

УДК

316.774:305-055.2(497.11)"2009/2015"
305:070(497.11)"2009/2015"

DOI

10.18485/KNJIZ.2015.1.24

Оригинални научни рад

Цветана Крстев, Универзитет у Београду, Филолошки факултет¹

Милош Утвић, Универзитет у Београду, Филолошки факултет

Јелена Јаћимовић, Универзитет у Београду, Филолошки факултет и Стоматолошки факултет

Ако коза лаже, рог не лаже – где су и ко су жене у српској дневној штампи²

Сажетак: Полазећи од претпоставке да је присуство жена у дневној и недељној штампи слика њиховог положаја у друштву, у овом раду покушавамо да утврдимо колико и како су жене представљене у српској штампи користећи се методама корпусне и рачунарске лингвистике, као и методама екстракције информација. Разматрање је засновано на корпусу састављеном од извода из дневне штампе и електронских медија (*Политика, Блиц, Вечерње новости, Данас, Б92*) прикупљеном у периоду 2009 – 2015, мањем изводу из женске недељне штампе (*Базар*) и узорку недељника *Врањске*. Методе рачунарске лингвистике ослањају се на коришћење исцрпног електронског речника српског језика, док се екстракција информација заснива на детектовању именованих ентитета коришћењем граматика за плитку синтаксну анализу. Добијени резултати су ручно евалуирани што нам је омогућило да утврдимо висок квалитет нашег система у препознавању имена особа, док квалитет препознавања пола именованих особа треба побољшати. С друге стране, добијени резултати нам говоре да је заступљеност жена уопште у српској штампи данас (дневној и

¹ Резултати представљени у овом раду делом су настали у оквиру истраживања на пројекту 178006 „Српски језик и његови ресурси: теорија, опис и примене“ који финансира Министарство просвете, науке и технолошког развоја Републике Србије.

² Резултати представљени у овом раду делом су настали у оквиру истраживања на пројекту 178006 „Српски језик и његови ресурси: теорија, опис и примене“ који финансира Министарство просвете, науке и технолошког развоја Републике Србије.

недељној) знатно мања од заступљености мушкараца, али и да се најистакнутијим женским особама посвећује знатно мање простора него најистакнутијим мушким личностима.

Кључне речи: електронски новински корпуси, корпусна лингвистика, рачуарска лингвистика, жене, родна равноправност

Лице и тело штампе

Родна равноправност, као основ развоја отвореног, демократског и праведног друштва, подразумева једнаку заступљеност, моћ и учешће оба пола у свим сферама јавног и приватног живота. Као резултат злоупотребе моћи јављају се различити облици дискриминације (расне, родне, сексуалне, мањинских група итд), те појава друштвених неједнакости и неправди.³ Услед чврсто укореењеног система патријархалних хијерархијских односа који постоје између мушкараца и жена, у свакодневном друштвеном, професионалном и породичном окружењу дискриминација заснована на роду често остаје скривена, непрепознатљива и нејасна. Ипак, свеprisутни портрет родних идентитета, улога и односа моћи дат је у виду порука масовних медија,⁴ за које се може поставити питање да ли промовишу родну равноправност или родне стереотипе.

Улога медија, почев од штампе, па све до новијих облика електронских медија, у конфигурисању понашања, модификацији друштвених норми и конструисању друштвене реалности од фундаменталног је значаја.⁵ Медијске репрезентације утичу на формирање јавног мњења не само рефлектујући слике стварности, већ и репрезентујући и креирајући интерпретације, чија ће значења настала у овом процесу често постати учвршћена у друштву и култури у оквиру које постоје.⁶ На тај начин, својом способношћу да осликају, успоставе или одбаце идентитете, културне вредности и социјалне односе, медији служе као моћно средство у изградњи друштвених улога полова и њихових односа.⁷ У разматрању

³ Teun A. van Dijk, *Discourse and power* (Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2008), 19-20.

⁴ K. Bussey i A. Bandura, „Social Cognitive Theory of Gender Development and Functioning“, u *The psychology of gender*, ur. Alice H. Eagly, Beall Anne E. i Sternberg Robert J. (New York: Guilford, 2004), 95.

⁵ M.P. Matud, C. Rodríguez i I. Espinos, „Gender in Spanish daily newspapers“, *Sex Roles* 64(2011), 253.

⁶ S. Redman i J. Taylor, „Legitimate family violence as represented in print media: Textual analysis“, *Journal of Advanced Nursing* 56(2006), 163.

G.B. Cunningham i dr., „Gender representation in the NCAA News. Is the glass half full or half empty?“, *Sex Roles* 50(2004), 861.

P.J. Shoemaker i S.D. Reese, *Mediating the message: Theories of influences in mass media content* (New York: Longman, 1996), 30.

Smiljana Milinkov, „Kako prići ženi u radijskom etru Srbije? Pokušaj formatiranja novosadskog Radija 021“, u *Verodostojnost medija : dometi medijske tranzicije*, ur. Rade Veljanovski. (Beograd: Fakultet političkih nauka: Čigoja štampa, 2011), 201.

Snježana Milivojević, „Žene i mediji: strategije isključivanja“, *Genero* posebno izdanje (2004), 14.

⁷ Dragana Cvetković, „Vest je muškog roda - etički aspekti rodne ravnopravnosti u informativnom programu nacionalnih televizija u Srbiji“, *Kultura* 127 (2010), 157.

равноправности полова, осим mediјске присутности, битан је и појам mediјске одсутности, који заједно омогућавају уочавање постојећег односа моћи међу половима и стварање стереотипа као доминантних представа у некој друштвеној заједници.⁸

О родном јазу и стереотипима сведоче различита спроведена истраживања која се баве критичком анализом дискурса медијских садржаја који имају највише утицаја на аудиторијум. Разлике у приказивању полова најчешће су праћене кроз пропагандни програм, односно телевизијске рекламе у којима се јасно огледа матрица традиционалних родних улога.⁹ Ипак, стереотипи мушкараца као „јачег пола“, а жена као „лепше половине“ осликавају се и у радијским рекламама,¹⁰ информативном програму националних телевизија,¹¹ у текстовима и на фотографијама магазина¹² и новинске штампе.¹³ Медијска празнина у погледу репрезентовања жена одјекује страницама како светске тако и српске штампе и електронских медија. Жене су недовољно заступљене у различитим медијима, а и када су присутне, превасходно су оскудно обучене и приказане у стереотипним улогама.¹⁴ Заузимајући десети део медијског простора, о женама се обично говори у рубрикама црне

⁸ *Ibid.*

⁹ A.N. Arima, „Gender stereotypes in Japanese television advertisements“, *Sex Roles* 49 (2003):89.

D.L. Ganahl, T.J. Pensen i S.B. Netzley, „A content analysis of prime time commercials: A contextual framework of gender representation“, *Sex Roles* 49 (2003), 550.

K. Kim i D.T. Lowry, „Television commercials as a lagging social indicator: Gender role stereotypes in Korean television advertising“, *Sex Roles*, 53 (2005), 908.

N. Uray i S. Burnaz, „An analysis of the portrayal of gender roles in Turkish television advertisement“, *Gender Roles*, 48 (2003), 86.

F. Valls-Fernández i J.M. Martínez-Vicente, „Gender stereotypes in Spanish television commercials“, *Sex Roles* 56 (2007), 697.

¹⁰ A. Furnham i L. Thomson, „Gender role stereotyping in advertisements of two British radio stations“, *Sex Roles* 40 (1999), 163.

¹¹ D. Cvetković, nav. rad, 173.

¹² R. Hovland i dr., „Gender role portrayals in American and Korean advertisements“, *Sex Roles* 53(2005), 897.

T. Mastin i dr., „Product purchase decision-making behavior and gender role stereotypes: A content analysis of advertisements in *Essence* and *Ladies' Home Journal*, 1990–1999“, *The Howard Journal of Communication* 15 (2004), 229.

¹³ M. Len-Ríos i dr., „Representation of women in news and photos: Comparing content to perceptions“, *The Journal of Communication* 55 (2005), 165.

S. Rodgers i dr., „Stereotypical portrayals of emotionality in news photos“, *Mass Communication & Society* 10(2007), 138.

Zorica Mršević, „Žene u medijima: 2010 u Srbiji“, *Genero* 12 (2008), 80.

S. Milivojević, nav. rad, 17.

Smiljana Milinkov, „Siromaštvo žena - medijska isključenost“, *Godišnjak Filozofskog fakulteta u Novom Sadu* 37-I (2013), 315.

Jelena Višnjić, „‘Killing me softly’: izveštavanje štampanih medija o ženama žrtvama nasilja“, *Genero* 16 (2012), 143.

D. Cvetković, nav. rad, 173.

M.P. Matud, nav. rad, 262.

¹⁴ R.M. de Cabo i dr., „Perpetuating gender inequality via the internet? An analysis of women's presence in Spanish online newspapers“, *Sex Roles* 70 (2014), 69.

хронике, културе и естраде,¹⁵ док су пет до десет пута мање заступљене у односу на мушкарце када су у питању озбиљне политичке теме.¹⁶ Иако је у Србији стопа активности жена са високим образовањем виша од стопе активности мушкараца истог нивоа образовања (71% односно 65%),¹⁷ међу ауторитетима (стручњацима, директорима, представницима компанија и слично) који се позивају да јавно изнесу своја мишљења и који су важни саговорници на дату тему ретко су особе женског пола.¹⁸ Међутим, не може се рећи да је у питању било каква завера новинара против жена, већ Цветковић указује на чињеницу да је затечено стање заправо друштвени проблем којим се треба бавити. Позитивних примера жена у улогама које по значају не заостају од оних додељених мушкарцима ипак има, али су они ретки и недовољно видљиви.

Током последњих пар деценија Србија је прошла кроз низ политичких, друштвених и економских промена, превасходно важних за жене и њихове значајно боље могућности образовања и запошљавања. Законски основ за увођење и уређивање родне равноправности у Србији дат је, пре свега, *Уставом Републике Србије* (усвојеним 2006. године),¹⁹ којим се утврђује обавеза државе да јемчи равноправност жена и мушкараца и развија политику једнаких могућности. *Закон о равноправности полова* (донет 2009. године), осим што регулише области запошљавања, здравствене заштите, породичних односа, образовања, политичког и јавног живота, спорта итд, у члану 41 каже да „Информације путем средстава јавног информисања не смеју садржавати нити подстицати дискриминацију засновану на полу.“²⁰ О значају медија, као једној од кључних области за унапређивање родне равноправности, сведочи и *Национална стратегија за побољшање положаја жена и унапређивање родне равноправности*,²¹ усвојена фебруара месеца 2009. године. Насупрот наведеним позитивним законским регулативама, на питање да ли је остварена равноправност жена у српском друштву могу се чути различити одговори. Неки сматрају да су жене потпуно равноправне, неки мисле да је већ остварена равноправност изгубљена, док има и оних који сматрају да је никад није ни било и да је потребно још много како би се она постигла. Шта би у овом случају могао бити „рог“ који би помогао у долажењу до неког веродостојног одговора?

С циљем откривања информација на основу којих грађани формирају представу о родној равноправности, те и неусаглашености стварног стања и родне равноправности којој тежимо, предмет овог истраживања биће утврђивање мере и улога којима су жене заступљене у српској штампи. Посебан значај овога рада заправо лежи у примењеној методи којом се на иновативан начин омогућава спровођење анализе садржаја медијских

¹⁵ S. Milivojević, nav. rad, 15-18.

¹⁶ D. Cveticović, nav. rad, 164.

¹⁷ Тамара Врачаревић, прев., *Жене и мушкарци у Републици Србији* (Београд: Републички завод за статистику, 2014), 59.

¹⁸ D. Cveticović, nav. rad, 170.

¹⁹ Устав Републике Србије, чл. 15.

²⁰ Закон о равноправности полова („Сл. гласник РС“, бр. 104/2009), чл. 41.

²¹ Национална стратегија за побољшање положаја жена и унапређивање родне равноправности („Сл. гласник РС“, бр. 15/2009), 4.6. Уклањање родних стереотипа у средствима јавног информисања и промоција родне равноправности.

текстова. Имајући у виду појаву све веће количине машински читљивих садржаја над којима се може вршити анализа, постојање алата који омогућава аутоматско препознавање и обележавање текста различитим информацијама утицаће на спровођење већег броја ефикаснијих и јефтинијих истраживања. Различити алати у виду програма намењених обради природних језика су се пре више деценија показали као изузетно корисни на пољу емпиријских истраживања корпусне лингвистике. Међутим, ретка су истраживања која се баве анализом садржаја ради утврђивања родне пристрасности путем аутоматског препознавања карактеристика и улога полова.²²

Мерење присутности жена у штампи – корпусни приступ

У овом раду ћемо користити корпусни приступ и методе рачунарске лингвистике да бисмо дошли до неких веродостојних података о заступљености мушкараца и жена у савременој српској штампи.

За наш експеримент користићемо више корпуса новинских текстова који ће имати различите улоге. Први корпус ће се састојати од новинских текстова из дневне штампе и електронских медија објављених на интернету. Новински текстови су преузимани са званичних страница *Политике*²³, *Блица*²⁴, *Вечерњих новости*²⁵, *Данаса*²⁶ и *Б92*²⁷ у периоду од 2009-2015, те је тај корпус назван *Бизвора*. С обзиром на то да су за преузимање текстова, њихову корекцију и сређивање за даљу обраду били задужени студенти у виду семинарских радова,²⁸ већина текстова је објављена фебруара и марта месеца одговарајуће године, у време израде семинарских радова. Сваки студент је био задужен за један дан у фебруару или марту месецу, када је требало да преузме из ових пет извора два или три чланка која су обрађивала тему која је према одређеном извору обележила тај дан. Препорука је била да се преузимају дужи чланци и да се избегавају спортске теме. Овај корпус је подељен у седам поткорпуса, по годинама објављивања чланака, па је тако касније и коришћен.

Други корпус се састоји од извесног броја чланака објављених у двонедељнику *Базар*.²⁹ На насловној страници овог часописа стоји: „Истраживања су показала да ‘Базар’ читају високообразоване жене и оне са средњом стручном спремом, које су углавном запослене. Такође, више од једне петине читалаца су мушкарци. Старосна доб је између 28 и 55 година, она која је и најпожељнија, јер су економски самосталне. Испоставило се да у

²² R.M. de Cabo, nav. rad, 58.

²³ Политика новине и магазини, „Политика“, 2015, <http://politika.rs/>.

²⁴ Ringier Axel Springer d.o.o., „Blic“, 2015, <http://www.blic.rs/>.

²⁵ Компанија Новости, „Вечерње новости“, 2015, <http://www.novosti.rs/>.

²⁶ DAN GRAF d.o.o., „Danas“, 2015, <http://www.danas.rs/>.

²⁷ B92 a.d., „B92.net“, 2015, <http://www.b92.net/>.

²⁸ Овај задатак су обављали студенти редовних студија Филолошког факултета, групе за Библиотекарство и информатику у оквиру предмета Информатички практикум IV. Задатак студената био је да преузете текстове коригују, јер текстови објављени на интернет презентацијама одабраних медија обилују грешкама, а да их потом преведу у валидне XML документе.

²⁹ Политика а.д., „Базар“, 2015, <http://www.bazar.co.rs/>.

једној породици чак три генерације жена чита 'Базар'.³⁰ Из овога је јасно да *Базар* спада у категорију часописа која је намењена пре свега, ако не и искључиво, женама. Овај корпус је требало да буде управо корпус новинских чланака из медија намењених пре свега женама. Требало је да буде већи и разноврснији, али то није остварено из три разлога. Пре свега, идеја за његово прикупљање је настала знатно касније, а затим чланци из већине циљних часописа нису доступни на интернету или су доступни само мањи одломци („Илустрована политика“, „Вива“). Коначно, садржај чланака неких од часописа из ове групе је такав да би се аутори овог рада, универзитетски наставници, осећали веома нелагодно да их укључе у наставне активности („Блиц жена“, „Gloria“, „Скандал“, „Свет“, итд)

Трећи корпус се састоји од чланака из недељника *Врањске новине*³¹ који су објављени на веб презентацији ових новина током 2011. и 2012. године. Овај часопис је по оријентацији на изврстан начин сличан новинама и медијима представљеним у првом корпусу јер се такође бави „озбиљним“ темама, али се разликује по полемичном тону, коришћеном језику и коначно по томе што се не објављује у престоници већ у мањем провинцијском граду.

У табели I представљене су димензије ових корпуса и поткорпуса мерене бројем речи и бројем реченица. Текстови су сегментирани на реченице аутоматски (видети наредни одељак), па дате бројеве треба узети само као приближне вредности. Мерење величине корпуса бројем речи је уобичајено у корпусној лингвистици, али је због вишезначности појма „реч“ неопходно прецизно одредити његово значење. Штавише, да би се избегла другачија тумачења, уместо појма „реч“ корпусни лингвисти користе помоћне појмове „токен“ и „корпусна реч“.³² Токени се дефинишу посредно, избором скупа карактера чији се елементи називају граничници или сепаратори. Уобичајени скуп сепаратора садржи белине (размак, табулатор, знак за нови ред) и знаке интерпункције. За потребе овог рада сваки карактер који није ни слово ни цифра представља један сепаратор. Сваки појединачни сепаратор, изузимајући белине, јесте један токен. Такође, сваки непрекидан низ карактера текста који се налази између два сепаратора и сам не садржи сепараторе представља један токен. Корпусна реч је посебна врста токена која се састоји искључиво од словних карактера. На пример, име „Ана-Марија“ се састоји из три токена („Ана“, „-“ и „Марија“) и две корпусне речи („Ана“ и „Марија“).

Из табеле I се види да су поткорпуси новинских вести неравномерне величине, што потиче од варијације у броју студената по школским годинама.

Табела I

Шта тражимо у корпусу и како то добијамо

У прикупљеним корпусима тражићемо појављивања референци на особе мушког и женског пола преко појављивања њихових имена. Имена која тражимо ће бити или пуна

³⁰ Политика а.д., „Базар“, 2015, <http://politika-ad.com/cir/strane/bazar>.

³¹ НИП „Врањске“ ДОО, „Врањске“, 2015, <http://www.vranjske.co.rs/>.

³² Graeme D. Kennedy, *An introduction to corpus linguistics* (London: Longman, 1998), 206-207.

имена, а под тим подразумевамо појављивање бар једног личног имена и бар једног презимена, или појављивање само презимена. Дакле, случајеве у којима се особе у тексту именују само личним именом нећемо узимати у обзир, осим ако се ради о истакнутим особама које су познате само под једним именом које може бити и лично, на пример, припадници краљевских породица и црквени великодостојници. Код пуних имена водићемо рачуна о променљивом редоследу личног имена и презимена, о могућности појављивања два презимена, средњег имена (имена родитеља) или иницијала, надимка, професионалних титула и сл. Код коришћења само презимена, водићемо рачуна о могућности коришћења присвојног придева или именице изведених из презимена за особе женског пола.

Уз појављивање имена, било пуних било само презимена, тражићемо и назнаке функција, улога или неких других одређења која објашњавају ко је особа о којој се у тексту говори, а која се у тексту појављују пре или после имена у облику именске фразе. Сва идентификована имена ће у тексту бити адекватно обележена,³³ а уз свако идентификовано име ће стајати назнака да ли оно означава особу мушког или женског пола, или назнака да се пол особе на коју се име односи на основу имена и анализираног контекста није могао одредити.

Да би се овако постављен задатак могао обавити коришћене су напредне методе екстракције информације – подобласти обраде природних језика – које су у нашем случају подразумевале коришћење веома обимних електронских речника српског језика који садрже исцрпне информације (морфолошке, деривационе, семантичке и многе друге) које су потребне за обављање овог задатка. Ови речници укључују и велики број имена и презимена која се користе у Србији, као и транскрибованих страних имена, пре свега с енглеског говорног подручја. Осим речника, коришћене су комплексне граматике плитког парсирања (енгл. *shallow parsing*) које анализирајући локални контекст потенцијалних имена потврђују претпоставку, одређују границе имена и придружених улога (ако су наведене), као и пол особе на коју се име односи. Ове граматике локалног парсирања су део обухватнијег система за препознавање именованих ентитета у текстовима на српском језику (енгл. *named entities* – подразумевају се сви прости и вишечлани називи којима се препознају јединствени ентитети, као што су особе, организације, локације, и сл.). У овом раду нећемо се бавити описом коришћених метода које су детаљније представљене у другим радовима.³⁴

³³ За обележавање је коришћен XML језик; свако име, као и улоге особе на које се име односи су окружени почетном и завршном етикетом, што се види из датих примера.

³⁴ Сандра Гуцул-Милојевић, „Властита имена у екстракцији информација“, ИНФОтека 11(2010), 47-58.

Sandra Gucul-Milojević, Vanja Radulović i Cvetana Krstev, “A View on the Representation of Women in Serbian Newspaper Texts”, u *Application of Finite-State Language Processing: Selected Papers from the 2008 International Nooj Conference*, ur. Tamás Váradi, Judit Kuti i Max Silberztein (Newcastle upon Tyne: Cambridge Scholars Publishing, 2010),160-170.

Cvetana Krstev i dr, “A system for named entity recognition based on local grammars”, *J Logic Computation* 24(2014), 473-489.

Неколико наведених примера илуструје тражена појављивања имена и начин обележавања резултата.³⁵

1. Пуно име мушке особе:
<pers><role>Britanski ministar spoljnih poslova</role><m.persName.full>Vilijam Hejg</m.persName.full></pers>
2. Презиме мушке особе:
...reкао је <pers><m.persName.last>Mali</m.persName.last></pers>
3. Пуно име женске особе:
<pers><f.persName.full>Gorica Mitić</f.persName.full><role>, bivša direktorka Toplane</role></pers>
4. Презиме женске особе:
...smatra da <pers><f.persName.last>Markovićeva</f.persName.last></pers> ne bi trebalo da чека...
5. Име истакнуте особе (чији пол није одређен):
...izveo državni udar protiv <pers.spec>kralja<x.persName.name POL="M">Idrisa</x.persName.name></pers.spec>...
6. Пуно име особе за коју систем није одредио ког је пола:
<pers><role>bivši ministar socijalne politike</role><x.persName.full>Natalija Koroljevska</x.persName.full></pers>
7. Презиме особе за коју систем није одредио ког је пола:
<pers><role>Kancelarka</role><x.persName.last>Merkel</x.persName.last></pers> је u Berlinu rekla...
8. Име особе за коју систем није одредио ког је пола:
<pers.spec>para<x.persName.first>Franja</x.persName.first></pers.spec>

Зашто је потребна евалуација и како је спроводимо

Ниједан аутоматизовани систем за проналажење и екстракцију информација није савршен, то јест међу резултатима које производи јавља се одређен број грешака.³⁶ Резултати које систем производи се не могу користити ни у каквим даљим истраживањима нити за доказивање одређених претпоставки уколико се претходно не утврди степен и врста грешака. Грешке које се јављају се обично могу окарактерисати као *пропусти* – није препознато нешто што је требало, *уљези* – препознато је нешто што није требало и *непоклапање* – опсежи онога што је требало препознати и онога што је препознато се

³⁵ Сви текстови у свим коришћеним корпусима су записани латиничним писмом, независно од тога којим писмом су текстови били штампани и приказани на веб презентацијама. Стога ће и сви примери бити записани латиничним писмом. Треба напоменути да би све подједнако функционисало и да је коришћено ћирилично писмо.

³⁶ Уосталом, ако би исти посао радили људи и они би такође направили одређен број грешака, посебно ако им задаци нису довољно добро дефинисани или се ради о досадном послу у коме се стално понављају слични задаци. Ко од њих прави већи број грешака зависи од угла гледања: неки сматрају да „машина никад не греши“ док други верују да „машина не може да надмаши човека“. У сваком случају, аутоматски се многи задаци, као и овај који се састоји од обележавања имена, могу неупоредиво брже урадити.

делимично преклапају али не у потпуности. Извори грешака такође могу бити различити и потичу делимично од вишезначности облика у природном језику за чије решавање је потребна анализа ширег контекста. С тим у вези, ни модел језика који се користи у систему за препознавања често није довољно разрађен, а понекад ни технологија на којој се систем заснива не допушта већа побољшања. На пример, граматике плитког парсирања не могу да користе довољно широк контекст да би све вишезначности биле исправно разрешене. Коначно, неке грешке препознавања потичу од грешака у самом тексту (типографске, правописне, синтаксне грешке). У неким случајевима се ово и не сматра грешкама, јер зашто би се од система очекивало да препозна нешто што је погрешно записано. Но, с друге стране, како човек уме да препозна тражено (нпр. име) чак иако су оно или неке речи у његовом окружењу погрешно записани, има смисла да се то очекује и од аутоматизованог система.

У нашем случају извори грешака препознавања су сви они управо поменути. Што се тиче типова грешака, они се у случају нашег задатка могу, пре свега, овако одредити: имамо грешке у препознавању имена, грешке у препознавању улоге особе на коју се име односи и грешке у препознавању њеног пола. У случају прва два типа грешака јављају се пропусти, уљеци, непоклапања као и грешке због грешака. У случају препознавања пола могуће су три нежељене ситуације: пол је погрешно одређен, пол није уопште одређен (*пропуст*) и име је означено и као мушко и као женско (*контрадикција*).

Сва препознавања имена и пола у целокупном корпусу *Бизвора* су проверена, што значи да су у текст додаване етикете (код грешака пропуста), односно, да су додавани атрибути с знаком грешке у случајевима других врста грешака. Грешке у препознавању улога особа овом приликом нису провераване, јер нису од превеликог значаја за ово истраживање. Проверу су радили студенти на основу исцрпног упутства за рад, а њихов рад су контролисали наставници,³⁷ тако да се коначан резултат може сматрати доста поузданим.

Следе примери коректног обележавања имена, грешака и њиховог обележавања:

9. Име је добор означено, а и пол је добро одређен:
<persColl>Srpske teniserke <f.persName.full>Jelena Janković</f.persName.full> i <f.persName.full>Ana Ivanović</f.persName.full></persColl> zadržale su svoje pozicije na najnovijoj WTA listi.
10. Име је добро обележено, али пол није уопште одређен. У том случају студенти су додавали атрибут POL са исправном вредношћу пола:
<pers><role>Najbolji pojedinac utakmice,</role><x.persName.full POL="M">Sašo Ožbolt</x.persName.full></pers>, imao je priliku da u poslednjem napadu Olimpije...

³⁷ Овај задатак су обавили студенти редовних студија Филолошког факултета, групе за Библиотекарство и информатику у оквиру предмета Проналажење информација школске 2014/15. године. У оквиру овог предмета студенти се упознају са разним методама које се примењују у системима за проналажење и екстракцију информација, па се кроз овај практичан задатак могу најбоље упознати са њиховим донетима. Контролу рада студената је обавила Милена Милинковић, студент докторских студија на Филолошком факултету.

11. Име је добро обележено, али пол је погрешан. И у овом случају додаје се исти атрибут као и у претходном случају.
Zasedanje je otvorio predsednik Međunarodnog suda pravde <pers><role>slovački sudija</role><f.persName.full POL="M">Peter Tomka</f.persName.full></pers>.
12. Име је добро одређено, али је пол контрадикторан – име је добило етикете и за мушки и за женски пол. У овом случају се исправна етикета задржава, а у неисправну се додаје атрибут POL као и у претходна два случаја. Треба напоменути да се у процесу првере ништа од етикета не брише, управо да би се постигнути резултати могли коректно проценити.
Takođe je podsetio da je <pers><role>hrvatski predsednik</role><m.persName.full><f.persName.full POL="M">Ivo Josipović</f.persName.full></m.persName.full></pers> u decembru 2012...
13. Пол је добро одређен, али је име само делимично добро одређено (грешка непоклапања). У овом случају додаје се атрибут IME с вредношћу UOK.
Šef civilnog vazduhoplovstva Azharudin <m.persName.full IME="UOK">Abdul Rahman</m.persName.full> je uneo konfuziju...
14. Име је делимично добро означено, а ни пол није добро одређен. У овом случају је потребно додавање оба атрибута: IME и POL.
Serena <pers><m.persName.last IME="UOK">POL="F">Vilijams</m.persName.last></pers> ponovo je postala svetski reket broj jedan.
15. Име је у потпуности пропуштено. У овом случају треба ручно додати XML етикете с вредношћу MISS атрибута IME.
Policija je identifikovala treću osobu kao 21-godišnju <f.persName.full IME="MISS">Rosauru Ernandes-Barios</f.persName.full>
16. Означено је као име особе нешто што то уопште није (уљез). У овом случају треба додати ручно етикете с вредношћу NOK атрибута IME.
Prvi reket sveta, koji je bio član reprezentacije koja je osvojila <pers><m.persName.full IME="NOK">Dejvis Kup</m.persName.full></pers> 2004. godine, rekao je da jedva čeka meč...
17. Дошло је до пропуста у обележавању због грешке у самом тексту. У овом случају треба ручно додати XML етикете с вредношћу ERR атрибута IME.
<m.persName.full IME="ERR"> Branislav Cevtković</m.persName.full>, direktor fabrike "Tihos"...

Добијени резултати и шта нам они говоре

Резултати пребројавања који су добијени на материјалу који је прошао двоструку проверу (прво студенти, затим наставник) показују да се у корпусу *Бизвора* јавља укупно 11.177 имена особа (типова који су претходно описани), од тога 9.825 (87,9%) имена мушкараца према 1.352 (24,42%) имена жена. У много мањем корпусу *Базар* јавља се укупно 561 име, од тога 424 (75,58%) мушка према 137 (24,42%) женских имена. Ови односи су визуелно приказани на слици I.

Слика I

Табела II

Подаци из табеле II говоре да се презимена за именовање мушких особа користе у корпусу *Бизвора* скоро исто толико као и пуна имена, нешто мање у *Базару*. Такво именовање женских особа је значајно мање у корпусу *Бизвора*, а ретко у *Базару*. С обзиром на то да је препознавано само код истакнутих личности, коришћење само личног имена је мало очекивано у свим случајевима. С тога је однос пуних мушких, према пуним женским именима увек мањи, као што се и види на графикону на слици II.

Слика II

Однос мушких према женским именима у разним годиштима корпуса *Бизвора* је приказан на слици II. Може се уочити да не постоји никакав тренд ни раста ни опадања једних имена у односу на друга, премда је доста упадљиво значајно повећање учешћа мушких имена 2014. године.

У корпусу *Бизвора* се 2.125 пута појављује име мушке особе с назнаком улоге, функције или професије, насупрот 328 појављивања код женских особа. Примећује се да међу „топ 10“ улога (функција, професија) мушкараца нема представника, посланика, професора и супруга (које се јављају као „топ 10“ улоге жена у корпусу), а да међу „топ 10“ жена нема премијера, лидера, адвоката и саветника. Наш приступ нам је омогућио да утврдимо за два „најеминентнија“ занимања *председник* и *министар* на шта се односе. У случају мушкараца, председници су најчешће држава (140), политичких партија и синдиката (72), општина (10) и скупштине (16). Међу женама нема председника држава и општина, једна је председница синдиката, док је, изгледа, председник скупштине омиљена женска функција (22). Међу мушкарцима има министара свих ресора – укупно 23 – док се жене јављају као министарке у само 6 ресора: енергетике, правде, финансија, спољних послова,³⁸ здравља и трговине.

Табела III

Коначно, резултати које смо добили нам омогућавају да утврдимо који се мушкарци и жене најчешће помињу у корпусу *Бизвора* и резултати су представљени у табели IV. Приликом рачунања броја појављивања особа у корпусу спојена су појављивања пуних имена и презимена што уноси одређену непрецизност, јер издвојена из контекста презимена могу означавати и друге особе (у табели је означено која презимена носе ту опасност). Додатна провера је показала да само презимена *Николић* и *Јанковић* носе реалну опасност, јер се осим особа наведених у табели још доста њих с истим презименима појављује у корпусу.³⁹

Табела IV

³⁸ Односи се на Републику Хрватску.

³⁹ Занимљиво је напоменути да се међу тим другим особама с истим презименом доста често јављају блиски рођаци особа из табеле IV.

Резултати из Табеле IV нам говоре да не само да су жене уопште мање присутне у штампи од мушкараца, већ и када се гледају појединачно најистакнутијих шест мушких и женских особа видимо да је поредећи по рангу заступљеност мушке особе 4 (Вучић према Ештон и Путин према Бараћ) до 6 пута већа (Дачић према Михајловић).

Када је реч о језику, резултати из Табеле III такође показују да је у све чешћој употреби родно сензитиван начин ословљавања, односно да се код навођења функција и улога женских особа чешће користе именице женског рода (добијене моцијом рода) од именица мушког рода – изузетак је једино пар *портпарол/портпаролка*. Када се ово упореди с присутношћу жена у новинама, било укупно било појединачно, изгледа да коришћење назива функција женског рода није у корелацији са пажњом коју женама посвећује штампа, иако се у пракси у ословљавању жена заиста све више користи родно сензитиван језик.⁴⁰

Колико је добар наш приступ

Код коришћења било ког аутоматизованог система за проналажење и екстракцију информација веома је битно да се зна колико је тај систем добар. Да би се то утврдило мора се обавити његова евалуација. То се може урадити било коришћењем такозваног *златног стандарда* (енгл. *gold standard*) где се рад система упоређује са унапред познатим позитивним резултатом (који је раније утврђен), било ручном провером рада система на неком мањем узорку. У недостатку златног стандарда за српски језик, ми смо овог пута применили други приступ, при чему су мањи узорци корпуси *Бизвора* и *Базар*.⁴¹

Да би се утврдио квалитет система за проналажење и екстракцију информација треба, пре свега, утврдити шта се од њега очекује. Здрав разум нам говори да корисник од једног таквог система очекује да добије што је могуће више информација које тражи (ако је могуће све), а такође очекује да међу информацијама које је добио буде што мање оних које није тражио (ако је могуће, да их уопште нема). Овако постављен задатак нам говори да ништа не значи податак да је систем пронашао, рецимо, 1.000 одговора на тражену информацију ако се у колекцији која се претражује налази 100.000 корисних информација, или ако је међу тих 1.000 заправо 999 нерелевантних. У информатичком окружењу увек је пожељно да се процена квалитета система изрази једним бројем или с неколико бројева, јер то олакшава поређење различитих система. Стандардне мере које се већ дуги низ година користе за процену система за проналажење и екстракцију информација су *одзив* (R), који представља однос укупно пронађених релевантних информација према укупном броју релевантних информација и *прецизност* (P) која представља однос укупно пронађених релевантних информација према укупном броју пронађених информација. Ове две мере се своде на једну заједничку меру ($\square 1 = \frac{2 * \square * \square}{\square + \square}$) која је хармонијска средина одзива и

⁴⁰ D. Cvetković, nav. rad, 174.

⁴¹ Евалуирани корпуси *Бизвора* и *Базар*, евентуално још једном проверени, моћи ће се користити у будућности као златни стандард за неке друге системе.

прецизности и која увек даје предност мањој од њих. Коришћење $F1$ мере за поређење два или више система онемогућава фаворизовање било одзива било прецизности.

У нашем случају, раздвојили смо тачност препознавања имена, мушких и женских, независно од тога да ли је систем правилно (и да ли је уопште) утврдио ког је пола особа која то име носи, и тачност препознавања пола особе за имена која су тачно или делимично тачно препозната. У случају препознавања имена, прецизност и одзив смо рачунали под следећим претпоставкама:

- релевантне пронађене информације ($RelPro$) су потпуно тачно препозната имена и имена код којих има грешака непоклапања ($IME="UOK"$);
- све релевантне информације ($TotRel$) су потпуно тачно препозната имена, имена код којих има грешака непоклапања ($IME="UOK"$) и имена која су пропуштена у препознавању ($IME="MISS"$);
- све пронађене информације ($TotPro$) су потпуно тачно препозната имена, имена код којих има грешака непоклапања ($IME="UOK"$) и погрешна препознавања (*уљези*) ($IME="NOK"$).

Одзив се према томе рачуна као $\square = \frac{\square\square\square\square\square}{\square\square\square\square\square}$, а прецизност као $\square = \frac{\square\square\square\square}{\square\square\square\square}$.

Табела V

Резултати у табели V показују да је успешност нашег система у препознавању имена веома добра ($F1=0,900$). Прецизност је, може се рећи, изузетна ($P=0,981-0,992$), то јест, такорећи све што је препознато су заиста имена. Примећује се да је препознавање само имена и само презимена значајно лошије од препознавања пуних имена, што је логично, јер је контекст препознавања ужи. Такође, прецизност препознавања женских имена је већа, али је одзив мањи (посебно само за презимена и имена). У будућности треба радити на побољшању система у односу на ове уочене недостатке.

У случају препознавања пола особа, рачунали смо само однос тачно препознатог пола код имена која су тачно препозната и имена код којих има грешака непоклапања ($IME="UOK"$);

Табела VI

Резултати у табели VI показују да је успешност препознавања пола код успешно препознатих имена изузетно висока када су мушка имена у питању (88,82%), нешто нижа код женских имена (75,40%). Уочава се такође да је успешност препознавања пуних имена добра и скоро иста код мушких и женских имена (83,35% наспрам 84,60%). Док је препознавање пола код навођења само презимена изузетно успешно када су у питању особе мушког пола (97,11%), за особе женског пола успешност препознавања је свега 46,99%. Препознавање пола код навођења само имена је лоше без обзира на пол, али су ова појављивања релативно ретка у анализираном корусу па је њихов утицај на укупне резултате мали.

У табели V се уочава још једна занимљивост. Особе мушког пола се много чешће именују само презименом или само именом (45,39% од укупног броја имена мушких особа) него особе женског пола (24,24% од од укупног броја имена женских особа).

Резултате ове евалуације можемо употребити за процену појављивања женских имена у произвољном новинском тексту.⁴² У *Врањским* је аутоматски препознато 18.317 имена:

- мушких имена 16.554 (90,38%), од тога
 - пуних 7.866 (47,52%)
 - презимена 8.688 (52,48%)
- женских имена 1.763 (9,62%)
 - пуних 1.664 (94,38%)
 - презимена 99 (5,62%)

На основу препознавања презимена не би требало изводити никакве закључке, јер су резултати евалуације показали да препознавање пола особе именоване само презименом није успешно. Међутим, препознавање пола код именовања пуним именом можемо узети као успешно. Добијени однос мушких према женским пуним именима 4,72 се уклапа у резултате добијене на корпусу *Сизвора* (однос 5,30, видети слику II).

И штампа је мушког рода

Шта нам је рог саопштио овом приликом?

Резултати овог истраживања о медијском представљању жена у српској штампи и електронским медијима недвосмислено нас наводе на чињеницу да је озбиљна штампа и даље мушког рода. Документована неравноправност полова у медијским репрезентацијама је заиста упадљива и омогућава нам идентификовање образаца који леже у основи културно и друштвено усвојених ставова, што у први план ставља проблем који је потребно решавати – констатна изложеност великог аудиторијума садржајима које карактерише релативна одсутност жена. Чини се да, имајући у виду учесталост и контекст појављивања жена у српској штампи, савремени медијски свет заправо много више личи на реалност Србије с половине прошлог века.⁴³

У интерпретацији података добијених овим истраживањем, потребно је имати на уму да анализом није обухваћена сва постојећа српска штампа, и да у том смислу ови резултати не могу бити посматрани као комплетан одраз значења и коначан суд о родној (не)равноправности у Србији. Ипак, с друге стране, ово истраживање, осим што јасно указује на постојање проблема, обезбеђује основу за даљи рад усмерен на његово решавање. У том правцу, како би се омогућило спровођење дубље квалитативне анализе медијских садржаја, у оквиру будућих истраживања требало би узети у обзир шири контекст

⁴² Претпостављамо да је нови, непознати текст довољно сличан по садржају и начину писања са корпусом на коме је вршена евалуација да се резултати могу поредити.

⁴³ Miloš Macura (ur.), *Statistički godišnjak NR Srbije za 1952. god. Godina IV* (Београд: Zavod za statistiku i evidenciju NR Srbije, 1953), 175-176. Удео жена у радној снази Србије је 1952. године износио само 19% што је веома блиско уделу жена у савременим српским медијима.

репрезентације жена. С обзиром на то да је истакнутост у новинама индикована дужином чланка, појављивањем на непарним странама и у горњем делу стране, било би корисно посматрати учесталост појављивања жена и у односу на величину чланака и делове у којима се оне помињу. Осим тога, пол аутора новинских чланака такође може утицати на значење добијене слике родне заступљености. Ради стицања комплетније слике, пожељно би било да се аутоматско препознавање особа врши, не само на основу личних имена која се помињу у тексту, већ и на основу свих оних израза који указују на неку особу или њену улогу.

Литература

- Arima, A.N. „Gender stereotypes in Japanese television advertisements“. *Sex Roles* 49, 2003, 81–90.
- Bussey, Kay i Albert Bandura. „Social Cognitive Theory of Gender Development and Functioning“. U *The psychology of gender*. Urednici Alice H. Eagly, Beall Anne E. i Sternberg Robert J. New York: Guilford, 2004, 92–119.
- Valls-Fernández, F. i J.M. Martínez-Vicente. „Gender stereotypes in Spanish television commercials“. *Sex Roles* 56, 2007, 691–699.
- Višnjić, Jelena. „‘Killing me softly’: izveštavanje štampanih medija o ženama žrtvama nasilja“. *Genero* 16, 2012, 141–156.
- Врачаревић, Тамара. прев. *Жене и мушкарци у Републици Србији*. Београд: Републички завод за статистику, 2014.
- Ganahl, D.L., T.J. Pensen i S.B. Netzley. „A content analysis of prime time commercials: A contextual framework of gender representation“. *Sex Roles* 49, 2003, 545–551.
- Гуцул-Милојевић, Сандра. „Властита имена у екстракцији информација“. *ИНФОтека* 11, 2010, 47–58.
- Gucul-Milojević, Sandra, Vanja Radulović i Cvetana Krstev. „A View on the Representation of Women in Serbian Newspaper Texts“. U *Application of Finite-State Language Processing: Selected Papers from the 2008 International Nooj Conference*. Urednici Tamás Váradi, Judit Kuti i Max Silberstein. Newcastle upon Tyne: Cambridge Scholars Publishing, 2010, 160–170.
- De Cabo, R. M., R. Gimeno, M. Martínez i L. López. „Perpetuating gender inequality via the internet? An analysis of women’s presence in Spanish online newspapers“. *Sex roles* 70, 2014, 57–71.
- Dijk, Teun A. van. *Discourse and power*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2008.
- Kennedy, Graeme D. *An introduction to corpus linguistics*. London: Longman, 1998.
- Kim, K. i D.T. Lowry. „Television commercials as a lagging social indicator: Gender role stereotypes in Korean television advertising“. *Sex Roles* 53, 2005, 901–910.

- Krstev, Cvetana, Ivan Obradović, Miloš Utvić i Duško Vitas. „A system for named entity recognition based on local grammars”. *J Logic Computation* 24, 2014, 473–489.
- Cunningham, G. B., M. Sagas, M. L. Sartore, M. L. Amsden, i A. Schellhase. „Gender representation in the NCAA News. Is the glass half full or half empty?”. *Sex Roles* 50, 2004, 861–870.
- Len-Ríos, M., S. Rodgers, E. Thorson i D. Yoon. „Representation of women in news and photos: Comparing content to perceptions“. *The Journal of Communication* 55, 2005, 152–168.
- Macura, Miloš (ur.). *Statistički godišnjak NR Srbije za 1952. god. Godina IV*. Beograd: Zavod za statistiku i evidenciju NR Srbije, 1953.
- Mastin, T., A. Coe, S. Hamilton i S. Tarr. „Product purchase decision-making behavior and gender role stereotypes: A content analysis of advertisements in *Essence* and *Ladies' Home Journal*, 1990–1999“. *The Howard Journal of Communication* 15, 2004, 229–243.
- Matud, M. P., C. Rodríguez i I. Espinosa. „Gender in Spanish daily newspapers”, *Sex Roles* 64 2011, 253-264.
- Milinkov, Smiljana. „Kako prići ženi u radijskom etru Srbije? Pokušaj formatiranja novosadskog Radija 021“. U *Verodostojnost medija : dometi medijske tranzicije*. Urednik Rade Veljanovski. Beograd: Fakultet političkih nauka: Čigoja štampa, 2011, 195–203.
- Milinkov, Smiljana. „Siromaštvo žena - medijska isključenost“. *Godišnjak Filozofskog fakulteta u Novom Sadu* 37-I, 2013, 311–322.
- Milivojević, Snježana. „Žene i mediji: strategije isključivanja“. *Genero*, posebno izdanje, 2004, 11–24.
- Mršević, Zorica. „Žene u medijima: 2010 u Srbiji“. *Genero* 12, 2008, 69–93.
- Redman, S. i J. Taylor. „Legitimate family violence as represented in print media: Textual analysis“, *Journal of Advanced Nursing* 56, 2006, 157–165.
- Rodgers, S., L.J. Kenix i E. Thorson. „Stereotypical portrayals of emotionality in news photos“. *Mass Communication & Society* 10, 2007, 119–138.
- Uray, N. i S. Burnaz. „An analysis of the portrayal of gender roles in Turkish television advertisement“. *Gender Roles*, 48(2003), 77–87.
- Furnham, A. i L. Thomson. „Gender role stereotyping in advertisements of two British radio stations“. *Sex Roles* 40, 1999, 153–165.
- Hovland, R., C. McMahan, G. Lee, J. Hwang i J. Kim. „Gender role portrayals in American and Korean advertisements“. *Sex Roles* 53, 2005, 887–899.
- Cvetković, Dragana. „Vest je muškog roda – etički aspekti rodne ravnopravnosti u informativnom programu nacionalnih televizija u Srbiji“. *Kultura* 127, 2010, 156–175.

Shoemaker P.J. i S.D. Reese. *Mediating the message: Theories of influences in mass media content*. New York: Longman, 1996.

Извори коришћених корпуса

B92 a.d. „B92.net.“ 2015. <http://www.b92.net/>.

DAN GRAF d.o.o. „Danas.“ 2015. <http://www.danas.rs/>.

Компанија Новости. „Вечерње новости.“ 2015. <http://www.novosti.rs/>.

НИП „Врањске“ ДОО. „Врањске.“ 2015. <http://www.vranjske.co.rs/>.

Политика а.д. „Базар.“ 2015. <http://www.bazar.co.rs/>.

Политика новине и магазини. „Политика.“ 2015. <http://politika.rs/>.

Ringier Axel Springer d.o.o. „Blic.“ 2015. <http://www.blic.rs/>.

Прилози

Корпус	број реченица	број речи
5извора-2009	3,010	61,754
5извора-2010	1,697	45,124

Сизвора-2011	1,174	29,943
Сизвора-2012	2,631	65,155
Сизвора-2013	2,835	72,954
Сизвора-2014	4,559	117,890
Сизвора-2015	876	21,807
Сизвора-укупно	16,782	414,627
Базар	2.773	47,410
Врањске	39,096	782,304

Табела I. Величине корпуса коришћених у истраживању

корпус	мушка имена				женска имена			
	пуна	презиме	име	укупно	пуна	презиме	име	укупно
<i>Сизвора</i>	5.012	4.691	122	9825	945	399	9	1352
%	51,01	47,45	1,24	100,00	69,90	29,51	0,67	100,00
<i>Базар</i>	234	165	25	424	115	11	11	137
%	55,19	38,91	5,90	100,00	83,94	8,03	8,03	100,00

Табела II. Учешће типова имена у корпусима *Сизвора* и *Базар*

Мушке особе			Женске особе			
бр.	улога	фрек.	улога	ж-фрек.	м-фрек.	фрек.
1	председник	519	председница	42	4	46
2	министар	261	представница	27	8	35
3	премијер	185	министарка	25	7	32
4	директор	143	портпаролка	9	15	24
5	потпредседник	131	потпредседница	21	1	22
6	лидер	74	директорка	15	6	21
7	секретар	63	канцеларка	13	0	13
8	адвокат	37	секретарка	8	5	13
9	саветник	29	посланица	10	2	12
10	портпарол	23	професорка	7	0	7
			супруга	7	0	7

Табела III. „Топ 10“ улога (функција или професија) мушкараца и жена у корпусу *Сизвора*. Код жена, у колони ж-фрек. дата је учесталост појављивања улоге као именице женског рода, а у колони м-фрек. учесталост појављивања улоге као именице мушког рода.

особа	фрек.	особа	фрек.
Александар Вучић*	330	Кетрин Ештон	83
Ивица Дачић	258	Зорана Михајловић	43
Борис Тадић*	235	Јелена Јанковић*	42
Томислав Николић*	208	Славица Ђукић-Дејановић	37
Владимир Путин	145	Верица Бараћ	36

Дарко Шарић*	135	Верица Калановић	28
--------------	-----	------------------	----

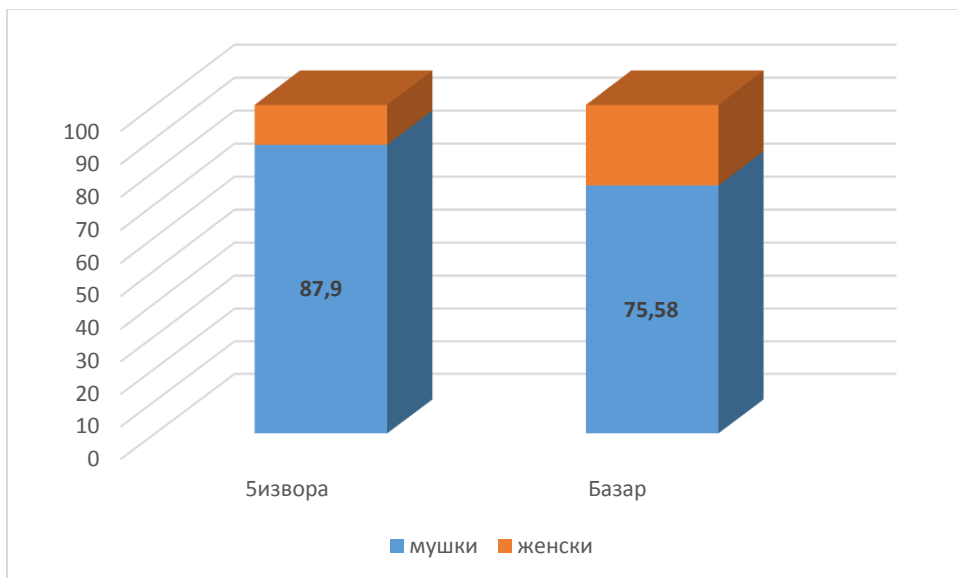
Табела IV. „Топ 6“ особа у корпусу *Бизвора*. Звездицом су означене особе чија презимена носе неке друге особе које се такође јављају у корпусу.

<i>Sizvora</i>	мушка имена				женска имена			
	<i>пуна</i>	<i>през.</i>	<i>име</i>	<i>укупно</i>	<i>пуна</i>	<i>през.</i>	<i>име</i>	<i>укупно</i>
<i>TotPro</i>	4780	4024	82	8886	847	272	4	1123
<i>RelPro</i>	4762	3877	81	8720	844	266	4	1114
<i>TotRel</i>	5012	4691	122	9825	945	399	9	1353
<i>P</i>	0,996	0,963	0,988	0,981	0,996	0,978	1,000	0,992
<i>R</i>	0,950	0,826	0,664	0,888	0,893	0,667	0,444	0,823
<i>F1</i>	0,973	0,890	0,794	0,932	0,942	0,793	0,615	0,900

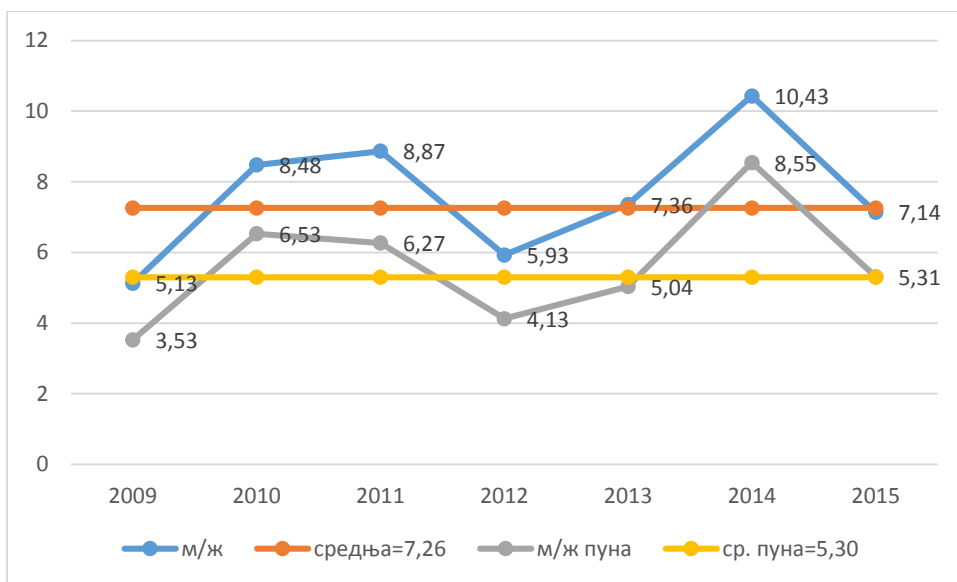
Табела V. Успешност препознавања мушких и женских имена.

Бизвора	мушка имена				женска имена			
	<i>пуна</i>	<i>презиме</i>	<i>име</i>	<i>укупно</i>	<i>пуна</i>	<i>презиме</i>	<i>име</i>	<i>укупно</i>
<i>Пол ОК</i>	3969	3765	11	7745	714	125	1	840
%	83,35	97,11	13,58	88,82	84,60	46,99	25,00	75,40
<i>Пол промењен</i>	71	4	0	75	34	132	0	166
%	1,49	0,10	0,00	0,86	4,03	49,62	0,00	14,90
<i>Пол непознат</i>	722	108	70	900	96	9	3	108
%	15,16	2,79	86,42	10,32	11,37	3,38	75,00	9,69
<i>Укупно</i>	4762	3877	81	8720	844	266	4	1114

Табела VI. Успешност препознавања пола именоване особе.



Слика I. Учешће мушких и женских имена у два корпуса



Слика II. Однос броја свих мушких према броју свих женских имена у корпусу *Бизвора* (м/ж) и однос пуних мушких имена према пуним женским именима у истом корпусу (м/ж пуна)

UDC

316.774:305-055.2(497.11)"2009/2015"
305:070(497.11)"2009/2015"

Cvetana KRSTEV
Miloš Utvić
Jelena Jaćimović
University of Belgrade

Original Scientific Article

Facts Are Stubborn Things – Women in Serbian Daily Press

Summary

There are various answers to the question whether gender equality has been achieved in contemporary Serbian society. Some consider women to be absolutely equal, others think that the achieved equality has been lost; there are those who believe that it never existed and that it will be a long time before it is achieved. What could be the stubborn fact which would help us determine what the plausible answer is? Assuming that the presence of women in daily and weekly press reflects their social position, we shall attempt to discover how much and in what way women are present in Serbian papers. The research relies on corpus and computational linguistics and extraction of information. The analysis will be based on a corpus consisting of portions taken from daily papers and electronic media (*Politika, Blic, Večernje novosti, Danas, B92*) which have more than 600.000 words (collected 2009-2014) and a smaller extract from weekly press (*Ilustrovana politika, Bazar*). The methods of computational linguistics will employ a comprehensive dictionary of the Serbian language and the extraction of information will be based on detecting the named entities with the help of grammars for shallow syntactic analysis. After evaluation, the obtained results will be examined in terms of quality and quantity.

Keywords: electronic newspaper corpus, corpus linguistic, computational linguistics, women, gender equality.