Primerjava slovenskih besednih vektorskih vložitev z vidika spola na analogijah poklicev

Anka Supej*, Matej Ulčar†, Marko Robnik-Šikonja†, Senja Pollak*

*Institut "Jožef Stefan"
Jamova cesta 39, 1000 Ljubljana
a.supej@gmail.com, senja.pollak@ijs.si

†Univerza v Ljubljani, Fakulteta za računalništvo in informatiko
Večna pot 113, 1000 Ljubljana
{matej.ulcar, marko.robnik}@fri.uni-lj.si

Povzetek

V zadnjih letih je uporaba globokih nevronskih mrež in gostih vektorskih vložitev za predstavitve besedil privedla do vrste odličnih rezultatov na področju računalniškega razumevanja naravnega jezika. Prav tako se je pokazalo, da vektorske vložitve besed pogosto zajemajo pristranosti z vidika spola, rase ipd. Ena izmed metod za ocenjevanje kakovosti vložitev so izračuni analogij. Prispevek se osredotoča na evalvacijo vektorskih vložitev besed v slovenščini z vidika spola. Sestavili smo seznam moških in ženskih ustreznic poklicev (dostopen prek repozitorija CLARIN) in preko analogij ovrednotili spolno pristranost modelov vložitev fastText, word2vec in ELMo z različnimi konfiguracijami in pristopi k računanju analogij.

Abstract

In recent years, the use of deep neural networks and dense vector embeddings for text representation have led to excellent results in the field of computational understanding of natural language. It has also been shown that word embeddings often capture gender, racial and other types of bias. One of the methods for assessing the quality of word embeddings are analogies calculations. The article focuses on the evaluation of Slovene word embeddings in terms of gender. We compiled a list of male and female equivalents of occupations and evaluated the gender bias of fastText, word2vec and ELMo embeddings.

1. Uvod

Raziskave na stičišču spola in jezika so metodološko različne. Sociolingvistične študije poročajo o načinih, po katerih se uporaba jezika med ženskami in moškimi razlikuje (npr. širše besedišče, milejše izražanje, uporaba tipičnih slovničnih struktur pri ženskah) (Lakoff, 1973; Tannen, 1990; Argamon et al., 2003). Opažanja, da se uporaba jezika med spoloma razlikuje, so navdihnila študije profiliranja avtorjev na besedilih različnih jezikov in tipov besedil (Koolen in van Cranenburgh, 2017; Pardo et al., 2015; Martinc et al., 2017), tudi za slovenščino ((Verhoeven et al., 2017; Škrjanec et al., 2018).

Dimenzija spola v korpusih ni prisotna le kot jezikovna variacija, temveč tudi v obliki večplastne pristranosti, tako v posameznih besedilih kot tudi v večjih korpusih. Sorodne raziskave ugotavljajo:

- da se pristranost odraža kot pomanjkanje omemb žensk: korpusi, ki se pogosto uporabljajo v raziskavah, vsebujejo znatno manj zaimkov ženskega spola (Zhao et al., 2018) ali drugih nanašalnic na ženske (Caldas-Coulhard in Moon, 2010; Baker, 2010);
- da so ženske manj pogosto avtorice ali urednice (Hill in Shaw (2013): le 16% urednic Wikipedije je žensk);
- da korpusi zajemajo spolno stereotipne kolokacije (Pearce, 2008), ki npr. predstavljajo ženske predvsem skozi reproduktivno funkcijo (Gorjanc, 2007) in jih ne povezujejo z (družbeno) močjo (Baker, 2010).

V zadnjih letih je razmah na področju obdelave naravnega jezika povezan predvsem z uporabo globokih nevronskih mrež, ki se uporabljajo tudi za učenje predstavitev besedil v obliki gostih vektorskih vložitev besed. Izkaže se, da tudi vektorske vložitve besed pogosto zajemajo pristranosti z vidika spola, rase ipd. Pristranost se v besednih vložitvah kaže preko semantičnih asociacij in posledične bližine v vektorskem prostoru (Mikolov et al., 2013b). Računsko jo lahko ovrednotimo npr. s kosinusno podobnostjo med vložitvami, ki opisujejo nek širši pojem (npr. spol), ter stereotipnimi koncepti (npr. v Caliskan et al. (2017): asociacija žensk in umetnosti ter moških in znanosti) ali preko izračuna analogij (Bolukbasi et al., 2016), ki predpostavljajo odnos: $\overrightarrow{moški} - \overrightarrow{poklic_M} \approx \overrightarrow{ženska} - \overrightarrow{poklic_Ž}$. Poleg študij, ki so pokazale na pristranost samih vložitev, so različni avtorji pokazali tudi prenos pristranosti v algoritme za različne naloge obdelave naravnega jezika, od strojnega prevajanja (Prates et al., 2020; Vanmassenhove et al., 2018) do študij sentimenta (Kiritchenko in Mohammad, 2018)). Na drugi strani pa nekateri avtorji (Nissim et al., 2019) opozarjajo tudi na pretirano poudarjanje pristranosti pri zasnovi raziskav z analogijami.

Glavni doprinos prispevka je evalvacija slovenskih modelov besednih vektorskih vložitev z vidika spola, ki še ni dovolj raziskano (izjema je npr. analiza slovenskega modela w2v v Supej et al. (2019)). Prispevek se osredotoča na kvantitativno evalvacijo in primerjavo širokega nabora izbora slovenskih vložitev ter različnih pristopov k evalvaciji, s čimer nagovarja predvsem razvijalce jezikovnotehnoloških orodij, ki vložitve uporabljajo. Kljub temu da s tem indirektno problematiziramo pristranost v jeziku ter pokažemo tudi na nekaj stereotipnih povezav, pa je podrobnejša kritična interpretacija izven fokusa tega prispevka.

V prispevku najprej predstavimo sorodna dela (2. razdelek). V 3. razdelku opišemo seznam moških in ženskih poklicev ter uporabljenih besednih vektorskih vložitev. V 4. in 5. razdelku predstavimo metodologijo in rezultate, nato zaključimo z diskusijo in načrti za nadaljnje delo.

2. Sorodna dela

Korpusi odražajo jezikovne variacije (vključno z različnimi vrstami pristranosti) v odnosu do družbenih dejavnikov. Orodja procesiranja naravnega jezika, ki se učijo na korpusih, lahko variacije in pristranosti podedujejo: nekatere študije prikažejo, da so orodja procesiranja naravnega jezika uspešnejša, ko tovrstne variacije upoštevajo (Volkova et al., 2013; Hovy, 2015). Študija Hovy (2015) pokaže, da vključitev informacij o starosti in spolu avtorjev izboljša uspešnost treh nalog v petih različnih jezikih. Pristranost v korpusih ima lahko tudi negativne posledice, kar lahko podkrepimo z nekaj primeri.

Pogosto uporabljeni korpusi vsebujejo pristranosti do te mere, da so orodja procesiranja naravnega jezika uspešnejša pri vhodnih podatkih, kjer je besedilo napisala starejša oseba (Hovy in Søgaard, 2015). Študija Garimella et al. (2019) pokaže, da sta oblikoslovni označevalnik in skladenjski razčlenjevalnik uspešna na tekstih, ki so jih napisale ženske, ne glede na to, na katerih podatkih sta bila naučena. Teksti moških avtorjev so bolje razčlenjeni/označeni, v kolikor je v učnih podatkih na voljo dovolj besedil, ki so jih napisali moški. Uspešnost orodij, kot so razčlenjevalniki, za tekste moških avtorjev je torej lahko posledica neuravnoteženosti množice učnih podatkov v prid moškemu avtorstvu. Pristranost v korpusih ima poleg negativnega vpliva na orodja obdelave naravnega jezika (Sun et al., 2019) tudi druge negativne posledice, kot je npr. nepravilna razrešitev koreferenc (Zhao et al., 2018).

Vektorske vložitve besed, prav tako naučene na korpusih, poleg sintaktičnih značilnosti besed ujamejo tudi njihove semantične relacije. To se izraža v geometriji prostora vektorskih vložitev: semantično povezane vložitve so si v vektorskem prostoru bližje in razporejene v podobnih smereh. Zato je z njimi mogoče računati tudi odnose, ki presegajo enostavno sorodnost besed, npr. preko analogij. Npr. odnos *Madrid:Španija* je podoben odnosu *Pariz:Francija* (Mikolov et al., 2013b). Vektorske vložitve besed so naučene na korpusih z različnimi algoritmi in tako kot korpusi vsebujejo pristranosti. Beseda, ki je npr. stereotipno povezana z določenim spolom, bo v vektorskem prostoru tako bližje vektorski vložitvi besed *ženska* ali *moški* (Garg et al. (2018) npr. pokažejo, da je pridevnik *časten* v angleščini bližje besedi *moški* kot besedi *ženska*), pristranosti pa se kažejo tudi preko stereotipnih rešitev analogij (npr. Bolukbasi et al. (2016): rešitev analogije *moški:programer::ženska:x* je *gospodinja*). Nissim et al. (2019) opozarjajo, da tovrstne raziskave pretirano poudarjajo pristranost.

Ker se lahko pristranost z uporabo orodij, ki uporabljajo vložitve, ojača (Zhao et al., 2017), se več raziskovalnih skupin ukvarja z metodami "razpristranjevanja" (angl. *debiasing*) vektorskih vložitev. Primeri teh postopkov so izenačevanja oddaljenosti med spolno zaznamovanimi besedami in poklici (Bolukbasi et al., 2016; Bordia in Bowman, 2019), vstavljanje dodatnih omejitev v učni korpus (npr. zagotavljanje enake razporeditve poklicnih aktivnosti med spoloma v učnih podatkih) (Zhao et al., 2017), odstranjevanja tekstov, ki povzročajo pristranost (Brunet et al., 2018), in učenje spolno nevtralnih vektorskih vložitev (Zhao et al., 2018). Gonen in Goldberg (2019) opozarjata, da mnogi postopki "razpristranjevanja" pristranost le zakrijejo, medtem ko ta dejansko ostane prisotna v vložitvah.

Študije na področju raziskovanja pristranosti v vektorskih vložitvah besed so pogosto zasnovane na analogijah poklicev. Ker je v slovenščini spol besed morfološko izražen, kot rezultat analogije pričakujemo žensko oz. moško obliko poklica. Predhodna raziskava na besednih vložitvah (word2vec) v slovenščini (Supej et al., 2019) je pokazala, da je natančnost iskanja analogij dokaj visoka tako pri iskanju moškega kot tudi ženskega poklica. Rezultati kljub temu odražajo spolne pristranosti: rezultat analogije *ženska:tajnica :: moški:x* da rezultat $x =$ *šef*, prvih 10 najbližjih rezultatov različnih analogij pa odseva več spolnih neenakosti: asociacija žensk z domačimi opravili, moških s poklici višjega statusa itd. V nasprotju s predhodno raziskavo se naš prispevek ne ukvarja s sociološko problematizacijo rezultatov analogij enega tipa besednih vložitev (tj. word2vec), temveč preko analogij poklicev ovrednoti različne modele vložitev, njihove konfiguracije in morebiten vpliv filtriranja podatkov na rezultate. V prispevku torej razširimo študijo (Supej et al., 2019) z obširno analizo razpoložljivih modelov slovenskih vektorskih vložitev besed na razširjenem seznamu poklicev.

3. Podatki

V tem razdelku predstavimo sestavljen seznam poklicev ter opišemo različne vektorske vložitve besed.

3.1. Seznam poklicev

Naš izbor poklicev temelji na standardni klasifikaciji poklicev (Vlada RS, 1997), katere osnova je *Mednarodna standardna klasifikacija poklicev*. Večina poklicev v klasifikaciji je večbesednih zvez (npr. *upravljalec/upravljalka metalurškega žerjava*), ki so zaradi svoje specifičnosti in obsežnosti manj primerne za računske naloge. Za potrebe izračuna analogij smo se omejili na enobesedne poklice. Celotni seznam enobesednih poklicev zajema 422 parov, ki jih omejimo še glede na naslednje kriterije:

(1) Poklic ima žensko in moško obliko (spolno nevtralne besede, npr. *pismonoša*, niso vključene). (2) Vsaj ena izmed oblik poklica se pojavi v Slovenskem oblikoslovnem leksikonu *Sloleks 2.0* (Dobrovoljc et al., 2019), ki omogoča razlikovanje med lastnimi in občnimi imeni (nekateri poklici so namreč tudi lastna imena; npr. *kovač*), ali pa se v referenčnem korpusu standardne slovenščine Gigafida 2.0 (2020) pojavi 500- ali večkrat. (3) V primerih, kjer ocenimo, da za poimenovanje v standardni klasifikaciji poklicev obstajajo bolj uveljavljene različice, naboru podatkov dodamo sopomenko z istim korenom (npr. za izraz *fotografka*) iz standardne klasifikacije dodamo sopomenko *fotografinja*). Pri izračunih za izhodiščne besede upoštevamo obliko, ki je v korpusih bolj pogosta, pri pravilnosti izračunanih analogij pa upoštevamo katerokoli različico. (4) Če standardna klasifikacija ne vključuje

ženske (npr. *dramatik*) ali moške oblike poklica (npr. *prostitutka*), smo ročno dodali ustrezno različico, v kolikor ta obstaja (npr. za *postreščka* in za *hosteso* uveljavljene ženske oz. moške oblike ni) in je prisotna v Gigafidi. (5) Iz nabora smo izključili poklice, kjer je ženska oz. moška varianta poklica homofon (npr. *strežnik*, *detektivka*), oz. kjer je poklic možno asociirati s poklici nepovezanim kontekstom (npr. *čarovnik/čarovnica*).

Končni seznam vsebuje 234 parov poklicev, ki bo prosto dostopen na repozitoriju CLARIN [1].

3.2. Modeli vektorskih vložitev

V eksperimentih smo uporabili več različnih konfiguracij znanih vektorskih vložitev:

- fastText (Bojanowski et al., 2016):
  - 100-dimenzionalni vektorji, naučeni tekom projekta EMBEDDIA[2] na Gigafidi 2.0,
  - 300-dimenzionalni vektorji, naučeni kot v prejšnjem primeru,
  - 100-dimenzionalni vektorji besed s portala Sketch Engine (word),
  - 100-dimenzionalni vektorji s portala Sketch Engine, kjer so vektorji vložitve lem (lemma),
  - 100-dimenzionalni vektorji CLARIN.SI-embed.sl (Ljubešić in Erjavec, 2018) in
  - 300-dimenzionalni vektorji s portala fasttext.cc;
- word2vec (Mikolov et al., 2013a): 256-dimenzionalni vektorji, ki so bili naučeni za potrebe portala Kontekst.io (Plahuta, 2020) in so na voljo po dogovoru[3];
- ELMo (Peters et al., 2018): 1024-dimenzionalni vektorji kontekstnih vložitev projekta *EMBEDDIA*, naučeni na Gigafidi (Ulčar, 2019), kjer so vzete povprečne vrednosti 200.000 najpogostejših besed (izračunano na podlagi slovenske Wikipedije). Uporabljenih je bilo več različnih vrst vektorjev:
  - vektorji z izhoda prvega (CNN) nivoja mreže, ki so kontekstno neodvisni (tj. *layer 0*),
  - vektorji z izhoda drugega (prvega LSTM) nivoja mreže, ki so kontekstno odvisni (tj. *layer 1*),
  - vektorji z izhoda tretjega (drugega LSTM) nivoja mreže, ki so kontekstno odvisni (tj. *layer 2*).

4. Metodologija evalvacije

Analogije poklicev smo za vsako izmed vložitev izračunali na štiri različne načine. Jedro pristopa je pri vseh načinih enako: za vsako moško obliko poklica (P_m) iščemo ustrezno žensko obliko (P_f). Izračunamo vektor

$$\overrightarrow{d} = \overrightarrow{P_m} - \overrightarrow{m} + \overrightarrow{f},$$

kjer je $\overrightarrow{m}$ moški vektor in $\overrightarrow{f}$ ženski vektor. V idealnem primeru bi bil vektor $\overrightarrow{d}$ enak $\overrightarrow{P_f}$. Vektorju $\overrightarrow{d}$ poiščemo vektorje N najbližjih besed glede na kosinusno razdaljo.

[1] http://hdl.handle.net/11356/1347
[2] http://embeddia.eu/
[3] https://kontekst.io/partnerstvo

m	f	m	f
moški	ženska	brat	sestra
gospod	gospa	oče	mati
fant	dekle	sin	hči
fant	punca	dedek	babica
deček	deklica	mož	žena
stric	teta	on	ona

Tabela 1: Pari inherentno moških in ženskih besed.

Pri iskanju najbližjih besed smo upoštevali vse besede, ki se nahajajo v vložitvah, razen besed *moški*, *ženska*, besede P_m ter vseh besed, ki vsebujejo nečrkovne simbole (številke, vezaje, druga ločila, itd.) Če se beseda P_f nahaja med N najbližjimi besedami, ta primer štejemo kot pravilno določenega, sicer kot napačnega. Pri tem smo ignorirali velike in male začetnice, na primer besede *Zdravnik*, *zdravnik* in *ZDRAVNIK* upoštevamo kot isto besedo.

Postopek ponovimo za vsako žensko obliko poklica (P_f), kjer iščemo ustrezno moško obliko (P_m). Vektor $\overrightarrow{d}$ v tem primeru izračunamo kot

$$\overrightarrow{d} = \overrightarrow{P_f} - \overrightarrow{f} + \overrightarrow{m},$$

pri iskanju najbližjih besed pa namesto besede P_m izpustimo besedo P_f. Končni rezultat predstavlja delež pravilno določenih primerov, oz. mera *natančnost pri* N (angl. *precision at* N oz. $P@N$).Višji N omogoča primerjavo v širši okolici zadetka v vektorskem prostoru.

Za določitev moškega vektorja $\overrightarrow{m}$ in ženskega vektorja $\overrightarrow{f}$ smo uporabili dva pristopa. V prvem je m kar beseda *moški* in f beseda *ženska*. V drugem pristopu razliko $\overrightarrow{f} - \overrightarrow{m}$, oz. $\overrightarrow{m} - \overrightarrow{f}$ podobno kot Bolukbasi et al. (2016) predstavimo s povprečno razliko vektorjev parov besed, ki se specifično nanašajo na žensko oz. moškega (Tabela 1).

Pri iskanju najbližjih N besed smo uporabili tudi alternativen pristop, kjer smo vse besede v vložitvah lematizirali z orodjem LemmaGen[4]. S tem smo izničili vpliv pregibanja besed; na primer, besedi *zdravnico* in *zdravnice* sta pri tem postopku enaki, saj imata isto lemo *zdravnica*.

5. Rezultati

Rezultate predstavimo za vsak pristop, opisan v 4. razdelku, z mero natančnost pri N, kjer je N enak 1, 5 in 10. Nekaterih poklicev z našega seznama ni v vseh vložitvah. Če iskane besede ni med N najbližjimi, je primer označen kot napačen, četudi te besede sploh ni med vložitvami. Primere, ko poklica, ki ga imamo na vhodu, ni med vložitvami in tako ne moremo izračunati vektorja $\overrightarrow{d}$, obravnavamo na dva načina. V prvem načinu (`all`) tak primer štejemo kot napačen, v drugem načinu (`covered`) pa ga izločimo iz primerov in na končni rezultat ne vpliva.

Rezultati, kjer imamo na vhodu moški poklic P_m in iščemo ustrezni ženski poklic P_f so v Tabeli 2. Rezultati, kjer za ženski poklic P_f na vhodu iščemo moški poklic P_m so v Tabeli 3. Pristop, pri katerem smo vse besede lematizirali, ima pripono `_lem`. Pristop, kjer smo za moški in ženski vektor oz. njuno razliko uporabili povprečne razlike vektorjev besed iz tabele 1, ima pripono `_avg`.

[4] https://github.com/vpodpecan/lemmagen3/

Vložitve	št. dimenzij in pristop	all P@1	all P@5	all P@10	covered P@1	covered P@5	covered P@10
ELMo Embeddia	1024D_l0_avg	0.128	0.278	0.299	0.166	0.359	0.387
	1024D_l0	0.162	0.291	0.308	0.210	0.376	0.398
	1024D_l0_lem_avg	0.286	0.308	0.312	0.370	0.398	0.403
	1024D_l0_lem	0.291	0.303	0.312	0.376	0.392	0.403
	1024D_l1_avg	0.295	0.303	0.308	0.381	0.392	0.398
	1024D_l1	0.291	0.303	0.303	0.376	0.392	0.392
	1024D_l1_lem_avg	0.295	0.303	0.308	0.381	0.392	0.398
	1024D_l1_lem	0.291	0.303	0.303	0.376	0.392	0.392
	1024D_l2_avg	0.291	0.308	0.308	0.376	0.398	0.398
	1024D_l2	0.286	0.308	0.308	0.370	0.398	0.398
	1024D_l2_lem_avg	0.291	0.308	0.308	0.376	0.398	0.398
	1024D_l2_lem	0.286	0.308	0.308	0.370	0.398	0.398
fastText.cc	300D_avg	0.594	0.722	0.735	0.607	0.738	0.751
	300D	0.436	0.688	0.718	0.445	0.703	0.734
	300D_lem_avg	0.641	0.739	0.748	0.655	0.755	0.764
	300D_lem	0.487	0.709	0.735	0.498	0.725	0.751
fastText Embeddia	100D_avg	0.667	0.709	0.714	0.672	0.716	0.720
	100D	0.632	0.709	0.714	0.638	0.716	0.720
	100D_lem_avg	0.671	0.714	0.718	0.677	0.720	0.724
	100D_lem	0.632	0.709	0.714	0.638	0.716	0.720
	300D_avg	0.662	0.709	0.718	0.668	0.716	0.724
	300D	0.679	0.714	0.714	0.685	0.720	0.720
	300D_lem_avg	0.679	0.714	0.718	0.685	0.720	0.724
	300D_lem	0.679	0.714	0.714	0.685	0.720	0.720
fastText CLARIN.SI-embed.sl	100D_avg	**0.761**	0.868	0.880	**0.761**	0.868	0.880
	100D	0.705	0.855	0.885	0.705	0.855	0.885
	100D_lem_avg	**0.761**	**0.880**	**0.902**	**0.761**	**0.880**	**0.902**
	100D_lem	0.709	0.859	0.885	0.709	0.859	0.885
fastText Sketch Engine (word)	100D_avg	0.714	0.765	0.774	0.717	0.768	0.777
	100D	0.688	0.765	0.774	0.691	0.768	0.777
	100D_lem_avg	0.722	0.778	0.782	0.725	0.781	0.785
	100D_lem	0.688	0.765	0.778	0.691	0.768	0.781
fastText Sketch Engine (lemma)	100D_avg	0.598	0.786	0.821	0.598	0.786	0.821
	100D	0.380	0.658	0.756	0.380	0.658	0.756
word2vec Kontekst.io	256D_avg	0.402	0.543	0.585	0.407	0.550	0.593
	256D	0.248	0.483	0.509	0.251	0.489	0.515
	256D_lem_avg	0.402	0.543	0.585	0.407	0.550	0.593
	256D_lem	0.248	0.483	0.513	0.251	0.489	0.519

Tabela 2: Rezultati za vse vložitve in variante, kjer je na vhodu moški poklic in iščemo ustrezen ženski poklic. Če za moški poklic na vhodu ne najdemo vložitve, tak primer štejemo kot napačno ugotovljen (all), oz. ga izpustimo iz rezultatov (covered). Najboljši rezultati v vsakem stolpcu so odebeljeni.

Rezultati kažejo, da dobimo boljše rezultate s fastText vložitvami, z izračunom, kjer namesto samega vektorja *moški* oz. *ženska* uporabimo povprečje besed z inherentno izraženim spolom ter z lematizacijo. Rezultate podrobneje razčlenimo v naslednji sekciji.

6. Diskusija

Vložitve, ki dosegajo največjo natančnost pri ugotavljanju analogij (z vhodnim moškim poklicem), so vložitve fastText CLARIN.SI-embed.sl (Tabela 2). Pri vhodnem ženskem poklicu dosegajo največjo natančnost, če upoštevamo le vložitve poklicev, ki so prisotni, fastText Embeddia, medtem ko so na vzorcu vseh vložitev najnatančnejše vložitve fastText CLARIN.SI-embed.sl (Tabela 3). V različnih modelih vložitev, pri različnih vhodnih podatkih velja, da lematizacija izhodnih podatkov in hkrati uporaba vektorja povprečne razlike med ženskimi in moškimi besedami (namesto uporabe le besed *ženska* oz. *moški*) izboljša natančnost analogije. Modeli, kjer je na vhodu poklic ženskega spola, v povprečju dosegajo višjo natančnost analogij v primeru `covered` rezultatov (če ženskega poklica ni v med vložitvami, tega ne štejemo kot napačno ugotovljeno analogijo). Rezultati `all` so podobni pri obeh tipih vhodnih podatkov.

Vložitve fastText Embeddia dosegajo zelo podobne rezultate s 100- in 300-dimenzionalnimi vložitvami, (glej Tabeli 2 in 3). (Druge vložitve so bile naučene na drugih jezikovnih virih, zato niso neposredno primerljive.) Vendar pa je iz Tabele 5 (vložitve fastText Embeddia) tudi razvidno, da igra dimenzionalnost veliko vlogo pri tem, kako pogosto je rezultat analogije sam vhodni poklic. Dimenzionalnost bi torej imela velik vpliv na nefiltrirane rezultate.

V vseh modelih vložitev je delež poklicev moškega spola večji kot delež poklicev ženskega spola (Tabela 4).

Vložitve	št. dimenzij in pristop	all P@1	all P@5	all P@10	covered P@1	covered P@5	covered P@10
ELMo Embeddia	1024D_l0_avg	0.226	0.299	0.303	0.707	0.933	0.947
	1024D_l0	0.137	0.295	0.303	0.427	0.920	0.947
	1024D_l0_lem_avg	0.291	0.299	0.303	0.907	0.933	0.947
	1024D_l0_lem	0.286	0.303	0.303	0.893	0.947	0.947
	1024D_l1_avg	0.291	0.303	0.303	0.907	0.947	0.947
	1024D_l1	0.282	0.303	0.303	0.880	0.947	0.947
	1024D_l1_lem_avg	0.291	0.303	0.303	0.907	0.947	0.947
	1024D_l1_lem	0.291	0.303	0.303	0.907	0.947	0.947
	1024D_l2_avg	0.282	0.299	0.299	0.880	0.933	0.933
	1024D_l2	0.274	0.295	0.299	0.853	0.920	0.933
	1024D_l2_lem_avg	0.282	0.299	0.299	0.880	0.933	0.933
	1024D_l2_lem	0.274	0.295	0.299	0.853	0.920	0.933
fastText.cc	300D_avg	0.291	0.590	0.675	0.393	0.798	0.913
	300D	0.111	0.415	0.585	0.150	0.561	0.792
	300D_lem_avg	0.453	0.654	0.701	0.613	0.884	0.948
	300D_lem	0.338	0.637	0.679	0.457	0.861	0.919
fastText Embeddia	100D_avg	0.654	0.705	0.709	0.900	0.971	0.976
	100D	0.342	0.632	0.658	0.471	0.871	0.906
	100D_lem_avg	0.658	0.705	0.709	0.906	0.971	0.976
	100D_lem	0.534	0.671	0.684	0.735	0.924	0.941
	300D_avg	0.607	0.705	0.709	0.835	0.971	0.976
	300D	0.239	0.624	0.697	0.329	0.859	0.959
	300D_lem_avg	0.688	0.709	0.714	**0.947**	**0.976**	**0.982**
	300D_lem	0.594	0.705	0.709	0.818	0.971	0.976
fastText CLARIN.SI-embed.sl	100D_avg	0.731	0.850	0.876	0.784	0.913	0.940
	100D	0.077	0.547	0.726	0.083	0.587	0.780
	100D_lem_avg	**0.782**	**0.876**	**0.885**	0.839	0.940	0.950
	100D_lem	0.607	0.821	0.855	0.651	0.881	0.917
fastText Sketch Engine (word)	100D_avg	0.701	0.761	0.769	0.886	0.962	0.973
	100D	0.167	0.598	0.718	0.211	0.757	0.908
	100D_lem_avg	0.735	0.761	0.769	0.930	0.962	0.973
	100D_lem	0.641	0.752	0.761	0.811	0.951	0.962
fastText Sketch Engine (lemma)	100D_avg	0.581	0.803	0.829	0.673	0.931	0.960
	100D	0.440	0.701	0.769	0.510	0.812	0.891
word2vec Kontekst.io	256D_avg	0.453	0.568	0.581	0.679	0.853	0.872
	256D	0.244	0.393	0.479	0.365	0.590	0.718
	256D_lem_avg	0.453	0.568	0.581	0.679	0.853	0.872
	256D_lem	0.342	0.457	0.530	0.513	0.686	0.795

Tabela 3: Rezultati za vse vložitve in variante, kjer je na vhodu ženski poklic in iščemo ustrezen moški poklic. Če za ženski poklic na vhodu ne najdemo vložitve, tak primer štejemo kot napačno ugotovljen (all), oz. ga izpustimo iz rezultatov (covered). Najboljši rezultati v vsakem stolpcu so odebeljeni.

Največja pokritost je v vložitvah fastText CLARIN.SI-embed.sl, pri vložitvah modela ELMo pa se na primer pojavi le 75 od 234 izbranih poklicev ženskega spola. Razlog za mnogo manjšo zastopanost poklicev pri modelu ELMo je, da smo se zaradi tehničnih razlogov omejili le na 200 tisoč najpogostejših besed v Wikipediji (ELMo vložitve so v osnovi kontekstualne vložitve in je proces povprečenja računsko zahteven). Pri drugih tehnologijah vložitev smo imeli približno milijon besed. Moški poklici, ki se ne pojavljajo v vložitvah, so običajno poklici, ki so tipično povezani z ženskim spolom (npr. *šiviljec* ali *kozmetik*). Tudi poklici ženskega spola, ki se ne pojavljajo v vložitvah, so npr. tradicionalno povezani z moškimi (v vložitvah različnih tipov na primer ni *avtomehaničarke, tesarke* itd.) ali pa gre za kulturno pogojene izključno moške poklice (npr. *nadškof*). Slabo zastopanost poklicev ženskega spola lahko povežemo tudi z drugimi faktorji – Zhao et al. (2018) poročajo, da se različne zvrsti tekstov pogosteje nanašajo na moške v okviru njihovega poklica kot pa je to pri ženskah.

Modeli vložitev v konfiguraciji `lem_avg` (uporaba vektorja povprečnih razlik med spolno zaznamovanimi besedami in lematiziranje izhodnih podatkov) dajejo zelo različne rezultate. Rezultati analogij pri modelih ELMo in word2vec so večinoma poklici. Pri vložitvah fastText Embeddia, CLARIN.SI-embed.sl in Sketch Engine (word) so rezultati poklici in ostale besede, sorodne vhodnemu poklicu, ter besede z istim korenom kot vhodni poklic. Rezultati modelov fastText.cc in Sketch Engine (lemma) so večinoma besede z istim korenom kot vhodni poklic.

Po mnenju Nissim et al. (2019) je interpretacija večine študij, ki povezujejo analogije s pristranostjo, pretirana. Računanje analogij je namreč zastavljeno tako, da se izključi vhodni poklic, četudi bi bil to dejanski rezultat z

najvišjo kosinusno podobnostjo. Kljub temu, da smo pri rezultatih izločili vhodne poklice, kar je za samo računanje analogij standarden postopek, smo analizirali tudi rezultate pred filtriranjem. Pri analizi teh rezultatov smo opazili, da je med rezultati analogije z najvišjo kosinusno podobnostjo pogosto sam vhodni poklic (Tabela 5), kar pa zelo variira med posameznimi modeli.

Rezultati analogij so zanimivi z vidika semantike. Prva rezultata analogij (vložitev fT Embeddia 100D_lem_avg) *ženska:krojačica :: moški:x krojač* in *ženska:šivilja :: moški:x krojač* ponazarjata, da besedne vektorske vložitve upoštevajo tako slovnične kot tudi semantične elemente (vektorske vložitve besede *šiviljec* ni, *krojač* pa je semantično povezana beseda). Rešitve nekaterih analogij (predvsem v modelu w2v Kontekst.io lem_avg) z vhodnim poklicem niso povezane ali so stereotipne. Npr., rešitve analogije *moški:rudar :: ženska:x* v modelu w2v Kontekst.io lem_avg so npr.: *barbika, klovnesa, čarovnica, lutka, prostitutka, akrobatka, najstnica, opica, princeska, striptizeta*. Na stereotipne analogije v modelu w2v opozorijo tudi v Supej et al. (2019).

V okviru analize smo naredili tudi skupni frekvenčni seznam rezultatov analogij za vse ženske oz. moške vhodne poklice za posamezen model vložitev (upoštevajoč le konfiguracije `lem_avg`) (cf. Tabela 6). Opazimo vzorec, da se pri modelih ELMo l2_lem_avg in w2v Kontekst.io lem_avg najbolj pogosti ženski poklici/besede pojavljajo pogosteje kot najpogostejši moški poklici. Ena od možnih interpretacij je, da izbrana modela v primerjavi z nekaterimi drugimi vsebujeta relativno manj vektorskih besednih vložitev (200.000 oz. približno 600.000 za posamezen model). Oba modela imata tudi manjšo zastopanost ženskih oblik poklicev med besednimi vložitvami. Poklici, ki se kljub temu pojavljajo med vložitvami, se zato ponovijo večkrat. Poklicev moškega spola je v besednih vložitvah več, zato se posamezni poklici ne pojavljajo tako pogosto.

Med pogostimi ženskimi analogijami pri modelih ELMo l2_lem_avg in w2v Kontekst.io lem_avg zaznamo poklice nižjega družbenega statusa (*čistilka, perica, gospodinja*) ter zastarele poklice, kjer je bila ženska v podrejenem položaju (*služkinja*). Pri najpogostejših moških analogijah so poklici nižjega družbenega statusa izjemno redki.

Ugotavljamo tudi, da se nekatere besede (predvsem poklici ženskega spola) v rezultatih pojavljajo ne glede na semantično povezanost z vhodnim poklicem. V več primerih je rešitev analogije (predvsem ko gre za vhodni tipično moški poklic) nepovezana z vhodnim poklicem (npr. *bolničarka* kot prva rešitev analogije *moški:rudar :: ženska:x* in *šivilja* kot prva rešitev analogije *moški:avtomehanik :: ženska:x* v modelu fT Embeddia 100D_lem_avg). Možna razlaga, za potrditev katere bi bili potrebni dodatni testi, je, da so nekatere vektorske vložitve besed bolj 'centralne' od drugih in so najbližji sosed velikemu številu drugih besed, kar je v vektorskih vložitvah mogoč pojav. Možnost za nadaljnje delo je (delno) zmanjšati vpliv tovrstnih vložitev s pomočjo mere, alternativne kosinusni podobnosti, tj. CSLS (Conneau et al., 2018) oz. podobne mere, ki upošteva medsebojne razdalje najbližjih *n* sosedov).

vložitve	m	f
ELMo	0.774	0.321
fastText_cc	0.979	0.739
fastText Embeddia	0.991	0.726
fastText CLARIN.SI-embedd.sl	**1.000**	**0.932**
fastText Sketch Engine (word)	0.996	0.791
fastText Sketch Engine (lemma)	1.000	0.863
word2vec Kontekst.io	0.987	0.667

Tabela 4: Delež poklicev moškega (m) in ženskega (f) spola v vložitvah.

Vložitve	št. dim. in pristop	delež
ELMo Embeddia	1024D_l0_avg	0.547
	1024D_l0	0.547
	1024D_l0_lem_avg	0.547
	1024D_l0_lem	0.547
	1024D_l1_avg	0.423
	1024D_l1	0.483
	1024D_l1_lem_avg	0.423
	1024D_l1_lem	0.483
	1024D_l2_avg	0.064
	1024D_l2	0.088
	1024D_l2_lem_avg	0.064
	1024D_l2_lem	0.088
fT fastText.cc	300D_avg	0.831
	300D	0.825
	300D_lem_avg	0.831
	300D_lem	0.825
fT Embeddia	100D_avg	0.143
	100D	0.141
	100D_lem_avg	0.143
	100D_lem	0.141
	300D_avg	0.419
	300D	0.513
	300D_lem_avg	0.419
	300D_lem	0.513
fT CLARIN.SI-embed.sl	100D_avg	0.316
	100D	0.310
	100D_lem_avg	0.316
	100D_lem	0.310
fT Sketch Engine (word)	100D_avg	0.096
	100D	0.135
	100D_lem_avg	0.096
	100D_lem	0.135
fT Sketch Engine (lemma)	100D_avg	0.803
	100D	0.927
w2v Kontekst.io	256D_avg	0.483
	256D	0.718
	256D_lem_avg	0.483
	256D_lem	0.718

Tabela 5: Delež primerov, pri katerih je rezultat analogije z najvišjo kosinusno podobnostjo sam vhodni poklic (pred filtriranjem za računanje rezultatov v Tabelah 2 in 3). Št. vseh primerov je 468 iz 234 parov poklicev.

7. Zaključki in nadaljnje delo

V prispevku smo na nalogi analogij moških in ženskih poklicev ovrednotili različne slovenske vektorske vložitve (z različnimi konfiguracijami in pristopi k računanju analogij). Ugotovili smo, da dobimo najboljše rezultate s fastText vložitvami. Pri ženskih analogijah za moške poklice

ELMo Embeddia l2_lem_avg				fastText CLARIN.SI_lem_avg				word2vec Kontekst.io_lem_avg			
m vhod		f vhod		m vhod		f vhod		m vhod		f vhod	
Rezultat	n	Rezultat	n	Rezultat	n	Rezultat	n	Rezultat	n	Rezultat	n
bolničarka	47	geograf	9	šivilja	15	mizar	11	kuharica	44	ortoped	14
biokemičarka	39	politolog	8	ključavničarka	11	biolog	10	gospodinja	38	pisatelj	14
frizerka	39	biolog	7	inštalaterka	9	ključavničar	9	šivilja	33	kardiolog	13
trgovka	39	dramaturg	7	keramičarka	9	zgodovinar	9	frizerka	32	nevrolog	13
čistilka	34	književnik	7	filologinja	8	internist	8	kozmetičarka	30	urolog	13
znanstvenica	34	scenarist	7	oftalmologinja	8	režiser	8	čistilka	29	psihiater	12
kuharica	33	animator	6	filozofinja	7	arheolog	7	fotografinja	29	ekolog	11
geologinja	30	esejist	6	geofizičarka	7	natakar	7	zdravnica	29	hišnik	11
perica	28	etnolog	6	kmetica	7	pisatelj	7	služkinja	26	biolog	10
služkinja	28	fotograf	6	nevrokirurginja	7	primarij	7	trgovka	26	korenjak	10
biologinja	27	ilustrator	6	strugarka	7	stomatolog	7	slikarka	25	maneken	10
gospodinja	26	lutkar	6	geologinja	6	tesar	7	tajnica	25	režiser	10
matematičarka	26	paleontolog	6	hematologinja	6	fotoreporter	6	veterinarka	25	akademik	9
mikrobiologinja	26	pravnik	6	kardiologinja	6	gostilničar	6	znanstvenica	25	akademski_slikar	9
arheologinja	25	režiser	6	paleontologinja	6	kardiolog	6	socialna_delavka	24	glasbenik	9

Tabela 6: 15 najpogostejših besed, ki se pojavljajo med prvimi desetimi rezultati analogij v določenem modelu vložitev glede na vhodni poklic v analogiji (*m* ali *f*).

na vhodu se najbolje odreže model fastText CLARIN.SI-embed.sl, za ženske poklice na vhodu pa so to modeli fastText CLARIN.SI-embed.sl ter fastText Embeddia. Pristop, kjer za izračun namesto samega vektorja *moški* oz. *ženska* uporabimo povprečje besed z inherentno izraženim spolom, izboljša rezultate, enako velja za lematizacijo. Najboljši rezultati (P@10) so tako 0.885 za ženske iztočnice z modelom fastText CLARIN.SI-embed.sl-100D_lem_avg in 0.982 s fastText Embeddia 300D_lem_avg z upoštevanim pogojem, da so poklici v vložitvah. Za moške poklice na vhodu pa 0.902 z modelom fastText CLARIN.SI-embed.sl 100D_lem_avg (enak rezultat velja tudi za pogoj prisotnih vložitev). Modeli fastText CLARIN.SI-embed.sl imajo največji delež iskanih moških in ženskih poklicev. Med obravnavanimi vložitvami kažejo vložitve modela Kontekst.io pri kvalitativni analizi na največjo pristranost modela glede na spol (stereotipno ženski in moški poklici, ki se pojavljajo med analogijami ne glede na iztočnico). Prispevek se sicer osredotoča na kvantitativno evalvacijo in je s tem uporaben predvsem za razvijalce novih orodij, podrobnejši kvalitativni analizi in odnosu med vložitevami, jezikom in družbeno močjo pa se bomo posvetili v prihodnje. V nadaljnjem delu bomo obravnavali tudi kontekstne vložitve modela BERT, preizkusili metode za zmanjševanje vpliva vložitev, ki so bolj centralne od drugih, ter študijo razširili na druge jezike projekta EMBEDDIA. Poleg tega bomo preizkusili vpliv spolnih pristranosti v napovednih modelih na praktičnih nalogah, kot je analiza sentimenta.

8. Zahvala

Delo je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije v okviru programov P2-0103 (Tehnologije znanja) in P6-411 (Jezikovni viri in tehnologije za slovenski jezik) ter EU prek okvirnega programa za raziskave in inovacije Obzorje2020 - projekt EMBEDDIA (št. 825153).

9. Literatura

Gigafida 2.0. 2020. Gigafida 2.0: Korpus pisne standardne slovenščine. https://viri.cjvt.si/gigafida, 1. 5. 2020.

Shlomo Argamon, Moshe Koppel, Jonathan Fine in Anat Rachel Shimoni. 2003. Gender, genre, and writing style in formal written texts. *TEXT*, 23:321–346.

Paul Baker. 2010. Will Ms ever be as frequent as Mr? a corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender & Language*, 4(1):125–149.

Piotr Bojanowski, Edouard Grave, Armand Joulin in Tomas Mikolov. 2016. Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama in Adam Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. V: *30th Conference on Neural Information Processing Systems*.

Shikha Bordia in Samuel R. Bowman. 2019. Identifying and reducing gender bias in word-level language models. *CoRR*, abs/1904.03035.

Marc-Etienne Brunet, Colleen Alkalay-Houlihan, Ashton Anderson in Richard S. Zemel. 2018. Understanding the origins of bias in word embeddings. *CoRR*, abs/1810.03611.

Carmen Rosa Caldas-Coulhard in Rosamund Moon. 2010. 'curvy, hunky, kinky': Using corpora as tools for critical analysis. *Discourse & Society*, 21(2):99–133.

Aylin Caliskan, Joanna J. Bryson in Arvind Narayanan. 2017. Semantics derived automatically from language corpora necessarily contain human biases. *Science*, 356(6334):183–186.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer in Hervé Jégou. 2018. Word translation without parallel data. V: *Proc. of International Conference on Learning Representation (ICLR)*.

Kaja Dobrovoljc, Simon Krek, Peter Holozan, Tomaž Erjavec, Miro Romih, Špela Arhar Holdt, Jaka Čibej, Luka Krsnik in Marko Robnik-Šikonja. 2019. Morphological lexicon Sloleks 2.0. CLARIN.SI. http://hdl.handle.net/11356/1230.

Nikhil Garg, Londa Schiebinger, Dan Jurafsky in James Zou. 2018. Word embeddings quantify 100 years of gen-

der and ethnic stereotypes. *PNAS*, 115(16).

Aparna Garimella, Carmen Banea, Dirk Hovy in Rada Mihalcea. 2019. Women's syntactic resilience and men's grammatical luck: Gender-bias in part-of-speech tagging and dependency parsing. V: *Proc. of the 57th Annual Meeting of the ACL*, str. 3493–3498. ACL.

Hila Gonen in Yoav Goldberg. 2019. Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. V: *Proc. of NAACL-HLT 2019*, str. 609–614.

Vojko Gorjanc. 2007. Kontekstualizacija oseb ženskega in moškega spola v slovenskih tiskanih medijih. V: I. Novak-Popov, ur., *Stereotipi v slovenskem jeziku, literaturi in kulturi: zbornik predavanj 43. seminarja slovenskega jezika, literature in kulture*, str. 173–180. Center za slovenščino kot drugi/tuji jezik, Ljubljana.

Benjamin Hill in Aaron Shaw. 2013. The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PloS One*, 8.

Dirk Hovy in Anders Søgaard. 2015. Tagging performance correlates with author age. V: *Proc. of the 53rd Annual Meeting of the ACL and the 7th IJCNLP*, str. 483–488.

Dirk Hovy. 2015. Demographic factors improve classification performance. V: *Proc. of the 53rd Annual Meeting of the ACL and the 7th IJCNLP*, str. 752–762. ACL.

Svetlana Kiritchenko in Saif M. Mohammad. 2018. Examining gender and race bias in two hundred sentiment analysis systems. *CoRR*, abs/1805.04508.

Corina Koolen in Andreas van Cranenburgh. 2017. These are not the stereotypes you are looking for: Bias and fairness in authorial gender attribution. V: *Proc. of the First Ethics in NLP workshop*, str. 12–22. ACL.

Robin Lakoff. 1973. Language and woman's place. *Language in Society*, 2(1):45–80.

Nikola Ljubešić in Tomaž Erjavec. 2018. Word embeddings CLARIN.SI-embed.sl 1.0. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1204.

Matej Martinc, Iza Škrjanec, Katja Zupan in Senja Pollak. 2017. PAN 2017: Author profiling - gender and language variety prediction. V: *Working Notes of CLEF 2017*.

Tomas Mikolov, Greg S. Corrado, Kai Chen in Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. V: *International Conference on Learning Representations*, str. 1–12.

Tomas Mikolov, Wen-tau Yih in Geoffrey Zweig. 2013b. Linguistic regularities in continuous space word representations. V: *Proc. of the 2013 Conference of the North American Chapter of the ACL: Human Language Technologies*, str. 746–751. ACL.

Malvina Nissim, Rik Noord in Rob van der Goot. 2019. Fair is better than sensational: Man is to doctor as woman is to doctor. *Computational Linguistics*, str. 1–17.

Francisco M. Rangel Pardo, Fabio Celli, Paolo Rosso, Martin Potthast, Benno Stein in Walter Daelemans. 2015. Overview of the 3rd author profiling task at PAN 2015. V: L. Cappellato, N. Ferro, G. J. F. Jones in E. SanJuan, ur., *Working Notes of CLEF 2015*, zvezek 1391 iz *CEUR Workshop Proceedings*. CEUR-WS.org.

Michael Pearce. 2008. Investigating the collocational behaviour of man and woman in the BNC using Sketch Engine. *Corpora*, 3:1–29, 05.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee in Luke Zettlemoyer. 2018. Deep contextualized word representations. V: *Proc. of NAACL-HLT 2018*, str. 2227–2237.

Marko Plahuta. 2020. O slovarju. https://kontekst.io/o-slovarju.

Marcelo O. R. Prates, Pedro H. Avelar in Luís C. Lamb. 2020. Assessing gender bias in machine translation: A case study with Google Translate. *Neural Computing and Applications*, 32:6363–6381.

Tony Sun, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang in William Yang Wang. 2019. Mitigating gender bias in natural language processing: Literature review. V: *Proc. of the 57th Annual Meeting of the ACL*, str. 1630–1640. ACL.

Anka Supej, Marko Plahuta, Matthew Purver, Michael Mathioudakis in Senja Pollak. 2019. Gender, language, and society: Word embeddings as a reflection of social inequalities in linguistic corpora. V: *Zbornik Slovenskega sociološkega srečanja 2019 - Znanost in družbe prihodnosti*, str. 75–83.

Deborah Tannen. 1990. *You Just Don't Understand: Women and Men in Conversation*. Ballantine Books, NY.

Matej Ulčar. 2019. ELMo embeddings model, Slovenian. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1257.

Eva Vanmassenhove, Christian Hardmeier in Andy Way. 2018. Getting gender right in neural machine translation. V: *Proc. of the EMNLP*, str. 3003–3008. ACL.

Ben Verhoeven, Iza Škrjanec in Senja Pollak. 2017. Gender profiling for Slovene Twitter communication: The influence of gender marking, content and style. V: *Proc. of the 6th BSNLP Workshop*, str. 119–125. ACL.

Vlada RS. 1997. 1641. uredba o uvedbi in uporabi standardne klasifikacije poklicev. *Uradni list RS*, 28:2217. https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina?urlid=199728&stevilka=1641.

Svitlana Volkova, Theresa Wilson in David Yarowsky. 2013. Exploring demographic language variations to improve multilingual sentiment analysis in social media. V: *Proc. of the EMNLP*, str. 1815–1827. ACL.

Iza Škrjanec, Nada Lavrač in Senja Pollak. 2018. Napovedovanje spola slovenskih blogerk in blogerjev. V: D. Fišer, ur., *Viri, orodja in metode za analizo spletne slovenščine*, str. 356–373. Ljubljana: Znanstvena založba FF.

Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez in Kai-Wei Chang. 2017. Men also like shopping: Reducing gender bias amplification using corpus-level constraints. V: *Proc. of the EMNLP*, str. 2979–2989. ACL.

Tianlu Zhao, Jieyu fand Wang, Mark Yatskar, Vicente Ordonez in Kai-Wei Chang. 2018. Gender bias in coreference resolution: Evaluation and debiasing methods. V: *Proc. of the NAACL-HLT*, str. 15–20. ACL.