



6.2 IZROČKI – vaje

JANES (••)  

Od Novega mesta do Njujorka
Vaje

Tamara Mikolič Južnič

Ljubljana, 28. 8. 2015

JANES (••)  

I.

I. Kako zapisujemo tujejezične elemente?

1. Se pogosteje uporablja citatni zapis ali poslovenjene oblike?

- a. Odpremo NoSke, izberemo korpus JANES v. 0.3.
- b. V iskalno okence za enostavno iskanje vpišemo naslednje besede v citatnih oblikah in poiščemo njihovo pogostnost.
 - like
 - good
 - please

I. Kako zapisujemo tujejezične elemente?

2. Poleg citatnih oblik se, kot smo videli, uporabljajo tudi različne poslovenjene oblike.

- a. Kako lahko poiščemo poslovenjene oblike?
- b. Katere so po vašem mnenju poslovenjene oblike zgornjih besed?
- c. Napišite te oblike v okence za enostavno iskanje in poiščite njihovo frekvenco. Prepišite najdene vrednosti za citatne oblike (vse skupaj) in poslovenjene oblike (vse skupaj).
 - *please*
 - *good*
 - *like*

I. Kako zapisujemo tujejezične elemente?

3. Katere oblike torej prevladujejo pri izbranih besedah, citatne ali poslovenjene?

II.

II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

1. Najprej si pogledjmo, koliko je tujejezičnih elementov v korpusu.

a. V korpusu JANES so besede, ki jih je avtomatski označevalnik prepoznal kot tuje, vendar program ne ločuje med različnimi jeziki. Besede so preprosto označene s posebno, vedno enako kodo. Ko iščemo s CQL, lahko uporabimo to kodo in vse take besede izluščimo naenkrat.

- CQL: [tag="Nj"]
- Koliko je vseh takih konkordanc?

II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

b. Za primerjavo odprimo tokrat govorni korpus GOS in preverimo, koliko je tujejezičnih elementov v tem korpusu.

- Vtipkajmo enak iskalni ukaz v polje CQL.
- Zakaj prihaja do take razlike?

II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

- c. Ali se prisotnost tujejezičnih elementov spreminja glede na besedilni tip? Poiščite skupno število tujejezičnih elementov v vsakem izmed štirih podkorpusov JANESa.
- Na vrhu izberite podkorpus, nato kot vrsto iskanja izberite CQL in v ukazno polje vpišite [tag="Nj"].
 - Izberite posamezni podkorpus in preverite število konkordanc.
Blogs (JANES v0.3 Blog)
Komentarji (JANES v0.3.1 Comment)
Forumi (JANES v 0.3.1 Forum)
Tviti (JANES v0.3.3 Tweet)

II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

2. Videli smo, da se poleg angleščine, ki nedvomno prevladuje, v spletni slovenščini pojavljajo tudi besede iz drugih jezikov. Kako bi ugotovili, kateri jeziki se poleg angleščine še pojavljajo?
- Izberite si enega izmed podkorpusov, poiščite vse tujejezične elemente v tem podkorpusu.
 - Napravite frekvenčni seznam besed.
 - Nato na prvih 10 straneh frekvenčnega seznama »ročno« preverite, kateri jeziki se pojavljajo (poleg angleščine). Zapišite te jezike (in primere rabe, ki ste jih našli).

II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

3. Kako točni so naši rezultati?
 - a. Ker avtomatsko označevanje ni 100 %, se pojavljajo napake: med besedami, označenimi kot neslovenskimi, je tudi nekaj takih, ki to zagotovo niso. Poiščite jih 5.
 - b. Zakaj prihaja do takih napak?

III.

III. Kdaj in kako se uporablja tujejezične elemente?

1. Kako se pojavljajo tujejezični elementi? Kakšne vsebine bodo izražene v tujih jezikih?
 - a. Na kakšen način bi se lahko pojavljali tujejezični elementi v korpusu JANES?

III. Kdaj in kako se uporablja tujejezične elemente?

- b. Ustvarite seznam konkordanc vseh tujih besed v korpusu JANES.
 - Izberite korpus JANES v0.3.
 - Pri vrstah iskanj izberite CQL in v iskalno polje vpišite [tag="Nj"].
- c. Ali se primeri, ki ste jih našli, ujemajo z vašimi predvidevanji? Navedite vsaj 5 načinov, na katere se pojavljajo tujejezični elementi.

III. Kdaj in kako se uporablja tujejezične elemente?

- d. Kateri način vam je najbližji oz. kaj sami uporabljate v podobnih okoliščinah?
- Napišite tri besede/besedne zveze, ki jih najpogosteje uporabljate in primerjajte njihovo pogostnost v vseh štirih podkorpusih.
 - Besedo/besedno zvezo vpišite v polje za enostavno iskanje (izberite vsak podkorpus posebej).
 - Prepišite kolikokrat na milijon se pojavlja v posameznem podkorpusu.

III. Kdaj in kako se uporablja tujejezične elemente?

Beseda/ besedna zveza	Tweet	Forum	Blog	Comment

III. Kdaj in kako se uporablja tujejezične elemente?

- e. Kdaj se po vašem mnenju avtorji odločajo za preklapljanje?
- Poiščite vse tuje besede v podkorpusu Tweet.
 - Napišite, kateri so razlogi za preklapljanje v opazovanih primerih.
 - Ali avtorji izrazito preklapljajo na kakem posebnem mestu v svojih tvitih?

IV.

IV. Kdo uporablja tujejezične elemente?

1. Korpus JANES je označen tudi s podatki o tem, kdo so avtorji besedil, ki jih vključuje.
Zlasti podkorpus TWEET ponuja veliko informacij v tem smislu.
Poglejmo si, pri katerih uporabnikih se tujejezični elementi pojavljajo pogosteje.

IV. Kdo uporablja tujejezične elemente?

- a. zasebni uporabniki = 'private'
podjetja, organizacije in drugi javni subjekti = 'corporate'.
Kateri tip uporabnikov uporablja več tujejezičnih elementov?
 - Vključimo lastnosti besedil pri iskanju.
 - Izberemo možnost 'corporate' pri izbiri TEXT.SOURCE.
 - Z iskanjem CQL poiščemo vse tujejezične elemente v izbranih tvitih.
 - Nato gremo nazaj na iskanje in pri TEXT.SOURCE izberemo možnost 'private'.
 - Z iskanjem CQL poiščemo vse tujejezične elemente v teh tvitih.

IV. Kdo uporablja tujejezične elemente?

2. Zabeležen je tudi spol uporabnikov.
 - a. Ali tujejezične elemente več uporabljajo ženske ali moški?
 - Pri lastnostih besedil izberemo spol avtorja.

IV. Kdo uporablja tujejezične elemente?

3. Kdo je v svojih tvitih uporabil absolutno največ tujejezičnih elementov?
 - a. Preverimo v podkorporusu TWEET
 - Izdelamo konkordance vseh tujejezičnih elementov v podkorporusu TWEET.
 - Izberemo Frekvence in v spodnjem kvadratu (Frekvenčna razporeditev po lastnostnih besedil) Izberemo prvo možnost (text.author).

IV. Kdo uporablja tujejezične elemente?

- b. Kaj pa če upoštevamo relativno frekvenco, kdo najpogosteje uporablja tujejezične elemente?
- Na zgoraj izdelanem frekvenčnem seznamu kliknemo na napis »Rel. frekvenca (%)«

IV. Kdo uporablja tujejezične elemente?

- c. Katere tujejezične elemente pa ta oseba uporablja?
- Kliknemo na »p« na začetku vrstice ob imenu avtorja, da odpremo konkordance vseh tvitov s tujejezičnimi elementi tega avtorja.
 - Izdelamo frekvenčni seznam tujejezičnih elementov za te konkordance.
 - Izberite še kakega drugega avtorja in preverite, ali so besede pri vrhu podobne.

V.

V. Stopnja (ne)standardnosti in tujejezični elementi

1. Podkorporus Tweet je označen tudi glede na stopnjo nestandardnosti besedila (oznake L1, L2 in L3).
Oglejmo si, ali višja stopnja nestandardnosti pomeni tudi pogostejšo uporabo tujejezičnih elementov.
Kakšna so vaša pričakovanja?

V. Stopnja (ne)standardnosti in tujejezični elementi

- a. Kje je po vašem mnenju največ tujejezičnih elementov?
- b. Preverimo, koliko je tujejezičnih elementov v besedilih z najnižjo stopnjo nestandardnosti (torej v besedilih, ki najbolj sledijo pravilom o standardni rabi slovenščine).
 - Izberemo podkorpus TWEET.
 - Pri Lastnostih besedil izberemo L1.
 - V iskalno polje CQL napišemo iskalni ukaz za tujejezične elemente: [tag="Nj"].

V. Stopnja (ne)standardnosti in tujejezični elementi

- c. Enako naredimo za besedila, označena z L2 in L3.
- d. Razmislimo o razlogih za take rezultate. Pobrskajte po konkordancah za posamezne oznake nestandardnosti in preverite, v kakšnem sobesedilu se kje pojavljajo tujejezični elementi.

V. Stopnja (ne)standardnosti in tujejezični elementi

2. Preverimo lahko tudi, kdo uporablja najbolj nestandarden jezik in največ tujejezičnih elementov.
 - a. Preverimo, kdo uporablja več tujejezičnih elementov v najbolj nestandardnih tvitih, moški ali ženske.
 - Izberemo podkorpus TWEET. Pri Lastnostih besedil izberemo L3.
 - Pri Lastnostih besedil izberemo spol avtorjev (enkrat ženski, enkrat moški).
 - V iskalno polje CQL napišemo iskalni ukaz za tujejezične elemente: [tag="Nj"].
 - Preverimo tujejezične elemente pri moških/ženskah.

V. Stopnja (ne)standardnosti in tujejezični elementi

- b. Poiščite tri avtorje moškega in tri ženskega spola, ki uporabljajo najbolj nestandarden jezik in hkrati največ tujejezičnih elementov.
 - Ko po zgoraj navedenem postopku dobimo konkordance, zgradimo frekvenčni seznam, pri čemer v oknu Frekvenčna razporeditev po lastnostih besedil izberemo »text.author«.