ANKA SUPEJ
Univerza v Ljubljani, Filozofska fakulteta, Ljubljana
MARKO PLAHUTA
podjetje Virostatik, Ljubljana
MATTHEW PURVER
Queen Mary University of London, London, UK
MICHAEL MATHIOUDAKIS
University of Helsinki, Helsinki, Finland
SENJA POLLAK
Jožef Stefan Institute, Ljubljana

GENDER, LANGUAGE, AND SOCIETY – WORD EMBEDDINGS AS A REFLECTION OF SOCIAL INEQUALITIES IN LINGUISTIC CORPORA

Abstract: *Research on language and gender has a long tradition, and large electronic text corpora and novel computational methods for representing word meaning have recently opened new directions. We explain how gender can be analysed using word embeddings: vector representations of words computationally derived from lexical context in large corpora and capturing a degree of semantics. Being derived from naturally-occurring text, these also capture human biases, stereotypes and reflect social inequalities. The relation between the English words man and programmer can correspond to that between woman and homemaker. In Slovene, the availability of male and female forms for many words for occupations means that such effects might be reduced; however, we study a range of such relations and show that some gender bias still persists (e.g. the relation between words woman and secretary is very similar to that between man and boss).*

Key words: *gender bias, word embeddings, occupations, language and society, natural language processing*

Introduction

Researchers have long been interested in the relationship between language and gender. What started as introspective research into how women and men are discussed and how their way of talking differs (Lakoff 1973), developed into sociolinguistic modelling of discourse styles and different kinds of statistical analyses, which, for example, explore words with which men or women are described. These approaches are now being increasingly complemented by advanced natural language processing (NLP) methods[1], among them *word embeddings* (see below), which can convey meaningful relationships between gender and language.

1. NLP methods are computational methods, designed to process and analyse large amounts of human (i.e. natural) language.

Language can be also understood as being one of the most powerful means through which sexism and gender discrimination are perpetrated and reproduced, via, for example, the content of gender stereotypes, as well as the language structures used (Menegatti and Rubini 2017). The stereotypes reproduced in the lexical choices of everyday communication are not neutral: they reflect the asymmetries of status and power in favour of the dominant social group, and affect recipients' cognition and behaviour (see Eagly et al. 2000, Maass and Arcuri 1996, Menegatti and Rubini 2017). On the structural level, the norm according to which the prototypical human being is male is reproduced in many languages (Silveira 1980); feminine terms usually derive from the corresponding masculine form; and masculine nouns and pronouns are often used with a generic function to refer to both men and women (Menegatti and Rubini 2017). Here, we focus on the relation between gender, language and occupations; and also in this domain, a large body of work addresses stereotype-consistent language use (e.g. Heilman 2001, Gaucher et al. 2011), as well as investigating the influence of gender-fair language use (currently initiating heated debates in Slovenian professional and public spheres) in the context of job advertisements, or in societal perceptions of professions (Horvath and Sczesny 2016, Horvath et al. 2016).

Word embeddings

Word embeddings are vector representations of words: each word is assigned a vector of (typically) several hundred dimensions. These are usually obtained via training algorithms such as *word2vec* (Mikolov et al. 2013a) and *GloVe* (Pennington et al. 2014), which characterize the word based on the lexical context in which it appears. These representations improve performance in a wide range of automated text processing tasks, partly because they capture a degree of semantics: words that are similar or semantically related are closer together in vector space. They can also capture regularities beyond simple relatedness, such as analogies (Mikolov et al. 2013b); for example, the vector-space relation between *Madrid* and *Spain* is very similar to that between *Paris* and *France*.

This provides a way to analyse complicated concepts like gender. If we examine words which differ systematically in gender (e.g. *man*:*woman*; *son*:*daughter*), we expect the vector difference to be approximately the same (Pennington et al. 2014). We can discover gender correspondences via gender-based "analogies" (e.g. testing which word X is to *woman* as *king* is to *man*) by simple vector addition and subtraction (e.g. *king* – *man* + *woman* *queen*).

Word embeddings and biases

Being derived from naturally-occurring text, word embeddings also capture human biases, stereotypes and reflect social inequalities (Caliskan et al. 2017). Research on English word embeddings has shown examples of this effect: for example, the word *submissive* can be closer to *woman*, with *honourable* closer to *man* (Garg et al. 2017). This can be both because we often refer to men as being *honourable* directly, and because we refer to them in contexts in which we typically describe honourable things. Bolukbasi et al. (2016) showed that while this sometimes leads to rational outputs (e.g. in the analogy task *man*:*king* : : *woman*:*x;* the closest *x* corresponds to the vector of *queen*), it sometimes shows bias (e.g. *man*:*computer programmer* : : *woman*:*homemaker*). Caliskan et al. (2017) further demonstrated that embeddings contain biased associations (e.g. between math/arts and female/male terms), while Garg et al. (2017) used them to analyse gender stereotypes over time. Biases in word embeddings also influence

automated tools: Kiritchenko and Mohammad (2018) found that the majority of sentiment analysis systems tend to assign higher positivity to sentences involving some genders/races than others. Recently, efforts to decrease bias in embeddings have been made (e.g. Bolukbasi et al. 2016) - however, bias still persists to some extent (Gonen and Goldberg 2019). On the other hand, Nissim et al. (2019) warn that many studies may over-estimate bias.

Experiment with word embeddings in Slovene

Experimental setup

Inspired by the findings with English word embeddings described above, we also focus on occupations. In Slovene, gender for occupations is frequently expressed in morphology, e.g. *sociolog* (male) and *sociologinja* (female form) that we translate as $sociologist_{M}$ and $sociologist_{F}$, respectively.[2] Formulated as an analogy task, we look for gender analogies of occupations in both directions, finding the closest word embedding *x* for $woman{:}manager_{F}$: : *man*:*x* and vice versa for $man{:}manager_{M}$: : *woman*:*x*. The working hypothesis is that *x* should be the male or female version of the occupation, respectively, i.e. ženska:*menedžerka* : : *moški*:*menedžer* and *moški*:*menedžer* : : ženska:*menedžerka*. Slovene word embeddings were trained using word2vec on around 15 Gb of text (academic, news, books etc.).[3]

The female- and male-specific words for occupations, used in the experiment were taken from *the 1641st Regulation on the Introduction and Use of the Standard Classification of Occupations* (ULRS 28/1997), out of which we selected two groups of occupations where men and women had the highest quantitative hourly wage difference: (1) *Legislators, senior officials, managers* and (2) *Experts*, but also included occupations from the group with the smallest difference, i.e. *Officials* (Eurostat and SURS 2018, reporting data from 2014). Some occupations have only one version for both men and women (e.g. *vodja*) – these were treated as gender-neutral. Note that even if words for occupations have several synonyms (e.g. *dekanja*, *dekanica*, *dekanka*) – we used the one provided in the Regulation. From the initial 48 selected occupation pairs, for quantitative evaluation we removed the two gender-neutral pairs, as well as *corrector* (sl. *korektor, korektorica*) since the male form is a homograph for make-up corrector, resulting in 45 pairs. Two of the occupations (namely, *sekretar*/*sekretarka* and *tajnik*/*tajnica*) translate as *secretary* in English – we marked the higher-ranking occupation (sl. *sekretar* or *sekretarka*) as *secretary** and the lower ranking as *secretary*.

In experiments, the task was to find *x* in setting $man{:}occupation_{M}$: : *woman*:*x* (and vice versa), where *x* is the most similar word embedding (with the highest cosine similarity score). For each analogy, we included top 10 words or phrases.

Experimental results and discussion

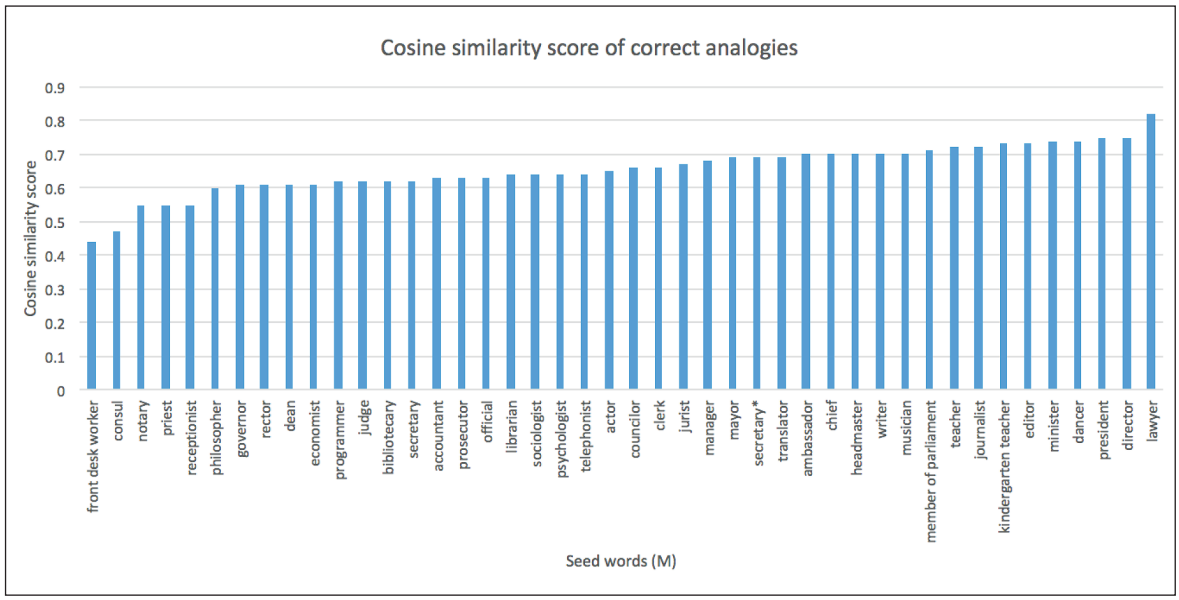
In general, the analogies followed the expected pattern. From 45 occupation word pairs, with female professions as seed words, male analogies were correct as the first hit in 71% and appeared in top 10 hits in 96% of cases. For the reverse task, the analogies were correct as the first hit in 87% of cases and appeared in top 10 hits in 98% of cases.
The correct match did not appear within the first 10 matches for two female word seeds—

2. In this paper, alternative word forms *(e.g. sociolog/inja or sociolog_inja)* are not taken into account.

3. The embeddings are the basis of kontekst.io (Plahuta 2019) and accessible upon request: https://kontekst.io/partnerstvo

receptionist$_F$ (sl. *recepcionistka*) and *front desk worker*$_F$ (sl. *informatorka*)—and once for male word seed (*attaché*). Examples when the match was not the first hit but was found in top 10 candidates include *secretary*$_F$ (sl. *tajnica*), where the first match for male equivalent was *boss*$_M$ (sl. šef), *priest*$_M$ (sl. *duhovnik*), where the first match was *nun* (sl. *nuna*), as well as *consul*$_M$, *notary*$_M$ (sl. *notar*) and *front desk worker*$_M$. The analogy *secretary*$_F$: *boss*$_M$ clearly stands out as an example, where the gender analogy expresses a hierarchical relation, and therefore reflects societal inequalities.

Figure 1. Cosine similarity score for correct female analogies for male occupation seed words[4].



For the correct matches in the analogy task, such as the pair *president*$_M$ (sl. *predsednik*) : *president*$_F$ (sl. *predsednica*), we computed the vector distances in similarity scores. For male specific occupations as seed words (Figure 1), the highest similarity score is observed for the occupations *lawyer* and *director*, while *front desk worker, consul, notary* and *priest* have the lowest score. It is interesting to observe that for two professions from the legal domain, *lawyer* is among the highest scored analogies, while *notary* is among the lowest; intuitively, this tells us that there are more differences in usage (and therefore perception) between *notary*$_M$ *and notary*$_F$ than there are between *lawyer*$_M$ and *lawyer*$_F$. In further work, it would be interesting to investigate in more detail where these differences lie and what they reflect; for this, (co-)occurrence corpus analysis of male and female forms and their contexts could be very informative. But even if the interpretation of these differences is not yet clear, it can serve as a starting point for investigating societal data. For example, according to the study *Mapping the Representation of Women and Men in Legal Professions Across the EU*, the distribution of notaries in Slovenia is imbalanced (cca. 40:60) in favour of women (Galligan et al. 2017,

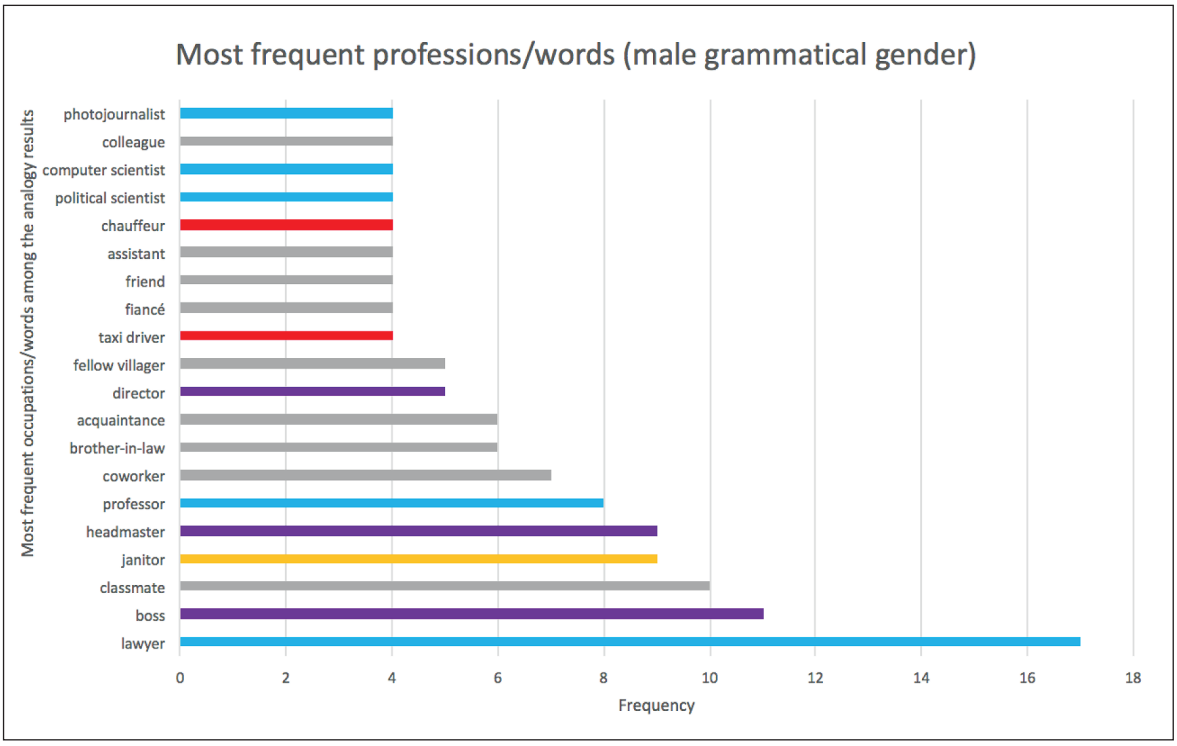
4. Occupation names in Slovene (as appearing in Figure 1): informator, konzul, notar, duhovnik, receptor, filozof, guverner, rektor, dekan, ekonomist, programer, sodnik, bibliotekar, tajnik, računovodja, tožilec, uradnik, knjižničar, sociolog, psiholog, telefonist, igralec, svetnik, referent, pravnik, menedžer, župan, sekretar, prevajalec, veleposlanik, načelnik, ravnatelj, pisatelj, glasbenik, poslanec, učitelj, novinar, vzgojitelj, urednik, minister, plesalec, predsednik, direktor, odvetnik.

69), as in the majority of former communist countries (a possible explanation being that the functions, prestige and income of a notary under communism was rather low and thus very different from the functions of a notary in a Western civil law country). On the other hand, the proportion of lawyers is imbalanced in favour of men (ibid., 64). However, distribution is certainly not the only factor, as for example, highly scored results also included occupations commonly associated with women (e.g. *kindergarten teacher* and *dancer*).

Not only first or correct matches, but also other analogues are interesting to analyse. For example, in analogues for *member of parliament* and *minister* more male proper names (politicians) occur. Also, for both directions, many words not related to the seed occupation were observed within the first 10 matches (e.g. *janitor*, *mechanic*, and *taxi driver* for males and *maid*, *housewife*, *servant*, *secretary*, *nurse*, *carer, cook* for females). Some of them correspond to popular occupations (see Vrabič Kek et al. 2016) that are mostly taken up by men (e.g. *mechanic*) or women (e.g. *nurse, secretary*). We therefore also analysed the top 20 male/female-specific words that appear within the first 10 matches of all analogies (see Figures 2 and 3). For males, there were many occupations that imply high social status (e.g. *lawyer*, two synonyms for *boss*, *director*, *headmaster*, *professor*, amounting to 50 counts altogether). Similar words appeared among the female-specific words (e.g. *lawyer*, *councillor*, two synonyms for *boss*, *vice-president*), but make up only 26 counts. The most common occupations (or words) among the male analogues were *lawyer* (sl. *odvetnik*) (17 examples), *boss* (sl. šef) (11), *classmate*-not an occupation (sl. *sošolec*) (10), *janitor* (sl. *hišnik*) (9), *headmaster* (sl. *ravnatelj*) (9). While *janitor* is nearly an exclusively male occupation, the other three are professions with high societal status, and belong to the categories with the highest wage difference per hour (above 2 eur). On the female side, the most common terms are *secretary* (sl. *tajnica*), *official* (sl. *uradnica*), *homemaker/housewife* (sl. *gospodinja*), *employee* (sl. *uslužbenka*) and *lawyer* (sl. *odvetnica*); here, with the exception of *lawyer,* all are occupations and roles with lower societal status and relatively small wage differences. The case of *housewife* is interesting, since it can mean both the occupation (*homemaker*; also found in the aforementioned regulation ULRS 28/1997) or can describe a stay-at-home woman. Given the presence of other words connected to house chores and care within the list (e.g. *maid*, *servant*$_{F}$, *hospital/care home worker*$_{F}$), even though none of our tasks in fact required analogies of these occupations, we can conclude that the connection between women and house chores was very much present in the original corpus on which the embeddings were trained.

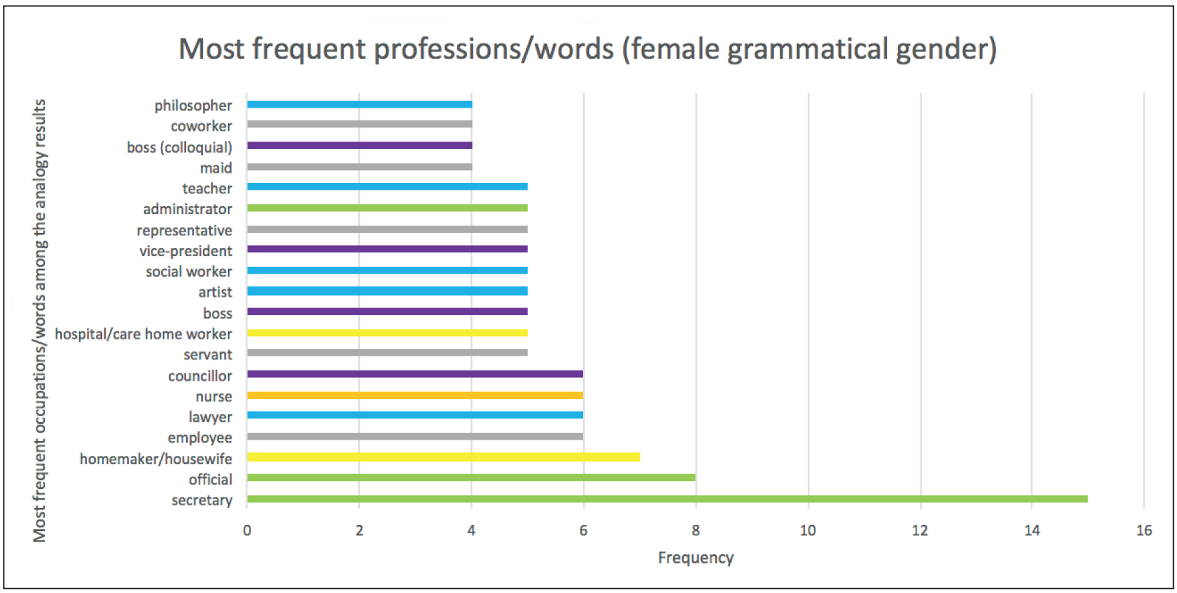
We also observed a few examples with stereotypical or even offensive analogies such as *stripper* (sl. *striptizeta*) for seed word *dancer*$_{M}$, or *gypsy* (sl. pej. *ciganka*) for *postman* (sl. *pismonoša*); the latter was not counted in quantitative results as it is a gender-neutral form.

Figure 2: Top 20 male specific words appearing within the first 10 matches of all analogies for female seed words[5].
Colour legend: (green – quantitative difference in wage per hour up to 0.49 eur; yellow – difference between 0.50 and 0.99 eur; orange – difference between 1.00 and 1.49 eur; red – difference between 1.50 and 1.99 eur; blue – difference between 2,00 and 2.49 eur; purple – difference over 2.50 eur) according to data from 2014 (Eurostat and SURS 2018). Words that represent non-specific professions *(e.g. assistant* (sl. pomočnik)) or not representing professions *(e.g. friend)* are marked with grey.



5. Occupation names in Slovene (as appearing in Figure 2): fotoreporter, stanovski kolega, računalničar, politolog, šofer, pomočnik, prijatelj, zaročenec, taksist, sovaščan, direktor, znanec, svak, sodelavec, profesor, ravnatelj, hišnik, sošolec, šef, odvetnik.

Figure 3: Top 20 female specific words appearing within the first 10 matches of all analogies for male seed words[6].
Colour legend refers to quantitative difference in wage per hour (see caption of Figure 2).



We have presented selected findings on gender bias in English word embeddings, and performed similar experiments on gender roles and occupations on Slovene.

By setting up a suitable analogy task – finding the female (or male) equivalent of a specified male (or female) profession – we show that a standard word embedding space for Slovene does exhibit gender regularities: in general, accuracy on the task is high. As expected, though, we also find that these regularities also capture stereotypes reflecting societal gender inequalities: the closest male analogue to $secretary_F$ (sl. *tajnica*) is found to be $boss_M$ (sl. šef); and the candidates for female analogue to $dancer_M$ (sl. *plesalec*) include *stripper* (sl. *striptizeta*). We also discovered that the most frequent close neighbours to the target occupation words seem to reflect similar stereotypes, with *nurse* closer to *woman* than to *man*, and with neighbours for male terms being more often high-status occupations, while those for female terms more often relate to low-status housework chores.

While these differences can be concretely measured, the interpretations thereof are currently rather more speculative; we expect this situation to improve with future developments of interpretability in NLP. However, we believe that these preliminary analyses clearly show the potential for embeddings-based analysis of gender as reflected in language and society.

Acknowledgements

This paper is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No. 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media). The results of this paper

6. Occupation names in Slovene (as appearing in Figure 3): filozofinja, sodelavka, šefica, služkinja, učiteljica, administratorka, predstavnica, podpredsednica, socialna delavka, umetnica, šefinja, strežnica, služabnica, svetnica, medicinska sestra, odvetnica, uslužbenka, gospodinja, uradnica, tajnica.

reflect only the authors‘ view and the Commission is not responsible for any use that may be made of the information it contains.

References

Bolukbasi, Tolga, Chang, Kai-Wei, Zou, James, Saligrama, Venkatesh, and Kalai, Adam (2016): Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In Proceedings of NIPS 2016.

Caliskan, Aylin, Bryson, Joanna J., and Narayanan, Arvind (2017): Semantics derived automatically from language corpora contain human-like biases. Science, 356 (6334): 183-186.

Eagly, Alice H., Wood, Wendy, and Diekman, Amanda B. (2000): Social role theory of sex differences and similarities: A current appraisal. In T. Eckes and H. M. Trautner (Eds.): The developmental social psychology of gender: 123-174. Mahwah: Lawrence Erlbaum.

Eurostat and SURS. (2018): 2.4 Plače. In Življenje moških in žensk v Evropi – statistični portret. Available at: https://stat.si/womenmen/bloc-2d.html (1. 6. 2019).

Galligan, Yvonne, Haupfleisch, Renate, Irvine, Lisa, Korolkova, Katja, Natter, Monika, Schultz, Ulrike, and Wheeler, Sally (2017): Mapping the Representation of Women and Men in Legal Professions Across the EU. Brussels: European Parliament.

Garg, Nikhil, Schiebinger, Londa, Jurafsky, Dan, and Zou, James (2017): Word embeddings quantify 100 years of gender and ethnic stereotypes. PNAS, 115 (16).

Gaucher, Danielle, Friesen, Justin P., and Kay, Aaron C. (2011): Evidence that gendered wording in job advertisments exists and sustains gender inequality. Journal of Personality and Social Psychology, 101: 109-128.

Gonen, Hila, and Goldberg, Yoav (2019): Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them. CoRR, abs/1903.03862.

Heilman, Madeline E. (2001): Description and prescription: How gender stereotypes prevent women‘s ascent up the organizational ladder. Journal of Social Issues, 57 (4): 657-674.

Horvath, Lisa K., Merkel, Elisa F., Maass, Anne, and Sczesny, Sabine (2016): Does Gender-Fair Language Pay Off? The Social Perception of Professions from a Cross-Linguistic Perspective. Frontiers in Psychology, 6: 2018.

Horvath, Lisa K., and Sczesny, Sabine (2016): Reducing women’s lack of fit with leadership? Effects of the wording of job advertisements. European Journal of Work and Organizational Psychology, 25: 316–328.

Kiritchenko, Svetlana, ad Mohammad, Saif M. (2018): Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. CoRR, abs/1805.04508.

Lakoff, Robin (1973): Language and Woman‘s Place. Language in Society, 2 (1): 45-80.

Maass, Anne, and Arcuri, Luciano (1996): Language and stereotyping. In C. N. Macrae, C. Strangor, and M. Hewstone (Eds.): Stereotypes and stereotyping: 193-226. New York: Guilford.

Menegatti, Michela, and Rubini, Monica (2017): Gender Bias and Sexism in Language. Oxford Research Encyclopedia of Communication. Available at: https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-470 (5. 9. 2019).

Mikolov, Tomas, Chen, Kai, Corrado, Greg, and Dean, Jeffrey (2013a): Efficient Estimation of Word Representations in Vector Space. In Proceedings of ICLR 2013.

Mikolov, Tomas, Yih, Wen-tau, and Zweig, Geoffrey (2013b): Linguistic Regularities in Continuous Space Word Representations. In Proceedings of NAACL-HLT 2013.

Nissim, Malvina, van Noord, Rik, and van der Goot, Rob (2019): Fair is Better than Sensational: Man is to Doctor as Woman is to Doctor. CoRR, abs/1905.09866.

Pennington, Jeffrey, Socher, Richard, and Manning, Christopher D. (2014): GloVe: Global Vectors for Word Representation. In Proceedings of EMNLP.

Plahuta, Marko (2019). O slovarju. Available at: https://kontekst.io/o-slovarju (1. 6. 2019).

Silveira, Jeanette (1980): Generic masculine words and thinking. In C. Kramarae (Ed.): The voices and words of women and men: 165-178. Oxford: Pergamon.

ULRS 28/1997. Uradni list Republike Slovenije (št. 28/1997): 1641. Uredba o uvedbi in uporabi standardne klasifikacije poklicev. Available at: https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina?urlid=199728&stevilka=1641 (5. 6. 2019).

Vrabič Kek, Brigita, Šter, Darja, and Žnidaršič, Tina (2016): Kako sva si različna: ženske in moški od otroštva do starosti. Ljubljana: SURS.

slovensko sociološko društvo

ZNANOST IN DRUŽBE PRIHODNOSTI

SLOVENSKO SOCIOLOŠKO SREČANJE

Bled, 18. – 19. oktober 2019

Ljubljana, 2019

Izdajatelj:
Slovensko sociološko društvo
Kardeljeva ploščad 5, 1000 Ljubljana

Uredniki:
Miroljub Ignjatović, Aleksandra Kanjuo Mrčela, Roman Kuhar

Tehnični urednik:
Igor Jurekovič

Programski odbor:
Predsedstvo Slovenskega sociološkega društva

Recenzentke:
Anja Zalta, Alenka Švab in Veronika Tašner

Oblikovanje in prelom:
Polonca Mesec Kurdija

Korekture:
avtorji

Tisk:
Demat, d.o.o., Stegne 3, Ljubljana

Naklada:
150 izvodov

Prvi natis

Publikacija je dostopna tudi na elektronskem naslovu:
http://www.sociolosko-drustvo.si/.

Ljubljana, 2019

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

316(497.4)(082)

SLOVENSKO sociološko srečanje (2019 ; Bled)
Znanost in družbe prihodnosti / Slovensko sociološko srečanje, Bled 18.-19. oktober 2019 ; [uredniki Miroljub Ignjatović, Aleksandra Kanjuo Mrčela, Roman Kuhar]. - 1. natis. - Ljubljana : Slovensko sociološko društvo, 2019

ISBN 978-961-94302-3-1
1. Gl. stv. nasl. 2. Ignjatović, Miroljub
COBISS.SI-ID 302109696