

4.4 Detektiv – delavnica Odkrivanje govornih elementov na Twitterju

Izdelaj seznam ključnih besednih oblik korpusa Twitter glede na korpus Kres.

Na spletni strani http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04 izberi možnost Sezname in izberi korpus Janes v 0.4 Tweet. Pri Možnostih filtriranja izberi Izloči besede in izberi datoteko tabor-tuje.txt, ki jo dobiš na spletni strani <http://nl.ijs.si/janes/dogodki/tabor-2016/>. Pri Možnostih izpisa izberi opcijo Ključne besede, za referenčni korpus pa izberi korpus Kres.

1. Ali na tem seznamu najdemo besede *eee*, *mhm*, *eem*? Zakaj?
2. Oglej si prvih 20 besed na seznamu. Kaj opaziš?
3. Na seznamu poišči besede *jst*, *dons*, *jujr*. Po čem so si te besede podobne? Izpiši še tri take besede. Zdaj te besede poišči v korpusu Gos. Kaj lahko sklepaš o profilu avtorja besedila, ki jih uporablja?
4. Na seznamu poišči besede *fora*, *valda* in *ful*. Kaj je njihova skupna lastnost? Izpiši še tri take besede. Zdaj te besede poišči v korpusu Gos. V katerem tipu diskurza prevladujejo? Ali bi te besede kdaj zapisal/a v šolskem spisu? Odgovor pokomentiraj.
5. Na seznamu ključnih besed poišči besede *videoposnetek*, *twit* in *btw*. Jih najdeš v korpusu Gos? Zakaj?
6. Seznam ključnih besed korpusa Gos in podkorpusa tвитov predstavi v besednem oblaku.

Vsaka skupina v okviru izbrane tematike pripravi predstavitev z naslednjo strukturo:

- I. Raziskovalno vprašanje: v kolikšni meri drži, da “na Twitterju pišemo, kot govorimo”?
- II. Potek raziskave: utemeljitev izbora vsebinskega sklopa, izbor gradiva oz. korpusov in orodja za njihovo analizo in opis metodologije.
- III. Raziskava in beleženje ugotovitev. Predstavitev besednega oblaka in ugotovitev na podlagi točk 1, 2, 3 in 4. Ali neformalne besede in izgovoru podoben zapis odsevata jezikovno neznanje avtorjev besedila? Zakaj? Kaj na podlagi podatkov iz korpusov lahko povemo o spolu, starosti ali regiji avtorja besedila?
- IV. Zaključki glede na zastavljeno raziskovalno vprašanje.