SLOVENE AND CROATIAN WORD EMBEDDINGS IN TERMS OF GENDER OCCUPATIONAL ANALOGIES

Matej ULČAR
Faculty of Computer and Information Science, University of Ljubljana

Anka SUPEJ
Jožef Stefan Institute

Marko ROBNIK-ŠIKONJA
Faculty of Computer and Information Science, University of Ljubljana

Senja POLLAK
Jožef Stefan Institute

Ulčar, M., Supej, A., Robnik-Šikonja, M., Pollak, S. (2021): Slovene and Croatian word embeddings in terms of gender occupational analogies. Slovenščina 2.0, 9(1): 26–59.

DOI: https://doi.org/10.4312/slo2.0.2021.1.26-59

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. The article focuses on evaluating Slovene and Croatian word embeddings in terms of gender bias using word analogy calculations. We compiled a list of masculine and feminine nouns for occupations in Slovene and evaluated the gender bias of fastText, word2vec and ELMo embeddings with different configurations and different approaches to analogy calculations. The lowest occupational gender bias was observed with the fastText embeddings. Similarly, we compared different fastText embeddings on Croatian occupational analogies.

Keywords: word embeddings, gender bias, word analogy task, occupations, natural language processing

1 INTRODUCTION

Gender biases in language are studied from many different perspectives. Sociolinguistic studies report how language use differs between men and women (e.g., women tend to have a richer vocabulary, use typical grammatical structures, and express themselves more moderately) (Lakoff, 1973; Tannen, 1990; Argamon et al., 2003). Observations that language use varies between the genders inspired author profiling studies on texts in different languages and of different genres (Koolen and van Cranenburgh, 2017; Pardo et al., 2015; Martinc et al., 2017), also in Slovene (Verhoeven et al., 2017; Škrjanec et al., 2018).[1]

The gender dimension is present as a linguistic variation in corpora and in the form of multi-layered bias, both in individual texts and in larger corpora. Research suggests that:

- The bias is manifested as lack of mentions of women: corpora often used in research contain significantly fewer female pronouns (Zhao et al., 2018) or other references to women (Caldas-Coulhard and Moon, 2010; Baker, 2010).
- Women are less often authors or editors (Hill and Shaw, 2013): only 16% of Wikipedia editors are female.
- Corpora capture stereotypical collocations (Pearce, 2008), which refer to women primarily through their reproductive function (Gorjanc, 2007) and do not associate them with (social) power (Baker, 2010).

Recent rapid developments in natural language processing (NLP) are primarily associated with the use of deep neural networks. Their use requires a representation of text in the form of numeric vectors, called word embeddings. The relations between words are expressed in the geometry of the embedded vector space: semantically related embeddings lie close in the vector space and are arranged in similar directions. This enables the study of relations beyond superficial similarities between words, e.g. through analogies such as the

[1] Note that in these studies non-binary identities are not considered. Male or female gender is assigned based on, for example, author's username on social media platforms or based on other grammatical markers.

relationship *Madrid:Spain* being analogous to the relationship *Paris:France* (Mikolov et al., 2013b).

As it turns out, word embeddings often contain bias, be it gender, race, or other types. Biases in word embeddings manifest through semantic associations and consequent proximities in the vector space (Mikolov et al., 2013b). Biases can be numerically evaluated by, for example, calculating cosine similarity between embeddings that describe a specific concept (e.g. gender) and potentially biased concepts. For example, Caliskan et al. (2017) show that word embeddings associate women with arts and men with science. Utilizing the aforementioned cosine similarity, a powerful approach to demonstrate potential bias in word embeddings is through a calculation of occupational analogies (Bolukbasi et al., 2016). Denoting a vector of word *w* with *v(w)*, this approach checks the existence of the following relationships between male and female word vectors: *v(man) - v(male occupation) ≈ v(woman) - v(female occupation)*. An example for Slovene is *v(moški) - v(učitelj) ≈ v(ženska) - v(učiteljica)*, where *učitelj* and *učiteljica* correspond to the masculine and feminine form of the noun for the concept (occupation) *teacher*, while *moški* and *ženska* denote *man* and *woman (the gender concept),* respectively. In case of no gender bias, the relationship between vectors for man and the masculine form of occupation and between the vector for woman and the feminine form of the same occupation would be approximately the same, as illustrated in Figure 1. However, being derived from naturally occurring text, it is not unexpected that human biases and social positions are captured in embeddings.

The illustration shows a simplified depiction of a few examples with 2-dimensional vectors. The arrows represent the difference between vectors *v(f)* and *v(m)*. The end points of arrows originating in masculine nouns for occupations represent the expected positions of equivalent feminine nouns if there were no bias.

In addition to studies that have shown the bias in word embeddings, different biases can be transferred onto algorithms for different NLP tasks, from machine translation (Prates et al., 2020; Vanmassenhove et al., 2018) to sentiment analysis (Kiritchenko and Mohammad, 2018). On the other hand, some authors (Nissim et al., 2019) warn that the analogy task's design may excessively emphasise biases.

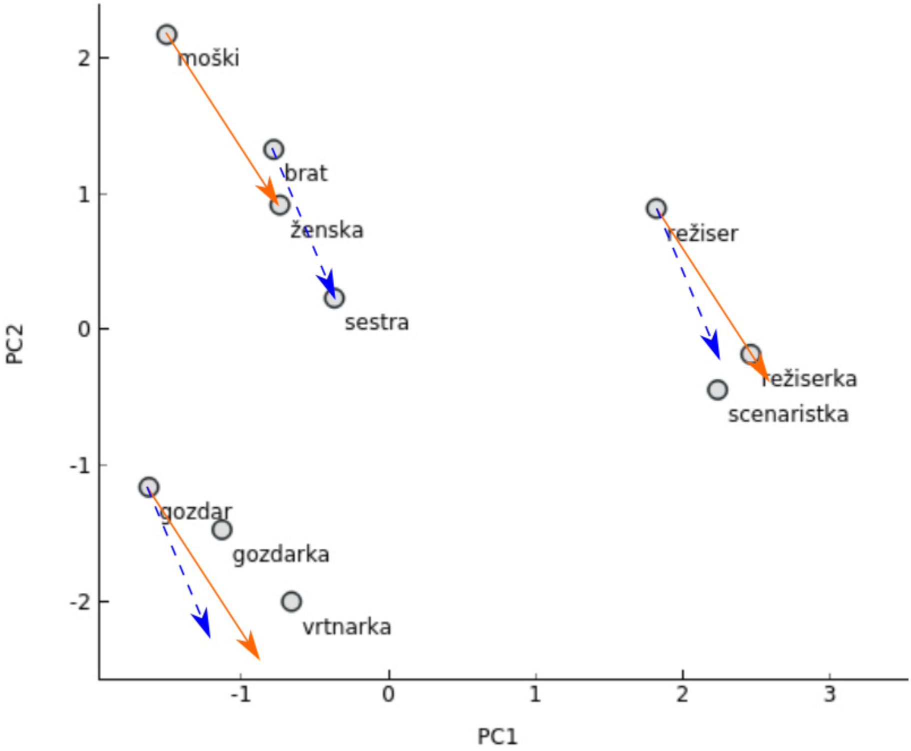

Figure 1: A simplified depiction of word vectors. The orange full arrow represents the difference between vectors for ženska [woman] and moški [man]. The blue dashed arrow represents the difference between vectors for sestra [sister] and brat [brother]. These two arrows indicate the expected (non-biased) gender difference vectors. For two male occupations, režiser [film director$_M$] and gozdar [forester$_M$], we add the gender difference vectors, and depict the resulting nearest female occupations (analogies), i.e. (gozdarka [forester$_F$] and vrtnarka [gardener$_F$]; režiserka [film director$_F$] and scenaristka [scriptwriter$_F$]). The difference to the expected non-biased point is larger for the gozdar - gozdarka pair.

Our study makes certain simplifications. First, we are not paying attention to non-binary expressions of gender, for example we do not specifically address the references such as *on/ona* or a newly proposed form introduced to be more inclusive of nonbinary gender identities *on_a* (Kern and Dobrovoljc, 2017) or noun writings of type *učitelj/učiteljica* (and *učitelj_ica*). Next, for many professions, the male form can be used as a general reference for a profession regardless of gender and we do not make any distinction between mentions of occupations when relating to a male representative or using a general mention (note also that unmarkedness of the masculine form in terms of gender is not anymore universally accepted (Kern and Dobrovoljc, 2017; Popič and

Gorjanc, 2018)). As we analyse and compare the gender bias between different embedding models, these are not severe limitations, as all the embedding models are treated equally. Moreover, similar studies on languages where the gender of a noun is not expressed morphologically can run into more serious problems (see the warnings by Nissim et al. (2019)).

The main contribution of the paper is the evaluation of Slovene and Croatian word embedding models in terms of gender, which has not yet been sufficiently researched (the exception being the analysis of the Slovene w2v model in Supej et al. (2019) and Croatian evaluation of embeddings in Svoboda and Beliga (2018)). The paper extends our work (Supej et al., 2020), where we focused on quantitative evaluation and comparison of a wide range of Slovene models and different approaches to evaluation, while in this paper, we extend the work and also compare Croatian word embeddings models. The focus of the paper is to draw the attention of the developers of linguistic and technological tools (which are based on word embeddings) to the implications the usage of biased embeddings might have. Despite indirectly problematising language bias and pointing out several stereotypical associations, a detailed critical interpretation falls out of this paper's scope.

The paper is divided into further six sections. We first present related work (Section 2). Section 3 describes Slovene and Croatian lists of male and female occupations and specifies the word embedding models used. In Sections 4 and 5, methodology and results are addressed, followed by a discussion in Section 6, and conclusions with plans for further work in Section 7.

2 RELATED WORK

Language corpora and datasets reflect linguistic variations (including different types of bias) in relation to social factors. NLP tools are trained on these data and can inherit the contained variations and biases. The bias in corpora can negatively impact NLP tools (Sun et al., 2019) and can perpetuate biases held towards certain groups. Word embeddings are trained on large corpora to capture syntactic and semantic relations between words and capture the expressed biases.

For instance, it has been shown that standard training data sets for part-of-speech perform better on older people's language (Hovy and Søgaard, 2015).

Garimella et al. (2019) show that a part-of-speech tagger and a dependency parser perform successfully on texts written by women, regardless of what data they had been trained on initially. On the other hand, male authors' texts are better tagged/parsed when the training data contained enough texts written by men. The success of tools such as parsers on male authors' texts may be due to the imbalances in the training data favouring male authorship. It has also been shown that NLP tools are more effective when demographic variations are considered (Volkova et al., 2013; Hovy, 2015). Hovy (2015) shows that including the information on the age and gender of authors improves the performance of three tasks in five different languages.

Biases can have negative consequences in the coreference resolution task (Zhao et al., 2018) and can perpetuate biases held towards certain groups (see examples in Zhao et al., 2017). In the context of texts on mental illness, Hutchinson et al. (2020) note that topics such as gun violence, homelessness, and addiction are over-represented, leading to disability topics receiving particularly negative scores in sentiment analysis tasks. Besides the aspects above, some authors call the attention to the effect biases can have on detection tools. For example, misogyny detection models may attribute high scores to non-misogynous texts simply because the latter contain the so-called identity terms, i.e. terms associated with misogyny (Nozza et al., 2019). In sum, the interplay of bias and NLP is an important and interesting field receiving increasing attention, notably regarding word embeddings, as explained next.

In terms of word embeddings, researchers have studied bias by investigating the proximity of gender-related words to other words in the vector space. For example, Garg et al. (2018) show that the adjective *honourable* lies closer to the word *man* than to the word *woman*. Second, biases are reflected in analogies, e.g. Bolukbasi et al. (2016) show that the embedding space solution of the analogy *man:computer programmer ≈ woman:x* is *x = homemaker*. Nissim et al. (2019) warn that such analogies overemphasise the practical impact of the biases.

As already mentioned, gender bias in word embeddings is often studied on analogies of occupations, which is also our study's case. In morphologically rich languages, such as Slovene and Croatian, the gender of words is expressed morphologically. Therefore, the result of the gender analogy is expected to be the female form of the male variant of the occupation (and vice versa). Svoboda and Beliga (2018) included masculine and feminine versions of job positions in Croatian as one of the evaluation aspects of Croatian word2vec and fastText word embeddings. Preliminary research on word2vec embeddings in Slovene (Supej et al., 2019) showed that the analogy task's accuracy is reasonably high both when attempting to find the female and the male equivalent of an occupation. Results nevertheless reflect gender biases: the first result of the analogy *woman:secretary ≈ man:x* is *x = boss*, while the first ten results of different analogies indicate other gender inequalities: the association of women with house chores and men with occupations of a higher status etc. In the work of Supej et al. (2020) that we extend in this paper, different word2vec, fastText and ELMo embeddings are compared on Slovene pairs of male and female occupations.

As tools based on biased word embeddings may reinforce biases (Zhao et al., 2017), many research groups focused on *debiasing* word embeddings: the main goal of such algorithms is to prevent language models from reproducing racist, sexist or in other ways harmful content. Debiasing also has other advantages – it has been shown that debiasing contributes to correct coreference resolution (Zhao et al., 2018). Some examples of these methods are equalising the distances between gender-specific words and occupations (Bolukbasi et al., 2016; Bordia and Bowman, 2019), inserting additional restrictions into the training corpus (e.g. ensuring equal representation of occupational activities between the genders in the training data) (Zhao et al., 2017), removing texts that cause bias (Brunet et al., 2019), and training gender-neutral word embeddings (Zhao et al., 2018). Schick et al. (2021) recently proposed a self-diagnosis and self-debiasing model where large language models examine their outputs regarding the potential presence of undesirable attributes. They introduced a debiasing algorithm that reduces the likelihood of a model producing biased text. Moreover, researchers recently also focused on methods for debiasing sentence representations, addressing the difficulty of retraining models that are often proposed in debiasing research (retraining models like BERT and ELMo often proves infeasible in practice) (Liang et al., 2020). Gonen and Goldberg (2019) caution that many debiasing methods only conceal bias, which continues to be present in the embeddings, and that many metrics used in the debiasing research have only positive predictive ability (i.e. they can detect the presence of bias but not its absence). On the other hand, studies such as Hirasawa and Komachi (2019) show that debiasing improves multimodal machine translation, thereby underlining the promising future of this research field. In our study, we do not aim to debias embeddings but only compare different embedding approaches in Slovene and Croatian concerning their gender bias.

3 DATA

In this section, we first present the lists of occupations in Slovene and Croatian we used to analyse gender biases, followed by the embedding models.

3.1 List of occupations

We first describe the list of occupations we collected for Slovene, followed by its equivalent in Croatian. Our selection of occupations in Slovene is based on the Standard Classification of Occupations (Vlada RS, 1997), based on the *International Standard Classification of Occupations*. Most occupations in this classification are multi-word expressions (e.g. *upravljalec/upravljalka metalurškega žerjava* [en. *metallurgical crane operator*]), which are less suitable for computation with embeddings due to their specificity and length. To calculate analogies, we limit our approach to single-word occupations. The complete list of single-word occupations in Slovene includes 422 male/female occupation pairs, further reduced in line with the following criteria:

1. An occupation has to exist both in female and male grammatical gender (gender-neutral words such as *pismonoša* [en. *postman*] are not included in the list).

2. An occupation as a common noun occurs at least 500 times in the Corpus of Written Standard Slovene *Gigafida 2.0* (2020).

3. When a more established version of the occupation exists, we manually add a synonym with the same root (e.g. in the case of *fotografka,* an arguably more established *fotografinja* was added [en. *photographer*]). When calculating analogies, the form more frequent in the corpora is inserted at the input, but all synonyms (if they appear among the results) are considered a correctly solved analogy.

4. If the standard classification does not include the female (e.g. *dramatik* [en. *playwright*]) or male variant (e.g. *prostitutka* [en. *prostitute*]) of the occupation, the missing version is manually added if it exists and appears in the Gigafida corpus (e.g. there are no established words for female and male versions of *postreščček* [en. *porter*] and *hostesa* [en. *hostess*], respectively).

5. Occupations where either the female or the male occupation variant is a homograph (e.g. *detektivka* [en. *detective*] also denotes a detective novel) or where an occupation could be associated with a context unrelated to occupations (e.g. *čarovnik/čarovnica* [en. *wizard/witch*]), were excluded from the final set of occupations. Likewise, we filtered out occupations that are also proper names, such as *kovač* [en. *blacksmith*]; for differentiating between common nouns and proper names Sloleks 2.0 (Dobrovoljc et al., 2019) was used. The final list contains 234 occupation pairs and is freely accessible in the CLARIN repository[2].

For Croatian, we compiled a list of occupations from two existing sources. The first source contains occupations from the word analogy dataset by Svoboda and Beliga (2018). It consists of 109 pairs of single-word occupations. The second source is ESCO (European Skills, Competences, Qualifications and Occupations)[3] and lists 2942 occupations in male and female form. Similar to the Slovene list of occupations, most of the classifications from ESCO are multi-word expressions, e.g. *špediterski službenik / špediterska službenica za uvoz i izvoz riba, rakova i mekušaca* [en. *import-export specialist in fish, crustaceans and molluscs*]. After removing all multi-word occupations, the ESCO source contains 309 pairs of single-word occupations. The final, combined list from both sources, filtered to remove duplicates, contains 375 occupation pairs.

3.2 Word embedding models

Different configurations of word embeddings for Slovenian and Croatian were used in the experimental phase. We first list the Slovene embedding models followed by the Croatian ones.

[2] http://hdl.handle.net/11356/1347

[3] https://ec.europa.eu/esco/portal

3.2.1 Slovene word embedding modelS

We analyse two non-contextual embedding models, fastText and word2vec, and the ELMo contextual model.

- fastText (Bojanowski et al., 2017):
  - 100-dimensional vectors, trained on Gigafida 2.0 in the EU EMBEDDIA[4] project,
  - 300-dimensional vectors, trained as above,
  - 100-dimensional word vectors from the Sketch Engine portal (*word*),
  - 100-dimensional word vectors from the Sketch Engine portal, where vectors are embeddings of word lemmas,
  - 100-dimensional CLARIN.SI-embed.sl vectors (Ljubešić and Erjavec, 2018), and
  - 300-dimensional vectors from the fastText.cc portal;
- word2vec (Mikolov et al., 2013a): 256-dimensional vectors, trained for the needs of the Kontekst.io portal (Plahuta, 2020); available at request[5];
- ELMo (Peters et al., 2018): 1024-dimensional vectors, contextual embeddings built in the EU EMBEDDIA project, trained on Gigafida (Ulčar, 2019). Contextual embeddings produce a different vector for each occurrence of the word based on its context. We computed word vectors from sentences in Slovene Wikipedia. To get a single representation for each word, comparable to other embeddings, for each of the 200,000 most common words, we calculated the centroid vector of all word occurrences. Several different types of vectors were used:
  - vectors from the output of the first (CNN) layer of the network that is context-independent (i.e. *layer 0*),

[4] http://embeddia.eu/

[5] https://kontekst.io/kontakt

  - vectors from the output of the second (first LSTM) layer of the network that is context-dependent (i.e. *layer 1*),
  - vectors from the output of the third (second LSTM) layer of the network that is context-dependent (i.e. *layer 2*).

3.2.2 Croatian word embedding model

For the Croatian language, we analyse several non-contextual embedding models:

- fastText (Bojanowski et al., 2017):
  - 100-dimensional vectors, trained in the EU EMBEDDIA project,
  - 300-dimensional vectors, trained as above,
  - 100-dimensional CLARIN.SI-embed.hr vectors of words and lemmas (Ljubešić, 2018),
  - 300-dimensional vectors from the fastText.cc portal.

4 EVALUATION METHODOLOGY

To assess the gender bias for each of the embedding models and each occupation, we calculated occupational analogies in four ways. However, the core analogy computation is the same in all cases: for every occupation of a masculine grammatical gender O_m, we search for a feminine noun equivalent O_f. The following vector is calculated:

$$v(d) = v(O_m) - v(m) + v(f),$$

where $v(m)$ is the male vector, and $v(f)$ is the female vector. If there were no gender biases, $v(d)$ would be equal or very similar to $v(O_f)$. For every vector $v(d)$, we find N closest word vectors according to the cosine similarity (we use $N = 1$, 5, or 10). When searching for closest words, all words appearing in the embeddings are considered, except for the words *man*, *woman*, the word O_m, and the words containing non-alphabetic characters (numbers, hyphens, punctuation etc.). If the word O_f is located among the N-closest words, we consider the analogy correct; else it is marked as incorrect. We convert all letters to lowercase: e.g. the words *Zdravnik*, *zdravnik* and *ZDRAVNIK* are all converted to *zdravnik* and thus considered the same word. The process is repeated for each female variant of an occupation O_f where we look for the male equivalent O_m. Here, the vector *v(d)* is calculated as:

$$v(d) = v(O_f) - v(f) + v(m).$$

When looking for closest words, O_f is omitted from the set of words, just as O_m was ignored before. The final result represents the proportion of correctly determined cases. The metric is called *precision at N* (*P@N*). A higher *N* allows for finding additional closest hits in the vector space.

Two approaches were used to determine the baseline male vector *v(m)* and female vector *v(f)*:

- The first approach defines *m* simply as the word *man* and *f* as *woman* (in Slovene corresponding to *moški* and *ženska* and in Croatian to *muškarac* and *žena).*
- In the second approach, similarly to Bolukbasi et al. (2016), the difference *v(f) −v(m)* or *v(m) −v(f)* is defined as the average difference of vectors of word pairs which refer specifically to a woman or man (Table 1).

Table 1: *Inherently male-female word pairs in Slovene (left) and Croatian (right)*

Slovene male-female word pairs		**Croatian male-female word pairs**	
m	***f***	***m***	***f***
moški [man]	ženska [woman]	muškarac [man]	žena [woman]
gospod [sir]	gospa [madam]	gospodin [sir]	gosopođa [madam]
fant [boy]	dekle [girl]	momak [boy]	djevojka [girl]
deček [boy]	deklica [girl]	dječak [boy]	djevojčica [girl]
brat [brother]	sestra [sister]	brat [brother]	sestra [sister]
oče [father]	mati [mother]	otac [father]	majka [mother]
sin [son]	hči [daughter]	sin [son]	kći [daughter]
dedek [grandfather]	babica [grandmother]	djed [grandfather]	baka [grandmother]
mož [husband]	žena [wife]	suprug [husband]	supruga [wife]
on [he]	ona [she]	on [he]	ona [she]
fant [boy]	punca [girl]	tata [dad]	mama [mum]
stric [uncle]	teta [aunt]		

When searching for the N closest words, we also tested lemmatisation's influence: in this case, all words in word embeddings were lemmatised using the LemmaGen[6] tool. By doing so, the effect of different word forms stemming from, e.g. conjugation and declination, was offset: for example, word forms *zdravnico* and *zdravnice* are considered a single near word since they share the same lemma *zdravnica* [doctor$_F$].

5 RESULTS

We present the results showing biases in all embeddings described in Section 3. We use the *P@N* measure*,* where *N* equals 1, 5, or 10. Some of the occupations from our list are not covered by all word embeddings, i.e. there is no word vector for them. Any example where the searched-for word is not among the top *N* closest words is counted as incorrect, even if the searched-for word does not appear in the embeddings. In cases where the embeddings do not cover the input occupation, and we cannot calculate the vector *v(d),* we dismiss all such examples so that they do not affect the final result. The reader, interested in the results where non-covered examples are also considered, is referred to our conference paper (Supej et al., 2020).

The results for Slovene analogies are presented in Table 2 and for the Croatian analogies in Table 3. Results for experiments where we have a masculine expression for the occupation O_m as the input, and we search for the equivalent feminine expression of the same occupation O_f, are shown in the rightmost columns (*m* input) for each language. Results, where we have O_f as the input and search for O_m, are shown in leftmost columns (*f* input) for each language. As explained in Section 4, we tested different approaches. The approaches where we lemmatised all the words or used the average difference of vectors of pairs of words from Table 1 generally perform better (i.e. they express lower gender bias). These two options have the suffixes *lem* and *avg* appended in the tables, respectively. In this section, we only show the results for applying both of these options (we do not apply lemmatisation to fastText (lemma) embeddings as they are already lemmatised). Full results are presented in Appendix A in Table 8 for Slovenian and in Table 9 for Croatian.

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

Table 2: *Results for all Slovenian embeddings*

Slovene word embeddings	dimensions and approach	***f*** **input**			***m*** **input**		
		P@1	P@5	P@10	P@1	P@5	P@10
ELMo Embeddia	1024D l0 lem avg	0.907	0.933	0.947	0.370	0.398	0.403
	1024D l1 lem avg	0.907	0.947	0.947	0.381	0.392	0.398
	1024D l2 lem avg	0.880	0.933	0.933	0.376	0.398	0.398
fastText.cc	300D lem avg	0.613	0.884	0.948	0.655	0.755	0.764
fastText Embeddia	100D lem avg	0.906	0.971	0.976	0.677	0.720	0.724
	300D lem avg	**0.947**	**0.976**	**0.982**	0.685	0.720	0.724
fastText CLARIN.SI-embed.sl	100D lem avg	0.839	0.940	0.950	**0.761**	**0.880**	**0.902**
fastText Sketch Engine (word)	100D lem avg	0.930	0.962	0.973	0.725	0.781	0.785
fastText Sketch Engine (lemma)	100D avg	0.673	0.931	0.960	0.598	0.786	0.821
word2vec Kontekst.io	256D lem avg	0.679	0.853	0.872	0.407	0.550	0.593

Note. Results for each approach, where we have a feminine word for occupation on the input (*f* input), and we search for the equivalent masculine term, and where we have a masculine word for occupation on the input (*m* input), and we search for the equivalent feminine term. The examples where the embeddings do not cover the input occupation were dismissed. The best result in each column is in bold.

Table 3: *Results for all Croatian embeddings*

Croatian word embeddings	dimensions and approach	***f*** **input**			***m*** **input**		
		P@1	P@5	P@10	P@1	P@5	P@10
fastText.cc	300D lem avg	0.731	0.939	0.954	0.546	0.637	0.644
fastText Embeddia	100D lem avg	0.905	0.941	0.968	0.625	0.666	0.672
	300D lem avg	**0.923**	**0.982**	**0.986**	0.631	0.675	0.678
fastText CLARIN.SI-embed.hr (word)	100D lem avg	0.907	0.930	0.944	**0.673**	**0.746**	**0.754**
fastText CLARIN.SI-embed.hr (lemma)	100D avg	0.244	0.678	0.826	0.266	0.521	0.588

Note. For each approach, where we have a feminine word for occupation on the input (*f* input) and we search for the equivalent masculine term, and where we have a masculine word for occupation on the input (*m* input) and we search for the equivalent feminine term. The examples where the embeddings do not cover the input occupation were dismissed. The best result in each column is in bold.

The results show that both lemmatisation of the words and using the average of several inherently male or female words for male and female vectors improve the reported scores. Applying both approaches gives the best results in most cases. For finding the closest *N* words, we have also tried the CSLS measure (Cross-Domain Similarity Local Scaling) (Conneau et al., 2018) instead of the cosine similarity. This measure avoids the problem of hubness in the search for nearest neighbours. Namely, some words (called hubs in the nearest neighbour graph representation) may be nearest neighbours of many other words, while others are nearest neighbours of no other word (outliers). CSLS computes nearest neighbours in both directions and largely avoids the problem of hubness. For the experiments with O_f on the input and searching for O_m, there is no significant difference in results between the cosine similarity and CSLS. For the experiments with O_m on the input and searching for O_f, using CSLS gives lower precision than the cosine similarity. This is especially the case where we used the words "man" and "woman" for vectors *v(m)* and *v(f)*. When using averages of several inherently male and female words for vectors *v(m)* and *v(f)*, the difference in precision between the cosine similarity and CSLS is smaller, but the cosine similarity still outperforms CSLS.

We give a more detailed discussion of the results for each approach in the next section. We only present the results of the cosine similarity measure.

6 DISCUSSION

In the case of Slovene word embeddings, the fastText CLARIN.SI-embed.sl embeddings reach the highest precision in the analogy task for male versions of occupations at the input (Table 2). When there are female versions of occupations at the input, the embedding model reaching the highest precision is fastText Embeddia. Similar results are observed for Croatian embeddings (Table 3). Lemmatisation of the output and averaging several inherently male and female words for vectors *v(m)* and *v(f)* (instead of using only the embeddings for *woman* or *man*) improves the precision in the analogy task for different models and different input data. As described in Section 5, we dismiss the examples where the embeddings do not cover the input occupation. If we do not dismiss these examples but instead count them as incorrect, the share of occupations covered by the embeddings has the largest effect on the score. The results for Slovene can be found in our paper (Supej et al., 2020). The fastText CLARIN.SI embeddings would then score the best, as these embeddings cover the occupations best. This is especially important for the female occupations since they have much lower coverage than male occupations.

Results in Table 2 and Table 3 have been filtered, so that the words *man, woman* and the occupation on the input are removed from the list of analogy results, as explained in Section 4. With unfiltered results, the input occupation is often the result of the analogy task (Table 4). For more detailed results (not only with lemmatisation and using several inherently male and female words for *v(m)* and *v(f)*) see Table 10 in Appendix A.

With the fastText Embeddia model, we reach similar results using 100- and 300-dimensional vectors (see Table 2 and Table 3). Other embeddings are not directly comparable with regards to dimensionality as they were trained on different resources. However, corpora used to train the embeddings play a more important role than the number of dimensions. The FastText Embeddia model in Table 4 shows that dimensionality plays a role in determining how often the input occupation is the result of the analogy. In a different setup, when considering the occupations that are not covered in the embeddings, dimensionality strongly influences the results (Supej et al., 2020).

Table 4: *Share of cases where the result of the analogy with the highest cosine similarity is the input occupation itself - before filtering is done to produce the results in Table 2 and Table 3 (both male to female and female to male analogies)*

Slovene word embeddings	Dimensions and approach	Share of outputs equal to inputs	Croatian word embeddings	Dimensions and approach	Share of outputs equal to inputs
ELMo Embeddia	1024D l0 lem avg	0.547			
	1024D l1 lem avg	0.423			
	1024D l2 lem avg	**0.064**			
fT fastText.cc	300D lem avg	0.831	fT fastText.cc	300D lem avg	0.672
fT Embeddia	100D lem avg	0.143	fT Embeddia	100D lem avg	**0.094**
	300D lem avg	0.419		300D lem avg	0.352
fT CLARIN.SI-embed.sl (word)	100D lem avg	0.316	fT CLARIN.SI-embed.hr (word)	100D lem avg	0.103
fT Sketch Engine (word)	100D lem avg	0.096			
fT Sketch Engine (lemma)	100D avg	0.803	fT CLARIN.SI-embed.hr (lemma)	100D avg	0.837
w2v Kontekst.io	256D lem avg	0.483			

Note. The number of all cases is 468 (from 234 occupation pairs) for Slovene and 750 (from 375 occupation pairs) for Croatian.

The coverage of masculine occupations is higher than that of feminine occupations in all word embedding models (Table 5). FastText CLARIN.SI-embed.sl word embeddings achieve the highest coverage of female occupations, while ELMo word embeddings contained only 75 of the 234 female occupations. As explained in Section 3.2.1, ELMo embeddings are limited to only 200,000 most common words in Wikipedia; therefore, we have significantly lower coverage of occupations for ELMo. For comparison, other word embedding models cover around 1 million words. Masculine occupations that do not appear in the embeddings are typically occupations associated with women (e.g. male variants of *seamstress* and *cosmetician*, in Slovene *šiviljec* and *kozmetik*, respectively). Likewise, feminine occupations not present in the embeddings are traditionally male occupations (e.g. embedding models do not contain female variants of occupations like *auto mechanic* and *carpenter* (in Slovene *avtomehaničarka* and *tesarka*, respectively), or occupations that have been culturally taken up exclusively by men, e.g., *nadškof* (en. *archbishop*). Poor representation of female occupations can also be attributed to other factors – Zhao et al. (2018) report that the mentions referring to men are more likely to contain a job title compared to female mentions.

Table 5: *Coverage of male (m) and female (f) occupations from the list in different embeddings as a ratio between covered occupations and all occupations*

Slovene embeddings	*m*	*f*	Croatian embeddings	*m*	*f*
ELMo	0.774	0.321			
fastText cc	0.979	0.739	fastText cc	0.848	0.527
fastText Embeddia	0.991	0.726	fastText Embeddia	0.856	0.594
fastText CLARIN.SI-embedd.sl	**1.000**	**0.932**	fastText CLARIN.SI-embedd.hr (word)	0.914	**0.722**
fastText Sketch Engine (word)	0.996	0.791	fastText CLARIN.si-embedd.hr (lemma)	**0.955**	**0.722**
fastText Sketch Engine (lemma)	**1.000**	0.863			
word2vec Kontekst.io	0.987	0.667			

Nissim et al. (2019) claim that most studies exaggerate biases pointed out by analogy tasks. The design of these studies excludes the input occupation from the possible results, even if the calculations could lead to this exact occupation to have the highest cosine similarity and hence appear in the results. This criticism is more relevant for English studies as in Slovene the gender in occupations is for the most part expressed by word morphology. Even though we omitted the input occupations from the results, which is a standard practice when calculating analogies, we analysed the results before this filtering. Analysis of the results showed that the input occupation is indeed often the result with the highest cosine similarity (Table 4), varying significantly between different models.

When manually comparing the results of different models from Tables 2 and 3, we also notice several differences between the models. In the case of ELMo and word2vec models, the outputs are largely occupations. The results of the analogy task in the case of fastText Embeddia, CLARIN.SI-embed.sl and Sketch Engine (word) are occupations, as well as words related to the occupation on the input, or words that share the same root as the input occupation. Results of the fastText.cc and Sketch Engine (lemma) models are typically words sharing the root with the input occupation.

Analogy results are interesting from a semantic point of view. The first results of the analogy task (Slovene "fastText Embeddia 100D lem avg") *ženska:krojačica :: moški:x* being *x=krojač* [en. *woman:tailor$_F$ :: man:tailor$_M$*] and *ženska:šivilja :: moški:x* being *x=krojač* [en. *woman:seamstress :: man:tailor*] are interesting. For example, while word embedding of *šiviljec* [en. *seamster*] is not available, *krojač* [en. *tailor*], a semantically linked one, from another morphological word family is. Another interesting element is illustrated by one of the results of the analogy: *ženska:manekenka :: moški:x* where *x=nogometaš* [en. *woman:model :: man:footballer*] (Croatian "fastText Embeddia 100D lem avg"). While *model* and *footballer* are not corresponding to the same professions, this result is an indication that female models and male footballers appear in similar textual contexts. It would be interesting to investigate those contexts further (e.g. both occupations represent desirable identities, such as being beautiful, rich, famous, successful).

There are indeed more examples where results of certain analogies (especially in the case of "word2vec Kontekst.io lem avg model") are not linked to the input occupation or are stereotypical. For example, the results of the analogy *moški:rudar :: ženska:x* in the aforementioned w2v model are, e.g. *barbika* [en. *barbie*], *klovnesa* [en. *clown$_F$*], *čarovnica* [en. *witch*], *lutka* [en. *doll*], *prostitutka* [en. *prostitute$_F$*], *akrobatka* [en. *acrobat$_F$*], *najstnica* [en.

teenager$_F$], *opica* [en. *monkey*], *princeska* [en. *princess*], *striptizeta* [en. *stripper$_F$*]. The case of stereotypical analogies in the w2v model is pointed out by Supej et al. (2019).

As part of the analysis, a frequency list of analogy results for female and male input occupations was compiled for each word embedding model (only the *lem avg* configuration of the models was taken into account) (see Table 6 for Slovene and Table 7 for Croatian).

The most frequently occurring words mostly follow the pattern that for a male occupation on the input, a female occupation is expected on the output. Presented Slovene embedding models follow this pattern; in the case of the Croatian embeddings, there are several examples among the frequently occurring words that do not follow the pattern: in the "fastText cc lem avg" with a female occupation on the input, there are several frequently occurring female occupation variants also on the output, e.g. *ethicist*, *biologist* (*etičarka*, *biologinja*, respectively). For *etičarka,* it is possible that this result is influenced by other similar words (e.g. *kozmetičarka*), as fastText models consider subword information. The most frequently occurring words are primarily occupations but not always – for example, female Scottish national (*Škotkinja*) and *father* (*otac*) frequently appear in the Croatian "fastText cc lem avg" model while one of the frequent words in the Slovene "word2vec Kontekst.io lem avg" is *korenjak* (denoting a brave man).

In Slovene word embeddings, we notice a pattern of the most frequently occurring feminine occupations/words appearing more often than the most frequently occurring male occupations in the "ELMo l2 lem avg" and "w2v Kontekst.io lem avg" models. Similar is observed for Croatian models presented in Table 7; however, the most frequently occurring words appear less often than in the Slovene embeddings. One possible explanation is that the models mentioned above contain fewer word embeddings than some other models (200,000 or approximately 600,000 for each model). Both models exhibit a lower representation of the female versions of occupations in the embeddings. Occupations that nevertheless appear in the embeddings, therefore, reappear more often. There are overall more male occupations in the embeddings, possibly causing individual male occupations to come up less frequently than female ones.

Table 6: *Most common words that appear among the top 10 results of the analogy task (that is, among the 10 closest words to the searched-for term, based on the cosine similarity measure) for selected Slovene embedding models*

ELMo Embeddia l2 lem avg				fastText CLARIN.SI lem avg				word2vec Kontekst.io lem avg			
***m* input**		***f* input**		***m* input**		***f* input**		***m* input**		***f* input**	
Result	***n***	**Result**	***n***	**Result**	***n***	**Result**	***n***	**Result**	***n***	**Result**	***n***
bolničarka [nurse]	47	geograf [geographer$_M$]	9	šivilja [seamstress]	15	mizar [carpenter$_M$]	11	kuharica [cook$_F$]	44	ortoped [orthopedist$_M$]	14
biokemičarka [biochemist$_F$]	39	politolog [political scientist$_M$]	8	ključavničarka [locksmith$_F$]	11	biology [biologist$_M$]	10	gospodinja [homemaker$_F$]	38	pisatelj [writer$_M$]	14
frizerka [hairdresser$_F$]	39	biolog [biologist$_M$]	7	inštalaterka [installer$_F$]	9	ključavničar [locksmith$_M$]	9	šivilja [seamstress]	33	kardiolog [cardiologist$_M$]	13
trgovka [salesperson$_F$]	39	dramaturg [playwright$_M$]	7	keramičarka [ceramist$_F$]	9	zgodovinar [historian$_M$]	9	frizerka [hairdresser$_F$]	32	nevrolog [neurologist$_M$]	13
čistilka [cleaner$_F$]	34	književnik [writer$_M$]	7	filologinja [philologist$_F$]	8	internist [internist$_M$]	8	kozmetičarka [cosmetician$_F$]	30	urolog [urologist$_M$]	13
znanstvenica [scientist$_F$]	34	scenarist [screenwriter$_M$]	7	oftalmologinja [ophthalmologist$_F$]	8	režiser [director$_M$]	8	čistilka [cleaner$_F$]	29	psihiater [psychiatrist$_M$]	12
kuharica [cook$_F$]	33	animator [animator$_M$]	6	filozofinja [philosopher$_F$]	7	arheolog [archeologist$_M$]	7	fotografinja [photographer$_F$]	29	ekolog [ecologist$_M$]	11
geologinja [geologist$_F$]	30	esejist [essayist$_M$]	6	geofizičarka [geophysicist$_F$]	7	natakar [waiter$_M$]	7	zdravnica [doctor$_F$]	29	hišnik [janitor$_M$]	11
perica [laundress]	28	etnolog [ethnologist$_M$]	6	kmetica [farmer$_F$]	7	pisatelj [writer$_M$]	7	služkinja [maid]	26	biolog [biologist$_M$]	10
služkinja [maid]	28	fotograf [photographer$_M$]	6	nevrokirurginja [neurosurgeon$_F$]	7	primarij [senior doctor$_M$]	7	trgovka [salesperson$_F$]	26	korenjak [brave man]	10
biologinja [biologist$_F$]	27	illustrator [illustrator$_M$]	6	strugarka [worker using a planer machine$_F$]	7	stomatolog [stomatologist$_M$]	7	slikarka [painter$_F$]	25	maneken [model$_M$]	10
gospodinja [homemaker$_F$]	26	lutkar [puppeteer$_M$]	6	geologinja [geologist$_F$]	6	tesar [carpenter$_M$]	7	tajnica [secretary$_F$]	25	režiser [director$_M$]	10
matematičarka [mathematician$_F$]	26	paleontolog [paleontologist$_M$]	6	hematologinja [hematologist$_F$]	6	fotoreporter [photojournalist$_M$]	6	veterinarka [veterinarian$_F$]	25	akademik [academic$_M$]	9
mikrobiologinja [microbiologist$_F$]	26	pravnik [jurist$_M$]	6	kardiologinja [cardiologist$_F$]	6	gostilničar [innkeeper$_M$]	6	znanstvenica [scientist$_F$]	25	akademski slikar [academic painter$_M$]	9
arheologinja [archeologist$_F$]	25	režiser [director$_M$]	6	paleontologinja [paleontologist$_F$]	6	kardiolog [cardiologist$_M$]	6	socialna delavka [social worker$_F$]	24	glasbenik [musician$_M$]	9

Table 7: *15 most common words that appear among the top 10 results of the analogy task (that is, among the 10 closest words to the searched-for term, based on the cosine similarity measure) for selected Croatian embedding models*

ELMo Embeddia l2 lem avg				**fastText cc lem avg**				**fastText CLARIN.SI-embedd.hr (word) lem avg**			
***m* input**		***f* input**		***m* input**		***f* input**		***m* input**		***f* input**	
Result	***n***	**Result**	***n***	**Result**	***n***	**Result**	***n***	**Result**	***n***	**Result**	***n***
krojačica [tailor$_{F}$]	34	povjesničar [historian$_{M}$]	10	kemičarka [chemist$_{F}$]	12	etičarka [ethicist$_{F}$]	8	krojačica [tailor$_{F}$]	31	znanstvenik [scientist$_{M}$]	16
automehaničarka [auto mechanic$_{F}$]	29	konobar [waiter$_{M}$]	10	vještakinja [expert$_{F}$]	11	otfamologinja [ophthalmologist $_{F}$]	7	automehaničarka [auto mechanic$_{F}$]	23	biology [biologist$_{M}$]	16
zavarivačica [welder$_{F}$]	20	biolog [biologist$_{M}$]	9	fizičarka [physicist$_{F}$]	10	redatelj [director$_{M}$]	6	zavarivačica [welder$_{F}$]	22	profesor [professor$_{M}$]	9
keramičarka [ceramist$_{F}$]	16	umjetnik [artist$_{M}$]	8	biokemičarka [biochemist$_{F}$]	10	glumac [actor$_{M}$]	6	šivačica [seamstress]	18	povjesničar [historian$_{M}$]	9
kemičarka [chemist$_{F}$]	15	sociolog [sociologist$_{M}$]	8	vozačica [driver$_{F}$]	9	biologinja [biologist$_{F}$]	6	keramičarka [ceramist$_{F}$]	18	konobar [waiter$_{M}$]	9
biokemičarka [biochemist$_{F}$]	15	fizioterapeut [physiotherapist$_{M}$]	8	pravnica [jurist$_{F}$]	9	paleografkinja [paleographer$_{F}$]	5	soboslikarica [painter-decorator$_{F}$]	17	genetičar [geneticist$_{M}$]	9
šivačica [seamstress]	14	redatelj [director$_{M}$]	7	frizerka [hairdresser$_{F}$]	9	ihtiologinja [ichthyologist$_{F}$]	5	biokemičarka [biochemist$_{F}$]	16	redatelj [director$_{M}$]	8
spremačica [maid]	14	poslovođa [manager$_{F/M}$]	7	masažerka [massage therapist$_{F}$]	8	suscenarist [co-screenwriter$_{M}$]	4	kemičarka [chemist$_{F}$]	15	poslovođa [manager$_{F/M}$]	8
čistačica [cleaner$_{F}$]	13	paleontolog [paleontologist$_{M}$]	7	tehničarka [technician$_{F}$]	7	scenografkinja [scenographer$_{F}$]	4	genetičarka [geneticist$_{F}$]	13	policajac [police officer$_{M}$]	8
genetičarka [geneticist$_{F}$]	13	književnik [writer$_{M}$]	7	političarka [politician$_{F}$]	7	otac [father]	4	cvjećarka [florist$_{F}$]	12	zaposlenik [employee$_{M}$]	7
fizičarka [physicist$_{F}$]	13	geologinja [geologist$_{F}$]	7	matematičarka [mathematician$_{F}$]	7	književnik [writer$_{M}$]	4	biofizičarka [biophysicist$_{F}$]	12	umjetnik [artist$_{M}$]	7
astrofizičarka [astrophysicist$_{F}$]	13	dramaturg [playwright$_{M}$]	7	lutkarica [puppeteer$_{F}$]	7	dopukovnik [lieutenant colonel$_{M}$]	4	znanstvenica [scientist$_{F}$]	11	sociolog [sociologist$_{M}$]	7
šnajderica [seamstress]	12	znanstvenik [scientist$_{M}$]	6	glumica [actor$_{F}$]	7	daktilografkinja [typist$_{F}$]	4	geologinja [geologist$_{F}$]	11	snimatelj [cameraman]	7
mehaničarka [mechanic$_{F}$]	12	zaštitar [security guard$_{M}$]	6	trgovkinja [salesperson$_{F}$]	6	astrobiologinja [astrobiologist$_{F}$]	4	tehničarka [technician$_{F}$]	10	satnik [captain$_{M}$]	7
informatičarka [computer scientist$_{F}$]	12	sociologinja [sociologist$_{F}$]	6	terapeutkinja [therapist$_{F}$]	6	škotkinja [Scottish national$_{F}$]	3	mehaničarka [mechanic$_{F}$]	10	porter [doorkeeper$_{M}$]	7

In the case of the Slovene "ELMo l2 lem avg" and "w2v Kontekst.io lem avg" models, occupations of a lower social class (*čistilka* [en. *cleaner*$_{\mathrm{F}}$], *perica* [en. *laundress*], *gospodinja* [en. *homemaker*$_{\mathrm{F}}$]), as well as archaic occupations with women in inferior roles (*služkinja* [en. *maid*]) are observed among the frequent analogy results of female grammatical gender. Socially inferior occupations are rare among the most frequent male analogies. There are less socially inferior occupations observed among the Croatian results (exceptions being, e. g., the female variants of *cleaner* and *maid* (*čistačica* and *spremačica,* respectively) in the "ELMo Embeddia l2 lem avg" model).

We observed that certain words (especially female occupations) appear among the results despite being semantically unrelated to the input occupation. Several analogy results (especially in the case of a typical male occupation on the input) are unrelated to the input occupation (e.g. *bolničarka* [en. nurse$_{\mathrm{F}}$] is the first result of the analogy *moški:rudar :: ženska:x* [en. *man:miner :: woman:x*] and *šivilja* [en. *seamstress*] the first result of the analogy *moški:avtomehanik :: ženska:x* [en. *man:auto mechanic :: woman:x*] in the Slovene model "fastText Embeddia 100D lem avg"). One explanation is that certain word embeddings are more "central" than the others and, therefore, the closest neighbour of many other words. To check if this explanation is true, instead of the cosine similarity measure, we used the CSLS measure (Conneau et al., 2018) that considers the shared distances of *N* closest neighbours. We observed that the precision is worse when using the CSLS measure than the cosine similarity (Section 5), and therefore we do not report these results. However, when observing the most common words, returned as the analogy task results (Table 6 and Table 7), the distribution of the most common words is more uniform when using the CSLS measure.

Direct comparison of models between Croatian and Slovene is not possible, as the embeddings are trained on different text corpora, and the professions used for analogy calculations are not the same. However, we can notice that in Croatian the occupational gender bias in tested embeddings is slightly higher. Interestingly, the statistical data shows that the employment gap and the pay gap between women and men are lower in Slovenia compared to Croatia (Eurostat, 2021). In future, it would be interesting to study if the female employment rate and gap, as well as the gap in salaries for the same professions between countries, is correlated with the gender bias in embeddings models trained on the corresponding national languages and the changes of this correlation through time.

7 CONCLUSIONS AND FURTHER WORK

We evaluated different Slovene and Croatian word embeddings on analogies of male and female occupations (using different configurations and approaches to calculate analogies). Our focus is on the quantitative evaluation, and the results may be informative for developers of NLP tools. The lowest gender bias was obtained using the fastText embeddings. In finding female analogies (male occupation on the input), the best performing models proved to be fastText CLARIN.SI-embed.sl and fastText CLARIN.SI-embed.hr for Slovene and Croatian, respectively, while the best performing models for finding male analogies (female occupation on the input) were the respective fastText Embeddia models. The approach where averages of several inherently male and female words were used instead of using only the embeddings for woman or man improved the results. Lemmatization likewise improves the precision. With female occupations at the input, the best results (P@10) of 0.982 and 0.986 are achieved using the "fastText Embeddia 300D lem avg" models for Slovene and Croatian, respectively (the examples where the embeddings do not cover the input occupation were dismissed). With male occupations on the input, the best results of 0.902 and 0.754 are produced by the "fastText CLARIN.SI-embed.sl 100D lem avg" and "fastText CLARIN.SI-embed.hr 100D (lem) avg" (cases where the input occupation is not present among the embeddings were likewise dismissed). Lowest results for male input reflect lower coverage of female occupation equivalents in the embeddings model. The "fastText CLARIN.SI-embed.sl" and "fastText CLARIN.si-embedd.hr (lemma)" models contain the highest ratio of searched-for female and male occupations. The qualitative analysis identifies the word2vec Kontekst.io model as the model with the highest degree of gender bias in the results (stereotypically male/female occupations appearing among the results regardless of the grammatical gender of the input occupation).

In future work, we will focus on a detailed qualitative analysis and the relationship between word embeddings, language, and social power. Moreover, we will align occupations in Slovene and Croatian. Further work will also encompass an evaluation of BERT contextual embeddings and experiments in other languages. The impact of the gender bias will be tested in predictive models on practical tasks such as the sentiment analysis.

Acknowledgments

The research was supported by the Slovene Research Agency through research core funding no. P6-0411 and P2-103, as well as project no. J6-2581. This paper is supported by European Union’s Horizon 2020 Programme project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media, grant no. 825153). The results of this paper reflect only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *TEXT*, *23*, 321–346.

Baker, P. (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.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS’16)* (pp. 4356–4364).

Bordia, S., & Bowman, S. (2019). Identifying and Reducing Gender Bias in Word-Level Language Models. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop,* (pp. 7–15).

Brunet, M. E., Alkalay-Houlihan, C., Anderson, A., & Zemel, R. S. (2019). Understanding the Origins of Bias in Word Embeddings. *Proceedings of International Conference on Machine Learning (ICML 2019).*

Caldas-Coulhard, C. R., & Moon, R. (2010). ‘Curvy, hunky, kinky’: Using corpora as tools for critical analysis. *Discourse & Society*, *21*(2), 99–133.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora necessarily contain human biases. *Science*, *356*(6334), 183–186.

Conneau, A., Lample, G., Ranzato, M., Denoyer, L., & Jegou, H. (2018). Word translation without parallel data. *Proceedings of the International Conference on Learning Representation (ICLR).*

Dobrovoljc, K., Krek, S., Holozan, P., Erjavec, T., Romih, T., Arhar Holdt, Š., Čibej, J., Krsnik L., & Robnik-Šikonja, M. (2019). Morphological lexicon Sloleks 2.0. CLARIN.SI. http://hdl.handle.net/11356/1230

Eurostat (2021). Gender statistics. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_statistics#Labour_market

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *PNAS*, *115*(16).

Garimella, A., Banea, C., Hovy, D., & Mihalcea, R. (2019). Women's syntactic resilience and men's grammatical luck: Gender-bias in part-of-speech tagging and dependency parsing. *Proceedings of the 57th Annual Meeting of the ACL* (pp. 3493–3498).

Gigafida 2.0. Retrieved from https://viri.cjvt.si/gigafida

Gonen, H., & Goldberg, Y. (2019). Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. *Proceedings of NAACL-HLT 2019* (pp. 609–614).

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

Hill, B., & Shaw, A. (2013). The Wikipedia gender gap revisited: Characterising survey response bias with propensity score estimation. *PloS One*, 8.

Hirasawa, T., & Komachi, M. (2019). Debiasing Word Embeddings Improves Multimodal Machine Translation. *Proceedings of Machine Translation Summit XVII, Vol. 1* (pp. 32–42).

Hovy, D., & Søgaard, A. (2015). Tagging performance correlates with author age. *Proceedings of the 53rd Annual Meeting of the ACL and the 7th IJC-NLP* (pp. 483–488).

Hovy, D. (2015). Demographic factors improve classification performance. *Proceedings of the 53rd Annual Meeting of the ACL and the 7th IJCNLP* (pp. 752–762).

Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., & Denuyl, S. (2020). Social Biases in NLP Models as Barriers for Persons with Disabilities. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5491–5501).

Kern, B., & Dobrovoljc, H. (2017). Pisanje moških in ženskih oblik in uporaba podčrtaja za izražanje »spolne nebinarnosti«. Jezikovna svetovalnica. Retrieved from https://svetovalnica.zrc-sazu.si/topic/2247/pisanje-mo%C5%A1kih-in-%C5%BEenskih-oblik-in-uporaba-pod%C4%8Drtaja-za-izra%C5%BEanje-spolne-nebinarnosti

Kiritchenko, S., & Mohammad, S., (2018). Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics* (pp. 43–53).

Koolen, C., & van Cranenburgh, A. (2017). These are not the stereotypes you are looking for: Bias and fairness in authorial gender attribution. *Proceedings of the First Ethics in NLP workshop* (pp. 12–22).

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

Liang, P. P, Li, I. M., Zheng, E., Lim, Y. C., Salakhutdinov, R., & Morency, L. (2020). Towards Debiasing Sentence Representations. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5502–5515).

Ljubešić, N., & Erjavec, T. (2018). Word embeddings CLARIN.SI-embed.sl 1.0. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1204

Ljubešić, N. (2018). Word embeddings CLARIN.SI-embed.hr 1.0, Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1205

Martinc, M., Škrjanec, I., Zupan, K., & Pollak, S. (2017). PAN 2017: Author profiling - gender and language variety prediction: notebook for PAN at CLEF 2017. *Proceedings of the Conference and Labs of the Evaluation Forum*.

Mikolov, T., Corrado, G. S., Chen, K., & Dean, J. (2013a). Efficient estimation of word representations in vector space. *Proceedings of the International Conference on Learning Representations* (pp. 1–12).

Mikolov, T., Yih, W-t., & Zweig, G. (2013b). Linguistic regularities in continuous space word representations. *Proceedings of the 2013 Conference of the North American Chapter of the ACL: Human Language Technologies* (pp. 746–751).

Nozza, D., Volpetti, C., & Fersini, E. (2019). Unintended Bias in Misogyny Detection. *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 149–155).

Nissim, M., van Noord, R., & van der Goot, R. (2019). Fair is better than sensational: Man is to doctor as woman is to doctor. *Computational Linguistics*, *46*(3), 487–497.

Pearce, M. (2008). Investigating the collocational behaviour of man and woman in the BNC using Sketch Engine. *Corpora*, *3*(1), 1–29.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualised word representations. *Proceedings of NAACL-HLT 2018* (pp. 2227–2237).

Plahuta, M. (2020). O slovarju. Retrieved from https://kontekst.io/oslovarju

Popič, D., & Gorjanc, V. (2018). Challenges of adopting gender-inclusive language in Slovene. *Suvremena lingvistika, 44*(86), 329–350.

Prates, M. O. R., Avelar, P. H., & Lamb, L. C. (2020). Assessing gender bias in machine translation: A case study with Google Translate. *Neural Computing and Applications*, *32*, 6363–6381.

Rangel, F., Celli, F., Rosso, P., Potthast, M., Stein, B., & Daelemans, W. (2015). Overview of the 3rd author profiling task at PAN 2015. In L. Cappellato, N. Ferro, G. J. F. Jones in E. SanJuan (Eds.), *CLEF 2015 Labs and Workshops, Notebook Papers*.

Schick, T., Udupa, S., & Schütze, H. (2021). Self-Diagnosis and Self-Debiasing: A Proposal for Reducing Corpus-Based Bias in NLP. *arXiv preprint arXiv:2103.00453*.

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K-W., & Wang, W. Y. (2019). Mitigating gender bias in natural language processing: Literature review. *Proceedings of the 57th Annual Meeting of the ACL* (pp. 1630–1640).

Supej, A., Plahuta, M., Purver, M., Mathioudakis, M., & Pollak, S. (2019). Gender, language, and society: Word embeddings as a reflection of social inequalities in linguistic corpora. *Proceedings of the Slovensko sociološko srečanje 2019 – Znanost in družbe prihodnosti* (pp. 75–83).

Supej, A., Ulčar, M., Robnik-Šikonja, M., & Pollak, S. (2020). Primerjava slovenskih besednih vektorskih vložitev z vidika spola na analogijah poklicev. *Proceedings of the Conference on Language Technologies & Digital Humanities 2020* (pp. 93–100).

Svoboda, L., & Beliga, S. (2018). Evaluation of Croatian Word Embeddings. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 1512–1518).

Škrjanec, I., Lavrač, N., & Pollak, S. (2018). Napovedovanje spola slovenskih blogerk in blogerjev. In D. Fišer (Ed.), *Viri, orodja in metode za analizo spletne slovenščine* (pp. 356–373). Ljubljana: Znanstvena založba FF.

Tannen, D. (1990). *You Just Don't Understand: Women and Men in Conversation*. New York: Ballantine Books.

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

Vanmassenhove, E., Hardmeier, C., & Way, A. (2018). Getting gender right in neural machine translation. *Proceedings of the EMNLP* (pp. 3003–3008).

Verhoeven, B., Škrjanec, I., & Pollak, S. (2017). Gender profiling for Slovene Twitter communication: The influence of gender marking, content and style. *Proceedings of the 6th BSNLP Workshop* (pp. 119–125).

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

Volkova, S., Wilson, T., & Yarowsky, D. (2013). Exploring demographic language variations to improve multilingual sentiment analysis in social media. *Proceedings of the EMNLP* (pp. 1815–1827).

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K-W. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. *Proceedings of the EMNLP* (pp. 2979–2989).

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K-W. (2018). Gender bias in coreference resolution: Evaluation and debiasing methods. *Proceedings of the NAACL-HLT* (pp. 15–20).

PRIMERJAVA SLOVENSKIH IN HRVAŠKIH BESEDNIH VEKTORSKIH VLOŽITEV Z VIDIKA SPOLA NA ANALOGIJAH POKLICEV

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. Prispevek se osredotoča na evalvacijo vektorskih vložitev besed v slovenščini in hrvaščini z vidika spola z uporabo besednih analogij. Sestavili smo seznam moških in ženskih samostalnikov za poklice v slovenščini in ovrednotili spolno pristranost modelov vložitev fastText, word2vec in ELMo z različnimi konfiguracijami in pristopi k računanju analogij. Izkazalo se je, da najmanjšo poklicno spolno pristranost vsebujejo vložitve fastText. Tudi za hrvaško evalvacijo smo uporabili sezname poklicev in primerjali različne fastText vložitve.

Ključne besede: besedne vložitve, spolna pristranost, besedne analogije, poklici, obdelava naravnega jezika

APPENDIX 1

We present the results, comparing different approaches described in Section 4 and Section 5. The approach where we lemmatised all the words has the suffix *lem* appended in the tables. The approach where we used the average difference of vectors of pairs of words from Table 1 has the suffix *avg* appended in the tables. The results for Slovene word embeddings are shown in Table 8, the results for Croatian word embeddings in Table 9 and the share of cases, where the input occupation is the result of the analogy task, in Table 10.

Table 8: *Results for Slovenian embeddings*

Slovene word embeddings	**dimensions and approach**	***f* input**			***m* input**		
		P@1	P@5	P@10	P@1	P@5	P@10
ELMo Embeddia	1024D l0 avg	0.707	0.933	0.947	0.166	0.359	0.387
	1024D l0	0.427	0.920	0.947	0.210	0.376	0.398
	1024D l0 lem avg	0.907	0.933	0.947	0.370	0.398	0.403
	1024D l0 lem	0.893	0.947	0.947	0.376	0.392	0.403
	1024D l1 avg	0.907	0.947	0.947	0.381	0.392	0.398
	1024D l1	0.880	0.947	0.947	0.376	0.392	0.392
	1024D l1 lem avg	0.907	0.947	0.947	0.381	0.392	0.398
	1024D l1 lem	0.907	0.947	0.947	0.376	0.392	0.392
	1024D l2 avg	0.880	0.933	0.933	0.376	0.398	0.398
	1024D l2	0.853	0.920	0.933	0.370	0.398	0.398
	1024D l2 lem avg	0.880	0.933	0.933	0.376	0.398	0.398
	1024D l2 lem	0.853	0.920	0.933	0.370	0.398	0.398
fastText.cc	300D avg	0.393	0.798	0.913	0.607	0.738	0.751
	300D	0.150	0.561	0.792	0.445	0.703	0.734
	300D lem avg	0.613	0.884	0.948	0.655	0.755	0.764
	300D lem	0.457	0.861	0.919	0.498	0.725	0.751
fastText Embeddia	100D avg	0.900	0.971	0.976	0.672	0.716	0.720
	100D	0.471	0.871	0.906	0.638	0.716	0.720
	100D lem avg	0.906	0.971	0.976	0.677	0.720	0.724
	100D lem	0.735	0.924	0.941	0.638	0.716	0.720
	300D avg	0.835	0.971	0.976	0.668	0.716	0.724
	300D	0.329	0.859	0.959	0.685	0.720	0.720
	300D lem avg	**0.947**	**0.976**	**0.982**	0.685	0.720	0.724
	300D lem	0.818	0.971	0.976	0.685	0.720	0.720

Slovene word embeddings	**dimensions and approach**	***f* input**			***m* input**		
		P@1	P@5	P@10	P@1	P@5	P@10
fastText CLARIN.SI-embed.sl	100D avg	0.784	0.913	0.940	**0.761**	0.868	0.880
	100D	0.083	0.587	0.780	0.705	0.855	0.885
	100D lem avg	0.839	0.940	0.950	**0.761**	**0.880**	**0.902**
	100D lem	0.651	0.881	0.917	0.709	0.859	0.885
fastText Sketch Engine (word)	100D avg	0.886	0.962	0.973	0.717	0.768	0.777
	100D	0.211	0.757	0.908	0.691	0.768	0.777
	100D lem avg	0.930	0.962	0.973	0.725	0.781	0.785
	100D lem	0.811	0.951	0.962	0.691	0.768	0.781
fastText Sketch Engine (lemma)	100D avg	0.673	0.931	0.960	0.598	0.786	0.821
	100D	0.510	0.812	0.891	0.380	0.658	0.756
word2vec Kontekst.io	256D avg	0.679	0.853	0.872	0.407	0.550	0.593
	256D	0.365	0.590	0.718	0.251	0.489	0.515
	256D lem avg	0.679	0.853	0.872	0.407	0.550	0.593
	256D lem	0.513	0.686	0.795	0.251	0.489	0.519

Note. For each approach, where we have a feminine word for occupation on the input (*f* input) and we search for the equivalent masculine term, and where we have a masculine word for occupation on the input (*m* input) and we search for the equivalent feminine term. The examples where the embeddings do not cover the input occupation were dismissed. The best result in each column is in bold.

Table 9: *Results for Croatian embeddings*

Croatian word embeddings	**dimensions and approach**	***f* input**			***m* input**		
		P@1	P@5	P@10	P@1	P@5	P@10
fastText.cc	300D avg	0.604	0.883	0.944	0.536	0.603	0.609
	300D	0.452	0.838	0.914	0.429	0.599	0.606
	300D lem avg	0.731	0.939	0.954	0.546	0.637	0.644
	300D lem	0.660	0.924	0.954	0.508	0.618	0.634
fastText Embeddia	100D avg	0.896	0.941	0.959	0.625	0.669	0.672
	100D	0.797	0.928	0.937	0.459	0.634	0.656
	100D lem avg	0.905	0.941	0.968	0.625	0.666	0.672
	100D lem	0.833	0.932	0.941	0.503	0.641	0.662
	300D avg	0.829	0.937	0.973	0.616	0.675	0.675
	300D	0.703	0.914	0.950	0.431	0.662	0.672
	300D lem avg	**0.923**	**0.982**	**0.986**	0.631	0.675	0.678
	300D lem	0.865	0.950	0.964	0.578	0.672	0.675

Croatian word embeddings	**dimensions and approach**	***f* input**			***m* input**		
		P@1	P@5	P@10	P@1	P@5	P@10
fastText CLARIN.SI-embed.hr (word)	100D avg	0.896	0.933	0.941	0.670	**0.749**	**0.754**
	100D	0.778	0.904	0.919	0.491	0.699	0.740
	100D lem avg	0.907	0.930	0.944	**0.673**	0.746	**0.754**
	100D lem	0.815	0.904	0.915	0.550	0.711	0.746
fastText CLARIN.SI-embed.hr (lemma)	100D avg	0.244	0.678	0.826	0.266	0.521	0.588
	100D	0.278	0.593	0.693	0.126	0.336	0.406

Note. For each approach, where we have a feminine word for occupation on the input (*f* input) and we search for the equivalent masculine term, and where we have a masculine word for occupation on the input (*m* input) and we search for the equivalent feminine term. The examples where the embeddings do not cover the input occupation were dismissed. The best result in each column is in bold.

Table 10: *Share of cases where the result of the analogy with the highest cosine similarity is the input occupation itself - before filtering is done to produce the results of Tables 2 and 3 (both male to female and female to male analogies)*

Slovene word embeddings	**Dimensions and approach**	**Share of outputs equal to inputs**	**Croatian word embeddings**	**Dimensions and approach**	**Share of outputs equal to inputs**
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	fT fastText.cc	300D avg	0.672
	300D	0.825		300D	0.664
	300D lem avg	0.831		300D lem avg	0.672
	300D lem	0.825		300D lem	0.664

Slovene word embeddings	**Dimensions and approach**	**Share of outputs equal to inputs**	**Croatian word embeddings**	**Dimensions and approach**	**Share of outputs equal to inputs**
fT Embeddia	100D avg	0.143	ft Embeddia	100D avg	**0.094**
	100D	0.141		100D	**0.094**
	100D lem avg	0.143		100D lem avg	**0.094**
	100D lem	0.141		100D lem	**0.094**
	300D avg	0.419		300D avg	0.352
	300D	0.513		300D	0.441
	300D lem avg	0.419		300D lem avg	0.352
	300D lem	0.513		300D lem	0.441
fT CLARIN.SI-embed.sl (word)	100D avg	0.316	fT CLARIN.SI-embed.hr (word)	100D avg	0.103
	100D	0.310		100D	0.114
	100D lem avg	0.316		100D lem avg	0.103
	100D lem	0.310		100D lem	0.114
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	fT CLARIN.SI-embed.hr (lemma)	100D avg	0.837
	100D	0.927		100D	0.771
w2v Kontekst.io	256D avg	0.483			
	256D	0.718			
	256D lem avg	0.483			
	256D lem	0.718			

Note. The number of all cases is 468 (from 234 occupation pairs) for Slovene and 750 (from 375 occupation pairs) for Croatian.