

KOST

KORPUS **S**LOVENŠČINE
KOT **T**UJEGA JEZIKA

Priročnik za označevanje napak,
delovna verzija

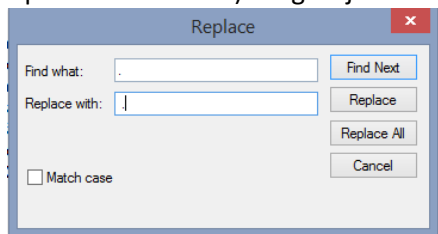
Mojca Stritar Kučuk

april 2022

POSTOPEK OZNAČEVANJA

Priprava besedila (txt)

1. Urednica KOST-a (= Mojca Stritar Kučuk) ti **pošlje** (deli?) **paket besedil**. Vsako besedilo je v ločeni **.txt datoteki**, ki ima ime sestavljeno iz kode, npr. L-2021-035.txt.
2. Delaš z vsako datoteko **posebej**.
3. Datoteko **odpreš** (na mojem računalniku se to avtomatsko odpre v Beležnici).
4. **Ločila** v besedilu morajo biti pisana **nestično** (grozno, ampak resnično; eno ločilo se namreč šteje za en znak), zato najprej poskrbiš za to. V Beležnici lahko greš na zavihek **Edit > Replace** (lahko uporabiš bližnjico na tipkovnici CTRL + H). V zgornje okence vpišeš **piko**, v spodnje pa **presledek in piko** in klikneš **Replace All**.



5. Isto lahko ponoviš za **vejico**.
6. Za **ostala ločila** (podpičje, dvopičje, oklepaj) tega ponavadi ni treba početi, ker jih ni veliko.
7. Besedilo razdeliš na **smiselne odstavke**. Gre za to, koliko sobesedila se bo videlo v konkordančniku korpusa, potem ko boš kliknil na izbrano konkordanco; prikaže se namreč samo en odstavek. To torej nima jezikovne teže, ampak gre bolj za tehnično stvar. Če je besedilo zelo kratko (nekaj povedi), lahko ostane v enem odstavku.

Označevanje napak (Svala)

8. V brskalniku odpreš <https://svala.cjvt.si/>.
9. Zgoraj desno klikneš na gumb **Show options** in izbereš **Switch to correction annotation KOST** (privzeta je namreč taksonomija napak za korpus domačih govorcev Šolar).
10. V istem meniju na vrhu izbereš **Show source text**. Zdaj se na vrhu pojavi okno **Source text**, v katerega **skopiraš** besedilo, ki ga je treba označiti. Pod tem oknom imaš okno **Target text**, v katerega pride popravljena verzija besedila.
11. Preveri, ali v besedilu ni še **drugih ločil**, pri katerih je treba popraviti stičnost, ali kakega osebnega podatka, ki ponesreči ni bil zakrit. Ko vse to storiš, klikneš na gumb **Copy to target** pod okencem Source text. S tem se izvirno besedilo skopira v spodnje okence.

12. Prebereš **target text** in v njem **popraviš napake** - napišeš pravilne oblike. Če česa ne znaš popraviti, namesto pravilne oblike napišeš [???].

Source text:

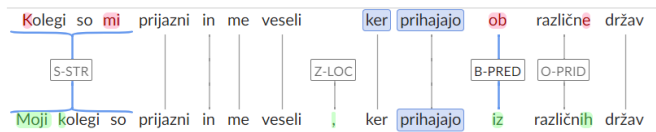
Zdaj živim v Ljubljani in študiram ekonomijo na Ekonomski fakulteti . Slovenija je zelo lepa in čista . Ljubljana nije prevelika , ampak je internacionalno mesto in všeč mi je . Ljudje so gostoljubni . Študije , grad in ljudje su mi všeč . Semafori nisu mi všeč , ker rdeča luč predolgo traje . Hrana je raznovrsta , dobra in poceni . Kolegi so mi prijazni in me veseli ker prihajajo ob različne držav .

copy to target

Target text:

Zdaj živim v Ljubljani in študiram ekonomijo na Ekonomski fakulteti . Slovenija je zelo lepa in čista . Ljubljana ni prevelika , ampak je internacionalno mesto in všeč mi je . Ljudje so gostoljubni . Študij , mesto in ljudje so mi všeč . Semaforji mi niso všeč , ker rdeča luč predolgo traja . Hrana je raznovrstna , dobra in poceni . Moji kolegi so prijazni in me veseli , ker prihajajo iz različnih držav .

Spaghetti mode disable



13. Pod obema okencema vidiš še tretje okence, v katerem sta **sopostavljeni izvirno in popravljeno besedilo**, in sicer po posameznih povedih. Besede so med seboj povezane s črtami, t. i. **špageti**.
14. S klikom na špaget se ti na levi strani pokaže **meni napak**. Ko torej klikneš na špaget, na levi izbereš ustrezno kategorijo napak. Izbereš lahko tudi več oznak naenkrat.
15. Če želiš dati eno oznako več besedam, jih označiš (klikneš in držiš Shift ali Ctrl), potem pa levo zgoraj izbereš gumb **group**.
16. Če je sistem avtomatsko povezal več besed, pa jih želiš ločiti, klikneš na špaget in na levi izbereš gumb **orphan**.

previous next

prev mod next mod

group orphan

auto

Enter filter text

+ Zapis

- Besedišče

Samostalnik

Glagol

Zaimek

Pridevnik

Prislov

Predlog

Veznik

Ostalo

+ Oblika

+ Skladnja

+ Povezani popravki

Source text:

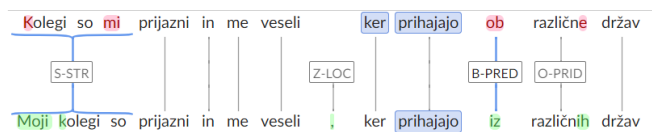
Zdaj živim v Ljubljani in študiram ekonomijo na Ekonomski fakulteti . Slovenija je zelo lepa in čista . Ljubljana nije prevelika , ampak je internacionalno mesto in všeč mi je . Ljudje so gostoljubni . Študije , grad in ljudje su mi všeč . Semafori nisu mi všeč , ker rdeča luč predolgo traje . Hrana je raznovrsta , dobra in poceni . Kolegi so mi prijazni in me veseli ker prihajajo ob različne držav .

copy to target

Target text:

Zdaj živim v Ljubljani in študiram ekonomijo na Ekonomski fakulteti . Slovenija je zelo lepa in čista . Ljubljana ni prevelika , ampak je internacionalno mesto in všeč mi je . Ljudje so gostoljubni . Študij , mesto in ljudje so mi všeč . Semaforji mi niso všeč , ker rdeča luč predolgo traja . Hrana je raznovrstna , dobra in poceni . Moji kolegi so prijazni in me veseli , ker prihajajo iz različnih držav .

Spaghetti mode disable



Shranjevanje označenih besedil

17. Ko besedilo v celoti označiš, v meniju na desni klikneš na ukaz **export graph**. Odpre se ti mapa, v katero se shranjujejo dokumenti. Avtomatsko so shranjeni s končnico **.json**. Vsak dokument torej poimenuješ z enakim imenom kot v izvirniku, a z drugo končnico: L-2021-035.txt > **L-2021-035.json**. Če je dokument razbit na več odstavkov, vsakega shraniš v poseben dokument in imenu dodaš še eno številko: L-2021-035-1.json, L-2021-035-2.json, L-2021-035-3.json itn.
18. Če ugotoviš, da je v katerem od že izvoženih dokumentov napaka, ga uvoziš (**import graph**) v Svalo, popraviš in spet izvoziš.
19. Paket označenih besedil **vrneš urednici**.

Problemi

20. Če pri označevanju naletiš na problem, napako (pa ne take sorte, da so jo naredili tvorci besedil) ali kar koli drugega, si to zabeležiš in **javiš urednici**.

KLASIFIKACIJA NAPAK

Težimo k temu, da čim manj posegamo v besedilo in popravljamo čim manj napak. Predvsem popravljamo zapis, besedišče in obliko besed, skladnjo pa skušamo popravljati čim manj.

Zapis

Zapis / Ločilo (Z-LOC)

- **vejica:**
 - **odveč** - v "target text" napišeš brez vejice, daš kot "orphan" in označiš "ločilo": *V zimskem času , večina ljudi uporablja avtomobile*
 - **manjka** - v "target text" vpišeš vejico, daš kot "orphan" in označiš "ločilo"
 - manjka **pred odvisnikom**: *Všeč mi je ker je hiša in ni majhna*
 - manjka **pred priredjem**: *Ni mi všeč plavanje v morju ampak mi je zelo všeč ta okolica*
 - manjka **za vrinenim stavkom**: *Ena izmed najboljših idej je prepovedati vozila, ki hodijo na gas in zamenjati jih*
- **pika:**
 - manjkajoča pika pri kratlici: *itn > itn .* ("zapis": "ločilo")
 - položaj pike pri **premем govoru**: *sem en mikrob". > sem en mikrob."* (povežemo in damo oznako Z-LOC)
 - izražanje **ure**: *okoli 9 > okoli 9.00*

Zapis / Črkovanje (Z-CRK)

- **z > s, s > z**: *ne ispušča plinov*
- premena **po zvenečnosti** na koncu besede: *kod > kot*
- **izpust samoglasnika** (pogovorno): *da se vedno zgubi > izgubi; tut > tudi*
- izpust **t**: *korisno > koristno*
- očitni **lapsus**:
 - *vse okoli je moje > morje*
 - *se veliko uporablja > uporablja*
 - *mora imetu > mora imeti*
 - *aku > ali*
- **podvojene črke**: *oodih > oddih; surrealizmom > surrealizmom*
- napačna **črka**
- odvečna oznaka **naglasa**: *prizadéte > prizadetih* (tudi napaka oblike)
- zapis **tujih besed**: *Ridli Scot > Ridley Scott*
- izražanje **ure**: *ob 10ih > ob desetih*
- **števniki**: *od 66h skladišč > iz 66 skladišč; 01. 01. 2022 > 1. 1. 2022*

Zapis / Skupaj/narazen (Z-SN)

- glagol **biti**: **ni sem** poskusil > nisem poskusil
- **stopnjevanje** oblike pridevnika: **naj bolj** > najbolj
- **simboli**: 86% > **86 %**
- tudi raba **vezaja**: **86 letna** > 86-letna, **e učilnici** > e-učilnici; v Buenos Aires-u > Airesu

Zapis / Mala/velika začetnica (Z-MV)

- tudi primeri, v katerih je napako verjetno zakrivil **program**: v naslednjem času tj. **Med** poletnimi počitnicami
- razni čudni primeri: sončna energija **in** paneli

Zapis / Krajšave (Z_KR)

in dr. > idr.

Besedišče

Besedišče / Samostalnik (B-SAM)

- "klasični" primeri: **prirodi** > naravi; **kadiranje** > kajenje
- napačna črka v **korenu** besede: **jedelnik** > jedilnik; hidroelektrane > hidroelektra**rne**
- napačna končnica: **kolokvijume** > kolokvije
- beseda napačnega **spola**: **ščitimo našo planeto**
- primeri, ki bi jih lahko uvrstili med **črkovanje**, vendar gre za obstoječe besede: **na cvetu** > na **svetu**
- podobna **samostalniška besedna zveza** - popraviš samostalnik, napaka zaimka je pa tudi napaka besedišča + povezan popravek: **mi je zelo všeč ta okolica** > **to okolje**
- **samostalnik**, čeprav je druga **besedna vrsta**: v omejenih **prostih** > **prostorih**
- **strešice**, če obstaja beseda v drugem jeziku: **student** > **študent**
- tuja **imena**: **Skoplje** > Skopje, **Graz** > Gradec
- kadar gre za nekakšen **povedkovnik**, označimo kot samostalnik: **ni me stid** > **sram**

Besedišče / Glagol (B-GLAG)

- "klasični" primeri: **vozila, ki hodijo** na gas; **sem se zelo težko naučila** na mir
- napačna **beseda**, pa čeprav je **podobna**: **zabavljali** > zabavali
- glagolski **vid**: **vzemala** dih > jemala dih
- glagolske **predpone**: **prenašala** > prinašala
- napačna črka v **nedoločniku**: **mora imati** > imeti; **odlučiti** > odločiti, **iči** > iti
- **ima** > **je**: **ima** malo ljudi > je malo ljudi
- **negacija** (zaradi veznika): **vse do takrat, ko ni** prišel koronavirus

Besedišče / Pridevnik (B-PRID)

- **napačno napisana beseda**, če v slovenščini obe obstajata: V enem delu hiše živijo moj stric, teta in **nujna** dvojčka
- zamenjave **ski, ški**: turske > turške
- **pripona**: **vetrene** elektrane > **vetrne** elektrane; hitar > hiter; plačen > plačan

Besedišče / Zaimek (B-ZAIM)

- **dvojina**: onidva > onadva
- **dolga, kratka oblika zaimka**: uredila sem **sebi** življenje > si
- **svojilni zaimek**: zakaj toliko ljudi radi izberejo Piran za **njihove** počitnice > svoje
- **en > neki**: spaval pri **enem** prijatelju
- **oziralni zaimek kateri, ki**: **kateri** sem ogledala > ki sem si ga ogledala; **kim** omogočajo > ki jim
- **vprašalnice**: vprašanje je **kje** je zdaj s tem telefonom > kaj
- **kazalni zaimek**: **taj** > ta
- **raba dolge oz. kratke oblike zaimka**: Jaz **sebi** rad kupim majice > Jaz **si** rad kupim majice

Besedišče / Prislov (B-PRISL)

- **toliko, tako**: **toliko** velik problem > **tako** velik problem
- **doma > domov**
- **jeziki**: govori **srbski** > srbsko
- **besedne zveze**: in **tega slično** > in podobno
- **deležje**: zahvaljujoče > zahvaljujoč

Besedišče / Predlog (B-PRED)

- **pazi, pri menjavi s/z gre za napako črkovanja!**
- **sa > s/z**: **sa** enom prijateljicom
- **napačen predlog**, za katerim se zamenja **sklon** - predlog je primarna napaka, sklon pa popraviš in označiš kot povezano napako: se gre **o** čiščenju > se gre **za** čiščenje

Besedišče / Veznik (B-VEZ)

- **kdaj > ko**: **kdaj** sem obiskal Turčijo
- **kje > kjer**: Turčijo, **kje** sem spaval
- **od kdaj > odkar**

Besedišče / Ostalo (B-OST)

- **tudi** > poleg tega: **Tudi** je en euro bil osem turških lir
- **in** > **tudi**
- **po meni** > po mojem mnenju
- **se** > **si**
- **še** > **že**
- **krajšave**: **it.** > itd.; **UE** > **EU**
- prek meja **besedne vrste**: ker je takrat Jezus Kristus **mrtev** > umrl; vsi so za zdravo **jesti** > hrano
- **ne** > **ni**: **ne** nujno gledati > ni nujno gledati
- neprimeren pozdrav v mejlu: **Dober dan** > Pozdravljeni/Spoštovani
- **števniki**: **petindvajset** > petindvajset
- **simboli, krajšave**: **dg** hrane > dag hrane
- **edna** > **ena**

Oblika

Oblika / Samostalnik (O-SAM)

- napačen **sklon**, obstoječa končnica: *nagajajo ptic**ami*** > *pticam*
- napačen **sklon**, neobstoječa končnica: v *Slovenij**u***
- napačna končnica, ker ima beseda napačen **spol**: *veliko vrst**ov** živali* > *veliko vrst*
- manjka **j**: *inštruktori* > *inštruktor**ji***
- **število**: *jaz sem največ jedla rib**o** in sladice*
- **število**, kjer je zraven še več povezanih popravkov – samostalnik je osnovni popravek, ostali so povezani: *ker je prizor zelo lep* > *ker so prizori zelo lepi*
- težave z **j**: *objavljamo vide**eje*** > *videe*
- oblika samostalnika s **posebnostmi**: zdravo **teleso**
- palatalizacija: **Americi** > Ameriki

Oblika / Glagol (O-GLAG)

- **sam** > *sem*
- oblike za **dvojino**: *delam**o** sestra in jaz*
- **sedanjik**: *zato iš**kem*** > *iščem*
- **nije** > *ni*
- **slišati**: *sliš**am***
- napake pri besedi **rad**
- preteklik, prihodnjik, pogojniki – velja za **deležnik in pomožni glagol**: *rad bi im**al*** > *imel*; *pogovarj**ao*** > *pogovarjal*
- napačen glagolski **čas**: *potem bo lah**ko*** > *potem je bilo lah**ko***; *nisem mislila* > *ne mislim*
- glagolsko število ob **količini**: *zakaj toliko ljudi radi izberej**o*** > *rado izbere* (oboje glagol)
- **ruske** variante z glagoli: *vse stvari so del**a** skupaj* > *delajo*; *ljudje ne vedno dobri* > *niso*

- **modalni** glagoli, napake pri pregibanju: *htela* > *hotela*
- **namenilnik**: *takrat grem jesta* > *jest*
- **zdeti se** in podobne čudne oblike - označimo tudi izpust *je*: *mi se zdilo* > *zdelo se mi je*

Oblika / Pridevnik (O-PRID)

- **sklon** pridevnika - če je samostalniška besedna zveza v neustreznem sklonu, ki pa v slovenščini obstaja, se to označi kot osnovno napako pri samostalniku, oblika pridevnika pa je povezana napaka (dobi torej dve oznaki):
v zelo lep, urejen hotel > *v zelo lepem, urejenem hotelu*
- **sklon** pridevnika - če je pridevnik v neustreznem sklonu, ki je v slovenščini napačen v vseh kontekstih, tega ne označimo kot povezano, pač pa kot primarno napako: *kopale smo se v čudovite morje*
- manjkajoč vokal v priponi: *raztresni* > *raztreseni*
- **stopnjevanje**: *največi* > *največji*, *boljo* > *boljšo*
- **(ne)določna oblika**: *cel virtualen svet* > *ves virtualni svet*

Oblika / Zaimek (O-ZAIM)

- **skloni** pri vseh vrstah samostalniških in pridevniških zaimkov
- **osebni** zaimek: **ih** > *jih*
- **svojilni** zaimek: **vašego** odgovora > vašega
- **kazalni** zaimek: med **tistem** počitnicom > med *tistimi počitnicami*
- **nedoločni** zaimek: v **enih** od restavraciji > v *eni* od restavracij

Oblika / Prislov (O-PRISL)

- *najpogostejše* > *najpogosteje*
- *bolj* > *bolje*

Oblika / Ostalo (O-OST)

- **kratice**: predsednik **od UEFA** > predsednik *UEFE* (odvečni predlog + napaka oblike)
- **števniki**: **štirje** predavanja > *štiri* predavanja

Skladnja

Skladnja / Struktura (S-STR)

- strukture z **da**: Rada bi **da živim** sama
- struktura z **dajalnikom** namesto **svojilnega zaimka**: **bi mi na** jedilniku bila > **bi na mojem** jedilniku bila
- **modalni glagoli, treba**: **se jih vsak dan treba** čistiti > *jih je vsak dan treba* čistiti
- **modalni glagoli, lahko**: **moremo rečiti** > lahko rečemo
- **ne mara** > ima rad

Sem sodijo še razne čudne strukture, ki so prekomplicirane, da bi jih označevali kot napake besedišča:

- *sva odločitva* > *sva se odločili*
- *10. razred v srednjo šolo* > v *(izpust) 10. razred srednje šole*
- *ki jaz všečem zelo* > *ki mi je zelo všeč*
- *bom delala tukaj do bom upokojila* > bom delala tukaj **do takrat, ko se bom upokojila**

Skladnja / Besedni red (S-BR)

- **rad imam**: v poletnih mesecih **rada ima**
- problem **znotraj naslonskega niza**: uresničile **so mi se** otroške sanje
- naslonski niz za **veznikom**: da **komunikacija je** bistven element; ampak **mi je zelo všeč** ta okolica
- naslonski niz v **glavnem stavku**: Če bi sama, **dovolj bi mi bila** ena soba
- **pa**: **kopanje pa** v bazenih
- **tudi**: **Tudi** sem se zelo zblížil
- **leto**: od **1991 leto** > iz leta 1991 (napačen predlog, napaka oblike, pa še besedni red)

Skladnja / Odvečni podatki (S-ODV)

- **odveč beseda:** *in je* njihovo delo pa je podobno delu hidroelektrarn
- **odveč predlog:** *rad bi se udeležil na tečaj* > *udeležil tečaja*
- **deležnik glagola biti v prihodnjiku:** *ne bova bila* > *ne bova*

Skladnja / Izpuščeni podatki (S-IZP)

- **manjkajoče besede**
- **se:** *sem najbolj naučil* > *sem se najbolj naučil*
- **oziralni zaimki,** če manjka **osebni** zaimek: *ki sem jedla* > *ki sem ga jedla*

Povezani popravki (POV)

- **povezana oblika zaimka ali pridevnika:** *zelo mi je všeč ta okolica* > *to okolje* (povezana oblika zaimka)
- **povezana velika začetnica:** *Jaz rada imam* > *Rada imam*
- **popravek predloga,** ker se zamenja svojilni zaimek: *z mojim* > *s svojim*

Nejasen popravek

označen z [???