši, še najslabši rezultat je pri strojnem učenju, ampak tudi tam znaša natančnost več kot 89 %).



Slika 26: Končni rezultati za natančnost pri postavljanju vseh vejic



Slika 27: Končni rezultati za F1 pri postavljanju vseh vejic

Slika 27 prikazuje še F1, rezultati so si blizu, tako LanguageTool kot Besana imata malenkost boljši rezultat. Pri Šolarju je najboljši rezultat torej doseglo strojno učenje (ker pri Besani ne smemo upoštevati opozoril, da neznano kje manjka vejica).

7 Sklep

Priprava zbirke primerov rabe vejice v slovenščini se je pokazala kot močno orodje, ki je za začetek omogočilo ugotoviti, kako dobro dosedanja programa (LanguageTool in Besana) popravljata napačno postavljene vejice.

Zbirka je omogočila preizkus strojnega učenja za postavljanje vejic v slovenščini. Rezultat pri postavljanju vseh vejic je primerljiv z obstoječima programoma, popravljanje vejic pa še ni tako uspešno. Za postavljanje vseh vejic se je kot najbolj uspešna metoda pokazal ADTree z uporabo lem, oblikoskladenjskih oznak in rezultatov skladenjskega razčlenjevalnika, pomembno je tudi, da se uporabita označevalnik in razčlenjevalnik, naučena na učnih primerih, iz katerih so izločene vejice.

Za programa, ki uporabljata pravila za postavljanje oziroma popravljanje vejic, sta bili razviti dve metodi, ki omogočata ugotoviti, kako je treba dopolniti pravila. Prva je iskanje besed in okolic pri pravilno postavljenih vejicah, manjkajočih vejicah, odvečnih vejicah in mestih brez vejic. S pomočjo te metode smo dopolnili pravila za LanguageTool, kar je izboljšalo tako priklic kot natančnost, dodali pa smo tudi iskanje odvečnih vejic.

Druga metoda je bila razvita z uporabo analizatorja povedi, ki ga razvija podjetje Amebis. Pri tej metodi se vejicam pripiše oznaka, ki pove vzrok, zakaj vejica na tistem mestu mora biti. Ta metoda po eni strani omogoča, da se s pomočjo zbirke primerov ugotovi, katere vrste vejic delajo največ težav pišočim (kar je skupaj s prvo metodo uporabljeno v Holozan (2015)), po drugi strani pa nam omogoča ugotoviti tudi, katerih vrst vejic Besana ne popravlja uspešno. S pomočjo teh dveh metod nam je uspelo izboljšati tudi rezultat, ki ga pri popravljanju vejic doseže Besana, in sicer tako priklic kot natančnost.

Ker je zbirka primerov rabe vejice Vejica 1,0 prosto dostopna z licenco Creative commons, to omogoča, da jo uporabljajo tudi drugi raziskovalci in lahko primerjajo svoje rezultate z rezultati, ki so doseženi v pričujočem delu.

7.1 Možnosti nadaljnjega dela

Tako zbirka primerov kot vsi trije načini postavljanja vejic imajo še odprte smeri za izboljšave; tukaj bomo le nakazali nekatere.

7.1.1 Zbirka primerov rabe vejice

V prvem koraku je treba popraviti vse napake, najdene med preizkušanjem. S tem bomo zmanjšali moteči podatkovni šum pri preizkušanju in strojnem učenju. Dobro bi bilo tudi preveriti primere iz korpusa Lektor, in sicer predvsem zaradi napak pri razrezu na povedi in zaradi primerov v tujih jezikih, ki jih je treba izločiti, ker zgolj motijo preizkušanje in učenje.

Dodatno se je izkazalo, da je prišlo do nekaterih težav zaradi kopiranja zbirke v Excel (v primerih iz Šolarja sta na dveh mestih namesto presledka stala tabulatorja, tako da se je del stavka za tabulatorjem izgubil; izgubili so se tudi primeri, ki se začenjajo s pomišljajem). Teh napak ni toliko, da bi vplivale na rezultate, vseeno pa se splača ob drugih popravkih urediti še to.

Za nadaljnje delo pa bi bilo zanimivo v zbirko primerov dodati še primere iz korpusa SSJ500k, ki pa bi jih bilo treba ročno pregledati in označiti napake pri postavljanju vejic. S tem bi pridobili to, da bi pri strojnem učenju lahko uporabiti leme, oblikoskladenjske podatke in razčlembe povedi, ki so v korpusu SSJ500k ročno preverjeni. Na ta način bi pri strojnem učenju lahko ločili težave, ki so posledica nepopolnega označevanja.

7.1.2 Strojno učenje

Program WEKA podpira veliko število klasifikatorjev, tako da niso bili vsi preizkušeni, pa tudi pri tistih, ki so bili preizkušeni, je odprtih še veliko možnih poskusov s parametri klasifikatorjev. Problem je tudi čas (pri nekaterih klasifikatorjih tudi pomnilnik), potreben za izračunavanje, pri klasifikatorju ADTree se je pokazalo, da večanje drevesa izboljšuje rezultat, vendar zgornja meja ni bila dosežena, ker postane preračunavanje pri tako velikih drevesih prepočasno (najboljši rezultat se je mlel skoraj tri dni). Po drugi strani sama zahtevnost učenja v resnici ni tako problematična (ker jo moramo izvesti le enkrat), če je le potem uporaba naučenih pravil dovolj hitra.

Vsekakor je še veliko možnih kombinacij klasifikatorjev, parametrov, različnih atributov, oken, kjer bi bilo verjetno možno doseči še boljši rezultat. Nekaj zanimivih poskusov v tej smeri je že v Krajnc (2015), odprta licenca zbirke primerov Vejica pa odpira možnosti za še nadaljnje delo drugih raziskovalcev na tem področju.

Pri iskanju napak pri postavljanju vejic bi lahko uporabili ocene, ki jih klasifikator pripiše možnosti, da je na določenem mestu vejica. Tako bi lahko spremenili prag, pri katerem se odločimo, da gre za napako, s čimer bi lahko izboljšali natančnost na račun priklica (kar bi

bilo glede na trenutne rezultate še posebej koristno pri iskanju manjkajočih vejic, kjer imamo zdaj odličen priklic, vendar do neuporabnosti slabo natančnost). Za bolj temeljito preizkušanje pa je treba narediti tudi bolj avtomatiziran postopek, ki bi omogočal hitrejše delo in zmanjšal možnost napak zaradi ročnih postopkov. V ta namen bi se morda dalo uporabiti programski jezik R[26], s katerim se da uporabo statističnih metod in strojnega učenja dobro povezati po eni strani s paketom WEKA (RWeka[27]), po drugi strani pa z drugimi programskimi jeziki (C++, Perl).

7.1.3 LanguageTool

Obstoječih pravil na dosedanji način verjetno ni mogoče več kaj dosti izboljšati, vendar se odpira nova možnost z uporabo strojnega učenja.

Pri strojnem učenju zgradi ADTree odločitveno drevo, ki ga je preprosto interpretirati, da dobimo verjetnost vejice na določenem mestu. Pri strojnem učenju je bilo sicer uporabljeno oblikoskladenjsko označeno in skladenjsko razčlenjeno besedilo, kar je relativno zahtevna operacija, zato bi to lahko bila težava pri praktični uporabi. Vendar tudi Besana uporablja analizator povedi, ki izvaja vse to, in vseeno deluje zadovoljivo hitro.

Druga možnost pa je, da se poskusi zgraditi odločitveno drevo le s pomočjo atributov, ki jih je lažje dobiti, morda celo samo iz samih besed. Rezultat tega postopka bo verjetno sicer slabši (oziroma bo potrebno veliko večje drevo, da se doseže primerljivo dobro delovanje), vendar bi bilo podporo za samo odločitveno drevo najbrž veliko lažje dodati v LanguageTool.

7.1.4 Besana

Možnosti za izboljšave v Besani so v veliki meri povezane z izboljševanjem analizatorja povedi, saj ta v zelo veliki meri vpliva na rezultate. Dodatna težava so še nepodprte slovnične strukture, ki jih Besana ne more uporabiti v svojih pravilih za določanje vejic (tak primer so vmesni odvisniki sredi naštevanja).

Glede na to, da se je pri preverjanju primerov, pri katerih je Besana dodala odvečne vejice, našlo kar nekaj napak že v sami zbirki primerov rabe vejice, je priporočljivo sistematično preveriti vse te primere in popraviti najdene napake v zbirki primerov, da ne bodo motili rezultatov. V naslednjem koraku je treba z novo zbirko primerov ponoviti postopke iz točke 5.3 in tako postopoma popravljati Besano.

---

[26] https://en.wikipedia.org/wiki/R_%28programming_language%29
[27] https://cran.r-project.org/web/packages/RWeka/index.html

Druga smer, v katero bo po potrebi možno razvijati analizator, pa je dodatna podpora za primer, ko želimo postaviti vse vejice v besedilu. Na primer: zdaj analizator ne predvideva, da manjkajo vejice v naštevanju, ker te vejice pisci zelo redko pozabijo, tako opozarjanje pa bi lahko povzročilo težave pri samostalniških sklopih tipa »človek žaba«. Zato je smiselno iskanje takih vejic dodati s posebno nastavitvijo, ki analizatorju (in samim pravilom za določanje manjkajočih vejic) pove, da iščemo vse vejice, pri normalni uporabi pa se ta pravila ne uporabljajo.

8 Viri in literatura

Alegria, Iñaki, Arrieta, Bertol, de Ilarraza, Arantza Diaz, Izagirre, Eli, Maritxalar, Montse, 2006: *Using Machine Learning Techniques to Build a Comma Checker for Basque*. V Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions. Sydney: Association for Computer Linguistics. 1–8.

Arhar, Špela, Holozan, Peter, 2009: ASES – leksikalna podatkovna zbirka za razvoj slovenskih jezikovnih tehnologij. V Mikolič, Vesna (ur.), *Jezikovni korpusi v medkulturni komunikaciji*. Koper: Založba Annales.

Bajec, Anton, Kolarič, Rudolf, Rupel, Mirko, 1968: *Slovenska slovnica.* Ljubljana: Državna založba Slovenije.

Beeferman, Doug, Berger, Adam, Lafferty, John, 1998: *Cyberpunc: A lightweight punctuation annotation system for speech*. IEEE Conference on Acoustics, Speech and Signal Processing. Seattle, WA, USA.

Blue, Tina, 2008: *Using Commas with Dates*. Pridobljeno 16. 4. 2012, iz http://grammartips.homestead.com/datecommas.html

Breznik, Anton, 1921: *Slovenska slovnica za srednje šole*. Prevalje: Družba sv. Mohorja.

Breznik, Anton, Ramovš, Fran, 1935: *Slovenski pravopis*. Ljubljana: Znanstveno društvo.

Cazinkić, Robert, 2000: Oziralni prilastkovi odvisniki. *Jezik in slovstvo* 46/1. 29-40.

Cedilnik, Danica, 1995: *Cikcak po pravopisu*. Žalec: Sledi.

Dobrovoljc, Helena, 2004: *Pravopisje na Slovenskem*. Ljubljana: Založba ZRC.

Dobrovoljc, Helena, Jakop, Nataša, 2011: *Sodobni pravopisni priročnik med normo in prakso*. Ljubljana: Založba ZRC.

Freund, Y., Mason, L., 1999: *The Alternating Decision Tree Learning Algorithm*. V zborniku ICML '99 Proceedings of the Sixteenth International Conference on Machine Learning. Str. 124–133.

Gojčič, Aleksandra, 2009: *Ločila*. Seminarska naloga. Ljubljana: Filozofska fakulteta, Univerza v Ljubljani.

Gomboc, Mateja, 2009: *Mala slovnica slovenskega jezika*. Ljubljana: DZS.

Hara, Tadayoshi, Chen, Chen, Kano, Yoshinobu, Aizawa, Akiko, 2013: *Modeling Comma Placement in Chinese Text for Better Readability using Linguistic Features and Gaze Information*. V zborniku 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations. Sofija, Bolgarija, 49–58.

Hardt, Daniel, 2001: *Comma checking in Danish*. Paper presented at Corpus Linguistics 2001 conference: Lancaster University (UK), 266–271.

Helfrich, Antje, Music, Bradley, 2000: *Design and evaluation of grammar checkers in multiple languages*. V Proceedings of the 18th conference on Computational linguistics - Volume 2 (COLING '00), Vol. 2. Association for Computational Linguistics, Stroudsburg, PA, USA, 1036–1040.

Hillard, D., Huang, Z., Ji, H., Grishman, R., Hakkani-Tur, D., Harper, M., Ostendorf, M., Wang, W., 2006: *Impact of Automatic Comma Prediction on Pos/Name Tagging of Speech*. V zborniku IEEE/ACL 2006 Workshop on Spoken Language Technology.

Holozan, Peter, 2006: *Dodatne dvoumnosti zaradi popustljivosti analizatorja pri analizi slovenskih stavkov*. V Erjavec, Tomaž, Žganec Gros, Jerneja (ur.): Zbornik 5. slovenske in 1. mednarodne konference JEZIKOVNE TEHNOLOGIJE 2006. Ljubljana: IJS.

Holozan, Peter, 2011: *Samodejno izdelovanje besedilnih logičnih nalog v slovenščini.* Magistrska naloga. Ljubljana: Fakulteta za računalništvo in informatiko, Univerza v Ljubljani.

Holozan, Peter, 2012: *Kako dobro programi popravljajo vejice v slovenščini*. V Erjavec, Tomaž, Žganec Gros, Jerneja (ur.): Jezikovne tehnologije: Zbornik C 15. mednarodne multikonference Informacijska družba IS 2012, 8. do 12. oktober 2012. Ljubljana: Institut Jožef Stefan. 101–106.

Holozan, Peter, 2013: Uporaba strojnega učenja za postavljanje vejic v slovenščini. *Uporabna informatika* XXI/3. 196–209.

Holozan, Peter, 2015: Možnosti uporabe jezikovnih tehnologij za določanje težav pri rabi vejice. V: Dobrovoljc, Helena, Lengar Verovnik, Tina (ur.), *Pravopisna razpotja – razprave o pravopisnih vprašanjih*. Ljubljana: Založba ZRC. 77–92.

Huang, Jing, Zweig, Geoffrey, 2002: *Maximum Entropy Model for Punctuation Annotation from Speech,*. V zborniku *ICSLP-2002*. Denver.

Israel, Ross, Tetreault, Joel, Chodorow, Martin, 2012: *Correcting Comma Errors in Learner Essays, and Restoring Commas in Newswire Text*. 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Montreal, Canada, June 3-8, 2012. 284–294.

Jakubíček, Miloš, Horák, Aleš, 2010: Punctuation Detection with Full Syntactic Parsing. V Gelbukh, Alexander (ur.), *Research in Computing Science 46, Special issue: Natural Language Processing and its Applications*. Mexico:Instituto Politécnico Nacional. 335–343.

Kocjan-Barle, Marta, 1992: *Abeceda pravopisa, vaje*. Ljubljana: Državna založba Slovenije.

Komac, Polona, 1986: *Angleška slovnica po naše*, Četrta izdaja. Ljubljana: Cankarjeva založba.

Korošec, Tomo, 2003: K pravilom za skladenjsko vejico v Slovenskem pravopisu 2001. *Slavistična revija* 51/2. 247–264.

Kovář, Vojtěch, 2014: *Partial Grammar Checking for Czech Using the SET Parser*. V zborniku 17th International Conference, TSD 2014. Berlin Heidelberg: Springer Verlag. 308–314.

Krajnc, Anja, 2015: *Postavljanje vejic v slovenščini s pomočjo strojnega učenja.* Diplomsko delo. Ljubljana: Fakulteta za računalništvo in informatiko, Univerza v Ljubljani.

Krajnc, Anja, Robnik-Šikonja, Marko, 2015: *Postavljanje vejic v slovenščini s pomočjo strojnega učenja in izboljšanega korpusa Šolar*. V Fišer, Darja (ur.): Zbornik konference Slovenščina na spletu in v novih medijih. Ljubljana. 38–43.

Lečič, Rada, 2010: *Raba vejice v slovenščini, gradivo za seminar za prevajalce*. Ljubljana: Veris.

Levec, Fran, 1899: *Slovenski pravopis*. Dunaj: Cesarsko kraljeva zaloga šolskih knjig.

Muster-Čenčur, Monika, 1985: *Nemška slovnica po naše*. Ljubljana: Cankarjeva založba.

Piškur, Karin, 2015: *Postavljanje vejic v slovenskih besedilih z orodjem LanguageTool*. Diplomsko delo. Ljubljana: Fakulteta za računalništvo in informatiko, Univerza v Ljubljani.

Popič, Damjan, 2014: *Korpusnojezikovna analiza vplivov na slovenska prevodna besedila*. Doktorska disertacija. Ljubljana: Filozofska fakulteta, Univerza v Ljubljani.

Robertson, Grant, 2006: *Comma quirk irks Rogers*. The Globe and Mail, 6. avgust.

Rozman, Tadeja, Stritar, Mojca, Krapš Vodopivec, Irena, Kosem, Iztok, Krek, Simon. *Nova didaktika poučevanja slovenskega jezika : sporazumevanje v slovenskem jeziku*. Ljubljana: Ministrstvo za šolstvo in šport: Amebis, 2010. 664 str., pril. http://www.slovenscina.eu/Media/Kazalniki/Kazalnik15/Nova_didaktika_Sporazumevanje.pdf.

Shieber, Stuart M., Tao, Xiaopeng, 2003: *Comma restoration using constituency information*. V Proceedings of the 2003 Human Language Technology Conference and Conference of the North American Chapter of the Association for Computational Linguistics. 142–148

Skaza, Jože, 2003: *Pravopis, Priročnik z vajami*. Ljubljana: Založništvo Jutro.

Slomšek, Anton Martin, 2006: *Blaže in Nežica v nedelskej šoli*. Celje: Društvo Mohorjeva družba.

Stritar, Mojca, 2006: *Oblikovanje korpusa usvajanja slovenščine kot tujega jezika*, Erjavec, Tomaž, Žganec Gros, Jerneja (ur.): Zbornik 5. slovenske in 1. mednarodne konference Jezikovne tehnologije. Ljubljana: Institut "Jožef Stefan".

Šek Mertük, Polonca, 2011: *Vejica premalo ali preveč pri študentih razrednega pouka*. Revija za elementarno izobraževanje, letnik 4, št. 1–2, str 123–146.

Šolar, Jakob, 1959: *Veznik in*. Jezik in slovstvo, letnik 5, številka 2, str. 14-18. URN:NBN:SI:DOC-SASZPN6Q from http://www.dlib.si

Šuman, Josip, 1884: *Slovenska slovnica za srednje šole*. Celovec: Družba sv. Mohora. URN:NBN:SI:DOC-GTGYH0QY from http://www.dlib.si

Thomson, A. J., Martinet, A. V., 1986: *A Practical English Grammar*, Fourth edition. Oxford: Oxford University Press.

Toporišič, Jože, 1975: *Slovenski knjižni jezik 1*. Maribor: Založba Obzorja.

Toporišič, Jože, 1984: *Slovenska slovnica*. Ljubljana: Založba Obzorja.

Toporišič, Jože, 2001: *Slovenski pravopis*. Ljubljana: SAZU.

Verovnik, Tina, 2005: *Jezikovni obronki*. Ljubljana: GV založba.

Vogel, Jerca, Kastelic, Silva, Hodak, Marjana, 2010: *Slovenščina 4: z besedo do besede*. Ljubljana: Mladinska knjiga.

Vodnik, Valentin, Eger, Janez Leopold, 1811. *Pismenost ali gramatika sa perve shole*. URN:NBN:SI:DOC-DU5DNOB1 from http://www.dlib.si

Weiss, Peter, 2002: *Pregled ločil v slovenskih besedilih protestantskih piscev 16. stoletja*. Med dialektologijo in zgodovino slovenskega jezika. Maribor: Zora 18. 250-270.

Zhang, Zhu, Gamon, Michael, Corston-Oliver, Simon, Ringger, Eric, 2002: *Intra-sentence punctuation insertion in natural language generation*. Tehnično poročilo MSR-TR-2002-58. Microsoft Research.

Žagar, France, 1987: Pouk slovenske slovnice in pravopisa v višjih razredih osnovne šole. Maribor: Založba Obzorja.

Žagar, France, 1991: Slovenska slovnica in jezikovna vadnica. Maribor: Založba Obzorja.

Žele, Andreja, 2003: Glagolska vezljivost: iz teorije v slovar. Ljubljana: Založba ZRC.

Žibert, Živa, 2006: Slovenska vejica: balast ali skladenjska nujnost slovenskega knjižnega jezika?. Diplomska naloga. Ljubljana: Fakulteta za družbene vede, Univerza v Ljubljani.

več avtorjev, 1950: *Slovenski pravopis*. Ljubljana: DZS.

več avtorjev, 1990: *Slovenski pravopis: 1, Pravila*. Ljubljana: Državna založba Slovenije.

9 Priloge

9.1 Pravila za rabo vejice v SP 2001

Vejica

SKLADENJSKA RABA

Vejico (,) pišemo: 1. med enakovrednimi deli proste ali zložene povedi; 2. med nadrednim in odvisnim stavkom (tudi ko gre za primerjalni odvisnik s *kot, kakor* ipd.); 3. med polstavkom in preostalim besedilom povedi; 4. med izpostavljenim stavčnim členom ali dostavkom in preostalim besedilom povedi; 5. med izrazi in stavki, ki niso stavčni členi ali deli priredja oz. podredja.

Med skladenjsko enakovrednimi deli proste ali zložene povedi

Teh delov je več vrst.

Nestavčni enakovredni deli

Ti so:

– vezalni (naštevanje): Polje, vinograd, gora, morje, ruda, kupčija tebe rede. – *Naši ljudje so padali v Krškem, v Dragi, v Dachauu, v Celju, na Rabu*. – Večja slovenska *mesta so Ljubljana, Maribor, Celje, Kranj, Nova Gorica, Koper*. – Bila je močna, bistra, oblastna ženska. – Psiček laja hov, hov, hov. – Verjel v prestolov trhlo sem oblast, v junaštvo šlev, v pismarjev vseh modrost, v pobeljenih grobov svetost, v glumačev, zvodnikov, skrivalcev čast. – Slovenci so se izseljevali deloma v Nemčijo, deloma v Francijo, deloma v Ame*riko*;

– stopnjevalni: *Ne le ti, tudi tvoj brat naj pride*. – Bil je talentiran, celo genialen učenec. – *To ni bilo samo grdo, ampak tudi nepošteno*;

– ločni: Bi kavo, čaj?;

– protivni: *Lepa, vendar nevarna pot ga je mikala*. – *Fant je malo* počasen, pa zanesljiv;

– vzročni in posledični: *Neizkušena, kajti mlada, je odšla v svet*. – *Mlada, torej neizkušena, je odšla v svet*. – *Slab, zato pa poceni izdelek so nam ponujali*;

– pojasnjevalni: *Prišel je pozno, šele ob treh*. – *V Ljubljani, 9. II. 1973*. – *Trst, dne 11. oktobra 1983*. – *Ivan Cankar, Na klancu*. – *Ivan Grohar, Sejalec*. – *Zbolel je za hudo, nalezljivo boleznijo* (tj. *za hudo, in sicer nalezljivo*). – *Predava dvakrat na teden, in sicer v torek in petek*.

Posebnosti

298 1. Kadar levi pridevnik določa celo preostalo besedno zvezo, začenjajočo se s pridevnikom, za njim ne pišemo vejice: *dolgi kodrasti lasje, huda nalezljiva bolezen*.

299 2. Zveze kot *v nedeljo, 9. maja,* (*bo*) ... se redkeje pišejo tudi brez vejic, tj. *v nedeljo 9. maja* (*bo*). Tip kot *v nedeljo ob 9. uri* (*bo*) … se navadno piše brez vejic. Če prvi in drugi primer združimo, se navadno piše takole: *v nedeljo, 20. maja, ob 17. uri* (*bo*).

300 3. Če enakovredne naštevane enote, ki niso povedi, navajamo v stolpcih, vsako v posebni vrsti, vejice med posameznimi enotami lahko opuščamo (na koncu pa tudi končno ločilo), npr.

Prošnji je treba dodati:	*Obvestijo se:*
1. spričevalo 4. *razreda srednje šole*	*predsednik*
2. izpisek iz rojstne matične knjige	podpredsednik
*3. zdravniš*ko spričevalo	člani
4. življenjepis	računovodstvo
5. potrdilo o opravljeni praksi	*arhiv*

301 Pristavčni deli

Vejica je na obeh straneh: Pri Prešernu, največjem slovenskem pesniku, so se učili vsi kasnejši pesniški rodovi. – *Od mene, svojega* prijatelja, tega ne pričakuješ. – Največji dramatik vseh časov, Shakespeare, se je rodil v Stratfordu na Angleškem. – Nekateri ptiči, npr. (ali *recimo, denimo*) *lastovke, štorklje in škorci, se na jesen selijo*. – *Svetovni slovenski kongres, Konferenca za Slovenijo, prireja posvetovanje o izseljenstvu*. – Slovensko narodno gledališče, Opera in balet, ima na sporedu tri nove predstave. – *Podjetju Mavrica, d. d., grozi stavka*.

302 Pomni

Na koncu povedi druge vejice pristavka ne pišemo, temveč samo ustrezno končno ločilo: *To povest je napisala Zofka Kveder, por. Jelovšek*. – Oglasil se je Stane Suhadolc, po domače Kovačev.

O pisanju podobnih primerov brez vejice prim. § 316–324, 326–332, 344, 345, 357.

303 Posebnost
Če je lastno ime določilo pred njim stoječe besedne zveze, ga od nje ločimo z vejico, kadar domnevamo, da zveza naslovniku tega imena ne napoveduje samoumevno: Največji slovenski pesnik, France Prešeren, se je rodil v Vrbi, sicer pa ne: Največji slovenski pesnik France Prešeren se je rodil v Vrbi.

304 Stavčni enakovredni deli

Ti so:

305 – vezalni: *Pod nebom ptice letajo, ribe v vodi plavajo, v hostah zamira glas sekire, s polja sejalec gre domov*. – *Na nebu zvezde sevajo, na vasi fantje pevajo*. – (*In vedela sva,*) *kje kosi mladijo, kje drozdi*, kod divji mo. hodi, kje vile se skrivajo v gozdi. – (*In spomnil sem se,*) *da je davno ni*, da veter ji čez daljni grob šumi;
– stopnjevalni: *Ni le zabavljal, ampak se je tudi pretepal*. – *Ni se niti* zahvalil, kaj šele da bi bil plačal;
– ločni: *Zdaj je jokal, zdaj se je smejal*;
– protivni: *Vse ima, vendar ni zadovoljen*. – *Mimo mene si šel, in* (= vendar) *se nisi oglasil*. – *Veliko si je prizadeval, ali uspeha ni bilo*. – Rad ima družbo, še rajši pa (*ima*) *knjigo*. – *Trubar je bil prvi, Prešeren* (*je*) *edini*. – *Danes* (*si*) človek, jutri (*boš*) črna zemlja;
– vzročni: *Ne hodi po progi, je nevarno*. – *Odhiteli so pospravljat seno, kajti pripravljalo se je k nevihti*;
– posledični: *Prosili so me, pa sem jim ustregel* (= zato). – Večerilo se je, zato so gnali črede domov. – *Bilo nas je devet, in* (= zato) *eden je moral od doma*. – *Zatri knjigo, in tema* se *zgrne nad nami* (= Če nam zatre. knjigo, se tema zgrne nad nami);
– pojasnjevalni: *Iz samostana je zvon zavabil, to molijo nune*;
– v bibliografskih zapisih med naštevanimi enotami: *Milko Kos, Srednjeveška Ljubljana, Topografski opis, 1955, 79 strani*. (Prim. § 235.)

Pomni
O pisanju podobnih primerov brez vejice gl. 298–300, 316–324.
Z vejico se ločijo tudi vrinjeni stavki: *Da*leč, ne vem kje, se je oglasil žvižg. – *Jure, ne bodi len, je pobral kost in jim namlatil rebra*. – *Prešeren je bil, kakor je splošno znano, po poklicu odvetnik*. – *Moj prijatelj, zakaj ga ne bi imel tudi jaz, ima novo hišo*. – Cankar, to vemo, je živel dalj časa na Dunaju.

Nestavčno-stavčni enakovredni deli

Pozimi, vendar tudi kadar je mraz, kurimo. – *Previdno, toda ne da bi se obotavljali, smo šli za njimi*. – *Pozimi in kadar je mraz, moramo zakuriti*.

Posebnosti
Vejice med skladenjsko enakovrednimi deli ne pišemo:
1. pred nekaterimi prirednimi vezniki: vezalnimi *in, pa, ter*, ločnimi *ali, oziroma, bodi*(*si*) in pred drugim delom dvodelnih vezniških zvez *ne – ne, niti – niti, tako – kakor* (vezalne *in, pa, ter* in ločni *ali* spoznaš po tem, da jih ni mogoče zamenjati z vezniki tipa *vendar, zato, kajti* ipd.):
a) vezalni: Oče in mati sta mu že davno umrla. – *Ivan in Pavla sta brat pa sestra*. – *Hana prinese steklenico in kozarce ter šunko*. – *Poješ mi pesem vetra in vej in trave in sonca na travi*. – *Sem dolgo upal in se bal*. – Zavriskaj v lepi svet in utrgaj kresni čarni cvet in praprotnih semen nastrezi in na mah pod bukev lezi! – (*Premišljeval je,*) kako ga ima rada in kako mu streže. – (*Rekel je,*) *da pride in prinese kruha*. – (*Hitro je nataknila rokavice,*) stoječ med vrati dnevne sobe in razgledujoč se okoli sebe. – Poiskal je vžigalice ter prižgal svečo. – Z rokami se je oprl na stol pa se počasi dvignil;
b) stopnjevalni: *Nisem ne tat ne ubijalec*. – *Društvo se je ohranilo do današnjega dne pa tudi njegova pravila še veljajo*;
c) ločni: *Stopi ven ali noter*. – *Ali delaj doma ali pojdi v svet*. – Ali ne more ali pa noče. – (*Prejkone je bila v dvomu,*) *ali naj me tika ali vika*. – *Bo ali ne bo?* – *Prišel bo predsednik oziroma njegov namestnik*. – Ta poročila so bodi(*si*) resnična bodi(*si*) *zlagana*. – *Prebivalci so se* bodisi poskrili bodisi razbežali. – *Naš gost tega ni opazil oziroma tega ni pokazal*;

Pomni

320 1. Pred naštevalnim *ali, ne* oz. *niti* se vejica lahko piše: *Kdo lase* ti razčesava, ali mati, ali sestra, ali vila iz goščav? – Nisem ji rekel ne zvečer, ne drugi dan, ne *ob slovesu*. – Niti mati, niti sestra, niti vila iz goščav (mi ne razčesava las). – Ne kral nisem, ne ubijal, ne prešuštvoval, čista je moja vest. (Povsod več kot po dva veznika.)

321 2. Pred vezalnimi *in, pa, ter*, ločnimi *ali, oziroma, bodi*(*si*) ali pred stopnjevalnimi *ne – ne, niti – niti, tako – kakor* pišemo vejico, če jo zahteva vrinjeni ali vmesni stavek oz. dostavek: Moj prijatelj, tudi jaz imam prijatelja, in njegova žena živita v Kranju. – *Da pride, je rekel, in da prinese kruha*. – Povej mi, kje si bil ves ta čas, in pojdi potem hitro na postajo. – Stal sem tam in vse videl z lastnimi očmi. Vejico pred vezalnim *in* lahko pišemo tudi tedaj, če sta osebka prirednih stavkov slabo družljiva:
Komaj je zatisnil oči, se je zbudil, in dan je bil.

322 2. kadar posamezne izraze (lahko tudi okrepljene) ponavljamo, ne da bi jih v govoru ločili s premori ali z načinom izgovora: Beli so beli češnje cvetovi. – Daj daj, povej še enkr*at*. – *Perilo je belo belo*. – *Stopi stopi k sosedu in ga prosi za lopato*. – Beži beži, kdo ti bo to verjel! – *Priden priden, fantek moj!* – *Joj joj, kako boli!* – *Sultan je zalajal hov hov hov*. – Hodil je hodil celo noč, vendar ni nikamor prišel. – *Rjuhe so bele prebele*. – *Sam*

samcat je prišel; če s ponovitvijo stopnjujemo ali poudarjamo isti pomen, vejico pišemo: *Daj, daj vendar*. – *Sam, samcat je*;

323 3. med prirednimi stalnimi besednimi zvezami z izpuščenim veznikom *in, ali*, npr. *dva tri* (*dni*), (*hodi*) *gor dol, na vrat na nos*, križem kražem, hočeš nočeš, hočeš nočeš moraš;

324 4. neobvezno med prirednimi istovrstnimi naštevanimi prvinami, kjer bi vejice samo motile: 25 latiničnih črk: a b c č . – Ločila so grafična ali pisna znamenja, npr. *.,;!? itd*. – *Končnice* -a -e -i -o so naglašene ali nenaglašene.

325 Med nadrednim in odvisnim stavkom

Ta vejica se stavi ne glede na to, ali je nadredni stavek pred odvisnim, za njim ali pa ga obdaja:

Največja sreča za človeka je, če je zdrav. – *Strašna mu je bila misel, da bo tepen*. – *Hitreje govori, kakor misli*. – Ker je deževalo, so ostali doma. – Daj mu nageljček, da bo tudi on vedel, da je prvi maj. – *Dejal sem jim, da ga ni doma, in jih odslovil*. – Strah, ali se mi stvar posreči, mi je jemal spanec. – *Novica, da* bomo kmalu rešeni, nam je vlila novih moči. – Prej, ko še ni bilo ceste, smo hodili čez vrtove. To vejico pišemo tudi, ko gre za primerjalni odvisnik s *kot, kakor* ipd.: Ko sem se vrnil domov, sem spoznal, da je vse kaj drugega, če človek stoji doma, kot č*e misli na dom*.

Ni jasno, kaj misli o tem. – *Radovednost, kdo je v hiši, jo je prignala na obisk*. – Povej mi, s kom občuješ, in povem ti, kdo si. – Ne vedel bi, kako se v strup preobrača vse, kar srce si sladkega obeta.

Kdor za smolo prime, se osmoli. – *Koder hodi, mu je s trnjem pot posuta*. – *Poberi se, od koder si prišel!* – Kolikor jezikov znaš, toliko mož veljaš. – Med rožami, ki ne rasto pri nas, sem videl polzasenčen, tih obraz.

Vejico pišemo tudi tedaj, kadar je povedek (ali vsaj osebna glagolska oblika) kakšnega stavka izpuščen (tu zmeraj v oklepaju): še zmeraj kaj dela, čeprav (*je*) že v letih. – *Kolikor bo dal on, toliko* (*bom dal*) *jaz*. – Pomeniva se prijateljsko, če (*se*) *ne* (*bova*), *takoj odidem*. – *Sosedje niso hodili k nam zato, ker je bila naša hiša lepša, ampak* (*so hodili*) *zato, ker je babica znala toliko lepega povedati*. – Plačal bom, kadar (*boš*) *ti*. – *Odslovil sem ga, kakor* (*sem to storil*) *nerad*.

Pomni

Pred členkom *da* vejice ne pišemo: *Ti da tega ne veš? – Baje da pije*. – *Iz Blatnega Dola da ste?*

Posebnosti

1. Kadar prideta skupaj dve podredni vezniški besedi, med njima vejico opuščamo: Le piši materi, da če ne bo denarja, ne bo. več stanoval pri nas. – *Skusili ste, da kdor se iz prvega prenagli, kmalu odneha*. Tako kopičenje veznikov ni priporočljivo. V starejših besedilih se je tu vejica pogosto pisala: žive naj vsi narodi, ki hrepene dočakat dan, da, koder sonce hodi, prepir iz sveta bo pregnan.

2. Kadar prideta skupaj priredni in podredni veznik pred vmesnim odvisnikom, pišemo vejico samo pred prvim veznikom: *Pridem sam, in če bo le mogoče, pripeljem še koga*. – *Votlina je bila suha, in ker je ne doseže nobena sapica, tudi topla*. – *Medveda je priklenil k stebru, in ko je spet prišel divji mož, sta se spoprijela*. Vejico pišemo pred prvo in pred drugo vezniško besedo, kadar priredni veznik zastopa opuščeni nadredni stavek: Nekaj pa vam obljubim, namreč (*to obljubim*), *da kmalu spet pridem*.

3. Vejice ne pišemo med deli večbesednega veznika: Namesto da bi se učil, je lenaril. – Lenaril je, namesto da bi se učil. – že ko sem ga prvič videl, sem ga spregledal. – *Kljub temu da je bil bolan, je šel na delo*. – To smo odločili, češ da je še čas. – *Medtem ko ste spali, je bilo hudo neurje*. – *Rad ga imam, zato ker je tako miren*. – *Zelo* zgodaj sem se odpravil na pot, zato da mi ne bi bilo treba hoditi v vročini. – *Prej ko mine let in dan, od ljubce pride list poslan*. Taki vezniški izrazi so še: *tako da, toliko da, potem ko, vtem ko, še ko*, brž ko, *šele ko, s tem da, posebno ko, zlasti če …*

V nekaterih primerih razmerje med nadrednim in odvisnim stavkom lahko izrazimo tako, da prvi del prvotno večbesednega veznika prenesemo v nadredni stavek, za njim pa seveda pišemo vejico: *Utihnil je zato, ker je bil mrtev*. – *Rada ga imam zato, ker je tako miren*. – *Zeblo ga je tako, da se je ves tresel*.

4. Pred primerjalnimi *kakor, kot, ko* in stopnjevalnim *tako – kakor* vejice ne pišemo, če veznik ne uvaja odvisnika z izraženo osebno glagolsko obliko: Ne kaže drugega kakor molčati. – *Bolje je delati kot govoriti*. – Več je govoril kot delal. – *Bila je bolj pridna kakor nadarjena*. – *Za narod* se *briga kakor za lanski sneg*. – *Bil je tako evropski kakor svetovni prvak*. Vendar: Moral bi bil več delati, kakor je.

5. Za vmesnim stavkom pred besedo, ki nadomešča opuščeni odvisnik, enak prvemu delu priredja, pišemo vejico: *Prišel je, sam ni vedel, zakaj* (*je prišel*). Večkrat je to mogoče pisati tudi brez vejice.

6. Tudi med izrazi, nastalimi iz stavkov, ki na začetku povedi ali pred kakšnim notranjim stavkom izražajo razmerje govorečega do vsebine povedanega: *Skratka, uspeh je presegel vsa* pričakovanja. – *Sicer pa, kaj bi vam še naprej govoril*. – *Najprej, tole mora ostati med nami*. – *Kratko in malo, do konca naslednjega tedna mora biti rokopis pripravljen*. – Uspeh je presegel vsa pričakovanja, skratka, vsi so bili zadovoljni.

Med polstavkom in drugim delom povedi

Kadar je jedro polstavka deležje, deležnik, pridevnik, samostalnik:

Želeč jim lahko noč, so se poslovili. – *Prijazno meketaje vrh skale, sta kozi vabili ovce k sebi*. – Slednjič ga je zavrnil, rekoč: »Kaj pa to tebi mar!« – Ne misleč na posledice, je vse povedal. – Hodil je po sobi, pogledujoč na vse strani. – Pri teti sem prebil vsake počitnice, začenši od druge šole. – *Slovenec mile ne ljubi matere, vanj* upajoče. – *Od doma se mi zdiš poslanka,* nesoča mnog mi ljub pozdrav. – (Dokler ptič po drevju letaš) ga jed, v tičnici nastavljena, *prijazno k sebi vabi*. – Iz moje mladosti, vse polne ponižanja, se svetijo tiho tvoje oči. – *Fant pa se je, velik navihanec, pri tem samo muzal*.

Če je pridevniški polstavek levo od odnosnice, ga ne ločimo z vejico: Iz moje ponižanja polne mladosti se tiho svetijo tvoje oči. Samostalniških zvez, nastalih iz odvisnika, z vejico ne ločimo od okolja: *Ob slovesni pod*elitvi bralnih značk osnovnošolcem (= Ko so slovesno podelili bralne značke osnovnošolcem,) *je govoril tudi šolski ravnatelj*.

Posebnosti

334 1. Kadar je deležje golo, brez dopolnila, ob njem vejico pišemo, če ga razumemo kot polstavek, drugače pa ne: *B*oječ se, se je deklica približala skupini na trati ali Boječ se se je deklica približala skupini na trati.

335 2. Deležij po nastanku ne ločimo z vejico od drugega besedila, kadar so prislovi: *Opotekaje se je hodil po sobi*. – Molče je odšel. – *Skrivši jih je obiskovala*. – Odšli so omahujoč.

Prim. prislov iz deležnika: Boječe se je deklica približala skupini na trati.

336 3. Deležniških in pridevniških polstavkov ne ločimo z vejico od preostalega besedila, če so navadni, tj. levi prilastki: Preteči prej *oblak* (na polja ulije dež krotak). – Po sreči hrepeneče (*srce je nemirno*). – (Dokler ptič po drevju leta, ga) v tičnici nastavljena (*jed prijazno k sebi vabi*). – (*Iz moje*) vsega ponižanja polne (mladosti se svetijo tiho tvoje oči).

337 Kadar nedoločniški polstavek lahko razvijemo v osebno glagolsko obliko, ga lahko ločimo z vejico ali pa ga pišemo brez nje: Imel je željo, postati zdravnik (= Imel je željo, da bi postal zdravnik). . Njegova želja je, postati zdravnik in Njegova želja je postati zdravnik. – Postati zdravnik, je njegova želja in Postati zdravnik je njegova želja. Prim. še *Lenariti ves dan, ga je sram* (= Da bi lenaril …). – Moja dolžnost je, govoriti resnico (= Da govorim resnico.). Če pa se nedoločniški polstavek rabi kot povedkovo določilo ali povedkov prilastek, se od sobesedila ne sme ločiti z vejico: *Ne moreš oditi kar tako*. – *Treba je pridno delati*. – Začelo je močno snežiti. – Slišala je ptičko peti.

338 Med izpostavkom ali dostavkom in preostalim delom povedi

339 a) Izpostavek: *Sin, ta ti bo šele zagodel! – Ali se ga bojiš, te reve! – Torej so ga, lisjaka zvitega, le ujeli!* – Vprašaj ga, vsevedneža. – V mojih mladih letih, takrat je bilo vse drugače. – Postati zdravnik, to je njegova želja;

340 b) dostavek: *Prinesi nam tudi kr*uha, črnega. – Prinesi mi sivi plašč, in rokavice. – *Visoko na vrhovih jamborov zastave veselo plapolajo, kot privid* (= in sicer kot privid). – *Sonce je sijalo kakor v mrzli vodi, veliki mirni reki tam gori*. – *Perje je vrabcu viselo navzdol, nerodno in zmršeno*. – Stanuje v eni ljubljanskih občin, bežigrajski. – Pridi jutri, sam, točno. – *Veronika je prišla domov, sama, in celo kmalu*.

341 Med pastavki in drugim besedilom

342 a) Zvalniki in ogovori: Oče, od medu mi dajte sat prelep. – *Pojdi, moj sinko, na pot!* – Gospa, kar po domače. – *Naprej, zastave Slave! – Mladosti leta, kmalu ste minila!* – Ti reva ti, ti bo. mene učil! – Odloči se že, mevža mevžasta!;

343 b) samostojno rabljeni medmeti in medmetne zveze: *Joj, kam bi del!* – Hudiča vendar, kaj pa n*orite! – Bum bum bum, je zagrmelo po vratih*. – *Lej lej, pa je le prišel*. – *Zdravo, Tine, ali me res ne poznaš? – O, Janez, kam pa ti?* – Dober dan, kako kaj živite? – Križ božji, kakšen pa si;

Pomni

1. Medmeti (ali medmetne zveze) niso rabljeni samostojno, tj. kot stavki, v primerih, ko so stavčni členi: *Miška pa smuk v luknjo*. – *Zunaj hej, znotraj fej*. – Krava reče mu. – *Štrbunk je reklo*. – *Fej te bodi! – Recite da in hajdi z menoj*. – *Pošast hodi škrab škrab škrab*. – *Pozdravil je z dober dan*.

2. Medmet se od zvalnika ne loči z vejico, kadar v govoru ni ločen od njega s premorom: *O mati, o domovina!*

c) samostojni členki: Da, vse je čisto. – *Ne, to ni samo površje*. – Tako je rečeno, *da*. – Pač, še ena rešitev je. – *Nikar, ne bo prav*. – *No, pa vstopite! – Na, vse ti zna! – Vsekakor, jutri bodo pripravljeni*. – *Je spet pijan, kajpak*. – Zares, take reči motijo. – *Kako, še tajil boš!* – Kaj, tudi ti si prišel tako daleč?! – *Le, samo previdno*. – *Seveda, tega ne tajim*. – *Nikakor, tega si ne dam odvzeti*. – Čez mesec dni ti vrnem, prav gotovo. – *Si bil tam, a?* – Porinil je fanta naprej, češ, tu ga imate. – Nekaj pa vam obljubim, namreč, da kmalu pridem.

Pomni

1. Pisanje ali opuščanje vejice pri nekaterih členkih je odvisno od stilnega poudarka takih izrazov: *Tega seveda ne tajim*. – *Take stvari se, seveda, ne povedo na glas*.

2. Zlasti če je poved zapletena in je v njej že več vejic, vrinjeni stavek ali več stavkov od drugega dela povedi rajši ločimo s pomišljajem: Komaj je zatisnil oči – človek bi mislil, da je minila le minuta – se je zbudil in se začel pripravljati na pot.

NESKLADENJSKA RABA VEJICE
Neskladenjsko se vejica rabi:
1. pri pisanju imen in večbesednih poimenovanj, ko jih iz tehničnih, bibliografskih in podobnih razlogov pišemo za priimkom, npr. v seznamih ali stvarnih kazalih; s tako vejico nakazujemo, da smo pravilno zaporedje sestavin (npr. za abecedni seznam) spremenili:
Molan, Ivan kruh, črni
Potočnik, Marko kruh, beli
Novak, Franc sir, bohinjski
2. kot znamenje za mejo med celim številom in decimalkami: *1,50* ali *1,572*;
3. včasih namesto presledka za označevanje milijonic (če so v istem zapisu tisočice zaznamovane s piko): *52,500.000.- SIT*. Prim. § 255.

9.2 Vejica v starejših slovnicah in pravopisih

Za primerjavo smo zbrali tudi zglede za rabo vejice iz starejših slovnic in pravopisov.

9.2.1 Slovenska slovnica za srednje šole (Josip Šuman, 1884)

Vejica loči: 1. v prostem stavku a) jednake člene, če niso z veznikom in, ter, ali, i – i, niti – niti zvezani, n. p. *Polje, vinograd, ruda, kupčija tebe rode. Stoji, stoji tam lipica, prelepa, hladna senčica.*
b) pristavek in zvalnik, n. p. *Sv. Ciril in Metod, slovanska aposteljna, sta postavila temelj slovenskemu slovstvu. Ura, enkrat zamujena, ne pride nobena. Lambergar, le molči ti!*
c) samostalne medmete, izklice, rotitve, če stoji glavna ločilnica za celim stavkom, n. p. *oj, kak se mi smiliš! Na, vzemi! Za boga, kaj delate!*
2. v sostaviju posamezne stavke, n. p. *Čas je veter, ki le pleve razpihava in le zrno pušča* Levst. *Koder se nebo razpenja, grad je pevca bez vratarja* Preš. *Zavpije Zora, zakriči, raz konja pade, omedli* n. ps.
Op. Vejica so opušča med prirednimi stavki, če imajo isti osebek in so med seboj zvezani z veznikom *in, ter, ali, i – i, niti – niti*, n. p., *lisičjo mast so nekdaj lovci drago prodajali in jo sem ter tje še prodajajo* Erj.

9.2.2 Slovenski pravopis (Fran Levec, 1899)

Vejica (,) nam rabi:
I. V prostih stavkih, in sicer:
1. Za zvalnikom in, kadar stoji zvalnik sredi stavka, tudi pred njim, da ogovorjeno osebo ali stvar loči od stavka, n. pr. *Pegam, pegam, pridi vun! – Lambergar, le molči ti! – Preveč te, mati, ljubil je Orest! – Čas, Črtomir, je vzeti orožje! – Kaj boš s sekiro, sin, začel? –*

Vam, ptice, odslej bom prepeval,
Vam, rože, srce razodeval! ...
Strune, milo se glasite,
Milo, pesemca, žaluj!

2. Pred pristavkom in za njim, kadar pristavek z besedo, ki jo pojasnjuje, ni zvezan v en pojem, n. pr. *Frančišek Prešeren, najslavnejši slovenski pesnik, je umrl l. 1849. – Rudolf Habsburški je premagal Otokarja, kralja češkega. – Na Notranjem stoji vas, Vrh po imenu. – V tej vasici je živel Martin Krpan, močan in silen človek. –*

Oj, sin, večerna zvezda mojih dni,
Tolažba moja, moj ponos! ...

Tam po nebu zvezda plava,
Svetla zvezda, vlažena danica ...

Toda: *Habsburžan Rudolf Utemeljitelj je ustanovil Novo mesto. – Francoski kralj Ludovik Sveti je umrl v Afriki.*

3. Pred besedami in kraticami, ki napovedujejo kakšno naštevanje (*kakor, in sicer, in to, n. pr.*), n. pr. *Zdanji svet tarejo razne nadloge, kakor vojska, kuga lakota. – Zgodovina pozna mnogo slavnih vojaških poveljnikov, n. pr. Aleksandra, Hanibala, Cezarja, Napoleona. – Jurčič je spisal več lepih povesti, in sicer: "Jurja Kozjaka", "Domna" in "Desetega brata". – Kranjsko delimo na tri dele, in to: v Gorenjsko, Dolenjsko in Notranjsko.*

4. Za samostojnimi medmeti, vzkliki in rotitvami, kadar stoji glavno ločilo na koncu stavka, n. pr. *Na, Krpan, pij in jej! – Da, romarica sem, gospod! – Oh, krasen pač čaka nas dar! – Gorje, kdor nima doma! – Sedanji svet zahteva mnogo, da, premnogo beriva. – Za Boga, kaj delate? – Ostrigli, oh, so mi peroti! –*

Krasnejše deže, nego je ta,
Ni, bogme, je, – to vam rečem ...
Ne, palice svoje ovčarske
Za žezlo kraljevo ne dam! ...
Res, hrabro srce ti imaš,
Res, vrl prevoznik si ti naš! ...
Oj, pašniki solnčni, lesovje temno,
Pri srcu ko nekdaj sta meni! ...
Joj, kam, bi del? – "I, kjer si vzel!" –
"Hu, to teži!" ...

Ne piše pa se vejica za takimi medmeti (*o, oj, i*), kadar nimajo posebnega poudarka, n. pr. *I kaj pa nosiš v tovoru? – Oj padi, padi, gosti mrak! – Oj kralj Matjaž, oj kralj Matjaž, prelepi kraljič ogrski! – O vrba, srečna, draga vas domača! – "O Brdavs, Brdavs!" vpije malo in veliko.*

II. V skrčenih stavkih, in sicer:

1. Med istovrtnimi stavkovnimi členi, ki niso zvezani z vezniki, n. pr. *Zavpije Zora, zakriči, raz konja pade, omedli. – Voda gine, pada, sahne. – Mi čvrsti Slovenci smo, gremo na boj za pravdo, za dom, za cesarja. – Godba odganja ljudem kalne misli, kroti neugnane strasti, povzdiguje potrto dušo proti nebu. – Le vprašaj gore, dole, griče, zvestobe naše vsi so priče; glej mesto, grad in trg in vas, povsod vdanosti dokaz! – Bil sem v Gorici, v Reki, v Trstu. – Trubar, Dalmatin, Bohorič so najstarejši slovenski pisatelji.*

Kadar stojita pred samostalnikom dva pridevnika kot prilastka, tedaj ju loči samo takrat vejica, če se moreta zvezati z veznikom ***in***, n. pr. *Sofoklej je bil sin mogočne, cvetoče domovine. – Tiha, skrivnostna noč je razprostrla svoja krila po dolini. – O Vrba, srečna, draga vas domala! – Ta puhli, izpačeni svet zapustiva! – Prijeten je zimski večer v mornem, gorkem kotu. – Ne šalite se s starim, nesrečnim možem!*

Kadar pa je drugi pridevnik spojen s samostalnikom v en pojem, in prvi pridevnik ta pojem pojasnjuje samo kot prilastek, tedaj med obema pridevnikoma ne sme stati vejica, n. pr. *Letošnji mali traven je mokroten. – Naš graščak ima jako svojivega lovskega psa. – Lavdon je bil znamenit avstrijski vojaški poveljnik. – Prekrasna Velika noč nam je prinesla grozni ljubljanski protres. – Najbolje je pod ljubo domačo streho.*

Časih premeni vejica v enem ter istem stavku misel, ki jo izraža stavek, n. pr. *Beri dobre slovenske knjige* (t. j. *Beri* ***dobre*** *knjige slovenske, a ne beri* ***slabih*** *slovenskih knjig*) ... *Beri dobre, slovenske knjige* (t. j. *Beri dobre knjige, in to* ***slovenske****, a ne beri knjig v tujem jeziku*) ...

2. Med istovrstnimi stavkovnimi členi, ki so zvezani med seboj z vezalnimi vezniki (*tudi, ni, niti, niti – niti, ne – ne, ne le – ampak tudi, potem nato vrhutega, zdaj – zdaj, časih – časih,*

nekaj – nekaj, nekoliko – nekoliko ...), n. pr. *Ni sebi, ni drugim se ne laži! – Pameten človek ne gleda samo na besede, ampak tudi na srce. – Lama je nekoliko podobna kozi, nekoliko pa velblodu. – Ob potresu se zemlja zdaj vzdiguje, zdaj upada.*

Kadar pa so takšni istovrstni stavkovi členi zvezani med seboj z vezniki *in, ter, ali, pa i – i, ravnotako – kakor*, tedaj se med njimi ne sme pisati vejica, n. pr. *Na Starem so trgu po lipo zeleno trobente in gosli in cimbale pele. – Bil godec je mlad in lep in vesel. – Vali se in hrka in vre in kipi. – Čuj njih vik in krik in stok in jok! – Kresilno gobo pa nekaj brusov sem naložil – Moške ali krepke rime se ne posrečijo vsakemu pesniku. – Rak si vsako leto sleče svoj oklep ter si preskrbi novega. – Iz zlata se kujejo novci pa tudi druge dragotine. – Konon je dal i pirejski i atenski zid popraviti. – Prešernove poezije so ravnotako umetne kakor preproste.*

Kadar sta dva z *in* zvezana skrčena stavka odvisna od podrednega veznika ali oziralnega zaimka, tedaj pred *in* ni pisati vejice, n. pr. *Ko je bil Krpan premagal velikana in od cesarja prejel zasluženo darilo, se je povrnil v svojo domovino. – Ko je bil Leonida izvedel, da so Perzijani prišli črez goro in ga mislijo prijeti za hrbtom, je odpustil svoje zaveznike. – Boj se tistih ljudi, ki se ti v obraz sladkajo in te črez mero hvalijo.*

Tudi kadar se v takšnih primerih podredni veznik ali oziralni zaimek ponavlja in sta tedaj oba stavka samostojna, se pred *in* ne piše vejica, n. pr. *Ko je bil Krpan premagal velikana in ko je bil od cesarja prejel zasluženo nagrado, se je povrnil v svojo domovino. – Ko je bil Leonida izvedel, da so Perzijani prišli črez goro in da ga mislijo prijeti za hrbtom, je odpustil svoje zaveznike. – Boj se tistih ljudi, ki se ti v obraz sladkajo in ki te črez mero hvalijo.*

Ravnotako pred primerjalnimi členicami *ko, kot, kakor, nego* ne pišemo vejice, kadar v nasebniku ali primerniku po skrčitvi primerjalnega stavka ne ostane nič drugega nego primerjani predmet, n. pr. *Berač je ravnotako tvoj bližnji kakor bogataš. – Hujša je žeja ko glad. – Boljše poštenje nego sramotno življenje. – Boljša pamet kakor žamet.*

III. V priredjih, in sicer:

1. Med krajšimi glavnimi stavki, iz katerih je zloženo priredje, n. pr. *Visoko vrh planin živim, v veselju rajskem tu živim. – Očetov zgled ti bodi pred očmi, in dika domovini boš enkrat. – Na Nanos se hodi, na Triglav pleza. – Žetev je sicer velika, ali delavcev je malo. – Ne sodimo dela po človeku, ampka človek se sodi po delu. – Niti bodi med, niti bodi jed (strup).*

2. Pred vgozdenimi kratkimi glavnimi stavki in za njimi, n. pr. *Bo, mislim, pisemce za te. – Blag, vem, te goni nagib. – Moj brat se, pravijo, še letos povrne od vojakov.* – Ravnotako tudi v podredjih, n. pr. *Če se postavimo na noge, to Vam pravim, potelj bo svet vedel, kaj smo Bohinjci. – Tukaj stanuje slovenski narod, lahko bi rekel, kar svet stoji.*

3. V premem (dobesednem) govoru pred vgozdenim in za vgozdenim napovednim (uvodnim) stavkom; ravnotako pred napovednim stavkom, ki stoji na koncu premega govora, n. pr. *"Nihče," je rekel Sokrat Krezu, "se pred smrtjo ne more srečen imenovati." – "Nihče se pred smrtjo ne more srečen imenovati," je rekel Sokrat Krezu. – "Ti si vrl mladenič," reče cesarica. – "Črnilo sem," zakliče, "pero in papir!"*

Kadar pa stoji napovedni stavek za premim vprašalnim ali pa klicalnim stavkom, tedaj se pred napovednim stavkom ne piše vejica, ampak vprašaj, oziroma klicaj, n. pr. *"I kaj pa nosiš v tovoru?" vpraša cesar Krpana. – "I kaj? Kresilno gobo pa nekaj brusov sem naložil, gospod!" odgovori Krpan. – "Ako so brusi, pokaj so pa v vrečah?" se začudi cesar. – "O Brdavs, Brdavs!" vpije malo in veliko.*

IV. V podredjih, da loči odvisne stavke od glavnih stavkov in ravnotako podredne odvisne stavke od nadrednih odvisnih stavkov, n. pr. *Čast je ledena gaz, ki hitro zvodeni. – Ne spravljajte si zakladov na zemlji, kjer jih rja in molj konča! – Kdor hoče visoko priti, mora trden v glavi biti. – Dobro mora biti podkovan, kdor hoče na Triglav. – Veselite se z menoj, ker našel sem ovco, ki sem jo bil izgubil! – Nevoščljivost ni sita, dokler ni pokončano, kar ona*

osvraži. – Da zdaj, ko že na Kranjskem vsak pisari, jaz tudi v trop, ki se poti in trudi, se vriniti želim, se mi ne čudi!

V posledičnih podretjih se piše vejica pred veznikom ***da*** ali pa pred ***tako da***, n. pr. *Oče je bil zamišljen tako, da ni precej vedel, kaj hoče sin. – Oče je bil zamišljen, tako da ni precej vedel, kaj hoče sin.* – Napačno: *Oče je bil zamišljen, tako, da ni precej vedel, kaj hoče sin.*

V. V skrajšanih stavkih, da skrajšani odvisnik loči od glavnega stavka, n. pr. *Ura, enkrat zamujena, ne vrne se nobena. – Dasi neznan, po rodu brat se, tvoj. – Usliši glase, k tebi vpijoče! – Ptiči so hoteli stržka ubiti, izvedevši njegovo prevaro. – Značajen sam, še v ljudstvu značaj obujaj! – Pizander, od odhajajočega Agezilaja zapuščen, napravi veliko brodovje, hoteč poizkusiti bojno srečo. – Cezar, udobivši za vožnjo ugodno vreme, je odjadral v Britanijo.*

Premagan pri Bohinjskem sam jezeru
Stoji, naslonjen na svoj meč krvavi ...
Na tujem, zgodaj ločen že od doma,
Živim med glasnim hrupom tihe dni.

VI. V periodah in med posameznimi stavki prva in druge polovice (proreka in poreka), n. pr. *Cezar je poslal Lingonom pismo in poročnike, naj Helvečanom ne pomagajo z žitom, niti z drugo stvarjo: ako jim bodo pomagali, da bode ž njimi enako ravnal kakor s Helvečani. – Ako primerjamo Prešernove poezije veličastni simfoniji, mnogovrstni kakor pesnikovo dušno življenje, katera zdaj hrepeni, zdaj se smehlja, zdaj srdi, zdaj se dviguje v jasne višave, zdaj pogreza v temno brezno: potem je "Krst pri Savici" tej simfoniji finale, v katerem se zlivajo in zlagajo vsi glasovi v mehko, otožno harmonijo.*

9.2.3 Slovenska slovnica za srednje šole (Anton Breznik, 1921)

§ 114. Vejica (,) nam rabi le sredi stavka, in sicer:

a) Loči **zvalnike**, samostojne **medmete** in **vzklike** od celotnega stavka, n. pr. *Tako si šla, mladost, tako si mi pobegnila, sreča, brez blagoslova!* (Cankar) – *Gorje, kdor nima doma!* (Jenko) – *Ostrigli, oh, so mi peroti!* (Greg.) – *Mislil sem si: majčkeno poženem, da bodo vsaj konji na suhem. Ali toliko, da sta pretegnila konja, pok, je že ustrelil Francoz name in, fiu, krogla mimo ušes pa, ššk, v vodo.* (Finžgar.)

Če medmeti nimajo posebnega poudarka (n. pr. *i, o, oj*), se ne piše vejica, n. pr. *O rodni dom, o hiše očetove streha ti!* (Žup.) – *Oj sijaj, sijaj solnce, oj solnce rumeno!* (nar. pes.)

b) Loči **pristavek** od besede, ki jo pojasnjuje, n. pr. *Primož Trubar, prvi slovenski pisatelj, je umrl l. 1586.*

Brez vejice pišemo: *Prvi slovenski pisatelj Primož Trubar je umrl l. 1586.*

c) Loči **nedoločnik**, kadar nadomešča odvisni stavek z *da* (*da bi, ne da bi, namesto da bi ...*), n. pr. *Bog ima pravico, grešnike kaznovati. – Tako ga bom premikastil, da se mu nikdar več ne bodo vrnile hudobne želje, po Dunaju razsajati.* (Levst.)

Brez vejice se piše n. pr. *Nemogoče mi je opaziti vse neznatne razločke,* (Žup.) – *Menda ne bo odveč tukaj našteti besede, ki imajo polglasnik v korenu.* (Škrabec) – *Napačno je med lastna imena in njih sklonila devati vezaj.* (Levec) – *Po toči zvoniti je prepozno. – Človeku ni dobro samemu biti.*

č) Loči take stavčne člene, ki se dado zvezati z veznikom "*in*", n. pr. *Iglica gre skozi oplen, blazino, soro in os. – Široka, rodovitna polja se razprostirajo okoli lepe, prijazne poti. – Človeško telo je ustvarjeno po gotovih pravilih: višina, obseg prsi, dolgost vratu, nog in rok, jakost kosti, velikost glave, širokost pleč in razvitost mišičevja morajo biti med seboj v določenem razmerju. – Ne bojim se meča, ne sulice, ne drugega velikanovega orožja.* (Levst.) – *Voda gine, pada, sahne.*

Take stavčne člene moremo zvezati z veznikom "*in*", n. pr.: *Iglica gre skozi oplen in blazino v soro in os. Široka in rodovitna polja se razprostirajo okoli lepe in prijazne vasi.* Pri členih, ki se ne dado zvezati z "in", se ne piše vejica, n. pr. *Blaž Koren je bil znamenit slovenski zemljepisec. – Najbolje je pod ljubo domačo streho.*

Kadar so **istovrstni** stavčni **členi** zvezani med seboj z vezniki **in, pa ,ter, ali, kakor i – i, tako – kakor, ni – ni (niti – niti), ni (niti),** se ne piše vejica, n. pr. *Zemlja se vrti okoli svoje osi in teka okoli solnca. – Otroci, pukajoč cvetna peresca, ugibljejo, pridejo li v pekel ali v vice ali v nebesa.* (Erjavec) – *Kakor Herkul na razpotju smo stali v tem odločilnem trenutku. – Še več grušča kakor Sava sama prinašajo njeni mnogi dotoki.* (Erj.) – *Na Prešernove pesmi smo Slovenci lahko bolj ponosni kakor na vse drugo, kar imamo. – Hrepenenje mu je pokazalo i pot i kraj.* (Žup.) – *Kaj nimate ni srca ni pametai? – Že dve leti je lakota v deželi in bo še pet let, v katerih se ne bo oralo in želo.* (Dalmatin) – *Niti sebi niti drugim (ni sebi ni drugim) se ne laži!*

Opomnja. Tudi pred veznikom **ne, ne – ne**, se ne piše vejica, če se rabi v pomenu ni, niti, ni – ni, niti – niti, n. pr. *Grdih napak ni konca ne kraja.* (Levst.) – *Kaj nimate ne srca ne pameti? – Ni povedal natanko ne ure ne dneva.*

Pri vseh drugih veznikih, n. pr. *ne le – ampak tudi, tudi, a, toda, vrh tega, potem, nato zdaj – zdaj, nekoliko – nekoliko* itd., stoji vejica, n. pr. *Bil je tak netek, da so ljudje zmerom jedli, pa vendar niso bili nikdar siti.* (Levst.) – *Zgodovina nam kaže ne le lepa in slavna dela, temveč tudi strasti in hudobije.*

d) Stoji pred besedami in kraticami, ki napovedujejo kako naštevanje, n. pr. *in sicer, in to, na primer, posebno, zlasti, kakor*, n. pr. *Slovenijo delimo na več delov, in sicer na Kranjsko, Koroško, Štajersko* itd.

e) Loči v **priredjih** in **podredjih** posamezne stavke med seboj, n. pr. *Tvoja roka, moj otec, zemljé mi lepoto je odpirala, ti kazal si njiv mi plodove, sprovajal me v les, razkladal skrivnosti glasov in družine dreves in vedela sva, kje kosi mladijo, kje drozdi, kod divji mož hodi, kje vile se skrivajo v gozdi.* (Žup.)

Posebej pomni: pred **in, pa, ter** se v priredjih ne piše vejica, n. pr. *Časi se izpreminjajo, doba prehiteva dobo in gorje mi, kdor ostane zadaj!* (Cankar.) – *Kralj je predse sklical može in jih vprašal, ali so kaj slišali o zlati hruški in ali rodi ta hruška kak sad.* (Milč.)

Vejica stoji pred temi vezniki le, ako jo zahteva spred stoječi stavek, n. pr. *Sava nastane iz dveh rek: iz Save Dolinke, ki izvira blizu Ráteč, in save Bohinjke, ki prihaja iz Savice. – Velik razloček je med zakoni, ki vladajo naše svobodne sile, ter zakoni, ki vladajo naše telesne sile.*

f) Loči **deležnike**, ki se rabijo namesto zavisnega stavka, od ostalih stavčnih členov, n. pr. *Točiš svetlicam v čaše medu, da pride bučela, gnana od tajne skrbi, spomladi na delo za božič.* (Žup.) – *Izgubivši pravo pot v življenju, smo zašli v nesrečo. – Bili smo sanjači, hrepeneči v neskončnost in v solnce zato, ker je bilo tedaj naše življenje tako tesno in temno.*

g) Loči **vrinjene, vmesne** ali **skrajšane stavke** v celotnem stavku, n. pr. *Jurko pa, ne bodi len, je pobral kost golenico in jim je mlatil rebra, da je kar treskalo.* (Milč.) – *Res prete, da bi tako ne, državljanom še večji davki.* (Levst.) – *Vsak človek ima trenotja, ko ga, hoti ali ne hoti, obidejo misli, prijetne, neprijetne, to ni v njegovi moči.* (Strit.)

h) Vejica stoji tudi v skrajšanih stavkih, n. pr. *Danes človek, jutri črna zemla. – Bog,*[28] *da si prišel! Še Bog, da je toliko! Pastir je v vse dovolil, še Bog, da je službo dobil.* (Valjavec.)

Opomnja. Vejica se ne piše pri sestavljenih prislovih, n. pr. *Bežale so Bog vedi kam in Bog vedi pred kom.* (Cankar.) – *Naločil sem nekaj polen v peč, kjer se je kuhalo kaj vem kaj.* (Žup.)

§ 115. Neslovenski je rabiti vejico pri prislovnih in povednih določilih, n. pr. *Mislil je na kovača, ki je bil obležal ob cesti s preklano glavo; bled je bil njegov obraz, od čela mu je kapala kri, v tankem curku, čez oko.* (Cankar; prav: *kri mi je kapala od čela čez oko v tankem curku.*) – *Zaspana vasica si je méla oči; ponekod so se budili petelini ter so prepevali, visoko na gredeh, z iztegnjenimi vratovi.* (Vl. Levstik; prav: *ter so prepevali visoko na gredeh z*

[28] Skrajšano za "Bog bodi hvaljen" ali kaj takega.

iztegnjenimi vratovi.) – Dalje: *Mati je pričakovala, da jej pade k nogam, skesan in spokorjen.* (Kraigher; prav: *da jej pade k nogam skesan in spokorjen*.) – *Bežala je z drobnimi koraki, vsa zasopla.* (Cankar; prav: *Bežala je z drobnimi koraki vsa zasopla.*)

9.2.4 Slovenski pravopis (Anton Breznik, Fran Ramovš, 1935)

Vejico (,) pišemo:
a) Za **zvalniki, medmeti** in **vzkliki**, n. pr. *Tako si šla, mladost, tako si mi pobegnila, sreča, brez blagoslova!* (Cankar) – *Gorje, kdor nima doma!* (Jenko) – *Ostrigli, oh, ostrigli so mi peroti!* (Gregorčič)
b) Loči **pristavek** od besede, ki jo pojasnjuje, če stoji za njo, n. pr. *Primož Trubar, prvi slovenski pisatelj, je umrl l. 1586.* – *Od mene, tvojega prijatelja, ne smeš pričakovati slabih nasvetov.*
c) Loči **nedoločnik**, kadar nadomešča odvisni stavek z *da* (*da bi, ne da bi, namesto da bi* ...), n. pr. *Bog ima pravico, grešnike kaznovati.* – *Tako ga bom premikastil, da se mu nikdar več ne bodo vrnile hudobne želje, po Dunaju razsajati.* (Levstik)
Opomba. Če je nedoločnik osebek v stavku, se **ne** piše vejica, n. pr. *Po toči zvoniti je prepozno.* – *Bob ob steno metati je nehvaležno delo.* – *Električne žice se dotikati je smrtno nevarno.*
Enako se **ne** piše vejica, če stoji nedoločnik kot dodatek *a)* za glagoli nepopolnega pomena ali *b)* kot dodatek pridevnikov, ki merijo na kako dejanje. *a) Z enakim ti vračati mi brani dober okus.* (Detela) – *Meni se je iz srca hotelo le-to jagnje z vami jesti.* (Trubar); *b) Prijatelji so bili pripravljeni zanj vse žrtvovati.* – *Ne bodi len bolnik obiskati.*
č) Loči take stavčne člene, ki se dado zvezati z veznikom "**in**", n. pr. *Iglica gre skozi oplen, blazino, soro in os.* – *Ne bojim se ne meča, ne sulice, ne drugega velikanovega orožja.* (Levstik) – *Voda gine, pada, sahne.*
d) Pred vezniki *ne le – ampak tudi, tudi, a, ali, ampak, toda, vrh teg, potem, nato, zdaj – zdaj, nekoliko – nekoliko, kajti, ker, ko, kadar, kjer, kamor, če, četudi, čeprav, temveč, marveč, da, dasi, pa, vendar* itd., n. pr. *Bil je tak netek, da so ljudje zmerom jedli, pa vendar niso bili nikdar siti.* (Levstik.) – *Zgodovina nam kaže ne le lepa in slavna dela, temveč tudi strasti in hudobije.* – *Žetev je velika, ali delavcev je malo.*
Opomba. Če si sledita dva veznika (ali členek in veznik), se piše vejica pred drugim veznikom če se z njim začenja nov stavek. N. pr. *V svoji neizkušeni poštenosti si je mislila da, če se sama ne briga po nepotrebnem za druge, jo bodo pustili tudi drugi pri miru.* (Detela.) – *Skusili ste, da, kdor se iz prvega prenagli, se kmalu upeha.* (Vodnik).
e) Loči v **priredjih** in **podredjih** posamezne stavke med seboj, n. pr. *Tvoja roka, moj otec, zemljé mi lepoto je odpirala, ti kazal si njiv mi plodove, sprovajal me v les, razkladal skrivnosti glasov in družine dreves in vedela sva, kje kosi mladijo, kje drozgi, kod divji mož hodi, kje vile se skrivajo v gozdi,* (Župančič.)
Pred *in, pa, ter* stoji vejica le, ako jo zahteva spred stoječi stavek, n. pr. *Sava nastane iz dveh rek: iz Save Dolinke, ki izvira blizu Ráteč,* ***in*** *Save Bohinjke, ki prihaja iz Savice.* – *Velik razloček je med zakoni, ki vladajo naše svobodne sile,* ***ter*** *zakoni, ki vladajo naše telesne sile.*
f) Loči **vrinjene, vmesne** ali **skrajšane stavke** v celotnem stavku, n. pr. *Jurko pa, ne bodi len, je pobral kost golenico in jim je mlatil rebra, da je kar treskalo.* (Milčinski.) – *Ciganka je zbolela sredi noči; cigan, kaj je hotel, je šel okrog sosedov prosit pomoči.* (Isti.) – *Res prete, da bi tako ne, državljanom še večji davki.* (Levstik.) – *Vsak človek ima trenotja, ko ga, hoti ali ne hoti, obidejo misli, prijetne, neprijetne, to ni v njegovi moči.* (Stritar.) – *Danes človek, jutri črna zemlja.* – *Še Bog, da je toliko!*

Opomba. Vejica se ne piše pri sestavljenih prislovih, n. pr. *Prestrašil se je sam Bog vedi zakaj. – Bežale so Bog vedi kam in Bog vedi pred kom.* (Cankar.) – *Pojasnil si hodijo iskati sam ne vem kam vsepovsod, edino k nam ne.* (Župančič.) – *Naložil sem nekaj polen v peč, kjer se je kuhalo kaj vem kaj. – Jaz nisem kdo ve koliko prispeval.* (Isti.)

9.3 Vejica v srednješolskih učbenikih

Zbrali so bili tudi primeri, kako je bila raba vejica opisana v srednješolskih učbenikih.

9.3.1 Slovenščina 3: Z besedo do besede

Učbenik za slovenščino – jezik v 3. letniku gimnazij in srednjih strokovnih šol
Mladinska knjiga, 2009

(stran 33)
Pastavek je samostojna poved ali pa je del daljše povedi – v tem primeru je od preostalega dela povedi ločen z vejico.

Pristavek je vedno del povedi. Od preostalega dela povedi je ločen z vejico. Kadar je pristavek sredi povedi, stoji vejica pred njim in za njim.

Polstavek je vedno del povedi. Od preostalega dela povedi je ločen z vejico. Kadar je polstavek sredi povedi, stoji vejica pred njim in za njim.

(stran 78)
Med odvisnim in glavnim stavkom vedno stoji vejica.

(stran 82)
(Prilastkov odvisnik) Z glavnim stavkom ga najpogosteje povezuje vezniška beseda ki/kateri, lahko pa tudi drugi oziralni zaimki (npr. *kjer*).

Če prilastkov odvisnik stoji sredi glavnega stavka, je od njega ločen z vejico na obeh straneh.

(stran 116)
Vejica v priredju
Osnovni in dopolnjevalni stavek v priredju večinoma ločuje vejica.
Le pred vezniki **in, pa, ter, ali – ali, ne – ne, niti – niti** VEJICE NE PIŠEMO.

(stran 117)
Vejica v priredno zloženi povedi z vrinjenim odvisnikom

Pred vezniki **in, pa, ter, ali – ali, ne – ne, niti – niti** vejice načeloma ne pišemo.
V nekaterih primerih VEJICO PIŠEMO tudi pred temi vezniki, in sicer takrat, kadar je **med prvim in drugim prirednim stavkom še odvisnik**.
Npr.:
V Ziljski dolini se je ohranila protestantska enklava, ki je preživela obdobje protireformacije[29] *in v eni izmed kmečkih družin so ohranili tudi Trubarjeva dela.*
Zbirke niso pridobili z enkratnim nakupom na londonskih avkcijah, ki so znane po svetu, niti je niso dobili kot dar v okviru katere od zasebnih bibliotek.

[29] Na tem mestu v učbeniku manjka vejica!

Pogosto so uporabljali rečne poti, ki so bile tisti čas bolj običajne kot danes, ali pa več dni več dni[30] *potovali z dragocenim tovorom po slabih cestah.*
Že pri podredno zloženih povedih smo omenili, da na meji neenakovrednih stavkov vedno stoji vejica. V zgornjih primerih sta enakovredna prvi in tretji stavek, odvisnik med njima pa je prvemu stavku podrejen. Njegovo neenakovrednost nakažemo tako, da **pred odvisnikom in za njim** postavimo vejico. Takšno zgradbo povedi lahko tudi grafično ponazorimo:
Poševnica nam pove, kje stoji vejica.

(stran 187–188)
Vejica je najpogostejše ločilo, ki stoji znotraj povedi. Z vejico členimo poved na posamezne enote, ki jih v govoru ločimo s premori in z njo zaznamujemo nekončno intonacijo.
Vejica ločuje:
a) enakovredne dele proste ali zložene povedi (npr. *V študijskem letu 2007/2008 so delovali slovenistični lektorati v Argentini, Avstriji, Belgiji, Bolgariji, Češki, Franciji, Hrvaški, Italiji, Japonski, Litvi, Madžarski, Makedoniji, Nemčiji, Nizozemski, Poljski, Portugalski, Romuniji, Rusiji, Slovaški, Srbiji./Najbolj razširjen tuji jezik v državah EU je angleščina, govori jo 51 % prebivalcev držav članic EU (38 % kot tuji jezik), sledita francoščina in nemščina.*);
b) nadredni stavek od podrednega (npr. *Vsaka država članica ob pristopu k tej povezavi določi, kateri jezik želi razglasiti za svoj uradni jezik v EU.*);
c) polstavek od preostalega besedila;
č) pristavčne dele (npr. *Lojze Kovačič. slovenski pisatelj, je razmišljal o pomenu jezika.*).

9.3.2 Slovenščina 4: Z besedo do besede

Učbenik za slovenščino – jezik v 4. letniku gimnazij in srednjih strokovnih šol
dr. Jerca Vogel, mag. Silva Kastelic, mag. Marjana Hodak
Mladinska knjiga, 2010

Vejica je najpogostejše ločilo, ki stoji sredi povedi. Z vejico členimo poved na posamezne enote, ki jih v govoru ločimo s premori, in z njo zaznamujemo nekončno intonacijo. Vejica ločuje:
a) enakovredne dele enostavčne povedi (npr. *Za splošno koristne namene se štejejo humanitarni nameni, nameni varstva pred naravnimi in drugimi nesrečami, invalidski, dobrodelni, ekološki, kulturni, športni, religiozni in drugi nameni.*);
b) enakovredne dele zložene povedi (npr. *Davčni zavezanec lahko posameznemu upravičencu nameni 0,1 %, 0,2 %, 0,3 %, 0,4 % ali 0,5 %, seštevek vseh pa ne more presegati 0,5 % dohodnine.*);
c) nadredni stavek od odvisnega (npr. *Zahteva velja do trenutka, ko davčni organ prejeme (SIC!) novo zahtevo ali preklic zahteve.*);
č) polstavek od preostalega besedila (npr. *V skladu s 142. členom Zakona o dohodnini – ZDoh-2 (Uradni list RS, št. 117/06, 10/08, 78/08) lahko davčni zavezanec rezident zahteva, da se do 0,5 % dohodnine, odmerjene po tem zakonu od dohodkov, ki se vštevajo v letno davčno osnovo, nameni za financiranje splošno-koristnih namenov in za financiranje političnih strank in reprezentativnih sindikatov.*);
d) pristavčne dele;
e) pastavek od preostalega dela besedila.

9.3.3 Na pragu besedila 3

Učbenik za slovenskih jezik v 3. letniku gimnazij, strokovnih in tehniških šol
Martina Križaj Ortar, Marja Bešter Turk, Marija Končina, Mojca Bavdek, Mojca Poznanovič Rokus, 2006

Opazujte rabo vejice v naslednjih povedih in odgovorite na vprašanja.

[30] V učbeniku se *več dni* ponovi dvakrat.

1. Počakaj, Tone, še na otroka!
2. Usoda nam lahko vzame bogastvo, duha pa ne more.
3. Čeprav me je prizadel s svojimi besedami, je ostal moj prijatelj.
4. Sosed je po nesreči povozil srno, težko skoraj trideset kilogramov.
5. Kuj železo, dokler je vroče.
6. Literarna ustvarjalnost Josipa Stritarja zajema pesmi, povesti, romane in esejistične spise.
7. Nekatere sosednje države, npr. Hrvaška in Italija, so za Slovence turistično zelo zanimive.
8. Če ne bo deževalo, bomo naredili, kar smo vam obljubili.
9. Sosedova Anja, ta je prava lisica, je prodajalca zelo osmešila.
10. Človek sem, nič človeškega mi ni tuje!
11. No, tako se pa nisva zmenila!
12. Povedala je, kar je želela, in takoj zapustila sestanek.
13. Koncert bo v petek, 15. oktobra, ob 20. uri v Gallusovi dvorani Cankarjevega doma.
14. Profesorica naj se opozorila, da bo ura že tri, ko bomo končali pouk.
15. Soteskanje vsebuje prvine alpinizma, jamarstva in kajakaštva.
16. Tisti, ki smo šli smučat, smo morali biti pred šolo že ob pol sedmih.
17. V knjigi, ležeči na mizi, je Peter našel izgubljeni listek.
18. Ne, to sploh ni res.
19. Bila sem zadovoljna, ker sem do večera prebrala knjigo in tudi napisala obnovo.
20. Bilo je ledeno, zato smo morali voziti zelo previdno.

Kaj ločuje vejica v 6. in 15. povedi?
V katerih povedih vejica ločuje
- enakovredne stavke,
- neenakovredne stavke,
- pastavek?
Kaj pa ločuje v 4., 7., 9., 13. in 17. povedi?

Vejica je nekončno ločilo, s katerim ločimo razne smiselne dele povedi, in sicer:
- naštevalne enote v enostavčni povedi,
- stavke v večstavčni povedi,
- zvalniški, členkovni in medmetni pastavek (če ga postavimo sredi povedi, ga obdamo z dvema vejicama),
- dodano polstavčno pojasnilo oz. opis in vrinjeno sporočevalčevo mnenje (če ju vrinemo sredi povedi, ju od drugih delov povedi ločimo z dvema vejicama, tj. na njunem začetku in na koncu).

V posebnih primerih pa vejice v navedenih položajih ne pišemo – oglejmo si nekaj posebnosti:
1. **Vejice ne pišemo <u>pred</u>**
- ponovljeno besedo, če v govoru ni ločena s premorom (npr. *Hitro <u>hitro</u> na pot!* 'zelo hitro'),
- vezalnimi vezniki *in, pa, ter* (npr. *Z rokami se je oprl na stol in se počasi dvignil; Na morju smo plavali, hodili na dolge izlete ter se veliko pogovarjali*),
- ločnim veznikom *ali, bodisi* (npr. *Tega ne veste ali nočete vedeti?*),
- drugim delom dvodelnih veznikov *ne – ne, niti – niti* (npr. *Niso nam pomagali ne z besedami <u>ne</u> z dejanji; Dobrih lastnosti ne more sreča nikomur niti dati <u>niti</u> vzeti*),
- primerjalnim veznikom *kakor, kot, ko*, če veznik ne uvaja odvisnika (če torej v nadaljevanju ni osebne glagolske oblike, prim. *Požrešnost je pokončala več ljudi kakor meč*).
2. **Vejice ne pišemo <u>med</u>**
- prirednimi deli stalnih besednih zvez z izpuščenim veznikom (npr. *<u>Na vrat na nos</u> se je odločil za nakup novega pohištva*),
- dvema podrednima veznikoma (npr. *Opozorite ga, <u>da če</u> ne pride, ne bo dobil svojega deleža*),
- prirednim in podrednim veznikom na začetku vmesnega odvisnika (npr. *Peter je že začel pisati nalogo, <u>in ker</u> ni znal, je kaj naprej pogledoval k sošolcu*),
- deli večbesednih veznikov (npr. *Kljub temu da ste veliko delali, morate snov pred izpitom dobro utrditi*).

9.3.4 Govorica jezika 2 in 3

Slovenščina za 2. in 3. letnik gimnazij
Klara Zajc Berzelak, Alenka Arbiter, dr. Nataša Hribar
Modrijan, 2008

Pravi, da sem črn in da je on bel.
Določite glavni in odvisni stavek ter utemeljite rabo vejice.

Na ekvatorju je krog polnih barv, na severu je bela točka, južni pol pa je črn.
Ko je z mešanjem treh barvnih svetlob odkrival barvno fotografijo, je imel komaj 30 let.
Negative so spremenili v diapozitive ter jih osvetlili z modro, zeleno in rdečo svetlobo.
Poskusi, kako urediti barve v sisteme, segajo že v 17. stoletje.
Delimo jih na čepke, ki zaznavajo barve, in paličke, ki se odzivajo na svetlobo katere koli valovne dolžine.
Britanski urad za barve je ugotovil, da zaznavamo 1400 modrih, 1375 rjavih, 1000 rdečih, 820 zelenih, 555 oranžnih, 360 vijoličnih, 50 sivih im 12 belih tonov.
James Clerk Maxwell je o svojem poskusu in "teoriji primarnih barv" spregovoril na srečanju Kraljevega društva v petek, 17. maja 1861, in tako vzpostavil barvni sistem RGB (Red, Green, Blue).

12. Utemeljite rabo vejice, tako da boste zapisali stavčno strukturo povedi.
13. V izhodiščnem besedilu poiščite še nekaj primerov, ki niso zajeti v zgornjih povedih, in utemeljite rabo vejice.

Vejica je nekončno ločilo. Z njo ločujemo stavke (priredne in podredne stavke, postavke, vrinjene stavke, pristavke, dostavke, pastavke), uporabljamo pa jo tudi za označevanje naštevalnih enot, decimalnih števil in še v nekaterih drugih primerih.

Kako pomembno ločilo je vejica, nam povesta naslednja primera.

Hvala lepa, Kaja.
Hvala, lepa Kaja.

Preglednice:

Priredno razmerje	v protivnem priredju (vezniki: *a, ali, pa, ampak, temveč, marveč, samo, le*) *hitro, a površno* *Prišel je na razgovor, toda nikogar ni navdušil.* v pojasnjevalnem priredju (vezniki: *kajti, saj, sicer*) *Morali se boste potruditi, kajti vse je odvisno od reklame.* v posledičnem in sklepalnem priredju (vezniki: *zato, torej, zatorej*) *Dokumentov ni priložil, zato njegove prošnje ne bomo obravnavali.* v stopnjevalnem priredju (vezniki: *ne samo – ampak tudi, ne le – tudi*) *Ni samo naporno, ampak tudi odgovorno.* če sta uporabljena veznika *ne – ne, niti – niti*, vejice ni *Ni ne telefoniral ne pisno sporočil.*
vejice ni	v vezalnem priredju (vezniki: *in, ter, pa*) *prošnja in življenjepis* *Za ugodno rešitev se vam že vnaprej zahvaljujemo in vas lepo pozdravljamo.* v ločnem priredju (vezniki: *ali, ali – ali, ali – ali pa*) *Lahko pošljete po pošti ali pa sporočite po telefonu.*
posebnosti	če se v priredja, v katerih vejice sicer ni, vrine stavek, ga na obeh straneh ločimo z vejico *Lahko bi odpisal, kar bi bilo najboljše, ali pa vsaj poklical.* če se vezniki *ne, niti, ali* večkrat ponovijo, gre za naštevanje – potem vejico lahko uporabljamo *Ni ne poslušal, ne odgovarjal in ne obžaloval.*
Podredno razmerje	vejica loči glavni stavek od odvisnika; najpogostejše vezniške besede so: *ker, da, ki, če, kar, kdor, kjer, kam, čeprav ...* *Poglej, kaj je narobe.* *Prošnje nismo upoštevali, ker ste zamudili rok.*

vejice ni	med dvema veznikoma *Pregledali smo vašo prošnjo, a ker niste priložili potrdila o šolanju, je žal ne moremo upoštevati.* med deli večbesednega veznika *Namesto da bi preveril podatke, se je takoj odločil.* pred vezniki *ko, kakor, kot,* kadar nastopajo v besedni zvezi *Ravnajo kot neodgovorni otroci.*
Drugi primeri	za označevanja pristavka *Obrnite se na gospoda Vojka Logarja, našega zastopnika, ki Vam bo lahko priskrbel vse potrebno.* *Rok Fras, rojen 16. 2. 1986, /.../* za označevanje vrinjenega stavka *Vstopil je, v rokah je še vedno držal tisti papir, in sedel k mizi.* (v takem primeru lahko uporabimo tudi oklepaj, pomišljaj, tri pike) za označevanje polstavka *Odgovor najdete v sklepu št. 4, sprejetem na prejšnjem sestanku.* za označevanja dostavka *Sestanke bo v stavbi upravne enote, soba št. 6.* za označevanje pastavka (na primer členki, medmeti) *Ne, ne bo ga.* za ločevanje dobesednega navedka v premem govoru (namesto pike) *"Ja, seveda bom tu," je zaključil.* neobvezno v abecednih seznamih, kjer je priimek na prvem mestu *Horvat, Tanja* med celim številom in decimalkami *6,35 %* namesto presledka ali piki za označevanje milijonic *2,246.800*

9.3.5 Besede 1

Učbenik za slovenski jezik v 1. letniku triletnih poklicnih šol
Mateja Gomboc
DZS, 2005

Vejica ima v povedi pomembno mesto, saj je **nekončno ločilo** in nam pomaga razčlenjevati poved na posamezne dele: jih našteva, nadomešča priredne veznike in, ter, pa, ločuje več stavkov v povedi.
V govoru se kaže kot premor in z rahlim dvigom glasu.

Vejico pišemo:
• **med enakovrednimi stavčnimi deli** – besedami in besednimi zvezami:
Za pripravo novoletnega plesa so odgovorni naš Tine, njegova prijateljica Maja, Mateja iz sosednjega razreda in Andraž.

Pomni
Če naštevamo večje enote delov povedi in jih pišemo v posameznih vrsticah, na koncu vsakega dela napišemo vejico, za zadnjim pa piko.
Postregli bomo:

- *narezek z domačo salamo in sirom ter skutni namaz s hrenom,*
- *govejo ali gobovo juho,*
- *svinjsko pečenko, nadevano z jajci in s panceto,*
- *pečeno kislo zelje s krompirjem in smetano,*
- *solato z vrta,*
- *sadni kruh.*

Pravilen je tudi zapis brez vejic.

• **v večstavčni povedi** – stavki v povedi so med seboj lahko v enakovrednem razmerju (*Srečala sva se po dolgem času in povabil sem ga na sok.*) ali v neenakovrednem (*Srečal sem prijatelja, ki ga že dolgo nisem videl.*).

Vejica vedno stoji:
• pred podrednimi vezniki – da, ko, ker, če, kjer, kadar, ki, čeprav ...
Ne bi mogla reči, da imaš prav.
Pozvonilo je, ko sem vstopil v razred.
Tega ne vem, ker me ni bilo pri pouku.
Povej, če veš.

Pomni
- Tudi če stavka obrnemo in postavimo veznik na začetek povedi, med njima ostane vejica: *Ko sem vstopil v razred, je pozvonilo.*
- Če sta v povedi dva veznika skupaj, zapišemo vejico pred prvim: *Samo veš, da če se ne boš sproti učila, si ne boš mogla vsega zapomniti.* Vendar se je priporočljivo izogniti takemu kopičenju veznikov: *Sama veš, da si ne boš mogla vsega zapomniti, če se ne boš sproti učila.*
- Vejice ne pišemo med deli večbesednega veznika: *Šli bomo na ekskurzijo, kljub temu da je napovedano močno deževje. Že danes bom napisal vso nalogo, tako da bom potem imel več časa zate.*

• pred prirednimi vezniki – toda, ampak, saj, zato, temveč, vendar ...

Tekla sem, toda bila sem prepozna.
Pripravila sem si plašč, saj postaja hladno.
Mračilo se je že, vendar jo je kar čakal in čakal.

• med vrinjenim stavkom, pristavkom in preostalo povedjo:

Pred leti, takrat še ni bilo interneta, smo si pisali pisma dvakrat na teden,
Ker bodo letošnji novoletni prazniki, ki se jih še posebej veselimo, dolgi, lahko gremo na daljše počitnice.

Pomni
- Vejice ne zapišemo pred vezniki in, ter, pa, razen če pred njimi ne stoji vrinjen stavek: *Katarina se igra s kockami in si zraven poje. Katarina se igra s kockami, ki jih je dobila za rojstni dan, in si zraven poje.*
- Prav tako vejice ne pišemo pred dvodelnimi vezniki ne – ne, niti – niti, bodisi – bodisi, ali – ali: *Niti hrane se ne dotakne niti ne spi.*

• med pastavkom, polstavkom in preostalo povedjo: *Dober dan, ste že doma? Ti povej, Stana. Da, seveda bom. Stala je tam, vsa nesrečna in sama. Misleč, da te ni doma, te nisem niti poklicala. Sestanek bo v sredo, 17. januarja, ob 18. uri, v prostorih društva.*

• med izpostavljenim stavčnim členom in preostalo povedjo: *Tone Pavček, naš pesnik, je lani praznoval visok jubilej.*

• med navedbo kraja in datuma: Ankaran, *14. junija 2000.*

Razmislite
- V srednjem veku je sel vojskovodji na bojišče prinesel kraljev ukaz: *Vse ujetnike pomilostiti ne ubiti.* Kaj je po vašem mnenju storil vojskovodja?

Pomni
Vejico pišemo tudi med številkami, in sicer brez presledka:
- med celimi in decimalnimi števili: 10,25

- včasih tudi za zaznamovanje milijonic: 10,234.000 SIT

9.3.6 Slovenska jezikovna vadnica

za poklicne šole
Jože Kocbek
Založba Obzorja Maribor, 1966

Vejico rabimo za razčlenjevanje stavka po sestavinah na pomenske in oblikovane enote. Opozarja na miselno povezavo besed in s tem na stavčno melodijo. Olajšuje na branje. Nakazuje kratek odmor, pred njim pa zvišanje glasu.
V slovenščini stavimo vejico nekaj po pomenu, nekaj pa po zgradbi stavka. Drugačno vlogo ima v nezloženem, drugačno v zloženem stavku.
Loči:
– besede in posamezne stavčne člene,
– posamezne stavke,
– vrinjene in okrajšane stavke od celotnega stavka.

A) Vejica v nezloženih stavkih
1. Loči od celotnega stavka samostojne izraze, ki niso pravi stavčni členi:
a) Zvalnike, poudarjene medmete, samostojne miselne prilove (glej o miselnih prislovih na str. 93) in vprašalnice (glej o vprašalnicah na str. 93), npr.: *Tone*, odpri radio! *Da*, to se bo zgodilo! *Joj*, kako je hudo! *Ne*, tega ne smeš storiti! *Pač*, še ena rešitev je. *Nikar*, ne bo prav! *Seveda*, tega ne tajim. Čez mesec dni se vrnem, *prav gotovo*. *Kaj*, tudi ti? *Joj*, kam bi del! *No*, pa naj bo! *O*, saj vemo, kako je!
Pomni
Če ti izrazi niso miselno samostojni, ne pišemo vejice, npr.: Nikakor si tega ne dam vzeti. Prav gotovo ti čez mesec dni vrnem. Seveda ne tajim tega.

b) Izraze, ki izrečeno misel povzemajo, popravljajo, dopolnjujejo ali pojasnjujejo, npr.: *Skratka*, uspeh je bil viden. *Sicer pa*, kaj bi vam še govoril. *Kratko in malo*, do konca tedna mora biti delo opravljeno. Knjižnica je odprta vsak dan, *in sicer* dopoldne in popoldne. Letalo pristaja dvakrat na teden, *namreč* v torek in petek. Prešeren je uvedel nove pesniške oblike, kakor sonet, gazele, gloso.
2. Ločuje pristavek (glej o pristavku na str. 118) od besede, ki jo pojasnjuje, *če stoji za njo*, npr.: Ivan Cankar, *največji mojster slovenske besede*, se je rodil na Vrhniki. V nedeljo, *(dne) 9. maja*, bo obletnica osvoboditve.
3. Ločuje istovrstne stavčne člene:
a) Kadar naštevamo *stvari*, *lastnosti* ali *določila* brez vezalnega veznika: Polje, vinograd, gora morje, ruda, kupčija tebe rede. Prešernov oče Šimen je bil dobrosrčen, šaljiv in kreposten mož. Oglašali so se ptiči tukaj, tamkaj, vsepovsod.
b) Kadar z veznikom izražamo *nasprotno*, *dopustno* ali *kako drugo razmerje*, npr.: Lepa, *vendar* nevarna pot. Fant je zanesljiv delavec, *čeprav* malo počasen. (Glej o protivnem priredju na str. 126!)
Tudi ločimo z vejico: Jutri, pravzaprav danes, odpotujem. Meščane, posebno (zlasti, predvsem) izobražence, so močno nadzirali.
c) Kadar ponavljamo iste besede ali cele izraze, npr.: Stoji, stoji tam beli grad.
Vendar ne pišemo vejice, kadar s ponovitvijo iste besede le-to samo stopnjujemo, npr. Tako hudo hudo me boli. Zdaj zdaj bo tu!
Pomni
Vejice na stavljamo:
1. Kadar so istovrstni stavčni členi zvezani z vezalnimi vezniki *in*, *pa*, *ter*, *ali*, *ne – ne*, *niti – niti*, *ali – ali*, npr.: Oče in mati sta mu že davno umrla. Ivan in Pavla sta brat pa sestra. Stopi ven ali noter! Ni ga ne v šoli ne doma. (Glej o vezalnem priredju na str. 125!)
2. Prav tako ne pišemo vejice pred primerjalnima veznikoma *kakor* in *kot* (*ko*), če ne vežeta popolnih stavkov: Skrbel je zanj kakor pravi oče. Sinoči je pela ko slavček ljubo. Zdaj vem toliko kakor prej.

č) Kadar zaradi poudarka izpostavimo stavčni člen in ga v stavku ponovimo z zaimkom ali prislovom: Sin, *ta* ti bo šele zagodel. V mojih mladih letih, *takrat* je bilo vse drugače.
d) V datumu, kadar dokumentarno ugotavljamo kraj in čas, npr.: V Ljubljani, (dne) 17. julija 1965 ali: Ljubljana, 17. julija 1965. Datum sam lahko pišemo: 17. julija 1965 = 17. 7. 1965 = 17. VII. 1965 = 17/7-1965.
Vendar ne pišemo vejice v stavčni zvezi: Fran Erjavec se je rodil v Ljubljani 9. 9. 1834.

B) Vejica v zloženih stavkih

1. V priredju ločimo:

a) stavke brez veznika, nor.: Na nebu zvezde sevajo, na vasi fantje pevajo, pojo glasno, pojo lepo.

b) stavke protivnega, sklepalnega in vzročnega priredja z vezniki: Vse ima, *vendar* ni zadovoljen. Iz koče sem slišal ječanje, *pa* sem stopil noter. Sami ste videli, *torej* poznate zadevo. (Glej o priredju na str. 125!)

Pomni

Če so v vezalnem priredju stavki vezani z vezniki *in, pa, ter, ne – ne, niti – niti*, ne pišemo vejice.

Vendar "in" ni zmeraj vezalni veznik, zlasti v stavkih z različnimi osebki ima ponavadi protivni, vzročni ali sklepalni (posledični) pomen. Zato v takem primeru lahko pišemo vejico pred *in*, npr.: Mimo si šel, *in* se nisi oglasil. Toplo nam je bilo, *in* sneg se je tajal na strehah. Potem je zazvonilo, *in* iz šole so se usule gruče.

Kadar veznik "in" veže stavčni člen in istovrstni odvisnik, med njima ne pišemo vejice, npr.: Pozimi in kadar je mraz, kurimo. Po pobočjih in kjer se sonce upre, sneg hitro skopni.

2. V podredju ločimo:

a) odvisnik od glavnega stavka: Kdor za smolo primer, se osmoli. Če je odvisnik v sredi nadrednega stavka, ga ločimo na obeh straneh z vejico, npr.: Novica, da bomo kmalu rešeni, nam je vlila novih moči;

b) z vejico ločimo odvisnike iste stopnje, npr.: Venomer je tožil, da ga rana sklei, da je lačen, da je žejen.

Ne stavljamo pa vejice, kadar take stavke vežejo vezniki *in, pa, ter*, npr.: To priča, da imate tenek posluh *in da* čutite, kaj pomeni igralcu jezik;

c) z vejico ločimo odvisnike različnih stopenj med seboj, npr.: Ta hiša je preprost dom, iz kakršnega so izšli po vrsti odlični možje, ki se jim moramo zahvaliti, da smo izobražen narod;

č) kadar v podredju zadeneta skupaj veznik nadrednega in podrednega stavka, postavljamo vejico pred oba veznika, npr.: Žive naj vsi narodi, ki hrepene dočakat dan, *da, koder* sonce hodi, prepir iz sveta bo pregnan.

Pomni

1. Da se izognemo dveh zaporednih veznikov, stavke razporedimo raje enega za drugim. Ne bomo rekli: Učitelj je rekel, da, če bomo pridni, ne bomo šli na izlet. Skusili ste, da, kdor se iz prva prenagli, se kmalu upeha. Kar to knjigo vzemi in, kadar jo prebereš, jo vrni! – Pač pa bomo raje rekli: Učitelj je rekel, da ne bomo šli na izlet, če ne bomo pridni. Skusili ste, da se kmalu upeha, kdor se iz prva prenagli. Kar to knjigo vzemi in jo vrni, kadar jo prebereš! (Glej o podredju na str. 129!)

2. V nekaterih zvezah, zlasti na začetku stavka, so *sestavljeni vezniki* dobili pomen enotnih izrazov, zato jih pišemo brez vejice, npr.: *Še prej ko* (preden) se je zdanilo, smo krenili na pot. Jezil se je, *češ da* si mu storil krivico.

Taki sestavljeni vezniki so še: in ko, potem ko, medtem ko, še preden, kakor da, s tem da, kljub temu da, razen če, ne da bi, samo da, namesto da itd.

Kadar pa taki sestavljeni vezniki nimajo enotnega pomena, jih ločimo z vejico, npr.: Prišel je *prej, kot* sem ga pričakoval. *Prej, ko* je bil še mlad, je vse drugače govoril. Govoril je *tako, da* so ga vsi radi poslušali.

C) Vejica v vrinjenem stavku

Kadar v stavek vrinemo drug stavek tako, da se oblika prvega stavka ne spremeni, dobimo vrinjeni stavek. Tak vrinjeni stavek ločimo o drugega stavka z vejico, npr.: Pesnik je imel*, kakor smo slišali,* v mladosti čudno usodo. Po zunanjosti sem bil podoben*, tako mi je mati često dejala,* očetu.

Pomni

Glagolske izraze: kdo ve kako, bog si ga vedi ipd. občutimo kot prislove in ne pišemo vejice. Primer: Janez ni *kdo ve kako* (= posebno) nadarjen. *Bog si ga vedi* kod se potepa po svetu.

Če imajo glagoli v takih izrazih svoj samostojni pomen, jih seveda ločimo z vejico: *Kdo ve, kod* hodi naš Janez?

Č) Vejica v okrajšanem stavku

1. Z vejico ločimo *deležnik*, kadar ima poleg sebe prislovno ali predmetno dopolnilo in nadomešča odvisnik, npr.: Čakal je pred skalo, *boječ se zalezovalcev*. Hiša, *krita s slamo*, je še bolj izpostavljena nevarnosti ognja.

Vendar ne ločimo deležnikov z vejico, če so le-ti prilastki ali prosti prislovi, npr.: Čakal je pred skalo *boječ*. Mikale so ga *vabeče* maslenke. *Opotekaje se* je hodil po cesti.

2. Z vejico ločimo tudi *okrajšana priredja* in *podredja*: Mi smo včeraj bili na predavanju, vi pa ne. Čeprav v letih, še zmerom rad dela. Kolikor bo dal on, toliko jaz.

Z vejic *ne ločimo* nedoločnika, pa naj nadomešča predmetni ali prilastkov odvisnik, npr.: Imel je resno voljo *popraviti* škodo (= da bi popravil škodo). Smrtno nevarno je *dotikati se* električnih žic. Tonček je bil takoj pripravljen *oditi* po zdravnika. (Glej str. 145!)

9.3.7 Slovenski knjižni jezik 1

Jože Toporišič
Založba Obzorja Maribor, 1975

Vejica je najnavadnejše in najpogostejše ločilo v prostem ali zloženem stavku. Pišemo jo:

a) med istovrstnimi (priredno povezanimi ali prirejenimi) stavčnimi členi ali njihovimi deli in na meji posameznih stavkov priredja in podredja: *Polje, vinograd, gora morje, ruda, kupčija tebe rede. (V) – Oči imajo, pa ne vidijo. – Saj ni le neumen, ampak tudi hudoben. – Kam si rekel, da si dal čistit svoj plašč? – Čeprav je že dovolj star, je Peter še vedno kot otrok. – Čeprav dovolj star, je Peter še vedno kot otrok. – Pastir je hitel v stan in se s skrbjo oziral v težke oblake. – Pastir je hitel v stan, ozirajoč se s skrbjo v težke oblake.*
Kako je z vejico pred *in, pa, ter, ali, ali – ali, ne – ne, tako – kakor* glej niže.
b) ob delih sporočila, ki niso stavčni členi prostega stavka, temveč se mu samo pridružujejo na začetku ali koncu ali pa v sredini; vsi pojasnjujejo ali vrednotijo njegovo vsebino, jo zanikujejo ali ji pritrjujejo itd. Take dele sporočil imenujemo zvalnike, medmete, trdilnice, nikalnice, vprašalnice, vrivke, pristavke, vrinjene stavke ipd. Od ostalega sporočila so ločeni s premorom (samo zvalniki včasih ne). Prim.: *Fantje, kam jo mahnemo danes? – Joj, kam bi del? (A) – Dà, prav imate. – Nè, to res ni lepo. – Nikar, preveč bo! – Kaj, tudi ti si se me spomnila in me prišla obiskat? – Sicer pa, kaj bi vam pravila, kako me je bilo strah. – Premnog človek, med njimi tudi berač in cigan, je že prenočil pod to streho. – Trubar, začetnik naše književnosti, se je rodil na Rašici. – oho, to pa ne gre kar na ukaz. (Ž) – Največji dramatik vseh časov, Shakespeare, je živel burno življenje.*
c) pri datiranju po navedbi kraja: *Trbovlje, (dne) 11. X. 1926*. To doživljamo kot dva stavka in glas to tudi res nakazuje: dvigne se na koncu prvega člena. Tega dviga ni v primerih kot *Vrnil se je v Trbovlje 11. 10. 1926* ni, torej za *Trbovlje* tudi ne pišemo vejice.
Posebej si je glede stave vejice treba zapomniti naslednje primere:
a) pred vezniki *in, ter* in *pa* pišemo vejico samo tedaj, če pomenijo toliko kot *a, ali, vendar, zato* ..., sicer pa ne: *Gledal me je, in me ni videl.* (Varianta s *ter* je stilsko slaba.) *Tremo je imel, pa ni čudno, če pri višini 1,80 ni uspel.* Če protivni *pa* ne uvaja stavka, temveč stoji sredi njega, predenj seveda ne stavimo vejice: *Oče je garal noč in dan, sin pa ne pozna mere v zapravljanju.* Razume se tudi, da vejico pred temi vezniki pišemo, če je prejšnji del stavka tak, da ga ločimo od okolja z vejico:
Srečal sem strica Martina, ko se je vračal z lova, in Barbko. – France, moj prijatelj je to, in vaš Ludvik sta se šla kopat. – Popoldne je prišel inštalater, dovolj dolgo smo ga čakali, ter nas rešil poplave. Primer za odsotnost vejice pred temi vezniki: *Za praznike sta nas obiskala Tone in Reza. Pisala sta, da prideta s popoldanskim vlakom in naj nikar ne hodimo ponju na postajo. Kljub temu smo ju tam pričakali ter se potem veseli pripeljali k nam domov. To je bilo pozdravljanja, stiskanja rok in poljubljanja. Ata pa mama sta bila vsa iz sebe.*
b) pred *ali* pišemo vejico samo, kadar izraža nasprotje ali uvaja odvisni vprašalni stavek, ne pa tudi, kadar izraža izbiro: *Listal je res po knjigi, ali bral ni. – Vprašali so ga, ali bi bil pripravljen prevzeti službo. – Prinesi mi Cankarja ali Prešerna, oba rad berem. Povej mi lepo po vrsti ali pa me pusti pri miru.* Seveda tudi pred *ali*, ki izraža izbiro, vejico pišemo. če tako zahteva prejšnji del sporočila: *Rekel je, naj prinesem drv, ki jih je pred dnevi nacepil, ali pa še bolje trsak.*
c) pri vezniških parih *ne – ne* in *niti – niti* pišemo pred drugim *ne* ali *niti* vejico samo tedaj, če to zahteva prejšnji del sporočila: *Prinesel ni niti masla, ki sem si ga tako želel, niti sira. – Ta človek ni ne (niti) krop ne (niti) voda. Nisem ne prosil ne odklonil.*
č) pred *kakor* pišemo vejico načeloma samo, če uvaja stavek z osebno glagolsko obliko in če uvaja pristavek: *Na tekmi je bilo več ljudi, kot (kakor) so prireditelji pričakovali. – Daleč so bile Juretove misli, kakor v težkih in hudih sanjah. (C) – Več je pojedel kakor zaslužil.*
d) kadar prideta drug ob drugega nadredni in podredni veznik, pišemo vejico pred obema; če gre za vezniške pare, tj. za sestavljene veznike *in ko, prej ko, brž ko, medtem ko, potem ko, vtem ko, že ko, še ko* in *in da, namesto da, toliko da, kljub temu da, tako da, s tem da* ter *še preden, in če, in ker*, stavimo vejico samo pred prvim – seveda če to dopuščajo druga pravila: *Pregovor pravi, da, kdor ne seje, nima kaj žeti. – Le piši materi, da, če ne pride do sobote, sploh ni treba hoditi. – Rekel je, da pride sam, in če bo le mogoče, pripelje še sestro. – Njegov čas je potekel, in da naravnost povem, očetova volja tudi. – Radio je dober, in ker tudi ni drag, vam ga res priporočam.*
Nič hudega ni, če tudi v primerih, kjer naj pišemo vejico pred oba veznika, le-to postavljamo samo pred prvega: *Pregovor pravi, da kdor ne* ... Tudi ne bi bilo tako hudo, če bi pisali *Pozimi, in kadar je mraz, v sobah kurimo*, čeprav je to mogoče pisati sploh brez vejice. Vejico pišemo takrat, če je *in kadar* ... nekak pristavek, dodatek sporočilu.
e) kadar ponavljamo drugo za drugo isto besedo ali besedno zvezo, pišemo vejico pred drugo; samo med ponovljena izraza, katerih prvi člen lahko zamenjamo s kakšnim prislovom, vejice ne pišemo. Prim.: *Kam, kam gre tvoja pot? – Fantje, fantje, oh, kako je vam tam v dežju in snegu, v mrzli gošči tam? (Ž)* – Vendar: *Zdaj zdaj*

bo tu. Namesto tega lahko rečem *vsak čas bo tu*. Tako še *koj koj, brž brž, zelo zelo, hudo hudo, kmalu kmalu, veliko veliko, beži beži*. Prim. *Hudo hudo se motite = zelo hudo se motite.* Celo med ponovljena povedka v takih primerih ne stavljamo vejice: *Gazimo gazimo belo mečavo, kmalu snega bo do uhljev in čez (Bm)*, saj *gazimo gazimo* pomeni toliko kot *kar naprej gazimo*. – Če med prislovnimi dvojicami, za katere smo rekli, da njunih členov ne ločimo z vejico, v govoru delamo premor in imajo sploh samostojen intonacijski potek, stavimo vejico tudi mednju: *Koj, koj bo tu! Hudo, hudo me boli!* Take primere lahko razvežemo takole: *Hudo, zelo hudo me boli!* Tudi v tem primeru gre torej za stopnjevano čustvo govorečega. V književnosti res pogosto najdemo take razvezave: *Beli so, beli češnje cvetovi, temni, pretemni so talcev grobovi. (Kk) Lep, prelep je šopek ta. (M)*

9.3.8 Slovenski jezik

Učbenik za poklicne, tehniške in druge šole za gospodarstvo ter družbene službe
Vera Remic-Jager
Založba Obzorja Maribor, 1980

Vejica je zelo občutljivo ločilo, ki ne loči le pomenskih enot, temveč zaznamuje tudi besedni naglas, nakazuje melodijo in ritem stavka, dostikrat pa tudi čustveno razgibanost (npr. pri Cankarju).

Vejica v prostem stavku
Vejica v prostem stavku loči:
1. **Samostojne** izraze
a) Poudarjene zvalnike in vzklike:
Kuj me, življenje, kuj!
Tako si šla, mladost, tako si mi pobegnila, sreča, brez blagoslova!
Ostrigli, oh, so mi peruti.
b) Samostojne miselne prislove (vprašalnice, pritrdilnice in nikalnice):
Kaj, tudi ti? Kako, še lagati hočeš?
To je storil, da. Pač, še ena možnost je.
Ne, tega ti ne moremo oprostiti.
POMNI!
Ti izrazi pa niso vedno miselno samostojni; lahko so v sklopu z drugimi in skupno z njimi izražajo občutje, željo, zapoved, povabilo itd.
V takih primerih vejico opuščamo:
O domovina! Oh zlati časi! Seveda pojdi! Ne hodi! Glej ga no! Zares si priden.
c) Izraze, ki misel povzemajo, popravljajo, dopolnjujejo ali pojasnjujejo; te razlage pogosto uvajamo z besedami, kot so: na primer, in sicer, denimo, recimo, skratka itd.
Skratka, prekosili so sami sebe. Predavanje je dvakrat na teden, in sicer v torek in petek. Nekatere ptice, denimo vrabci, ščinkavci in sinice, se na jesen ne selijo. Sicer pa, kaj bi te še prepričeval. Letošnji program gledališča je usmerjen navzven, se pravi k zunanjemu, komedijskemu prikazovanju.
2. Priredno zložene stavčne člene:
a) Kadar naštevamo osebe, kraje, stvari in lastnosti, pa niso zvezani z vezniki:
Bral sem Levstika, Jurčiča, Stritarja, Kersnika, Tavčarja in Finžgarja. Jurčič je rad obiskoval vaške posebneže, rokovnjače, vojaške begunce, tihotapce in kmete.
Stritar je prepotoval Avstrijo, Nemčijo in Francijo. Doma je bil tu, tam, vsepovsod.
Tavčar je pisal romane, povesti, slike, črtice in novele.
Književnost delimo na epiko, liriko in dramatiko.
Orodje v desni, v levi meč.
POMNI!
Kadar so priredni stavčni členi (posamezne besede) združeni z vezalnimi vezniki: **in, pa, ter, ne – ne, niti – niti, tako – kakor**, ne pišemo vejice:
in – Polja in travnike je prekril sneg.
pa – Prinesla je pehar orehov pa košaro jabolk. Stari pa mladi se ne razumejo.
ter – Finžgar je v romanu Pod svobodnim soncem opisal Slovene ter Bizantince.
ne–ne – ne tega ne onega ni naredil. Ne hčere ne sina po meni ne bo.
niti–niti – Niti pismenega niti ustnega izpita ni opravil.
tako–kakor – Finžgar je pisal (tako) romane kakor povesti in drame.
Tudi pred primerjalnimi vezniki: kakor, kot, ko ne stavimo vejice, če ne vežejo ti vezniki celih stavkov:
Vleče se kakor megla.

Bela je kakor mleko.
Rdeč je ko kuhan rak.
Vejice tudi ne postavljamo med stavčne člene, ki pomenijo celoto, zato jih ne moremo in ne smemo ločiti (teh stavčnih členov ne moremo zvezati z veznikom in):
Zanimivi kemični poskusi; nenavaden fizikalni pojav; surov sovražnikov napad; pod ljubo domačo streho; naš petletni načrt; javna sodna razprava; imel je lep poslovilni govor.
b) Kadar **ponavljamo** iste besede in izraze:
Megla, megla iz jezera. O vi, vi z našim pismom, vi z našo besedo možje – kam, kam ste pahnili sebe in nas? Od štirih do ene, od štirih do ene. Žareči žeblji, žeblji v očeh ...
POMNI!
Kadar pa s ponovitvijo istih besed samo stopnjujemo, v takih primerih vejice **opuščamo**:
Takoj takoj pridem; že že grem; zdaj zdaj se bo ulilo; brž brž stopi; zelo zelo mi je žal; hudo hudo me je prizadelo; nič nič se ne izgovarjaj; hitro hitro vstani; malo malo je to; veliko veliko sem se naučil; strašno strašno se bojim.
c) Kadar **izpostavljamo** stavčni člen, ki ga v stavku nadomešča zaimek ali prislov:
Sin, ta mu je šele zrasel čez glavo!
Hči, ta je šele tičica. V mojih otroških letih, takrat smo bili otroci drugačni. V starih časih, takrat se je živelo! Poprosim ga, prijatelja. Povprašam ga, nevernika.
č) kadar izražamo z veznikom **nasprotno**, **dopustno** ali kako drugo razmerje:
Lepo, vendar težko in odgovorno nalogo je prevzel.
Ni zamerila odkriti, čeprav trdi besedi.
d) V **datumih**:
Ljubljana, 5. decembra 1957 ali Ljubljana, dne 5. decembra 1963.
V Ljubljani, 10. januarja 1957 ali v Ljubljani, dne 10. januarja 1957.

POMNI!
Ne ločimo pa z vejico kraja in datuma v besedilu:
Ta nesreča se je pripetila v Ljubljani 10. januarja 1957.
Umrl je v Ljubljani 5. decembra.

3. Vejica loči tudi **pristavek ali apozicijo** od odnosnice (to je besede, na katero se nanaša).
Pristavek (apozicija) je stavčni člen, pred katerega lahko postavimo: in sicer ali to je.
Apozicija stoji vedno **za odnosnico** in je od nje ločena z vejico.
Primož Trubar, **začetnik slovenske književnosti**, je bil doma na Raščici.
Matija Čop, **najboljši Prešernov prijatelj**, je utonil v deroči Savi.
V soboto, **dne 10. t. m.**, bo prva redovalna konferenca.
Apozicija je v zgornjih primerih sredi stavka, zato jo je treba staviti med vejici in s tem ločiti od stavka.
Vsi kasnejši rodovi so se učili pri Prešernu, **največjem slovenskem pesniku**.
Najlepši literarni spomenik je postavil slovenski materi Ivan Cankar, **največji slovenski pisatelj**.
Proslava bo v nedeljo, **dne 1. maja**.
Pristavek je v zgornjih primerih povsod na koncu stavka. Od samostalnikov, ki jih pojasnjuje (Prešeren, Cankar, nedelja), je ločen z vejico.

Vejica v zloženem stavku

Vejica v priredju

Stavke v priredju vedno ločimo z vejico, razen v tehle primerih:
1. V **vezalnem** priredju, kadar stavek zvežemo z vezniki in, ter, pa, niti – niti, ne – ne:
Sem dolgo upal in se bal.
Šivilja je zašila pa zlikala obleko.
Niti tega ne veš niti onega ne znaš.
Nisem ne prosil ne odklonil.
POMNI!
Vezniki in, ter, pa se včasih uporabljajo tudi v protivnem (lahko jih nadomestimo z veznikom toda) in sklepalnem priredju (lahko jih nadomestimo z veznikom zato); v tem primeru stoji pred njimi vejica:
Mimo je šel, in (toda) se ni oglasil.
Vsega ima na pretek, pa (toda) še ni zadovoljen.
Zdel se mi je sumljiv, pa (zato) sem stopil za njim.
Dolgo sem ga poznal, in (zato) sem mu zaupal.

2. V **ločnem** priredju:
Ali se uči ali obesi šolo na klin.
Ali pojdi takoj ali pa ostani doma.
POMNI!
Veznik ali stoji tudi v protivnem priredju namesto protivnih veznikov: toda, vendar in ampak. V tem primeru stavimo pred njim vejico:
Neurje je divjalo nad vasjo, ali drevja ni polomilo.

Vejica v okrajšanem priredju

Tudi v okrajšanem priredju ločimo stavke z vejico:
Danes meni, jutri tebi. (Danes meni zvoni, jutri bo tebi zvonilo.) Ti očeta do praga, sin tebe čez prag. Vsi so prišli, tudi teta. Ne obeduje doma, temveč v gostilni.

Vejica v podredju

Vejica v podredju loči:
1. **Odvisnik od glavnega stavka.** Odvisni stavek lahko stoji na **začetku**, na **koncu** zloženega stavka, lahko pa je vrinjen tudi v **sredo** pretrganega glavnega stavka. V takem primeru ga moramo omejiti z dvema vejicama.
Ker si len, me ne boš učil pridnosti. Ne boš me učil pridnosti, ker si len. Ti, ki si len, me ne boš učil pridnosti.

2. Odvisnike **iste stopnje**:
Bil je zadovoljen, ko je videl, kako skrbi za mladiče, kako jih neguje, kako trepeta zanje, kadar začuti v bližini sovražnika.
POMNI!
Vejice ne stavimo, kadar so odvisniki iste stopnje zvezani z vezalnimi vezniki in, pa, ter:
Samo to vem, da pred obličje nje ne smem in da ni mesta vrh zemlje, kjer bi pozabil to gorje.
Vejico tudi opuščamo, kadar veže veznik stavčni člen in istovrstni odvisnik:
Po pobočjih in kjer se sonce upre, sneg hitro skopni.

3. Odvisnike **različnih stopenj**:
Premlad si, da bi vedel, kako se služi kruh. Gospodar je bil razburjen, ker mu nihče ni znal pojasniti, kdo je moški, ki se potika okoli njegove domačije.

4. **Deležniške** zveze od **stavčne celote**:
Hodil je negotovo, oziraje se na vse strani.
Sodeč po vaši nalogi, snovi niste predelali.
POMNI!
Deležnikov ne ločimo z vejico, če so prilastki ali prosti prislovi:
Preteči prej oblak na polja ulije dež krotak.
Dano obljubo je treba izpolniti.
Na skupščini sprejeti zakoni stopijo v veljavo dne ...
Nevede ga je užalila.

5. **Okrajšana podredja**:
Prispeval sem toliko, kolikor ti.
Dasiravno bolehen, nikoli ne miruje.
POMNI!
a) Pred vezniki kakor, kot in ko ne stavimo vejice, kadar je odvisnik nepopolni stavek (ni izražen z osebno glagolsko obliko):
Laže bi ga nosil kot poslušal.
b) Tudi nedoločnika ne ločimo z vejico, kadar nadomešča odvisnik:
Branil se je hoditi v šolo.

Sestavljeni vezniki: brž ko, in ko, medtem ko, prej ko, potem ko, vtem ko, še ko, že ko, in ker, češ da, kljub temu da, namesto da, ne da, s tem da, tako da, še preden, ter da ipd. pomenijo celoto, zato stavimo vejico le pred **prvim** veznikom:
Brž ko moreš, pridi domov!
Brala je roman, **in ko** je bilo najbolj napeto, ga je morala odložiti.
Vreme je lepo, **in ker** ni vetra, tudi mirno.
Obdolžili so jo, **češ da** je pobrala vso zlatnino.
Pohajkuje, **namesto da** bi delal.

Ne da bi mu rekla, ji je za pot vse pripravil.
Še preden sem vstopil skozi vrata, se je usula ploha očitkov.
Že ko sem ga prvič videl, mi ni bil všeč.
Kljub temu da je bolan, je šel v šolo.
Včasih sta skupaj dva veznika, ki pa ne tvorita celote, temveč eden pripada glavnemu stavku (nadredni veznik), drugi pa vrinjenemu stavku (podredni veznik); v takih primerih postavimo vejico med oba stavka:
Videli ste, **da, kdor** se na začetku prenagli, se kmalu upeha.
Le huduj se name, **zakaj, kadar** si jezna, si mi najbolj všeč.
Nekateri slovničarji dopuščajo, da tudi v takih primerih stavimo vejico le pred prvim veznikom.

9.4 Spremembe pravil za LanguageTool

9.4.1 Popravljena pravila

Pravila, ki so bila spremenjena, so naslednja (označeni so spremenjeni deli):

```
<rulegroup name="Manjkajoča vejica pred 'temveč tudi'" id="TEMVEČ_TUDI_BREZ_VEJICE">
  <rule>
    <pattern case_sensitive="yes">
      <token regexp="yes" negate="yes">[,\(;\–]</token>
      <marker>
        <token>temveč</token>
        <token>tudi</token>
      </marker>
    </pattern>
    <message>Ponavadi je pred 'temveč tudi' kot drugim delom stopnjevalnega veznika 'ne le/samo - ampak/temveč tudi' vejica: <suggestion>, temveč tudi</suggestion>!</message>
    <short>Najbrž manjka vejica pred 'temveč tudi'</short>
    <example type="correct">Ne le vroče<marker>, temveč tudi</marker> vlažno je bilo.</example>
    <example type="incorrect">Ne le vroče<marker> temveč tudi</marker> vlažno je bilo.</example>
  </rule>
</rulegroup>

<rulegroup name="Manjkajoča vejica pred predlogom in 'kateri'" id="PREDLOG_KATERI_BREZ_VEJICE">
  <rule>
    <pattern case_sensitive="yes">
      <token regexp="yes" negate="yes">[Ii]n|[Aa]li|[Tt]er|[Bb]rez|[\(;:\–,]</token>
      <marker>
        <token regexp="yes">v|na|pred|nad|pri|ob|s|za|po|iz|o|skozi|pod|izza|okoli|vzdolž|do|proti|med|brez|od|zaradi</token>
        <token regexp="yes">kater(:?i|e|o|ih|ega|emu|im|ima|em)</token>
```

```
            </marker>
            <token negate="yes">koli</token>
        </pattern>
        <message>V podredjih je pred predlogom in 'kateri' običajno vejica: <suggestion>, <match no="2"/> <match no="3"/></suggestion>!</message>
        <short>Manjka vejica pred predlogom in 'kateri'</short>
        <example type="correct">Ni vedel<marker>, o kateri</marker> je govorila.</example>
        <example type="incorrect">Ni vedel<marker> o kateri</marker> je govorila.</example>
      </rule>
    </rulegroup>

    <rulegroup name="Manjkajoča vejica pred 'ko'" id="KO_BREZ_VEJICE">
      <rule>
        <pattern case_sensitive="yes">
            <token regexp="yes" negate="yes">[Ii]n|[Aa]li|[Tt]er|[Ll]e|[Ss]amo|[Tt]udi|[Pp]otem|[Zz]lasti|[Cc]elo|[Mm]edtem|[Aa]mpak|[Aa]|[Vv]endar|[Tt]oda|Vsakič|Takoj|[\(;:\–,]|[Dd]a|[Bb]olj|[Ss]aj|[Kk]er|tem</token>
            <marker>
              <token>ko</token>
            </marker>
        </pattern>
        <message>V podredjih je pred 'ko' vejica: <suggestion>, ko</suggestion>!</message>
        <short>Manjka vejica pred 'ko'</short>
        <example type="correct">Šla sva domov<marker>, ko</marker> je pričelo deževati.</example>
        <example type="incorrect">Šla sva domov<marker> ko</marker> je pričelo deževati.</example>
      </rule>
    </rulegroup>

    <rulegroup name="Manjkajoča vejica pred 'da'" id="DA_BREZ_VEJICE">
      <rule>
        <pattern case_sensitive="yes">
            <token regexp="yes" negate="yes">[,,,&quot;]|[Ii]n|[Aa]li|[Tt]er|[Tt]ako|[Ll]e|[Ss]amo|[Kk]ot|[Kk]akor|[Čč]eš|[Kk]o|[Ll]ahko|[Tt]emveč|[\(;:\–]|[Nn]e|[Aa]mpak|nam|jim|vam|mu|mi|ji|jima|vama|[Nn]amesto|[Aa]|[Tt]ko|[Ss]e|[Pp]a|[Dd]a|[Bb]rez|[Nn]amreč|temu</token>
            <marker>
              <token>da</token>
            </marker>
        </pattern>
        <message>V podredjih je pred 'da' vejica: <suggestion>, da</suggestion>!</message>
        <short>Manjka vejica pred 'da'</short>
```

```
        <example type="correct">Mislil je<marker>, da</marker> je konec sveta.</example>
        <example type="incorrect">Mislil je<marker> da</marker> je konec sveta.</example>
      </rule>
    </rulegroup>

    <rulegroup name="Manjkajoča vejica pred 'kdaj'" id="KDAJ_BREZ_VEJICE" default="off">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes" negate="yes">[Ii]n|[Aa]li|[Tt]er|[\(;:\–,]</token>
          <marker>
            <token>kdaj</token>
          </marker>
        </pattern>
        <message>V podredjih je pred 'kdaj' vejica: <suggestion>, kdaj</suggestion>!</message>
        <short>Najbrž manjka vejica pred 'kdaj'</short>
        <example type="correct">Pisalo je<marker>, kdaj</marker> je premiera.</example>
        <example type="incorrect">Pisalo je<marker> kdaj</marker> je premiera.</example>
      </rule>
    </rulegroup>
```

9.4.2 Dodana pravila

Pravila, ki so bila dodana, so naslednja:

```
    <rulegroup name="Manjkajoča vejica pred 'vendar'" id="VENDAR_BREZ_VEJICE">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes" negate="yes">vse|[Ii]n|[Aa]li|[Tt]er|[Aa]|[\(;:\–,]</token>
          <marker>
            <token>vendar</token>
          </marker>
        </pattern>
        <message>V podredjih je pred 'vendar' vejica: <suggestion>, vendar</suggestion>!</message>
        <short>Najbrž manjka vejica pred 'vendar'</short>
        <example type="correct">Slika je bila lepa<marker>, vendar</marker> je nisem kupil.</example>
        <example type="incorrect">Slika je bila lepa<marker> vendar</marker> je nisem kupil.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Manjkajoča vejica pred 'zakaj'" id="ZAKAJ_BREZ_VEJICE">
```

```
        <rule>
          <pattern case_sensitive="yes">
            <token regexp="yes" negate="yes">vse|[Ii]n|[Aa]li|[Tt]er|[\(;:\–,]</token>
            <marker>
              <token>zakaj</token>
            </marker>
          </pattern>
          <message>V podredjih je pred 'zakaj' vejica: <suggestion>, zakaj</suggestion>!</message>
          <short>Najbrž manjka vejica pred 'zakaj'</short>
          <example type="correct">Povedal je<marker>, zakaj</marker> je prišel.</example>
          <example type="incorrect">Povedal je<marker> zakaj</marker> je prišel.</example>
        </rule>
      </rulegroup>
      <rulegroup name="Manjkajoča vejica pred 'koliko'" id="KOLIKO_BREZ_VEJICE">
        <rule>
          <pattern case_sensitive="yes">
            <token regexp="yes" negate="yes">vse|[Ii]n|[Aa]li|[Tt]er|[\(;:\–,]</token>
            <marker>
              <token>koliko</token>
            </marker>
          </pattern>
          <message>V podredjih je pred 'koliko' vejica: <suggestion>, koliko</suggestion>!</message>
          <short>Manjka vejica pred 'koliko'</short>
          <example type="correct">Povedala je<marker>, koliko</marker> je hotela.</example>
          <example type="incorrect">Povedala je<marker> koliko</marker> je hotela.</example>
        </rule>
      </rulegroup>
      <rulegroup name="Manjkajoča vejica pred 'temveč'" id="TEMVEČ_BREZ_VEJICE">
        <rule>
          <pattern case_sensitive="yes">
            <token regexp="yes" negate="yes">vse|[Ii]n|[Aa]li|[Tt]er|[\(;:\–,]</token>
            <marker>
              <token>temveč</token>
            </marker>
          </pattern>
          <message>V podredjih je pred 'temveč' vejica: <suggestion>, temveč</suggestion>!</message>
          <short>Najbrž manjka vejica pred 'temveč'</short>
          <example type="correct">Ni je videl<marker>, temveč</marker> je odšel.</example>
          <example type="incorrect">Ni je videl<marker> temveč</marker> je odšel.</example>
        </rule>
      </rulegroup>
```

```
<rulegroup name="Manjkajoča vejica pred 'a'" id="A_BREZ_VEJICE">
  <rule>
    <pattern case_sensitive="yes">
      <token regexp="yes" negate="yes">vse|[Ii]n|[Aa]li|[Tt]er|[\(;:\–,]</token>
      <marker>
        <token>a</token>
      </marker>
    </pattern>
    <message>V podredjih je pred 'a' vejica: <suggestion>, a</suggestion>!</message>
    <short>Najbrž manjka vejica pred 'temveč'</short>
    <example type="correct">Ni je videl<marker>, a</marker> je odšel.</example>
    <example type="incorrect">Ni je videl<marker> a</marker> je odšel.</example>
  </rule>
</rulegroup>
<rulegroup name="Manjkajoča vejica pred 'in sicer'" id="IN_SICER_BREZ_VEJICE">
  <rule>
    <pattern case_sensitive="yes">
      <token regexp="yes" negate="yes">[,\(;:\-]|in</token>
      <marker>
        <token>in</token>
        <token>sicer</token>
      </marker>
    </pattern>
    <message>Ponavadi je pred 'in sicer' vejica: <suggestion>, in sicer</suggestion>!</message>
    <short>Najbrž manjka vejica pred 'in sicer'</short>
    <example type="correct">Bilo je v petek<marker>, in sicer</marker> popoldne.</example>
    <example type="incorrect">Bilo je v petek<marker> in sicer</marker> popoldne.</example>
  </rule>
</rulegroup>
<rulegroup name="Manjkajoča vejica pred 'medtem ko'" id="MEDTEM_KO_BREZ_VEJICE">
  <rule>
    <pattern case_sensitive="yes">
      <token regexp="yes" negate="yes">vse|[Ii]n|[Aa]li|[Tt]er|[\(;:\–,]</token>
      <marker>
        <token>medtem</token>
        <token>ko</token>
      </marker>
    </pattern>
    <message>Ponavadi je pred 'medtem ko' vejica: <suggestion>, medtem ko</suggestion>!</message>
    <short>Najbrž manjka vejica pred 'medtem ko'</short>
    <example type="correct">Bilo je v petek<marker>, medtem ko</marker> sem bil doma.</example>
```

```
        <example type="incorrect">Bilo je v petek<marker> medtem ko</marker> sem bil doma.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Manjkajoča vejica pred 'naj'" id="NAJ_BREZ_VEJICE">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes">[Rr]ekel|[Pp]rosi|[Uu]kazal|[Pp]rosila</token>
          <marker>
            <token>naj</token>
          </marker>
        </pattern>
        <message>V podredjih je pred 'naj' vejica: <suggestion>, naj</suggestion>!</message>
        <short>Najbrž manjka vejica pred 'zakaj'</short>
        <example type="correct">Potem mu je rekel<marker>, naj</marker> pride jutri.</example>
        <example type="incorrect">Potem mu je rekel<marker> naj</marker> pride jutri.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Odvečna vejica pred 'oziroma'" id="OZIROMA_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <marker>
            <token>,</token>
            <token>oziroma</token>
          </marker>
        </pattern>
        <message>Pri naštevanju je pred 'oziroma' vejica največkrat odveč: <suggestion> oziroma</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'oziroma'</short>
        <example type="correct">Pridem v nedeljo<marker> oziroma</marker> torek.</example>
        <example type="incorrect">Pridem v nedeljo<marker>, oziroma</marker> torek.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Odvečna vejica pred 'oz.'" id="OZ_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <marker>
            <token>,</token>
            <token>oz</token>
            <token>.</token>
          </marker>
        </pattern>
        <message>Pri naštevanju je pred 'oz.' vejica največkrat odveč: <suggestion> oz.</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'oz.'</short>
```

```
        <example type="correct">Pridem v nedeljo<marker> oz.</marker> torek.</example>
        <example type="incorrect">Pridem v nedeljo<marker>, oz.</marker> torek.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Odvečna vejica pred 'itd.'" id="ITD_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <marker>
            <token>,</token>
            <token>itd</token>
            <token>.</token>
          </marker>
        </pattern>
        <message>Pri naštevanju je pred 'itd.' vejica največkrat odveč: <suggestion> itd.</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'itd.'</short>
        <example type="correct">Pridem v nedeljo, torek, sredo<marker> itd.</marker></example>
        <example type="incorrect">Pridem v nedeljo, torek, sredo<marker>, itd.</marker></example>
      </rule>
    </rulegroup>
    <rulegroup name="Odvečna vejica pred '...'" id="TRI_PIKE_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <marker>
            <token>,</token>
            <token>.</token>
            <token>.</token>
            <token>.</token>
          </marker>
        </pattern>
        <message>Pri naštevanju je pred '...' vejica največkrat odveč: <suggestion> ...</suggestion>!</message>
        <short>Najbrž odvečna vejica pred '...'</short>
        <example type="correct">Pridem v nedeljo, torek, sredo<marker> ...</marker></example>
        <example type="incorrect">Pridem v nedeljo, torek, sredo<marker>, ...</marker></example>
      </rule>
    </rulegroup>
    <rulegroup name="Napačna vejica pri 'in da'" id="IN_DA_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes" negate="yes">[\(;:\–,]</token>
          <marker>
            <token>in</token>
            <token>,</token>
```

```
                <token>da</token>
              </marker>
            </pattern>
          <message>Pri zvezi veznikov 'in da' mora biti vejica največkrat pred 'in': <suggestion>, in da</suggestion>!</message>
          <short>Najbrž odvečna vejica pred 'da' in manjkajoča pred 'in'</short>
          <example type="correct">Pridem v nedeljo<marker> in, da</marker> boš boljše volje, prinesem čokolado.</example>
          <example type="incorrect">Pridem v nedeljo<marker>, in da</marker> boš boljše volje, prinesem čokolado.</example>
        </rule>
      </rulegroup>
      <rulegroup name="Napačna vejica pri 'ter da'" id="TER_DA_Z_VEJICO">
        <rule>
          <pattern case_sensitive="yes">
              <token regexp="yes" negate="yes">[\(;:\–,]</token>
              <marker>
                <token>ter</token>
                <token>,</token>
                <token>da</token>
              </marker>
            </pattern>
          <message>Pri zvezi veznikov 'ter da' mora biti vejica največkrat pred 'ter': <suggestion>, ter da</suggestion>!</message>
          <short>Najbrž odvečna vejica pred 'da' in manjkajoča pred 'ter'</short>
          <example type="correct">Pridem v nedeljo<marker> ter, da</marker> boš boljše volje, prinesem čokolado.</example>
          <example type="incorrect">Pridem v nedeljo<marker>, ter da</marker> boš boljše volje, prinesem čokolado.</example>
        </rule>
      </rulegroup>
      <rulegroup name="Napačna vejica pri 'oziroma da'" id="OZIROMA_DA_Z_VEJICO">
        <rule>
          <pattern case_sensitive="yes">
              <token regexp="yes" negate="yes">[\(;:\–,]</token>
              <marker>
                <token>oziroma</token>
                <token>,</token>
                <token>da</token>
              </marker>
            </pattern>
          <message>Pri zvezi veznikov 'oziroma da' mora biti vejica največkrat pred 'oziroma': <suggestion>, oziroma da</suggestion>!</message>
          <short>Najbrž odvečna vejica pred 'da' in manjkajoča pred 'oziroma'</short>
          <example type="correct">Pridem v nedeljo<marker> oziroma, da</marker> boš boljše volje, prinesem čokolado.</example>
          <example type="incorrect">Pridem v nedeljo<marker>, oziroma da</marker> boš boljše volje, prinesem čokolado.</example>
        </rule>
```

```
    </rulegroup>
    <rulegroup name="Odvečna vejica pri 'če da'" id="ČE_DA_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <marker>
            <token>če</token>
            <token>,</token>
            <token>da</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 'če da' vejice pred 'če' največkrat ne sem biti:
<suggestion>če da</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'da'</short>
        <example type="correct">Prišel bo,<marker> če da</marker> bo boljše volje,
prineseš čokolado.</example>
        <example type="incorrect">Prišel bo,<marker> če, da</marker> bo boljše volje,
prineseš čokolado.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Napačna vejica pri 'medtem ko'" id="MEDTEM_KO_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes" negate="yes">[\(;:\–,]</token>
          <marker>
            <token>medtem</token>
            <token>,</token>
            <token>ko</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 'medtem ko' vejice največkrat ne sme biti vmes, ampak
na začetku: <suggestion>, medtem ko</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'ko' in manjkajoča pred 'medtem'</short>
        <example type="correct">Bilo je v petek<marker>, medtem ko</marker> sem bil
doma.</example>
        <example type="incorrect">Bilo je v petek<marker> medtem, ko</marker> sem bil
doma.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Odvečna vejica pri 'medtem ko'"
id="MEDTEM_KO_Z_VEJICO_2">
      <rule>
        <pattern case_sensitive="yes">
          <marker>
            <token>medtem</token>
            <token>,</token>
            <token>ko</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 'medtem ko' vejice največkrat ne sme biti vmes:
<suggestion>medtem ko</suggestion>!</message>
```

```
        <short>Najbrž odvečna vejica pred 'ko' za 'medtem'</short>
        <example type="correct">Bilo je v petek, <marker>medtem ko</marker> sem bil doma.</example>
        <example type="incorrect">Bilo je v petek, <marker>medtem, ko</marker> sem bil doma.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Napačna vejica pri 'in ko'" id="IN_KO_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes" negate="yes">[\(;:\–,]</token>
          <marker>
            <token>in</token>
            <token>,</token>
            <token>ko</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 'in ko' mora biti vejica največkrat pred 'in': <suggestion>, in ko</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'ko' in manjkajoča pred 'in'</short>
        <example type="correct">Pridem v nedeljo<marker> in, ko</marker> boš boljše volje, prinesem čokolado.</example>
        <example type="incorrect">Pridem v nedeljo<marker>, in ko</marker> boš boljše volje, prinesem čokolado.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Napačna vejica pri 'in če'" id="IN_ČE_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes" negate="yes">[\(;:\–,]</token>
          <marker>
            <token>in</token>
            <token>,</token>
            <token>če</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 'in če' mora biti vejica največkrat pred 'in': <suggestion>, in če</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'če' in manjkajoča pred 'in'</short>
        <example type="correct">Pridem v nedeljo<marker> in, če</marker> boš boljše volje, prinesem čokolado.</example>
        <example type="incorrect">Pridem v nedeljo<marker>, in če</marker> boš boljše volje, prinesem čokolado.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Napačna vejica pri 'in ker'" id="IN_KER_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes" negate="yes">[\(;:\–,]</token>
          <marker>
```

```
            <token>in</token>
            <token>,</token>
            <token>ker</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 'in ker' mora biti vejica največkrat pred 'in': <suggestion>, in ker</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'ker' in manjkajoča pred 'in'</short>
        <example type="correct">Pridem v nedeljo<marker> in, ker</marker> boš boljše volje, prinesem čokolado.</example>
        <example type="incorrect">Pridem v nedeljo<marker>, in ker</marker> boš boljše volje, prinesem čokolado.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Napačna vejica pri 'brez da'" id="BREZ_DA_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <token regexp="yes" negate="yes">[\(;:\–,]</token>
          <marker>
            <token>brez</token>
            <token>,</token>
            <token>da</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 'brez da' mora biti vejica največkrat pred 'brez': <suggestion>, brez da</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'da' in manjkajoča pred 'brez'</short>
        <example type="correct">Pridem v nedeljo<marker> brez, da</marker> prinesem čokolado.</example>
        <example type="incorrect">Pridem v nedeljo<marker>, brez da</marker> prinesem čokolado.</example>
      </rule>
    </rulegroup>
    <rulegroup name="Odvečna vejica pri 's tem ko'" id="S_TEM_KO_Z_VEJICO">
      <rule>
        <pattern case_sensitive="yes">
          <marker>
            <token>S</token>
            <token>tem</token>
            <token>,</token>
            <token>ko</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 's tem ko' na začetku povedi vejice največkrat ne sme biti vmes: <suggestion>s tem ko</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'ko' za 's tem'</short>
        <example type="correct"><marker>S tem ko</marker> je prišel, je že pol naredil.</example>
        <example type="incorrect"><marker>S tem, ko</marker> je prišel, je že pol naredil.</example>
```

```
        </rule>
    </rulegroup>
    <rulegroup name="Manjkajoča vejica pred 'kljub temu da'" id="KLJUB_TEMU_DA_BREZ_VEJICE">
        <rule>
            <pattern case_sensitive="yes">
                <token regexp="yes" negate="yes">vse|[Ii]n|[Aa]li|[Tt]er|[\(;:\–,]</token>
                <marker>
                    <token>kljub</token>
                    <token>temu</token>
                    <token>da</token>
                </marker>
            </pattern>
            <message>Ponavadi je pred 'kljub temu da' vejica: <suggestion>, kljub temu da</suggestion>!</message>
            <short>Najbrž manjka vejica pred 'kljub temu da'</short>
            <example type="correct">Prišel je<marker>, kljub temu da</marker> je deževalo.</example>
            <example type="incorrect">Prišel je<marker> kljub temu da</marker> je deževalo.</example>
        </rule>
    </rulegroup>
    <rulegroup name="Napačna vejica pri 'kljub temu da'" id="KLJUB_TEMU_DA_Z_VEJICO">
        <rule>
            <pattern case_sensitive="yes">
                <token regexp="yes" negate="yes">vse|[Ii]n|[Aa]li|[Tt]er|[\(;:\–,]</token>
                <marker>
                    <token>kljub</token>
                    <token>temu</token>
                    <token>,</token>
                    <token>da</token>
                </marker>
            </pattern>
            <message>Pri zvezi veznikov 'kljub temu da' vejice največkrat ne sme biti vmes, ampak na začetku: <suggestion>, kljub temu da</suggestion>!</message>
            <short>Najbrž odvečna vejica pred 'da' in manjkajoča pred 'kljub'</short>
            <example type="correct">Prišel je<marker>, kljub temu da</marker> je deževalo.</example>
            <example type="incorrect">Prišel je<marker> kljub temu, da</marker> je deževalo.</example>
        </rule>
    </rulegroup>
    <rulegroup name="Odvečna vejica pri 'kljub temu da'" id="KLJUB_TEMU_DA_Z_VEJICO_2">
        <rule>
            <pattern case_sensitive="yes">
                <marker>
                    <token regexp="yes">[Kk]ljub</token>
                    <token>temu</token>
```

```
            <token>,</token>
            <token>da</token>
          </marker>
        </pattern>
        <message>Pri zvezi veznikov 'kljub temu da' vejice največkrat ne sme biti vmes: <suggestion>kljub temu da</suggestion>!</message>
        <short>Najbrž odvečna vejica pred 'da' za 'kljub temu'</short>
        <example type="correct">Prišel je, <marker>kljub temu da</marker> je deževalo.</example>
        <example type="incorrect">Prišel je, <marker>kljub temu, da</marker> je deževalo.</example>
      </rule>
    </rulegroup>
```

10 Kazala

10.1 Kazalo tabel

Tabela 1: Sestava zbirke primerov Vejica 1,0 21
Tabela 2: Rezultati evalvacije za LanguageTool - dosedanji rezultati 31
Tabela 3: Rezultati evalvacije za Besano – dosedanji rezultati 32
Tabela 4: Možna stanja pri analizi napak..... 37
Tabela 5: Oznake vrst vejic..... 42
Tabela 6: Vpliv klasifikatorja in velikosti; Šolar, celoten MSD, brez skladenjskih atributov, okno -5+5 55
Tabela 7: Vpliv atributov; Šolar, okno -5+5 57
Tabela 8: Vpliv okna; Šolar, MSD + delni MSD 2 + skladnja 59
Tabela 9: Vpliv označevanja; Šolar, MSD + delni MSD2 + skladnja, okno -5+5 62
Tabela 10: Vpliv parametrov na ADTree; Šolar, MSD + delni MSD2 + skladnja, okno -5+5, označevanje brez vejic, naučeno brez vejic 63
Tabela 11: Najpogostejše besede za manjkajočimi vejicami, ki jih LanguageTool ne odkrije 67
Tabela 12: Najpogostejše besede, pred katere LanguageTool postavi odvečno vejico 70
Tabela 13: Besede, pred katerimi je največ odvečnih vejic, ki jih LanguageTool ne odkrije . 71
Tabela 14: Najpogostejše besede za manjkajočimi vejicami, ki jih Besana ne odkrije 86
Tabela 15: Najpogostejše besede, pred katere Besana postavi odvečno vejico 88
Tabela 16: Besede, pred katerimi je največ odvečnih vejic, ki jih Besana ne odkrije 92
Tabela 17: Rezultati za postavljanje vseh vejic na korpusu Šolar 95
Tabela 18: Rezultati strojnega učenja postavljanja vejic za različne jezike 95
Tabela 19: Primerjava rezultatov za isti klasifikator in enak nabor atributov, a različen korpus 95
Tabela 20: Rezultat popravljanja vejic s strojnim učenjem 96
Tabela 21: Rezultati evalvacije za LanguageTool - novi rezultati..... 98
Tabela 22: Rezultati evalvacije za Besano - novi rezultati 100

10.2 Kazalo slik

Slika 1: Priklic manjkajočih vejic – dosedanji rezultati 30
Slika 2: Natančnost pri manjkajočih vejicah – dosedanji rezultati 33
Slika 3: F1 pri manjkajočih vejicah – dosedanji rezultati 34
Slika 4: Uspešnost Besane pri odvečnih vejicah – dosedanji rezultati 34
Slika 5: Priklic vseh vejic – dosedanji rezultati 35
Slika 6: Natančnost pri iskanju vseh vejic – dosedanji rezultati 36
Slika 7: F1 pri iskanju vseh vejic – dosedanji rezultati 36
Slika 8: Rezultat skladenjskega razčlenjevalnika 53
Slika 9: Zapis označevanja in skladenjske razčlembe v formatu XML 54
Slika 10: Zapis začetka zgornjega primera v formatu ARFF z oknom -5+5 55
Slika 11: Odločitveno drevo za ADTree -B 12 -E -3 64
Slika 12: Rezultat primerjave rezultatov preizkušanja s podatki iz korpusa 65
Slika 13: Deleži pravilno postavljenih vejic po vrstah v korpusu Šolar 73
Slika 14: Deleži pravilno postavljenih vejicah po vrstah v korpusu Lektor 74
Slika 15: Delež popravljenih vejic po vrstah v korpusu Šolar 75
Slika 16: Delež popravljenih vejic po vrstah v korpusu Lektor 76
Slika 17: Deleži vrst nepopravljenih vejic v korpusu Šolar 76
Slika 18: Deleži vrst nepopravljenih vejic v korpusu Lektor 77
Slika 19: Končni rezultati priklica manjkajočih vejic 101
Slika 20: Končni rezultati za natančnost pri manjkajočih vejicah 102
Slika 21: Končni rezultati za F1 pri manjkajočih vejicah 102
Slika 22: Končni rezultati priklica odvečnih vejic 103
Slika 23: Končni rezultati za natančnost pri odvečnih vejicah 103
Slika 24: Končni rezultati za F1 pri odvečnih vejicah 104
Slika 25: Končni rezultati priklica pri postavljanju vseh vejic 105
Slika 26: Končni rezultati za natančnost pri postavljanju vseh vejic 106
Slika 27: Končni rezultati za F1 pri postavljanju vseh vejic 106

Predhodna objava

Izjavljam, da so bili rezultati obravnavane raziskave prehodno objavljeni/sprejeti v objavo v recenzirani reviji oziroma monografiji ali javno predstavljeni v naslednjih primerih:

Holozan, Peter, 2013: Uporaba strojnega učenja za postavljanje vejic v slovenščini. *Uporabna informatika* XXI/3. 196–209.

Holozan, Peter, 2015: Možnosti uporabe jezikovnih tehnologij za določanje težav pri rabi vejice. V: Dobrovoljc, Helena, Lengar Verovnik, Tina (urednici), *Pravopisna razpotja – razprave o pravopisnih vprašanjih*. Ljubljana: Založba ZRC. 77–92.

Del rezultatov je bil predstavljen tudi na mednarodni konferenci.

Potrjujem, da sem pridobil pisna dovoljenja vseh lastnikov avtorskih pravic, ki mi dovoljujejo vključitev zgoraj navedenega materiala v pričujočo disertacijo. Potrjujem, da zgoraj navedeni material opisuje rezultate raziskav, izvedenih v času mojega doktorskega študija na Univerzi v Ljubljani.

IZJAVA O AVTORSTVU DOKTORSKE DISERTACIJE

Podpisani Peter Holozan, z vpisno številko 18119087, rojen 17. 3. 1970 v kraju Ljubljana, sem avtor doktorske disertacije z naslovom: Računalniško postavljanje vejic v slovenščini.

S svojim podpisom zagotavljam, da:

- je predložena doktorska disertacija izključno rezultat mojega lastnega raziskovalnega dela;
- sem poskrbel, da so dela in mnenja drugih avtorjev oz. avtoric, ki jih uporabljam v predloženem delu, navedena v seznamu virov in so v delu citirana v skladu z mednarodnimi standardi in veljavno zakonodajo v RS na področju avtorskih in sorodnih pravic;
- je elektronska oblika identična s tiskano obliko doktorske disertacije;
- **soglašam / ne soglašam** *(ustrezno obkrožiti)* z objavo doktorske disertacije na spletnih straneh Filozofske fakultete Univerze v Ljubljani.

V Ljubljani, dne 4. marca 2016 Podpis avtorja:_____