https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

Gender Bias in Computer-generated Thesauri: The Case of the Serbian Section of *Kontekst.io*, a Thesaurus of Synonyms and Semantically Related Terms

Dragana Čarapić, *Faculty of Philology, University of Montenegro, Nikšić, Montenegro (draganac@ucg.ac.me)*
(*https://orcid.org/0000-0003-3375-0169*)
and
Milica Vuković-Stamatović, *Faculty of Philology, University of Montenegro, Nikšić, Montenegro (vmilica@ucg.ac.me)*
(*https://orcid.org/0000-0001-5497-1361*)

Abstract: This paper studies gender bias in the computer-generated thesaurus *Kontext.io*, which is a search portal of synonyms and semantically related terms in Serbian, Croatian and Slovenian. Its Serbian section, which is the focus here, is based on a natural language processing (NLP) technique called word embeddings and a large internet corpus of Serbian. Gender bias is uncovered in four selected entries of this thesaurus: *žena* (woman), *muškarac* (man), *d(j)evojka* (young woman) and *momak* (young man). The analysis is first conducted semantically and the terms found are grouped into various semantic fields. After that, in the vein of the earlier studies of gender bias in traditional dictionaries and critical discourse analysis, an analysis of gender bias in the selected entries is provided. The results show that gender bias is ubiquitous and that it extends deeper than the earlier studies of gender bias in word embeddings have shown. We then give recommendations for improving this lexicographic product based on the results.

Keywords: GENDER BIAS, COMPUTER-GENERATED THESAURUS, WORD EMBEDDINGS, *Kontekst.io*, SERBIAN, LEXICOGRAPHY

Opsomming: Geslagsvooroordeel in rekenaargegenereerde tesourusse: Die geval van die Serwiese afdeling van *Kontekst.io*, 'n tesourus van sinonieme en semanties verwante terme. In hierdie artikel word geslagsvooroordeel in die rekenaargegenereerde tesourus *Kontext.io*, 'n soekportaal van sinonieme en semanties verwante terme in Serwies, Kroaties en Sloweens, bestudeer. Die Serwiese afdeling, waarop daar hier gefokus word, is gebaseer op 'n natuurliketaalprosesseringstegniek (NTP-tegniek) genaamd woordinbedding en 'n groot internetkorpus van Serwies. Geslagsvooroordeel word in vier uitgesoekte inskrywings in hierdie tesourus blootgelê: *žena* (vrou), *muškarac* (man), *d(j)evojka* (jong vrou) en *momak* (jong man). Die ontleding word eers semanties uitgevoer en die terme wat gevind word, word in verskillende semantiese velde ingedeel. Daarna, in dieselfde trant as vroeëre studies van geslagsvooroordeel in

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

tradisionele woordeboeke en kritiese diskoersanalise, word 'n ontleding van geslagsvooroordeel in die uitgesoekte inskrywings verskaf. Die resultate toon dat geslagsvooroordeel alomteenwoordig is en dat dit verder strek as wat vroeëre studies van geslagsvooroordeel in woordinbedding aangetoon het. Aanbevelings wat op die resultate gebaseer is, word dan gemaak om hierdie leksikografiese produk te verbeter.

Sleutelwoorde: GESLAGSVOOROORDEEL, REKENAARGEGENEREERDE TESOURUS, WOORDINBEDDING, *Kontekst.io*, SERWIES, LEKSIKOGRAFIE

1. Introduction

The recent rise of machine learning and artificial intelligence (AI) has transformed many fields. Among these, Natural Language Processing (NLP) has gained prominence by enabling computers to process and generate human language with ever increasing sophistication. Central to NLP are word embeddings — vector representations of words which capture their semantic meaning based on the context in which they are used in large corpora (Lee 2020). Word embeddings, however, reflect, perpetuate and even amplify societal biases from the source data, i.e. corpora, including gender bias (Gonen and Goldberg 2019). For instance, one type of word embeddings (*word2vec*), trained on the *Google News* dataset, in answer to the following: "man is to computer programmer as woman is to x", provides that "x is a homemaker" (Bolukbasi et al. 2016). Gender bias was demonstrated as omnipresent and consistent across different types of word embeddings and the proposed methods of mitigating it have so far produced only a limited effect — as Gonen and Goldberg put it, it is the effect of putting "a lipstick on a pig" (2019).

One area where word embeddings are starting to find its application is the creation of thesauri, especially for lesser-resourced languages, with poorly developed lexical databases (Arppe et al. 2023). Lexicographic products reflect but also shape our understanding of language and, by extension, the society (cf. Grenon-Nyenhuis 2000). Traditional dictionaries have been found to reflect and reproduce gender stereotypes (Bergenholtz and Gouws 2006; Nübling 2009). The same is true, and much more so, of word embeddings and the products based on them, given that they reflect gender bias existing in the source data. This fact should be taken into account when making lexicographic products based on word embeddings, bearing in mind that lexicographers have a moral responsibility to avoid reinforcing harmful gender stereotypes (Müller-Spitzer 2023).

This paper investigates gender bias in a computer-generated thesaurus intended for public use, aiming to uncover and point to some of such biases in it. The thesaurus in question is *Kontext.io*, which contains semantically related words and synonyms in Serbian, Croatian and Slovenian. Our study will focus on the examples from Serbian. It should be noted that in this study we consider thesaurus as a type of a dictionary, following Gouws (2017: 134).

Gender bias in word embeddings has been studied relatively extensively so far (Zhao et al. 2019; Basta et al. 2019; Yang and Feng 2020; Basta et al. 2021;

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

Caliskan et al. 2022; etc.). However, as far as we know, the present study will be the first to offer a linguistic analysis of gender bias in a thesaurus based on word embeddings intended for general public use. In addition, the study of gender bias in word embeddings has mostly focused on the phenomenon of gender bias in occupation terms, while the present study will go beyond that — we will cover all synonyms and semantically related terms provided under several selected entries in Serbian (*žena* (woman), *muškarac* (man), *d(j)evojka* (young woman), and *momak* (young man)).

The current study builds on the body of research examining gender bias in traditional dictionaries and in word embeddings. The analysis will be performed within the vein of critical discourse analysis, given that we can understand dictionary as a text (cf. Fuertes-Olivera and Tarp 2022). In critical discourse analysis, ideology and power are key terms, and lexicography and its products are always subject to them, as they "are never value-free, apolitical or asocial" (Chen 2019: 1). Critical discourse analysts depart from the premise that language is not a mere mirror of social phenomena, but also their constitutive factor (Vuković-Stamatović 2022: 429) and that the analysis must simultaneously include the analysis of text, discourse practice, i.e. the origin, distribution and/or use of the text, and, finally, its social context (Fairclough 1992, 1995). Given that in this study we are dealing with a computer-generated thesaurus, we cannot criticise the lexicographer's intervention but we can criticise the lack thereof, as well as the data which are the source of bias. The approach to lexicography through critical discourse analysis belongs to critical lexicography in general (Kachru 1995; Chen 2019).

The background of the paper consists of two parts: Section 2 will cover the phenomenon of gender bias in traditional dictionaries, while Section 3 will present word embeddings and discuss them through the lens of gender bias. After that, in Section 4, we will present the thesaurus *Kontekst.io* and our method, following which the analysis is provided.

2. Gender bias in traditional dictionaries

In patriarchal societies, men are seen as the primary sources of power and moral authority. They also tend to hold prominent positions in decision-making, social privilege and property ownership. Certain cultures are far more patriarchal than others, even though most societies are patriarchal to some degree and assign gendered responsibilities (James 2010). This has traditionally had an impact, in particular, on women's status in these societies and has also been reflected in their languages. Dictionaries do not only reflect the sociolinguistic reality of certain speech communities — they also tend to propagate it, especially bearing in mind that they are frequently seen as authoritative sources which can shape how we understand words. As such, a dictionary can be "an agent of social impact" (Gouws 2022: 40).

Various research studies have highlighted the presence of gender bias in dictionaries. Some recent papers on the issue include: Norri 2019; Iversen 2021;

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

Solonets 2021; Pettini 2021; Müller-Spitzer and Rüdiger 2022; Vacalopoulou 2022; Fuertes-Olivera and Tarp 2022; Müller-Spitzer 2023, etc., indicating that gender bias in dictionaries is really not a thing of the past (cf. Fuertes-Olivera and Tarp 2022) and that modern dictionaries still feature it, despite the increasing efforts dictionary makers have been investing in reducing it.

Gender bias in dictionaries manifests in different ways. Some of these include asymmetrical definitions, stereotypical usage examples, and a choice of collocations which reflects stereotypes.

To illustrate asymmetrical definitions for different genders, we will use an example from the *Dictionary of the Serbian Language* (Matica Srpska 2007), where the entries *man* and *woman* are defined as follows:

— man – a person, human being of a male gender, a person who is of a gender opposite to woman (2007: 746),
— woman – 1. a human being who has the ability to give birth, of a gender opposite to man. 2. a person who is married, a wife. 3. a. an adult person of a female gender. b. a female person working in a house, a maid. 4. fig. pej. a weak person, a coward (when talking about men) • easy ~ a woman of low morale in her relation to men. *take a woman* – to get married (2007: 369).

The asymmetry in the two definitions above can be seen even physically — the entry related to *man* is much shorter to that for *woman*. *Man* is, first of all, a *person*, who does not need any more specific defining, whereas *woman* is firstly described through her role of giving birth (meaning 1). In addition, woman is stereotypically portrayed in meaning 2, where she is defined through her marital status; in meaning 3b, where she is defined through household chores and working for someone else; then in meaning 4, where a *woman* means *a weak* man; and in the phrase *easy woman*, with negative connotations, again in relation to men. The only neutral meaning for *woman* is, in fact, provided in 3a. The described gender difference is a result of the Serbian patriarchal society, which is reflected in the given dictionary entries. However, as said before, dictionaries do not just describe the sociolinguistic reality, but they propagate it (and, perhaps, co-construct it?), and one of the dilemmas before dictionary makers is how to intervene and to what extent.

As Müller-Spitzer (2023: 80) notes, lexicographers have a responsibility, given that dictionaries may contribute to exclusion and perpetuating gender stereotypes. In other words, "the representation of gender in dictionaries is a matter of both language use and lexicographic-moral responsibility" (Müller-Spitzer 2023: 83). Due to this, many dictionary makers have opted to invest effort into reducing gender bias as much as possible in their dictionaries. As a result, a stark asymmetry in the definitions of genders is more rarely seen in modern dictionaries, however, gender bias remains at more subtle levels. Müller-Spitzer and Rüdiger (2022: 130) analyse the choice of examples in the entries on *man, woman, boy* and *girl* in the *Cambridge Dictionary*, and quote the following examples from these entries:

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

"He plays baseball, drinks a lot of beer and generally acts like one of the boys."
"Steve can solve anything — the man's a genius."
"Who was that beautiful girl I saw you with last night?"
"Both girls compete for their father's attention."

The examples contain gender stereotypes — men and boys are associated with playing sports, drinking alcohol, and intelligence; women and girls are described in terms of appearance and as fighting for men's attention.

A similar situation may be noted with collocations. For instance, the *Online OXFORD Collocation Dictionary of English* (2019) (https://m.freecollocation.com) provides the following adjectival collocates for the entry *girl*:

— baby; little, small, young; adolescent, teenage; bubbly, happy, lively; lovely, nice; attractive, beautiful, good-looking, gorgeous, handsome, pretty, stunning; single, unmarried.

In contrast, consider the following adjectival collocates given by the same dictionary for the entry *boy*:

— big; little, small; young; elder, eldest, older; baby; adolescent, teenage; good; naughty; bright, clever.

As can be seen, some of the collocates for boys and girls are similar, especially those referring to age and size. However, considerable differences may be noted in other semantic fields — for instance, on the one hand, *girls* have six collocates for describing pleasant physical looks, four for describing pleasant disposition, and two for describing marital/partner status, as qualities considered important for females in anglophone societies, while boys are not described at all in terms of these three aspects. On the other hand, two collocates for boys refer to their intelligence, as a quality more associated with men in gender stereotypes, while such collocates are not found for girls. In addition, *boys* are described as *naughty*, also a quality more associated with men, as more obedience is expected from girls. The choice of the collocates certainly comes from the corpus that informs the *Online OXFORD Collocation Dictionary of English* and the same is true of the thesauri created from word embeddings.

In the literature, we have not come across examples treating gender bias reflected in a choice of synonyms (and semantically related terms) for certain entries, but, as our present analysis will show, gender bias can be uncovered through this lens as well.

3. Word embeddings and gender bias

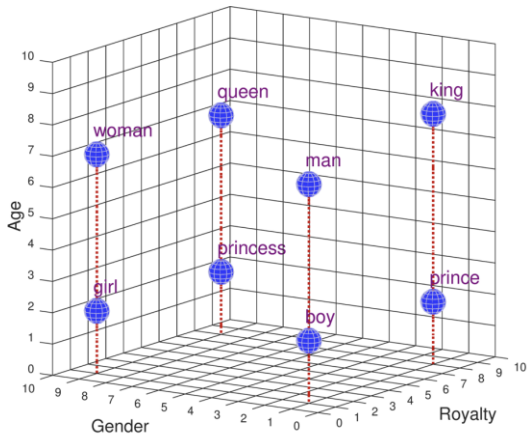
As suggested earlier, a word embedding is a numeric representation of a word in the form of a vector. Such representations are used in Natural Language Pro-

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

cessing (NLP) to allow for automatic text analysis and processing to complete tasks such as sentiment analysis (analysis of texts to determine if their emotional tone is positive, neutral, or negative), translation, web search, parsing through different kinds of texts, etc. They have also been used to inform and make some dictionaries, thesauri in particular (Morinaga and Yamaguchi 2018; Chaimae et al. 2020; Liang et al. 2023; Arppe et al. 2023), although this use of word embeddings has not been as frequent as the uses mentioned above.

When using the technique called word embeddings, each word is assigned a number format, i.e. a real-valued vector. Words used in similar contexts, i.e. similarly, are closer in vector space — for example, the words "tomato" and "potato" would be close together in the vector space because they have certain similarities, e.g. they are both edible domesticated plants, while "tomato" and "house" would be further apart because they are used in largely different contexts. Thus, the techniques involving word embeddings measure and classify semantic similarities between words by examining their distributional properties within large language data.

For easier understanding of what word embeddings actually are, below we provide a graphic visualisation for some words in a simple, three-dimensional vector space:



Word Coordinates

	Gender	Age	Royalty
man	[1,	7,	1]
woman	[9,	7,	1]
boy	[1,	2,	1]
girl	[9,	2,	1]
king	[1,	8,	8]
queen	[9,	7,	8]
prince	[1,	2,	8]
princess	[9,	2,	8]

Figure 1: An illustration of word embeddings in a 3D vector space (Touretzky 2024)

In Figure 1, the words *boy*, *girl*, *man*, *woman*, *prince*, *princess*, *king* and *queen* are represented along three semantic dimensions: gender (smaller values indicate male gender, while larger indicate female), age (smaller values indicate younger age), and royalty (greater values for royal status). The values assigned are not perfectly symmetrical — for instance, for age, *king* is assigned an 8 while a queen is assigned a 7. The reason for this is that often in the corpora there is more talk

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

of *old kings* than *old queens*. Each of the coordinates in this 3D space can be read using three numbers, one for each of the three dimensions involved — the list of these three numbers or values for all the words depicted in the figure is given on the right. These series of numbers for each word represent their vectors. Of course, as we said, this 3D representation is very simplified and words, in practice, are represented by a much higher number of dimensions — of the order of tens and even hundreds of dimensions; in addition, the numbers are more precise and are rarely whole but rather decimal numbers. Based on the numbers assigned to each word, words can be mathematically compared in terms of their semantic similarity or semantic distance, and analogies such as "a man is to a woman as a king is to a queen" can be derived. Based on this methodology, it is possible to derive a list of synonyms and semantically related terms from a corpus, and as we have seen, these can be used to form lexical databases, inform dictionaries, and make thesauri.

As suggested earlier, gender bias is rather frequent in word embeddings, primarily as a reflection of the source data which naturally contain it. Thus, word embeddings can yield analogies such as "a father is to a doctor as a mother is to a nurse" (Bolukbasi et al. 2016). However, word embeddings do not just reflect the source data and the biases in them; the algorithms which they are based on can even inadvertently amplify them (Bolukbasi et al. 2016; Basta et al. 2019). In response to this problem, different debiasing methods have been proposed — some of these focus on correcting the source data itself, some intervene in the algorithm, while some use human-made dictionaries for creating less biased word embeddings (Zhao et al. 2019; Yang and Feng 2020, etc.). However, the results are mixed and what is certain is that gender bias always remains in the output (Gonen and Goldberg 2019; Lee 2020). Most debiasing approaches focus on the words related to occupations, but gender bias extends much deeper and can be hidden across multiple features (Caliskan et al. 2022). So far, the debiasing methods have been successful at rather "hiding gender bias but not truly removing it", Ronchieri and Biagi (2023: 730) conclude.

4. Data and method

As suggested earlier, the data used in this study derives from the computer-generated thesaurus *Kontekst.io*. In the following text, we present *Kontekst.io* in more details, following which more light is shed on the method.

4.1 *Kontekst.io*

Kontekst.io is a search portal of synonyms and semantically related terms in Serbian, Croatian and Slovenian, based on word embeddings (Plahuta 2024). On its website, it is stated that it can primarily be used as "a dictionary of synonyms".

What the homepage of *Kontekst.io* looks like can be seen in Figure 2 below:

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

Figure 2: The homepage of *Kontekst.io*

As can be seen, the website's homepage is simple — the user first chooses one out of the three languages offered. The second step is to type in the search term or to choose one of the popular search terms offered below.

The output offers a list of words with similar distributional properties, i.e. semantically related words including synonyms and even antonyms, along with the normalised frequency of the word per 1 million words of the corpus, its score of semantic similarity with the entered term and the examples of its use in the source corpus. Below is part of the output for the entry *lijepa* (beautiful) from the Serbian section (Figure 3); we present only part of it, as the whole output contains some 50 synonyms and semantically related terms and cannot fit here.

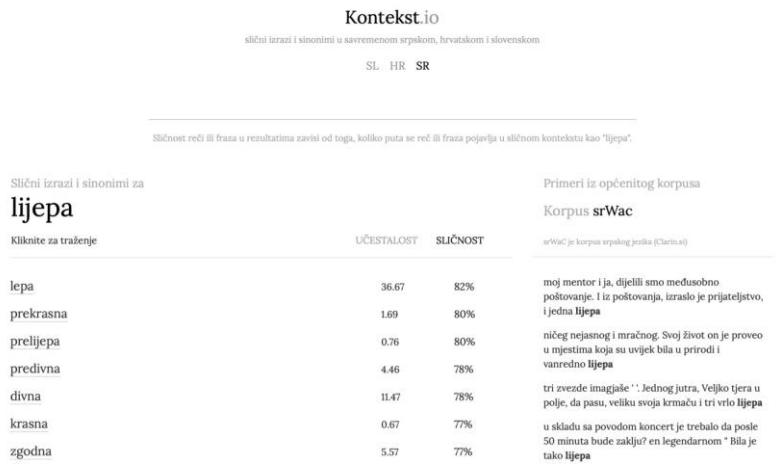


Figure 3: Some synonyms and semantically related terms for *lijepa* (beautiful) in *Kontekst.io* (Serbian)

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

The Serbian thesaurus is based on the Serbian web corpus *srWaC* (Ljubešić and Klubička 2016), which contains 554,627,647 running words (version 1.1), and was obtained by crawling the *.rs* web domain (the Serbian national web domain). Some of the advantages of using this corpus include its size, the fact that it is quite recent and the fact that it is available for public use. Additionally, there are not many viable corpus alternatives when it comes to corpora for Serbian, which is far less resourced than many other European languages. The disadvantage of the corpus is that all the texts come from the Internet.

As we find in Ulčar et al. (2021), *Kontekst.io* was made using the *word2vec* algorithm (Mikolov et al. 2013) and a 256-dimensional vector model trained for its needs. Using their mathematical approach, Ulčar et al. (2021) have also determined that *Kontekst.io* contains a high degree of gender bias in its entries referring to occupations, but they have not looked beyond that.

According to the data from *Similarweb.com*, a website tracking 100 million websites in 190 countries, *Kontekst.io* was the ninth most popular website in the category of dictionaries and encyclopedias in Serbia in April 2024. Most of the websites preceding it were global websites, such as *Wikipedia.org*. Based on this, we can say that *Kontekst.io* is a rather popular thesaurus in Serbia. Most of its traffic comes from the referrals from the *Google* search when users are looking for synonyms for certain words (again, according to *Similarweb.com*).

4.2 Method

In this study, we analyse the synonyms and semantically related terms provided under four selected entries: *žena* (woman), *muškarac* (man), *d(j)evojka* (young woman), and *momak* (young man), in the Serbian component of *Kontekst.io*. The number of the entries studied was limited by the space provided for this paper. The entries chosen are obviously gendered words, which is why it was expected that their analysis could most clearly point to the presence and types of gender bias in *Kontekst.io*.

To uncover and analyse gender bias in the selected entries, we use a method which encompasses comparative analysis, grouping the terms in separate semantic fields and applying the connotative analysis of the specific terms. The method embodies different levels of paradigmatic relations, starting from the hyponymy — superordinate term, and continuing with the synonymous expressions allocated to the respective semantic fields. For better visualisation, the results thus processed are presented in tables, comparatively for *woman* and *man*, as well as for *young woman* and *young man*. This level of analysis is purely semantic.

Further, the semantic method is complemented with critical discourse analysis (Fairclough 1992, 1995), given that the ideology underlying the selection of the terms presented as synonyms and semantically related terms, is uncovered and discussed. The selection is computer-produced but it reflects the bias existing in the source data which feed the thesaurus. More specifically, the analysis of the entries is in the vein of the analyses of gender bias in traditional dictionaries

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

presented in Section 2: our own examples and those of Müller-Spitzer (2023). Thus, the inequalities in the choice and the frequency of the different synonyms and semantically related terms (provided for the said male and female entries) are noted and discussed.

5. Results and analysis

Given that two pairs of entries are analysed in this paper, the analysis will be organised in two parts, each dedicated to one of these.

5.1 The entries for *žena* (woman) and *muškarac* (man) in Serbian *Kontekst.io*

The terms provided as synonyms and semantically related terms for the entries *woman* and *man* have first been grouped into separate semantic fields, as explained in the Method subsection. The results for Serbian *Kontekst.io* are presented in Table 1 below.

Table 1: Synonyms and semantically related terms for the entries *žena* (woman) and *muškarac* (man) in Serbian in *Kontekst.io*[1]

***Kontekst.io* (SERBIAN)**	
ŽENA (WOMAN)	**MUŠKARAC (MAN)**
SUPERORDINATE TERM: Osoba (person)	SUPERORDINATE TERM: Čovek, čovjek (human, man), stvor (creature)
AGE-RELATED: Beba (baby), devojčica, djevojčica (little girl), djevojka, devojka, devojaka (young woman), cura (lass), curica (lassy DIM.[2]), mlada dama (young lady), mlada devojka (young girl), starica (old woman), bakica (granny)	AGE-RELATED: Mališan (tot), dečak, dječak, dečko (boy), klinac (kid), mladić, momak (young man), tinejdžer (teenager), adolescent (adolescent), pedesetogodišnjak (fifty-year-old man), matorac (old man)
OCCUPATION-RELATED: Domaćica (housewife), dadilja (nanny), sluškinja (maid)	OCCUPATION-RELATED: Policajac (police officer), vojnik (soldier)
FAMILY-RELATED: Trudnica (pregnant woman), majka (mother), udovica (widow), porodica (family)	FAMILY-RELATED: Muž (husband), roditelj (parent)
ANIMAL-RELATED: Kuja (bitch PEJ.[3])	ANIMAL-RELATED: Mužjak (male), pas (dog), bik (bull), majmun (monkey), konj (horse)
EVIL-RELATED: Veštica (witch PEJ.)	EVIL-RELATED: Nasilnik (bully), vampire (vampire)

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

PATIENT/BENEFICIARY ROLES: Pacijentkinja (patient FEM.[4]), pacijentica (patient FEM.), zatvorenica (prisoner FEM.), mušterija (customer FEM.), klijentkinja (client FEM.)	PATIENT/BENEFICIARY ROLES: Pacijent (patient)
APPEARANCE-RELATED: Plavuša (blonde), crnkinja (black woman), belkinja (white woman), ženica (woman DIM.)	APPEARANCE-RELATED: Belac, bijelac (white man), crnac (black man)
FOREIGNER: Francuskinja (French woman)	FOREIGNER: Stranac (foreigner), Indijanac (Indian), Amerikanac (American)
MONEY-RELATED: Bogatašica (rich woman)	MONEY-RELATED: Bogataš (rich man), beskućnik (homeless man)
SOCIAL ROLE-RELATED: Dama (lady), komšinica (neighbour FEM.), prijateljica, drugarica (friend FEM.)	SOCIAL ROLE-RELATED: Džentlmen (gentleman)
OTHER: Duša (darling), seljančica (village girl)	SEXUALITY-RELATED: Frajer (hot shot), ljubavnik (lover)

The first thing which we can notice in Table 1 is that the superordinate term for *muškarac* (man) is *čov(j)ek* (human), whereas for *žena* (woman) it is *osoba* (person). Further, the following can also be noted (the observations are presented according to the semantic fields):

— **age:** more diminutives and the adjective *young* preceding the nouns designating women, e.g. *djevojka, devojka, devojaka* (young woman), *mlada dama* (young lady), *mlada devojka* (young girl), *curica* (lassy DIM.);

— **occupation:** gender bias detected under both the entries, e.g. *domaćica* (housewife), *dadilja* (nanny), and *sluškinja* (maid) under the entry *woman*, and *policajac* (police officer) and *vojnik* (soldier), under the entry *man*;

— **family:** there are more terms in this field for *woman* than *man*, e.g. *trudnica* (pregnant woman), *majka* (mother), *udovica* (widow), *porodica* (family), are all used for *woman*, while only *muž* (husband) and *roditelj* (parent) are used for *man.*

— **animal:** only one term with negative connotations is used for *woman*: *kuja* (bitch PEJ.); however, there are five terms in this field for *man*: *mužjak* (male), *pas* (dog), *bik* (bull), *majmun* (monkey), and *konj* (horse), some of which have negative connotations, too;

— **evil:** gender bias detected in the derogatory terms *veštica* (witch PEJ.) for *woman*, and *nasilnik* (bully) and *vampire* (vampire) for *man*;

— more **appearance**-related terms for women: *plavuša* (blonde), *crnkinja* (black woman), *belkinja* (white woman), *ženica* (woman DIM.);

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

— more **socially**- and **family**-defined **roles** for *woman*: *dama* (lady), *komšinica* (neighbour FEM.), *prijateljica, drugarica* (friend FEM.);

— more **money**-related characterisations for men, e.g. two depicting *man* (*bogataš* (rich man), and *beskućnik* (homeless man)) vs. one used for *woman* (*bogatašica* (rich woman));

— **sexuality**-related terms — only used for men, whereby both the terms in this field have a positive meaning: *frajer* (hot shot), *ljubavnik* (lover);

— **other terms** for *woman*: there is one endearing term in the data and it is used for *woman* (*duša* (darling)); the other term in this category is *seljančica* (village girl) and it is condescending.

We proceed with the results for the second pair of entries analysed.

5.2 The entries for *d(j)evojka* (young woman) and *momak* (young man) in Serbian *Kontekst.io*

In the same vein, we conduct the analysis for the entries: *d(j)evojka* (young woman) and *momak* (young man). The results are given in Table 2.

Table 2: Synonyms and semantically related terms for the entries *d(j)evojka* (young woman) and *momak* (young man) in Serbian in *Kontekst.io*

***Kontekst.io* (SERBIAN)**	
D(J)EVOJKA (YOUNG WOMAN)	**MOMAK (YOUNG MAN)**
SUPERORDINATE TERM: Dama (lady)	SUPERORDINATE TERM: Čova, čovek, čovjek (human, man), čovečuljak (man DIM.), muškarac (man)
AGE-RELATED: Bakica (granny DIM.), cura, curica (girl), devojčica, djevojčica, klinka, mala devojčica (little girl), mlada devojka (young girl), starica (old woman), tinejdžerka (teenage girl)	AGE-RELATED: Čikica (old man DIM.), dečak, dečkić (young boy DIM.), dečko, dječak, deran, klinac (young boy), matorac (old guy), mladić, momčić (young man DIM.)
OCCUPATION-RELATED: Dadilja (nanny), konobarica (waitress), služavka (maid), stažistica (intern FEM.)	OCCUPATION-RELATED: Pandur (cop), policajac (police officer)
FAMILY-RELATED: Majka (mother), sestra (sister)	FAMILY-RELATED: Muž (husband), roditelj (parent)
ANIMAL-RELATED: Kučka (bitch), mačka (cat)	ANIMAL-RELATED: Džukac (mutt), majmun (monkey), pas (dog)

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

EVIL-RELATED: Ludača (mad woman)	EVIL-RELATED: Gad (bastard), govnar (piece of shit DER.[5]), kučkin sin, (son of a bitch DER.), kurvin sin (son of a whore DER.), ludak (mad man), seronja (asshole DER., SLANG), tip (guy)
APPEARANCE-RELATED: Lepa devojka (pretty girl), lepotica (beauty), plavuša (blonde)	APPEARANCE-RELATED: Crnac (black man), debeljko (fat man), lepotan (handsome man), plavušan (blonde man)
SOCIAL ROLE-RELATED: Cimerka (roommate FEM.), drugarica (friend FEM.), komšinica (neighbour FEM.), nevesta (bride), poznanica (acquaintance FEM.), prijateljica (friend FEM.), verenica (fiancé FEM.)	SOCIAL ROLE-RELATED: Drugar (friend), dasa (hot shot), klipan (loon DER.), ortak (buddy)
SEXUALITY-RELATED: Fufa, fufica, kurva (hooker), prostituka (prostitute FEM.)	SEXUALITY-RELATED: Frajer (hot shot), jebač (fucker DER., SLANG)
MONEY-RELATED: Bogatašica (rich woman)	FOREIGNER: Indijanac (Indian), Amerikanac (American)

The following can be noted regarding the results in Table 2:

— **age:** more diminutive expressions are used preceding the noun for *young man*: *čovečuljak* (man DIM.), *čikica* (old man DIM.), *dečkić* (young boy DIM.), *momčić* (young man DIM.);

— **occupation:** gender bias is found under both *young woman* and *young man*, e.g. *dadilja* (nanny), *konobarica* (waitress), *služavka* (maid), *stažistica* (intern FEM.), in the former category, and *pandur* (cop) and *policajac* (police officer) in the latter;

— **family:** the same number of synonyms/semantically related terms are provided for both *young woman* and *young man*: *majka* (mother) and *sestra* (sister), vs. *muž* (husband) and *roditelj* (parent);

— **animal:** on the one hand, two terms are used for *young woman: kučka* (bitch), which has negative connotations, and *mačka* (cat), which has sexual connotations; on the other hand, three terms are used with negative connotations for *young man*: *džukac* (mutt), *majmun* (monkey), and *pas* (dog);

— **evil:** in this semantic field, there is one term for *young woman* (*ludača* (mad woman)), whereas there are seven derogatory terms for *young man*: *gad* (bastard DER.), *govnar* (piece of shit DER.), *kučkin sin* (son of a bitch DER.), *kurvin sin* (son of a whore, DER.), *ludak* (mad man), *seronja* (asshole DER., SLANG), and *tip* (guy);

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

— more **appearance**-related terms are used for *young man*, e.g. *crnac* (black man), *debeljko* (fat man), *lepotan* (handsome man), and *plavušan* (blonde man), than for *young woman*: *lepa devojka* (pretty girl), *lepotica* (beauty), and *plavuša* (blonde);

— there more **socially defined roles** under the entry *young woman*: *cimerka* (roommate FEM.), *drugarica* (friend FEM.), *komšinica* (neighbour FEM.), *nevesta* (bride), *poznanica* (acquaintance FEM.), *prijateljica* (friend FEM.), *verenica* (fiancé), than for *man*: *mladoženja* (bachelor), *razbojnik* (scourer), and *zatvorenik* (prisoner);

— there is one **money**-related characterisation for *young woman: bogatašica* (rich woman), and there are no such characterisations for *young man*;

— there are two expressions for **foreigner** under the entry *young man*: *Indijanac* (Indian) and *Amerikanac* (American);

— **sexuality**: four terms from this field are used for *young woman*: *fufa, fufica, kurva* (hooker) and *prostituka* (prostitute FEM.), while two such terms are used for *young man*: *frajer* (hot shot) and *jebač* (fucker DER., SLANG).

A discussion of these findings is presented in Section 6.

6. Discussion

In the analysis of the selected terms in *Kontekst.io* in Serbian, several elements of gender bias have been detected.

On the one hand, the superordinate term for both *muškarac* (man) and *momak* (young man) is *čov(j)ek* (human), whereas for *žena* (woman) it is *osoba* (person) and for *d(j)evojka* (young woman) it is *dama* (lady). The male terms are thus *more encompassing*, given that they can stand for the entire human kind, i.e. both men and women, unlike the female terms, which are much more specific and exclusive in the sense that they cannot extend their meaning beyond the female sex. Even more specifically, *person* stands for just one individual. Additionally, *lady* is a term which reflects a socially preferable role for women.

Another finding is that there are more diminutives under the entries for *woman*, as well as that the adjective *young* often precedes the nouns in this category. This has to do with the general notion of woman as a weaker sex in a patriarchal society. As such, woman requires more affection and protection through the use of diminutives.

We have also noticed that there is gender bias under all the analysed entries related to occupations. As said earlier, this phenomenon has been investigated earlier (for instance, Ulčar 2021) and it has been shown that there is substantial gender bias in word embeddings when it comes to the names of occupations.

In *Kontekst.io*, more family-related terms are provided under the entry *woman*

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

than *man*. This result is reflective of the social position of woman, who is largely defined through the role she has within her family in a patriarchal society. Furthermore, it is noteworthy that there is no mention of *father* as a semantically related term for either *man* or *young man*, whereas *mother* is identified as a semantically related term for both *woman* and *young woman*. The only family-related word that is provided for *man* and *young man* is *parent*, a gender-neutral designation that does not specify paternal identity. Such a disbalance is not seen, however, in the second pair of the entries, given that the additional family roles for females are assumed only after a certain age.

There are more terms in the semantic field *animal* for *man* and *young man*, than under the female entries. Many of these terms used for men have negative connotations, as they are associated with negative character traits: being *aggressive, violent* and *fierce.* This is also the result of gender bias existing in society and, consequently, in the source data used for the thesaurus.

Likewise, in the semantic field relating to *evil*, there are many more terms for *man* than *woman.* This result points to a similar notion of *man* as a sex more capable of inflicting violence and other evil acts and it also reveals that it is socially more acceptable to address other men rather than women using such terms. Further, of the two terms supplied in this field for women, *witch* particularly stands out and also reflects gender bias.

More appearance-related terms for the entry *woman* show that woman is more socially defined by her physical appearance than man. The terms relating to *young woman* and *young man*, however, are similar in number, suggesting that appearance becomes much more relevant for women only after a certain age, the same was noted for socially defined roles for women.

More money-related characterisations for men imply the importance of *money* in defining *man.* Also, in our data, foreigners tend to be men more often than women.

There are some more sexual terms for *young woman* than *young man*, but it should be noted that all the terms from this field have negative connotations. All of them are *derogatory* slang expressions and they imply negative character traits.

As we can see, gender bias is ubiquitous in all the entries studied. This is not just gender bias which is related to the female gender, which is usually more obvious — on the contrary, in the terms under the said entries we find reflections of many stereotypical characterisations of men, too. Thus, men are seen as being aggressive and violent, money is important for how they are socially defined, and they are seen to be working in certain professions, such as being a policeman or a soldier.

Some of the gender bias described here is more "innocuous" than the other. Perhaps on the more "innocuous" end of the spectrum are the terms for occupations, which have been the most frequent topic of interest for those seeking to address gender bias in word embeddings through improving their algorithms (Section 2). We believe that it is much more serious when the terms such

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

as *witch*, *prostitute*, *bitch*, and *mad woman*, for example, are provided as synonyms or semantically-related terms for *woman* or *young woman*. Most of the users of *Kontekst.io* will not be familiar with the discipline of semantics or the meaning of "semantically-related", especially in terms of semantic distribution, while many will understand what synonyms are and will perhaps see the terms provided as such. As suggested earlier, the website of *Kontekst.io* itself states that it can primarily be used as "a dictionary of synonyms", while most of its traffic comes from the *Google* search for synonyms. That being said, it does not help that this search portal does not contain instructions or a help page providing an explanation for common users on what these terms mean and what the results provided by the thesaurus actually show. We recommend that such guidance be provided visibly for the users.

Moreover, we have noticed that for derogatory terms and other terms with negative connotations, which were quite frequent in the results, no such designations were provided in order to point to users that these terms are not in common, everyday use. Providing the necessary flags, as is customary in dictionaries, should be the next step in improving *Kontekst.io*. Perhaps these can be imported from other dictionaries.

Additionally, it goes without saying that all products, including *Kontekst.io*, which are based on word embeddings, must be further improved towards eliminating more gender bias, whereby the process should not solely or primarily focus on the occupation terms. Lexicographic moral responsibility is not absolved if the thesaurus is created by a machine, i.e. an algorithm — people are always behind those algorithms, in the same way that they are behind the corpora informing the machines.

Gender bias can probably never be fully eliminated in any dictionary in general and in computer-generated ones in particular — we have demonstrated here how deep it can go in many different directions and for both the genders here studied. However, when making lexicographic products which are publicly available, one must be aware of the harms which such products may do, as they do not only reflect the existing gender bias but also may perpetuate and co-construct it.

7. Conclusion

This paper investigated gender bias in the computer-generated thesaurus *Kontext.io,* aiming to uncover examples of this bias in it and give recommendations for improving it as a lexicographic product. *Kontekst.io* is a search portal for synonyms and semantically related terms in Serbian, Croatian and Slovenian, and this paper focused on its Serbian section — specifically, the terms provided under the entries: *žena* (woman), *muškarac* (man), *d(j)evojka* (young woman), and *momak* (young man).

Gender bias in the studied entries was found to be ubiquitous and to run deeper than the earlier studies of this phenomenon in word embeddings have

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

shown. Earlier studies were mostly conducted by non-linguists who focused on the more obvious examples of gender bias and concentrated more on improving the algorithms than on detecting all types of gender bias in their products.

Based on the results, we gave some recommendations how *Kontekst.io* (and other similar lexicographic products) may be improved. These include providing clear definitions and instructions to users explaining what kind of terms are supplied in the entries, providing flags marking the use of certain words as derogatory, slang etc., and working further on the algorithms, thus eliminating more gender bias, especially beyond the occupation terms.

Ultimately, we believe that gender bias cannot be reduced to a certain satisfactory level without human intervention in lexicographic products such as the one studied here. However, we do understand the necessity of such products in lesser-resourced languages such as Serbian, but we hope that they can be additionally improved and the users more instructed into what they really are and what their limitations are, so as to avoid harmful consequences such as perpetuating and co-constructing gender bias.

Our analysis focused on the entries where gender bias might be more salient. Further investigation should focus on more entries and other language sections of *Kontekst.io* in order to uncover the more subtle forms of gender bias in them and produce further recommendations for improving this and other similar lexicographic products.

Endnotes

1. Note: Mistyped words have been excluded here.
2. Diminutive.
3. Pejorative.
4. Feminine.
5. Derogatory.

References

Arppe, A., A. Neitsch, D. Dacanay, J. Poulin, D. Hieber and A. Harrigan. 2023. Finding Words that Aren't There: Using Word Embeddings to Improve Dictionary Search for Low-resource Languages. *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*: 144-155. Toronto, Canada: Association for Computational Linguistics.

Basta, C., M.R. Costa-jussà and N. Casas. 2019. Evaluating the Underlying Gender Bias in Contextualized Word Embeddings. *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*: 33-39. Florence, Italy: Association for Computational Linguistics.

Basta, C., M.R. Costa-jussà and N. Casas. 2021. Extensive Study on the Underlying Gender Bias in Contextualized Word Embeddings. *Neural Computing and Applications* 33(8): 3371-3384.

Bergenholtz, H. and R. Gouws. 2006. How to Do Language Policy with Dictionaries. *Lexikos* 16: 13-45.

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

Bolukbasi, T., K.-W. Chang, J.Y. Zou, V. Saligrama and A.T. Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Lee, D., M. Sugiyama, U. Luxburg and I. Guyon and R. Garnett. 2016. *Advances in Neural Information Processing Systems* 29. *(30th Annual Conference on Neural Information Processing Systems 2016 (NIPS 2016), 5-10 December 2016, Barcelona, Spain)*: 4349-4357. La Jolla, CA, USA: NIPS Foundation.

Caliskan, A., P.P. Ajay, T. Charlesworth, R. Wolfe and M.R. Banaji. 2022. Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES 2022)*: 156-170. Oxford: AIES.

Chaimae, A., M. Rybinski, E.Y. Yacine and J.F. Aldana Montes. 2020. Comparative Study of Arabic Word Embeddings: Evaluation and Application. *International Journal of Computer Information Systems and Industrial Management Applications* 12: 349-362. https://cspub-ijcisim.org/index.php/ijcisim/article/view/469

Chen, W. 2019. Towards a Discourse Approach to Critical Lexicography. *International Journal of Lexicography* 32(3): 362-388.

Fairclough, N. 1992. *Discourse and Social Change*. Cambridge: Polity.

Fairclough, N. 1995. *Critical Discourse Analysis*. London/New York: Longman.

Fuertes-Olivera, P.A. and S. Tarp. 2022. Critical Lexicography at Work: Reflections and Proposals for Eliminating Gender Bias in General Dictionaries of Spanish. *Lexikos* 32(2): 105-132.

Gonen, H. and Y. Goldberg. 2019. Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings but Do not Remove Them. *arXiv preprint arXiv:1903.03862*

Gouws, R.H. 2017. Van tematiese na alfabetiese na tematiese ordening in woordeboeke — wisselwerking tussen teorie en praktyk. *Stellenbosch Papers in Linguistics Plus* 53: 133-148.

Gouws, R.H. 2022. Dictionaries as Instruments of Exclusion and Inclusion: Some South African Dictionaries as Case in Point. *Lexicographica* 38(1): 39-61. https://doi.org/10.1515/lex-2022-0003

Grenon-Nyenhuis, C. 2000. The Dictionary as a Cultural Institution. *Intercultural Communication Studies* 10(1): 159-166.

Iversen, S.H. 2021. The (Re)presentation of Knowledge about Gender in Children's Picture Dictionaries. Goga, N., S.H. Iversen and A.-S. Teigland (Eds.). 2021. *Verbal and Visual Strategies in Nonfiction Picturebooks: Theoretical and Analytical Approaches*: 67-79. Oslo: Scandinavian University Press.

James, K. 2010. Domestic Violence within Refugee Families: Intersecting Patriarchal Culture and the Refugee Experience. *Australian and New Zealand Journal of Family Therapy* 31(3): 275-284.

Kachru, B.B. 1995. Afterword: Directions and Challenges. Kachru, B. B. and H. Kahane (Eds.). 1995. *Cultures, Ideologies, and the Dictionary: Studies in Honor of Ladislav Zgusta*: 417-424. Tübingen: Max Niemeyer.

Lee, E. 2020. Gender Bias in Dictionary-derived Word Embeddings. Technical Report (CS230: Deep Learning, Fall 2020, Stanford University, CA). http://cs230.stanford.edu/projects_fall_2020/reports/55476615.pdf

Liang, H., Y.M.M. Ng and N.L. Tsang. 2023. Word Embedding Enrichment for Dictionary Construction: An Example of Incivility in Cantonese. *Computational Communication Research* 5(1): 1-26.

Ljubešić, N. and F. Klubička. 2016. *Serbian Web Corpus srWaC* 1.1, Slovenian Language Resource repository CLARIN.SI, ISSN 2820-4042. http://hdl.handle.net/11356/1063.

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

Mikolov, T., I. Sutskever, K. Chen, G. Corrado and J. Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality.
arXiv preprint arXiv:1310.4546

Morinaga, Y. and K. Yamaguchi. 2018. Improvement of Reverse Dictionary by Tuning Word Vectors and Category Inference. Robertas Damaševičius, R. and G. Vasiljevienė (Eds.). 2018. *Information and Software Technologies: 24th International Conference, ICIST 2018, Vilnius, Lithuania, October 4–6, 2018, Proceedings* 24: 533-545. New York: Springer.

Müller-Spitzer, C. 2023. Gender Stereotypes in Dictionaries: The Challenge of Reconciling Usage-based Lexicography with the Role of Dictionaries as Social Agents. *Lexikos* 33(2): 79-94.

Müller-Spitzer, C. and J.O. Rüdiger. 2022. The Influence of the Corpus on the Representation of Gender Stereotypes in the Dictionary. A Case Study of Corpus-based Dictionaries of German. Klosa-Kückelhaus, A., S. Engelberg, C. Möhrs and P. Storjohann (Eds.). 2022. *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*: 129-141. Mannheim: IDS-Verlag.

Norri, J. 2019. Gender in Dictionary Definitions: A Comparison of Five Learner's Dictionaries and their Different Editions. *English Studies* 100(7): 866-890.

Nübling, D. 2009. Zur lexikografischen Inszenierung von Geschlecht. Ein Streifzug durch die Einträge von Frau und Mann in neueren Wörterbu̇chern. *Zeitschrift fu̇r germanistische Linguistik* 37(3): 593-633.
https://doi.org/10.1515/ZGL.2009.037

***Online OXFORD Collocation Dictionary of English*.** 2019.
https://m.freecollocation.com

Pettini, S. 2021. One is a Woman, so That's Encouraging too. The Representation of Social Gender in "Powered by Oxford" Online Lexicography. *Lingue e Linguaggi* 44: 275-295.

Plahuta, M. 2024. *Kontekst.io*.
https://virostatiq.com/ [Accessed 20 July 2024]

Ronchieri, E. and C. Biagi. 2023. Comparing Methods for Mitigating Gender Bias in Word Embedding. Bui, T.X. (Ed.). 2023. *Proceedings of the 56th Hawaii International Conference on System Sciences, January 3–6, 2023*: 722-731. Honolulu: HICSS.
https://hdl.handle.net/10125/102720

***Similarweb*.** 2024. https://www.similarweb.com/website/blog.context.io/ [Accessed 10 July 2024]

Solonets, P.V. 2021. *Gender and Dictionary: Russian Perspective*. Unpublished Master's Thesis. Braga: University of Minho.

Touretzky, D. 2024. Word Embedding Demo (webpage), Carnegie Mellon University blogs.
https://www.cs.cmu.edu/~dst/WordEmbeddingDemo/tutorial.html

Ulčar, M., A. Supej, M. Robnik-Šikonja and S. Pollak. 2021. Slovene and Croatian Word Embeddings in Terms of Gender Occupational Analogies. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave* 9(1): 26-59.

Vacalopoulou, A. 2022. Gender Stereotypes in Greek Children's Dictionaries. *Dictionaries* 43(1): 167-192.

Vujanić, M., D. Gortan-Premk, M. Dešić, R. Dragićević, M. Nikolić, LJ. Nogo, V. Pavković, M. Radović-Tešić, N. Ramić, R. Stijović and E. Fekete. 2007. *Rečnik srpskog jezika* [Dictionary of the Serbian Language]. Novi Sad: Matica Srpska.

Vuković-Stamatović, M. 2022. "Accessing the EU Is Like Running on a Treadmill in the Gym": How the EU Accession Process is Metaphorically Presented in the Online Media of Serbia, Montenegro, and Bosnia and Herzegovina. *Annales. Series historia et sociologia* 32(3): 427-448.

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2068 (Article)

Yang, Z. and J. Feng. 2020. A Causal Inference Method for Reducing Gender Bias in Word Embedding Relations. *Proceedings of the AAAI Conference on Artificial Intelligence* 34(5): 9434-9441. https://ojs.aaai.org/index.php/AAAI/article/view/6486

Zhao, J., T. Wang, M. Yatskar, R. Cotterell, V. Ordonez and K.-W. Chang. 2019. Gender Bias in Contextualized Word Embeddings. *arXiv preprint arXiv:1904.03310*