Adjectival Attributes with the Nouns "čovek", "žena", "muškarac" and "muž"

UDC 811.163.41'322.2
DOI 10.18485/infotheca.2017.17.2.4

Marija Pantić
p.mmmmmm@gmail.com
University of Belgrade
Faculty of Philology

ABSTRACT: This research aims to produce a systematic account of the representation of men and women in the Serbian language by analysing adjectival attributes of the Serbian nouns ČOVEK (MAN), ŽENA (WOMAN/FEMALE/WIFE), MUŠKARAC (MALE) and MUŽ (HUSBAND). The analysis was performed on data obtained from the Corpus of Contemporary Serbian Language (SrpKor2013) in its advanced search mode, and it encompasses all adjectival words that occur with sufficient frequency with at least one of the four nouns. Selecting the adjectival words with an attributive function was done by reviewing the concordances, and the selected adjectival words were subjected to further analysis and classification.

KEYWORDS: adjectival attributes, adjectival words, corpus analysis, frequency lists, SrpKor2013, gender differences

PAPER SUBMITTED: 3 November 2017
PAPER ACCEPTED: 14 December 2017

1 Introduction

This paper deals with the similarities and differences in the representation of male and female persons in the Corpus of Contemporary Serbian Language through the analysis of the distribution of adjectival words occurring in the

grammatical role of adjectival attributes before the nouns ČOVEK (MAN), ŽENA (WOMAN/FEMALE/WIFE), MUŠKARAC (MALE) and MUŽ (HUSBAND).[1]

Visual culture plays an important part in establishing and maintaining sexist stereotypes. The results of a Google image search for the queries "muškarac" and "žena" match a modified, not the general, meaning of the nouns that constitute these queries. The modification is explicit in the headlines of the sources of the photographs that the search returns: the noun MUŠKARAC is featured as the head of the noun phrases *neodgovoran* MUŠKARAC (*an irresponsible* MAN), *pravi* MUŠKARAC (*a real* MAN), *savršen* MUŠKARAC (*a perfect* MAN), *idealan* MUŠKARAC (*an ideal* MAN), *zreo* MUŠKARAC (*a mature* MAN) и *odgovoran* MUŠKARAC (*a responsible* MAN), while the noun ŽENA is featured as the head of the noun phrases *super* ŽENA (*a super* WOMAN), *neodoljiva* ŽENA (*an irresistible* WOMAN), *idealna* ŽENA (*an ideal* WOMAN), *savršena* ŽENA (*a perfect* WOMAN) и *najlepša* ŽENA (*the most beautiful* WOMAN). All these noun phrases are examples of modification of the meaning of the nouns by adjectives in the role of adjectival attributes. The most obvious, however, is the implicit modification based on age: the meaning of these noun phrases is illustrated mostly by photographs of men and women between the ages of twenty and forty. The age criterion may not be explicit in the accompanying text either, but the photographs themselves illustrate the noun phrases in accordance with the presumed, expected or desirable qualities of the persons they refer to (Baker, 2014, 75).

This study does not deal with implicit modification, so the dominant expectations, the issue of desirability and the presumed values are out of its scope. The author's assumption is that the perception and the conception of the meaning of nouns is greatly influenced by adjectival words that modify them (Hoey, 2004), and the aim of this paper is to present the patterns of such modification that have been discovered for the selected nouns in the Corpus of Contemporary Serbian Language. It is the author's hope that other researchers with an interest in gender and language and who are working in other disciplines might be inspired to provide various perspectives on the socioeconomic, cultural, linguistic and psychological reasons behind the data presented here, as its interpretation is beyond the scope of this paper.

[1] All English translations are by the author and based on the primary meanings of the words in their contemporary usage. The English nouns FEMALE and MALE are approximate translations, with the latter allowing a distinction to be made between the Serbian nouns ČOVEK and MUŠKARAC, which would otherwise both be translated as MAN, while the former introduces a corresponding distinction between different meanings of the noun ŽENA.

Adjectival attributes are adjectival units, that is, adjectival words (adjectives, adjectival pronouns and ordinal numbers) and adjectival phrases, the meaning of which is attributed to the nominal concept denoted by the head of the noun phrase. This attribution is formally marked by gender, number and case agreement. The syntactic and semantic differences among adjectival attributes are reflected in how tightly or loosely they are tied to the nominal concept, that is, in their position: closer to the noun or further away from it (and closer to the beginning of the phrase) (Станојчић и Поповић, 2000). This research covers adjectival words[2] that function as adjectival attributes and, either independently or as part of adjectival phrases, occur immediately before the nouns ŽENA, MUŠKARAC, MUŽ and ČOVEK, that is, those adjectival words that are most tightly tied to the nouns under consideration. In Serbian, adjectival attributes usually occur before the noun, so adjectival words that occur in the same function after the noun are not analysed on this occasion.

The nouns ČOVEK, ŽENA and MUŠKARAC were selected for analysis as nouns that are primarily or to a significant extent used to refer to adult female and male persons. In Matica srpska's Serbo-Croatian dictionary *Rečniku srpskohrvatskog književnog jezika* these meanings are given in definitions 2. for the noun ČOVEK and 1.б. for the noun ŽENA, while no criteria of age exist in the definitions of the noun MUŠKARAC or the definition 1.a. of the noun ŽENA. The noun ŽENA is also defined in 2.a. as "a married partner[3] of the female[4]", and in 2.б. as "the one who is married or was married", while the meaning of the noun ČOVEK is defined in 5. as "a husband, a [male] spouse". The meaning of the noun MUŽ is defined in 1.a. as "a married man, a male married partner, a [male] spouse" and in 1.б. as "an adult male, a man". The usage of the noun MUŽ in the meaning 1.a. is parallel to the usage of the noun ŽENA in the meanings 2.a. and 2.б., so its inclusion in the set of the nouns to be examined allows for a partial but balanced coverage of the

[2] That is, those adjectival words generated by the regular expression `[pos="A"]`, which include adjectives, ordinal numbers and adjectival pronouns. Adjectival pronouns require different search queries and a different approach to the classification of the results, so they have been omitted from this study and can be the subject of a future one.

[3] The translations of dictionary definitions are adapted to English usage (for example, "bračni drug" translated as "a married friend" or "a married companion" would make little sense here).

[4] The grammatical gender of the Serbian nouns is made explicit here for convenience.

nouns that refer to male and female spouses. Different meanings of each of the four nouns are not treated separately in this study, and this has to be kept in mind when interpreting its results. There is a certain degree of synonymy between the nouns ČOVEK, MUŠKARAC and MUŽ, and most of their meanings are antonymous with different meanings of the noun ŽENA. The most important exception is the generic meaning of the noun ČOVEK, which can partly explain the difference in the number of results returned by the corpus search for the nouns ČOVEK, MUŠKARAC and MUŽ on the one hand (87,522 results) and the noun ŽENA on the other (56,402).

2 Some related research

An exhaustive overview of the research that has influenced the approach to and the treatment of this paper's topic is not possible here, but the most important ideas and findings ought to be mentioned.

Carmen Rosa Caldas-Coulthard and Rosamund Moon have had a decisive influence on affirming a corpus approach to gender inequalities in language. Their pioneering research, presented at a symposium dedicated to critical discourse analysis, which was held at the University of Birmingham in 1999, shows that the corpus search can yield solid evidence of discrimination, sexualisation and the construction of gender in the media. They initially analysed the adjectival attributes of the nouns MAN, WOMAN and GIRL in the British press (Caldas-Coulthard and Moon, 2010), but the expanded version of the original presentation, published in 2010, includes the noun BOY as well.

Using the entire Bank of English corpus, Moon has conducted research into gender and age stereotypes and prejudices by analysing adjectival premodification of noun phrases headed by the nouns MAN/MEN and WOMAN/WOMEN modified by adjectival attributes that clearly denote age, such as *young* and *old*. Her results show that young men and women are mostly described by adjectives with positive connotations, while adjectives used to describe older men and women often have negative connotations (Moon, 2014).

Cvetana Krstev and colleagues have applied methods of corpus and computational linguistics, as well as methods of information extraction, in two studies of the representation of women in the Serbian press (Gucul-Milojević et al., 2010; Крстев и др., 2015). At the centre of their research is the identification and classification of personal names. Their results show that women are much less represented in the Serbian press than

men and that prominent women get significantly less room than prominent men.

Michael Pearce has used Sketch Engine[5] to examine the grammatical roles of object and subject for the nouns MAN and WOMAN, as well as their adjectival attributes, in the British National Corpus.[6] He has relied primarily on Sketch differences, that is, an automatically generated overview of similarities and differences in the grammatical and collocational behaviour of two selected words, based on which he concludes that, in the corpus analysed, the stereotypical representations of men emphasise such traits as competitiveness, adventurousness, independence, rationality, aggression, physical strength and muscularity; while the stereotypical representations of women emphasise co-operativeness, gentleness, dependency, emotionality, sympathy and physical weakness. Pearce observes that certain patterns have an uneven distribution in the BNC. For example, prose literature contains 61.45% of the noun phrases in which adjectives denoting states of neuroticism modify the noun WOMAN, even though prose literature makes up only 16% of the word count of the BNC (Pearce, 2008).

Paul Baker has researched various manifestations of sexism in the Corpus of Historical American English (COHA)[7] and has discovered that words referring to men are more frequent than those referring to women, that nouns referring to men come first in co-ordinate noun phrases referring to men and women, and that the names of occupations with the suffix -MAN have a higher frequency than those with the suffix -WOMAN. He stresses that the noun MAN is still used with its generic meaning, despite numerous criticisms of such usage (Baker, 2014, 73–104).

The results of the application of corpus methods to the investigation of the complex and historically changeable relations between gender and language are often fascinating, and sometimes shocking, and they attract the attention of the wider public, especially when the subject of analysis is the mass media, which can show very sharp differences in the representation

[5] A commercial corpus management and corpus query software developed by Lexical Computing Ltd, a company that was founded by Adam Kilgarriff in 2003 (Kilgarriff et al., 2004).

[6] Available online. (Accessed 6 December, 2016).

[7] COHA is composed of more than 400 million words of text in more than 100,000 individual texts published between 1810 and 2009. All the texts are categorised according to year and decade, allowing the evolution of a word or phrase, in terms of its frequency or meaning, to be examined over time. Available online. (Accessed 10 December, 2016).

of men and women. Sexist patterns that are specific to various spheres of interest can be revealed by analysis of specialised corpora such as sports news[8] and film scenarios.[9]

3 SrpKor2013

3.1 Composition and structure

SrpKor2013 is an electronic, synchronic corpus of contemporary Serbian language that contains 122 million words. Balance in SrpKor can be expected in future, given its dynamic character and regular updating. The corpus includes 4,890 texts of the 5,058 that have been prepared for it, which have been classified into texts originally written in Serbian (4,545 texts) and translations into Serbian (513), and according to functional styles into literary (348), scientific and popular-scientific (188), newspaper (3,245), administrative (923) and other (354). The translations into Serbian cover the period from the eighteenth century to the present, while the texts written in Serbian cover the period from 1910 to 2012. With the exception of a small number of transcripts of spoken interviews, SrpKor consists of written texts, and in the process of their collection, preference was given to the already available electronic texts of various formats, while a relatively small number of non-electronic texts were digitised. Corpus texts are annotated with bibliographic information, and tokens of corpus texts are annotated with morphological information, such as part-of-speech and lemma, so that the corpus can be searched both as a collection of texts and according to information present in bibliographic and morphosyntactic annotation (Utvić, 2014, 257–259).

3.2 Alphabet

The unification of coding for the letters of the Serbian alphabets was achieved through the application of the modified version of the codes that were used in Duško Vitas's system AURORA (Vitas, 1982). Table 1 shows

[8] University of Cambridge Research. 2016. "Aesthetics over athletics when it comes to women in sport". University of Cambridge News, 12 August 2016. Available online. (Accessed 29 December, 2016).

[9] Hanah Anderson and Matt Daniels. 2016. "Film dialogue from 2,000 screenplays, broken down by gender and age". The Pudding, April 2016. Available online. (Accessed 23 November, 2016).

Diacritics				Digraphs			
Uppercase	AURORA code	Lowercase	AURORA code	Uppercase	AURORA code	Lowercase	AURORA code
Č	CY, Cy	č	cy	NJ, Nj	NX, Nx	nj	nx
Ć	CX, Cx	ć	cx	LJ, Lj	LX, Lx	lj	lx
Đ	DX, Dx	đ	dx	DŽ, Dž	DY, Dy	dž	dy
Š	SX, Sx	š	sx				
Ž	ZX, Zx	ž	zx				

Table 1. Serbian Latin alphabet diacritics and digraphs in AURORA code

how this code scheme maps the diacritics and digraphs used in the Serbian Latin alphabet (Utvić, 2014, 279–281).[10]

3.3 Search

In the process of indexing, the text of the corpus is treated not as a sequence of characters but as a sequence of token positions in the corpus, or corpus positions. This enables direct access to every token of an electronic corpus, which makes the search more efficient (Utvić, 2014, 152). All tokens of the corpus have the positional attributes **word** (a concrete realisation of a token in the text), **lemma** (assigned to a token through automatic annotation by TreeTagger), and **pos** (part-of-speech information assigned to a token through automatic annotation by TreeTagger) (Utvić, 2011).[11] The morphosyntactic notation applied in the morphological electronic dictionary of Serbian in the LADL/DELA format[12] was also used in the annotation of the corpus, with the tags selected restricted to the basic set, which includes ten tags for different parts of speech in Serbian and six tags for specific tokens that require special treatment.[13]

[10] Available online. (Accessed 22 March, 2017).

[11] Available online. (Accessed 22 March, 2017).

[12] The morphosyntactic description in the morphological electronic dictionary of Serbian uses the DELA formats (Dictionnaire électronique du LADL), developed by the French Laboratory for Automatic Documentation and Linguistics (Laboratoire d'automatique documentaire et linguistique, LADL) (Utvić, 2014; Krstev and Vitas, 2009).

[13] Available online. (Accessed 22 March, 2017).

The corpus is managed by IMS Corpus Workbench,[14] a collection of corpus tools, the main component of which is the query processor CQP (Corpus Query Processor). The web-based interface of SrpKor enables simple and advanced search, with the advanced search utilising most of the potential of the CQP query language CQL (CQP Query Language). The regular expression syntax that CQL is based on is a subset of the syntax of POSIX regular expressions (Utvić, 2014).

The simple search is the search of the corpus texts, not of annotation, and uses only the first positional attribute (**word**), so that it is sufficient, unlike in the advanced search, to use only the value of the positional attribute without the accompanying quotation marks, but the values of the positional attribute have to be encoded by the code AURORA. The advanced search enables a search of the morphological annotation of the corpus, that is, of the positional attributes **pos** and **lemma**. Regular expressions can be used in both the simple and advanced searches, in the first case within the value of the positional attribute **word**, and in the second as regular expressions over tokens. The corpus search generates results in the form of concordances that can be displayed in KWIC (Key Word in Context) and KWIP (Keyword in Paragraph) formats (Utvić, 2014, 308–313). The present research is based on the results generated by the advanced search and accessed in the KWIC format mostly with the default values. For a certain number of queries the results were examined page by page, as this was the only way to resolve ambiguity, polysemy and homonymy of forms.

3.4 Female authors in the corpus

Data on authorship is available for 713 texts of the 5,058 that have been prepared for the corpus, with 348 texts belonging to the literary style, 130 to the scientific and popular-scientific style, 67 to the newspaper style and 171 to other functional styles. The contribution of female authors is 52.5[15] texts in the scientific and popular-scientific style (40.4%), 23 texts in the newspaper style (34.3%), 30 texts in the literary style (8.6%) and 21.33[16] texts in other functional styles (12.5%).

[14] Available online. (Accessed 22 March, 2017).
[15] 0.5 stands for a text by two authors one of whom is female.
[16] 0.33 stands for a text by three authors one of whom is female.

4 The results

4.1 The initial set

The corpus search was conducted in several phases. The initial set of nouns did not include the noun MUŽ, so the initial tests of the hypothesis about the differences in the distribution of adjectival words in the role of adjectival attributes with the nouns that denote adult male and female persons were done for the noun phrases headed by the nouns ČOVEK, MUŠKARAC and ŽENA. As discussed in the introduction, the search was restricted to the adjectival words that occur before the nouns.

Eighty-five adjectival words were selected for examination, and most of them were classified into 35 antonym pairs or pairs of sets of synonyms and their antonyms. The search was performed using the queries that specify the relevant lemmas,[17] which meant that in the case of polysemous and homonymous lemmas a detailed review of the concordances was necessary to eliminate irrelevant results.

The analysis of the results has shown that two adjectival words from a particular synonym or antonym pair can have opposite frequency patterns, so that one of them occurs more frequently than the other with two of the three nouns. For example, in the antonym pair *mršav-debeo* (*thin-fat*), the adjective *debeo* has a higher frequency than its antonym with the noun ŽENA, while the reverse applies to the nouns ČOVEK and MUŠKARAC; in the pair *nag-go* (*nude-naked*), the adjective *nag* has a higher frequency than its synonym with the nouns MUŠKARAC and ŽENA, while the reverse applies to the noun ČOVEK; in the antonym pair *srećan-nesrećan* (*happy-unhappy*), the adjective *srećan* has a higher frequency than its antonym with the nouns ČOVEK and MUŠKARAC, while the reverse applies to the noun ŽENA.

However, adjectival words that were selected for this analysis were assumed to have wildly different patterns of distribution and frequencies, and even different meanings or connotations, with the nouns under consideration, and such an assumption in their selection introduces an important limitation for the generalisation of any conclusions that can be drawn from the data obtained. Objectivity could be achieved only through an exhaustive search of all the adjectival words in the relevant function and position, and by then selecting the most frequent among them for further analysis.

[17] For example the search queries for the noun phrases *uspešan* ČOVEK, *uspešan* MUŠKARAC and *uspešna* ŽENA are `[lemma="uspesxan"] [lemma="cyovek"]`, `[lemma="uspesxan"] [lemma="musxkarac"]` and `[lemma="uspesxan"] [lemma="zxena"]`.

NOUN	QUERY 1	MATCHES
ČOVEK	`[lemma="cyovek"]`	65,345
ŽENA	`[lemma="zxena"]`	56,402
MUŠKARAC	`[lemma="musxkarac"]`	15,183
MUŽ	`[lemma="muzx"]`	6,994

NOUN	QUERY 2	MATCHES
ČOVEK	`[pos="A"][lemma=cyovek]`	20,650
ŽENA	`[pos="A"][lemma="zxena"]`	11,606
MUŠKARAC	`[pos="A"][lemma="musxkarac"]`	2,890
MUŽ	`[pos="A"][lemma="muzx"]`	1,129

Table 2. Frequency of the nouns in the corpus and frequency of the adjectives preceding them.

4.2 Frequency lists

In the next phase, the corpus search results generated by the search queries `[pos="A"] [lemma="cyovek"]`, `[pos="A"] [lemma="zxena"]`, `[pos="A"] [lemma="musxkarac"]` and `[pos="A"] [lemma="muzx"]` were normalised to the canonical forms of the adjectival words and the nouns, that is, their lemmas, and ordered by frequency using the corpus processing suite Unitex.[18] The initial set of nouns is expanded by the last query to include the noun MUŽ, which, in its meaning "a married man, a male married partner, a [male] spouse", defined in 1.a., is parallel to the very common meaning of the noun ŽENA, so it was not necessary to exclude from the search results for the search query `[pos="A"] [lemma="zxena"]` those in which the noun ŽENA the meaning "a married partner of the female sex, a [female] spouse", defined in 2.a.. Additionally, the noun MUŽ also has the meaning "an adult male, a man", defined in 1.б. (PMC, 1967-1976).

A general insight into the frequency of the nouns under consideration and the adjectival words before them was obtained using the search queries given in Table 2.

In accordance with the annotation of the corpus, query 2 does not specify the forms of the noun and the adjectival word, nor does it require number, gender and case agreement, so the results containing the noun phrases sought are not identical with but are instead a subset of the results generated by

[18] http://unitexgramlab.org (Accessed 22 March, 2017).

this search, as they include all bigrams that consist of any adjectival word[19] and one of the four nouns, regardless of their grammatical relation.

The minimum frequency threshold for these bigrams was arbitrarily set to 5 occurrences in the whole corpus,[20] leaving aside the different frequencies of the nouns in them. In addition to the frequency of the bigrams in the whole corpus, their frequency in relation to the frequency of the nouns is also analysed. These values are from here on referred to as the frequencies of the adjectival words (or adjectives)[21] with the nouns.[22] According to the frequency lists, the minimum frequency threshold was satisfied by 335 of the bigrams containing the noun ČOVEK, 270 of those containing the noun ŽENA, 101 of those containing the noun MUŠKARAC and 22 of those containing the noun MUŽ, which gave a total of 474 adjectival words.

4.3 The review of the concordances and the classification of the results

Further analysis required that the concordances be generated for the purpose of determining if the bigrams in the frequency lists correspond to the noun phrases sought. The queries that specify the lemma of one of the 474 adjectival words followed by the lemma of one of the four nouns were used for this purpose (see footnote 17 for an example). Thus, the frequency lists served as an initial step in the selection of adjectival words, while the analysis of the concordances allowed for further examination of the morphological, syntactic and lexical-semantic properties and relations of the selected adjectival words and the four nouns under consideration. Partial incongruity of the frequency lists and the results generated by the corpus search is inevitable given the methodological differences between the two approaches. Those adjectival words that did not occur in a sufficient number of relevant results generated by the corpus search were excluded from further analysis.

[19] That is, any adjectival word generated by the regular expression `[pos="A"]`.

[20] Frequencies lower than 5 instances in the whole corpus were equated with 0.

[21] For an explanation of the difference between the two terms, see the introduction and footnote 1. The latter (narrower) term will be used wherever adjectives alone are discussed (not including participles).

[22] For example, the noun ŽENA occurs 56,402 times in the corpus, and the noun ČOVEK 65,345 times. An adjective that occurs 100 times with each of the nouns will have the frequency of 0.18% with the noun ŽENA and of 0.15% with the noun ČOVEK. Therefore, its frequency is higher with the noun ŽENA.

The analysis of the concordances has also shown that certain adjectival words , such as *divalj* (*wild*), *izgubljen* (*lost*), *mali* (*small*), *slobodan* (*free*) and *velik* (*big*), can have different meanings depending on which of the four nouns they modify. These adjectival words were assigned to a group of their own, the analysis of which is not presented in this paper.

All the other adjectival words are classified into the following groups:

1. Age;
2. Physical appearance;
3. State;
4. Predispositions;
5. Subjective evaluation;
6. Social and personal status;
7. The personality model;
8. Emphasis and distinction;
9. Belonging;
10. Relations.

All these groups were formed as lexical fields based on the results from the corpus specifically for the purpose of this research, and the selection of their members depended primarily on the author's intuition. They may partially overlap with different semantic, lexicological and computational models, but they are not based on them. Group 7 was formed in accordance with the five-factor personality model (Docs, 2017), which, considering its nature and purpose, could not supply the members of this group, so, given the approach to their selection, this group is not fundamentally different from the other groups. This classification is intended to facilitate a clearer representation of the data, but in interpreting the results for each group it is necessary to keep in mind which adjectival words it includes, given the inevitable subjectivity of their selection.

All adjectival words included in these groups have occurred in the corpus search with a meaning that corresponds to a given group and the function of an adjectival attribute in at least 5 results with at least one of the nouns. The queries for some of them, however, had generated the results that were not taken into consideration in cases of irrelevant functions, low frequencies of other meanings for polysemous adjectival words and wrong parts of speech for polysemous and homonymous lemmas. For example, the adjective *pun* (*full*) was included in the meaning 6.a. "the one who is moderately fat, fattened, large, plump" (PMC, 1967-1976), but it occurs with other meanings

in an insufficient number of results (example 1) some of which do not correspond to the function being sought (example 2). The query `[lemma="pun"] [lemma="cyovek"]` returns the adverb *puno* as well (example 3).

example 1 *...da bi ispunio misiju pravog i punog čoveka* (...to fulfil the mission of the real and complete man.).
example 2 *Posramljeni u čekaonici punoj žena...* (Embarrassed in the waiting room full [of] women)
example 3 *...kod nas ima puno žena slikara...*(...we have many women painters...)

Group 1: Age. This group includes the following adjectives:

Serbian	English
mlad, *star*,[23] *sredovečan*, *postariji*, *tridesetogodišnji*, *dvadesetpetogodišnji*, *četrdesetogodišnji*, *pedesetogodišnji*, *odrastao*, *punoletan*, *mator*, *ostareo*, *vremešan*, *zreo*, *godišnji*.	*young*, *old*, *middle-aged*, *elderly*, *thirty-year-old*, *twenty-five-year-old*, *forty-year-old*, *fifty-year-old*, *grown-up/adult*, *adult*, *age-old*, *aged*, *timeworn*, *mature*, *year-old*.

Adjectives belonging to group 1 (chart 1 in the Appendix) occur with all four nouns and have the highest frequency with the noun ŽENA (3%). In this group, the adjective that has the highest frequency with one of the nouns

[23] The Serbian adjective *star* is here included in the meaning 1.a. "who has lived for many years", but it can also denote a belonging to a sequence (group 9, subgroup 9.1), in its meanings 2.a. "the one that was in the past, former" and 2.б. "previous, earlier, past", but also a belonging to a culture (group 9, subgroup 9.2) in the meaning 4. "who, in characteristics or outlook, belongs to a previous time" and a social status derived from experience, importance or rank (group 6, subgroup 6.4) in the meaning 6.a. "long-standing, long known for something" and 8.в. "higher in rank, higher in some hierarchy". The search query `[lemma="star"] [lemma="cyovek"] ` generates less than 3% of results in which the adjective *star* has one of these meanings, while the search query `[lemma="star"] [lemma="zxena"]` returns only one result (out of 317) that does not have the meaning defined in 1.a. The search query `[lemma="star"] [lemma="muzx"]` also returns one result with the meaning 2.б., and additionally one result in which the meaning 1.a. is not necessarily the primary one. With the noun MUŠKARAC the adjective *star* occurs exclusively in the meaning 1.a. (PMC, 1967-1976).

is the adjective *mlad* with the noun ČOVEK (2.28%), and the most frequent noun phrases are *mlad* ČOVEK (1,488), *mlada* ŽENA (1,196) and *stara* ŽENA (317).

Group 2: Physical appearance. Depending on what aspect of physical appearance they describe, these adjectival words are classified into three subgroups:

Subgroup	Serbian	English
2.1. Colour	*plavokos*, *beo*, *crn*, *plav*, *plavook*, *sed*, *žut*, *bled*, *crnomanjast*, *tamnoput*;	*blond(e)-haired*, *white*, *black*, *blond(e)*, *blue-eyed*, *grey*, *yellow*, *pale*, *dark*, *dark-skinned*;
2.2. Figure	*visok*, *gojazan*, *punačak*, *vitak*, *debeo*, *mršav*, *krupan*, *nizak*, *onizak*, *ogroman*, *sitan*, *pun*, *razvijen*;	*tall*, *obese*, *chubby*, *slender*, *fat*, *thin*, *large*, *short*, *shortish*, *huge*, *small*, *full*, *muscular*
2.3. Current appearance	*nag*, *go*, *odeven*, *razgolićen*, *obučen*,[24] *zabrađen*, *doteran*, *naoružan*, *brkat*, *bradat*, *maskiran*.	*nude*, *naked*, *clothed*, *scantily dressed*, *dressed*, *head-scarved*, *well-groomed*, *armed*, *moustached*, *bearded*, *masked*.

Adjectival words belonging to group 2 (chart 2) do not occur with the noun MUŽ, and they have the highest frequency with the noun MUŠKARAC (1.4%). In this group, the adjectival word that has the highest frequency with one of the nouns is the participle *naoružan* with the noun MUŠKARAC (0.4%), while the most frequent noun phrases are *beo* ČOVEK (81), *visok* ČOVEK (76) and *naoružan* MUŠKARAC (61). The most frequent with nouns ČOVEK and ŽENA are adjectives of the subgroup 2.1, and with the noun MUŠKARAC adjectival words of the subgroup 2.3.

Group 3: State. Depending on the type of state that they describe, these adjectival words are classified into three subgroups.

[24] This is the passive participle of the verb *obući* (*to dress*), which is homographic with the passive participle of the verb *obučiti* (*to train*) in the corpus search, because of its lack of accents. The passive participle of the verb *obučiti* occurs once, with the noun ŽENA, and that result is not included here.

Subgroup	Serbian	English
3.1. Emotional	*srećan*, *nesrećan/nesretan*, *unesrećen*, *zaljubljen*, *razočaran*, *očajan*, *tužan*, *uplašen*, *uznemiren*;	*happy*, *unhappy*, *grief-stricken*, *in love*, *disappointed*, *desperate*, *sad*, *scared*, *upset*;
3.2. Permanent physical	*živ*, *mrtav/umrli*, *pokojni*, *preminuli*, *smrtan*;	*alive*, *dead*, *deceased*, *departed*, *mortal*;
3.3. Current psycho-physiological	*bolestan/oboleo*, *zaražen/inficiran*, *zdrav*, *gladan*, *umoran*, *trudan*, *noseći*, *bremenit*, *pijan*, *usamljen*, *uplakan*, *preporođen*.	*sick/ill*, *infected*, *healthy*, *hungry*, *tired*, *pregnant*, *carrying*, *bearing*, *drunk*, *lonely*, *crying/tearful*, *rejuvenated*.

Adjectival words belonging to group 3 (chart 3) occur with all four nouns and have the highest frequency with the noun ČOVEK (1.29%). In this group, the adjectival word that has the highest frequency with one of the nouns is the adjective *nesrećan* with the noun ŽENA (0.27%), and the most frequent noun phrases are *nesrećna* ŽENA (155), *srećan* ČOVEK (154) and *nesrećan* ČOVEK (126). The most frequent with the noun ČOVEK are adjectival words of the subgroup 3.1, with the noun ŽENA adjectival words of the subgroup 3.3, and with the nouns MUŠKARAC and MUŽ adjectival words of the subgroup 3.2.

Group 4: Predisposition. Depending on the type of predisposition that they denote, these adjectives are classified into two subgroups:

Subgroup	Serbian	English
4.1. Physical and psycho-physiological	*plodan*, *sterilan*, *levoruk*, *slep*, *jak/snažan*, *brz*, *slab*, *heteroseksualan*, *sposoban*, *feminiziran*;	*fertile*, *sterile*, *left-handed*, *blind*, *strong*, *fast*, *weak*, *heterosexual*, *capable*, *effeminate*;
4.2. Intellectual	*genijalan*, *inteligentan*, *talentovan*, *darovit*, *pametan*, *glup*.	*genius*, *intelligent*, *talented*, *gifted*, *smart*, *stupid*.

Adjectives belonging to group 4 (chart 4) do not occur with the noun MUŽ, and they have the highest frequency with the noun ČOVEK (0.83%). In this group, the adjective that has the highest frequency with one of the nouns is the adjective *pametan* with the noun ČOVEK (0.25%), and the most frequent noun phrases are *pametan* ČOVEK (166), *brz* ČOVEK (77) and *jak* ČOVEK (68). The most frequent with the nouns ČOVEK and MUŠKARAC are adjectives of the subgroup 4.1, and with the noun ŽENA adjectives of the subgroup 4.2.

Group 5: Subjective evaluation. Group 5 includes adjectives that convey a subjective evaluation, and, depending on what is being evaluated, they are classified into four subgroups:

Subgroup	Serbian	English
5.1. Subjective evaluation of effect	*privlačan/atraktivan*, *poželjan*, *fatalan*, *zanosan*, *prijatan*, *strašan*, *odvratan*, *zagonetan*, *šarmantan*, *opasan*, *drag*, *mio*, *čudan*, *čudnovat*, *dražestan*, *duhovit*, *zanimljiv*, *tajanstven*, *simpatičan*, *ljubak*, *misterioozan*, *mračan*, *otmen*, *prost/jednostavan*, *komplikovan*, *nezgodan*;	*attractive*, *desirable*, *fatal*,[25] *captivating*, *pleasant*, *frightening*, *disgusting*, *intriguing*, *charming*, *dangerous*, *dear*, *sweet*, *weird*, *peculiar*, *delightful*, *witty*, *interesting*, *secretive*, *appealing*, *lovely*, *mysterious*, *gloomy*, *decorous*, *simple*, *complicated*, *difficult*;
5.2. Subjective evaluation of appearance	*lep*, *prelep*, *ružan*, *elegantan*, *naočit*, *stasit*, *zgodan*;	*beautiful*, *gorgeous*, *ugly*, *elegant*, *eye-catching*, *well-built*, *shapely/handsome*;
5.3. Subjective evaluation of morals	*zao*, *dobar/valjan*, *loš/rđav*, *ispravan*, *grešan*, *moralan*, *svet*, *lak*, *pao*;	*evil*, *good*, *bad*, *upright*, *sinful*, *moral*, *saintly*, *easy*, *fallen*;
5.4. General subjective evaluation	*krasan/divan/predivan*, *idealan*, *savršen*, *sjajan*, *predivan*, *odličan*.	*great/admirable/wonderful*, *ideal*, *perfect*, *brilliant*, *great*.

Adjectives belonging to group 5 (chart 5) occur with all four nouns and have the highest frequency with the noun ŽENA (1.82%). In this group, the adjective that has the highest frequency with one of the nouns is the adjective *lepa* with the noun ŽENA (0.86%), and the most frequent noun phrases are

[25] Only in the sense of "femme fatale".

lepa ŽENA (487), *dobar* ČOVEK (366) and *dobra* ŽENA (78). The most frequent with the nouns ŽENA and MUŠKARAC are adjectives of the subgroup 5.2, and with the nouns ČOVEK and MUŽ adjectives of the subgroup 5.3.

Group 6: Social and personal status. Depending on the type of status that they denote, these adjectival words are classified into seven subgroups:

Subgroup	Serbian	English
6.1. Marital status	*udat/oženjen*, *bivši*, *budući*, *neudat/neoženjen*, *zakonit*, *razveden*, *venčan*, *nevenčan*, *sadašnji;*	*married*, *former*, *future*, *unmarried*, *lawful*, *divorced*, *wedded*, *unwedded*, *present*;
6.2. Educational and professional status	*zaposlen*, *nezaposlen*, *visokoobrazovan*, *obrazovan*, *neobrazovan*, *učen*, *školovan*, *stručan*, *nepismen*, *pismen*, *vešt*;	*employed*, *unemployed*, *highly educated*, *educated*, *uneducated*, *learned*, *schooled*, *professional*, *illiterate*, *literate*, *skillful*;
6.3. Financial status	*bogat*, *siromašan*, *imućan*, *prebogat*;	*rich*, *poor*, *wealthy*, *super-rich;*
6.4. General social or personal status	*biviši*,[26] *uspešan*, *moćan*, *nemoćan*, *uticajan*, *poznat*, *slavan*, *anoniman*, *znamenit*, *značajan*, *važan*, *ostvaren*, *ugledan*, *nevin/nedužan*, *nevidljiv*, *zaslužan*, *popularan*, *glavni/čelni*, *vodeći*, *uzoran*, *pobunjen*, *bespomoćan*, *jadan*, *sirot*, *iskusan*,[27] *ključni*;	*former*, *successful*, *powerful*, *powerless*, *influential*, *known*, *famous*, *anonymous*, *notable*, *significant*, *important*, *fulfilled*, *reputable*, *innocent/guiltless*, *invisible*, *deserving*, *popular*, *main/head*, *leading*, *exemplary*, *rebellious*, *helpless*, *miserable*, *unfortunate*, *experienced*, *key*;
6.5 Emotional status	*omiljen*, *voljen*, *ljubljen*	*favourite*, *loved*, *beloved*;

[26] The noun phrase *bivši* ČOVEK is an expression with the meaning "ruined man" (PMC, 1967-1976).

[27] The adjective *iskusan* with the noun ČOVEK has the meaning of professional and life experience, while with the noun ŽENA it most frequently refers to sexual experience.

6.6. Victimhood and criminality	*ubijen*, *silovan*, *zlostavljan*, *ranjen*, *pretučen*, *otet*, *vezan*, *prevaren*, *poražen*, *povređen*, *ugrožen*, *zarobljen*, *uhapšen*, *osuđen*;	*killed/murdered*, *raped*, *abused*, *wounded*, *beaten*, *kidnapped/abducted*, *tied*, *deceived/cheated*, *defeated*, *injured*, *endangered*, *imprisoned*, *arrested*, *convicted*;
6.7. Definition and a change of status	*pregledan*, *ispitan*, *anketiran*, *izložen*, *postavljen*, *tražen*, *neimenovan*, neidentifikovan, *izabran*.	*examined*, *questioned*, *surveyed*, *exposed*, *appointed*, *wanted*, *unnamed*, *unidentified*, *chosen/selected*.

Adjectival words belonging to group 6 (chart 6) occur with all four nouns and have the highest frequency with the noun ČOVEK (2.38%). In this group, the adjectival word that has the highest frequency with one of the nouns is the adjective *bivši* with the noun MUŽ (1.82%), and the most frequent noun phrases are *bogat* ČOVEK (238), *glavni* ČOVEK (155) and *čelni* ČOVEK (149). The most frequent with all four nouns are adjectival words of the subgroup 6.4.

Group 7: Personality model These adjectival words are classified according to the five-factor model of personality into the following five subgroups:

Subgroup	Serbian	English
7.1. Emotional stability	*depresivan*, *frigidan*, *miran*, *neurotični*, *obezglavljen*, *smiren*, *poremećen*, *zadovoljan*, *histeričan*, *ljubomoran*, *vatren*, *normalan*, *lud*, *osetljiv*;	*depressed*, *frigid*, *peaceful*, *neurotic*, *frantic*, *calm*, *disturbed*, *satisfied*, *hysterical*, *jealous*, *fiery*, *normal*, *crazy*, *sensitive*;
7.2. Extroversion	*aktivan*, *otresit*, *odlučan*, *vedar*, *nasmejan*, *društven*, *energičan*, *raspoložen*, *veseo*, *neustrašiv*, *povučen*, *privatan*, *tih*, *ćutljiv*, *zatvoren*, *stidljiv*, *preduzimljiv*, *ambiciozan*;	*active*, *bold*, *resolute*, *bright*, *smiling*, *sociable*, *energetic*, *cheerful*, *merry*, *fearless*, *withdrawn*, *private*, *quiet*, *silent*, *introverted*, *shy*, *enterprising*, *ambitious*;

7.3. Agreeableness	*častoljubiv*, *ljubazan*, *pristojan*, *pažljiv*, *kulturan*, *iskren*, *poverljiv*, *sujetan*, *samoljubiv*, *dobronameran*, *pošten*, *čestit/častan*, *vaspitan*, *plemenit*, *nasilan*, *agresivan*, *human*, *pravedan*, *nepravičan*, *surov*, *skroman*, *pravičan*, *tolerantan*, *prirodan*, *lukav*, *fin*;	*fame-seeking*, *kind*, *decent*, *attentive*, *cultured*, *sincere*, *confidential*, *vain*, *self-loving*, *well-meaning*, *honest*, *honourable*, [*well*] *bred*, *noble*, *violent*, *aggressive*, *humane*, *just*, *unfair*, *cruel*, *modest*, *fair*, *tolerant*, *natural*, *cunning*, *nice*;
7.4. Openness	*kreativan*, *strastan*, *svestran*, *hrabar*, *odvažan*, *samosvestan*, *prosvećen*, *emancipovan*, *blag*, *pragmatičan*, *praktičan*, *neupućen*, *nezavisan*, *naivan*, *zainteresovan*;	*creative*, *passionate*, *versatile*, *brave*, *bold*, *self-aware*, *enlightened*, *emancipated*, *mild*, *pragmatic*, *practical*, *uninformed*, *independent*, *naïve*, *interested*;
7.5 Conscientiousness	*umeren*, *odmeren*, *pouzdan*, *vredan/radan*, *odgovoran*, *razborit*, *realan*, *razuman*, *racionalan*, *mudar*, *misaon/misleći/uman*, *ozbiljan*, *civilizovan*, *principijelan*, *uporan*, *veran*, *neveran*, *porodičan*, *poslovan*, *dostojanstven*, *neodgovoran*, *lenj*, *oprezan*, *trezven*, *strog*.	*moderate*, *measured*, *reliable*, *industrious/diligent*, *responsible*, *sensible*, *realistic*, *reasonable*, *rational*, *wise*, *thoughtful/thinking*, *serious*, *civilised*, *principled*, *persistent*, *faithful*, *unfaithful*, *family*, *business*, *dignified*, *irresponsible*, *lazy*, *cautious*, *sober*, *strict*.

Adjectival words belonging to group 7 (chart 7) occur with all four nouns and have the highest frequency with the noun ČOVEK (3.62%). In this group, the adjectival word that has the highest frequency with one of the nouns is the adjective *poslovan* with the noun ČOVEK (0.48%), and the most frequent noun phrases are *poslovan* ČOVEK (313), *poslovna* ŽENA (169) and *pošten* ČOVEK (163). The most frequent with the nouns ČOVEK and ŽENA are adjectival words of the subgroup 7.5, with the noun MUŠKARAC adjectival words of the subgroups 7.5 and 7.3, and with the noun MUŽ adjectival words of the subgroup 7.1.

Group 8: Emphasis and distinction. Adjectives belonging to group 8 (chart 1) determine the extent to which a meaning of the noun that is common or dominant in a given social or personal context applies to the person the noun refers to.

Serbian	English
prav, *stvaran*, *istinski*, *poseban*, *izuzetan*, *prosečan*, *neponovljiv*, *običan*, *neobičan*, *redak*, *drugačiji*, *drukčiji*, *jedinstven*, *izvanredan*.	*real*, *actual*, *true*, *special*, *exceptional*, *average*, *unrepeatable*, *ordinary*, *unusual*, *rare*, *different*, *unique*, *extraordinary*.

These adjectives do not occur with the noun MUŽ, and they have the highest frequency with the noun ČOVEK (1.25%). In this group, the adjective that has the highest frequency with one of the nouns is the adjective *običan* with the noun ČOVEK (0.69%), and the most frequent noun phrases are *običan* ČOVEK (448), *pravi* ČOVEK (125) and *prosečan* ČOVEK (63).

Group 9: Belonging. Depending on the type of belonging that they attribute to the person the noun refers to, these adjectival words are classified into seven subgroups:

Subgroup	Serbian	English
9.1. Sequence	*prvi*, *drugi*, *treći*, *četvrti*, *peti*, *šesti*, *sedmi*, *deseti*, *dvanaesti*, *poslednji*;	*the first*, *the second*, *the third*, *the fourth*, *the fifth*, *the sixth*, *the seventh*, *the tenth*, *the twelfth*, *the last*;
9.2. Period and culture	*savremen*, *moderan*, *današnji*, *drevan*/*prvobitan*, *praistorijski*, *renesansni*, *antički*, *srednjovekovan*, *pećinski*, *starozavetni*, *primitivan*;	*contemporary*, *modern*, *today's*, *primeval*/*primordial*, *prehistoric*, *Renaissance*, *ancient*,[28] *medieval*, *cave*, *Old Testament*, *primitive*;
9.3. Place and culture	*ruski*, *zapadni*, *evropski*, *svetski*, *američki*, *gradski*, *srpski*, *lokalan*/*ovdašnji*/*tamošnji*, *sovjetski*, *šidski*, *narodski*/*narodni*, *urbani*, *seoski*, *turski*, *romski*, *albanski*, *balkanski*, *arapski*, *avganistanski*, *iranski*, *španski*, *bošnjački*, *britanski*, *nemački*;	*Russian*, *Western*, *European*, *world*, *American*, *city*, *Serbian*, *local* /[*of this place*]/[*of that place*], *Soviet*, *Šid's*, *people's*, *urban*, *rural*/*village*, *Turkish*, *Romani*, *Albanian*, *Balkan*, *Arab*, *Afghan*, *Iranian*, *Spanish*, *Bosniak*, *British*, *German*;

[28] Only with the sense of classical antiquity

9.4. Religion	*religiozan*, *pobožan*, *božji*, *crkven*, *verujući*, *duhovan*, *pravoslavan*, *muslimanski*, *jevrejski*;	*religious*, *devout*, *God's*, *Church's*, *believing*, *spiritual*, *Orthodox* [*Christian*], *Muslim*, *Jewish*;
9.5. Profession	*fudbalski*, *pozorišni*, *javan*;[29]	[*of*] *football*, [*of*] *theatre*, *public*;[30]
9.6. Political and state structures	*Miloševićev*, *partijski*, *oligarhijski*, *demokratski*, *Dinkićev*, *Tadićev*, *knežev*, *carski*, *politički*;	*Milošević's*, *party*, *oligarchic*, *democratic*, *Dinkić's*, *Tadić's*, *prince's*, *imperial*, *political*;
9.7. Family and partner	*Petrov*, *oficirski*, *Savin*, *lekarev*, *pokojnikov*, *rođen/vlastit/sopstven.*	*Petar's*, *officer's*, *Sava's*, *doctor's*, *the deceased's*, *own.*

Adjectival words belonging to group 9 (chart 8) occur with all four nouns and have the highest frequency is with the noun ČOVEK (8.95%). In this group, the adjectival word that has the highest frequency with one of the nouns is the ordinal number *prvi* with the noun ČOVEK (5.66%), and the most frequent noun phrases are *prvi* ČOVEK (3,696), *prva* ŽENA (670) and *drugi* ČOVEK (696). The most frequent with all four nouns are the adjectival words of the subgroup 9.1.

Group 10: Relations. According to the type of relation that they denote, these adjectival words are classified into three subgroups:

Subgroup	Serbian	English
10.1. Personal	*tuđ*, *nepoznat*, *stran*, *pogrešan*, *blizak*;	*foreign/another's*, *unknown/unfamiliar*, *unfamiliar/strange*, *wrong*, *close*;

[29] It is important to note here that in the corpus the noun phrase *javna* ŽENA occurs as an expression with the meaning "a woman of easy morals, a prostitue", while the adjective *javan* with the noun ČOVEK has the meaning 3. "who is associated with the sociopolitical life" and the noun phrase *javni* ČOVEK refers to a politician (PMC, 1967-1976).

[30] Public woman (Accessed 15 November 2017).

10.2. General	*konkretan*, *određen*, *pojedini*, *ijedan*, *nijedan*, *jedini*, *pojedinačan*, *nov*, *isti*, *brojni*, *ostali*, *mnogi*, *različit*, *prisutan*	*concrete*, *specific*, *individual*, *any one*, *not one*, *the only*, *singular*, *new*, *the same*, *numerous*, *other*, *many*, *various*, *present*;
10.3. Aspect or part	*spoljašnji*, *unutrašnji*, *ceo*, *telesni*.	*external*, *internal*, *whole*, *bodily*.

Adjectival words belonging to group 10 (chart 9) occur with all four nouns and have the highest frequency with the noun MUŠKARAC (2%). In this group, the adjectival word that has the highest frequency with one of the nouns is the adjective *nepoznat* with the noun MUŠKARAC (0.82%), and the most frequent noun phrases are *mnoge* ŽENE (208), *jedina* ŽENA (178) and *jedini* ČOVEK (175). The most frequent with all four nouns are adjectival words of the subgroup 10.2.

4.4 Overview of the results

The noun ČOVEK is most likely to be modified by the adjectival words that have been analysed, with their frequency with it in the relevant function being 25.47%, compared to 15.86% with the noun ŽENA, 13.32% with the noun MUŠKARAC and 8.62% with the noun MUŽ. (chart 10).

The noun MUŠKARAC is most frequently modified by adjectives denoting age (group 1), the nouns ČOVEK and ŽENA by adjectival words attributing some type of belonging to the person the noun refers to (group 9), and the noun MUŽ by adjectival words that attribute a certain social status to the person the noun refers to (group 6).

With the noun ŽENA the frequency of adjectives denoting age is nine times higher than with the noun MUŽ and by 3–4% higher than with the nouns ČOVEK and MUŠKARAC. Adjectival words that define an aspect of the physical appearance of the person the noun refers to (group 2) do not occur with the noun MUŽ, while their frequency with the noun MUŠKARAC is higher by about 65% than with the noun ČOVEK and almost two times higher than with the noun ŽENA. With the noun ČOVEK, the frequency of adjectival words denoting a state of the person the noun refers to (group 3) is four times higher than with the noun MUŠKARAC and by about 10% higher than with the noun ŽENA. Adjectives that denote a predisposition of the person the noun refers to (group 4) do not occur with the noun MUŽ, and their frequency with the noun ČOVEK is two-and-a-half to three times higher than with the nouns ŽENA and MUŠKARAC. With the noun ŽENA, the frequency of adjectives conveying subjective evaluation (group 5) is by about 2% higher than with the noun ČOVEK and three to three-and-a-half times higher than with the nouns MUŠKARAC

and MUŽ. With the noun MUŽ, the frequency of adjectival words that attribute a certain social status to the person the noun refers to (group 6) is higher by about 25% than with the noun ŽENA, by about 50% than with the noun ČOVEK, and it is twice as high as with the noun MUŠKARAC. With the noun ČOVEK, the frequency of adjectival words related to the five-factor model of personality is three times higher than with the nouns ŽENA and MUŠKARAC and four-and-a-half times higher than with the noun MUŽ. Adjectival words that determine the degree of conventionality of the person the noun refers to (group 8) do not occur with the noun MUŽ, and their frequency with the noun ČOVEK is five times higher than with the noun ŽENA and twice as high as with the noun MUŠKARAC. With the noun ČOVEK the frequency of adjectival words that attribute some type of belonging to the person the noun refers to (group 9) is about four-and-a-half times higher than with the noun MUŽ, almost four times higher than with the noun MUŠKARAC and almost three times higher than with the noun ŽENA. With the noun MUŠKARAC, the frequency of adjectival words that refer to an aspect or a part of the person the noun refers to, or to the relation of that person and some other persons (group 10), is almost four times higher than with the noun MUŽ, 40% higher than with the noun ČOVEK and about 55% higher than with the noun ŽENA.

To sum up, adjectival words belonging to the groups 3, 4, 7, 8 and 9 have the highest frequency with the noun ČOVEK, while the frequency of adjectival words belonging to the groups 1 and 5 is the highest with the noun ŽENA, of those belonging to the groups 2 and 10 with the noun MUŠKARAC, and of those belonging to the group 6 with the noun MUŽ.

ČOVEK		ŽENA	
Serbian	English	Serbian	English
prvi (3631)	*first* (3631)	*mlad* (1196)	*young* (1196)
mlad (1488)	*young* (1488)	*prvi* (635)	*first* (635)
drugi (562)	*second* (562)	*lep* (487)	*beautiful* (487)
običan (488)	*ordinary* (488)	*drugi* (392)	*seconf* (392)
dobar (366)	*good* (366)	*star* (317)	*old* (317)
poslovan (313)	*business* (313)	*mnogi* (188)	*many* (188)
savremen (267)	*contemporary* (267)	*poslovan* (169)	*business* (169)
star (262)	*old* (262)	*jedini* (159)	*only* (159)
pametan (166)	*smart* (166)	*nesrećan* (155)	*unhappy* (155)
pošten (163)	*honest* (163)	*udat* (138)	*married* (138)
		uspešan (138)	*successful* (138)

MUŠKARAC		MUŽ	
Serbian	English	Serbian	English
mlad (180)	*young* (180)	*bivši* (127)	*former* (127)
nepoznat (124)	*unknown* (124)	*pokojni* (66)	*deceased* (66)
stran (119)	*strange* (119)	*prvi* (50)	*first* (50)
star (118)	*old* (118)	*drugi* (38)	*second* (38)
drugi (114)	*second* (114)	*dobar* (27)	*good* (27)
mnogi (84)	*many* (84)	*ljubomoran* (24)	*jealous* (24)
godišnji (69)	*yearly* (69)	*nasilan* (16)	*violent* (16)
naoružan (61)	*armed* (61)	*slavan* (16)	*famous* (16)
odrastao (60)	*adult* (60)	*nov* (15)	*new* (15)
oženjen (50)	*married* (50)	*prevaren* (14)	*cheated* (14)
		bogat (14)	*rich* (14)

5 Further work

The four nouns examined in this study occur in the corpus with different meanings that can be taken into consideration in the analysis of the results. A complementary analysis of other nouns denoting male and female persons of different ages could be very informative. Different meanings and different connotations that certain adjectival words can have depending on the noun they modify also require further analysis. Analyses of the distribution of the degrees of adjectives and of antonym pairs would produce a more precise picture of the intensity and the frequency of affirmative and negative descriptions with each of the nouns, and reveal which adjectives are most often used in relative descriptions and with which nouns. For example, the adjective *star* (*old*) never occurs in its positive degree with the noun MUŠKARAC, which it modifies more frequently than the other three nouns.

A more detailed analysis of possessive adjectives would require a careful reading of the concordances and their classification based on the types of belonging and the relationships that these adjectives denote with a given noun. The same is true of possessive pronouns, which require a different search as well.

If adjectival words influence the general perception of the meaning of the nouns that they modify, it makes sense to assume that the reverse applies too, and so the frequency of the noun phrases examined here should be analysed in relation to the frequency of the adjectival words within them.

The results from SrpKor can be compared with the results of corresponding searches in the comparable corpora of other languages. As for other potential corpora of Serbian, a significantly different approach would be necessary for work with untagged corpora, but particularly interesting for historical and diachronic

insights into the representation of men and women would be the proverbs collected by Vuk Stefanović Karadžić[31] and Matica srpska's dictionary *Rečnik srpskohrvatskog književnog jezika*, as they are products of collective consciousness at a time of ideological change and mass emancipation. The potential of dictionary analysis for studies of language and gender was demonstrated by Fournier and Russell in an interesting research into gender stereotypes in the Oxford English Dictionary (Fournier and Russell, 2004). They have shown that gender stereotypes are very frequent in citations but that they also occur in definitions, thus revealing patterns of both historical and contemporary sexism. Even on a brief examination, the aforementioned dictionary of the Serbo-Croatian language offers great inspiration for a similar undertaking. For example, for the meaning 1.a. of the noun MUŽ (HUSBAND), the definition of which is "a married man, a male married partner, a [male] spouse", the following citation is given: "He is a husband and therefore has no work at home". The citation, "We women are born destined to listen to men, even if they are stupid", illustrates the meaning 2. of the conjunction AKO (EVEN IF, ALTHOUGH, THOUGH). A number of entries, such as those for the nouns ŽENSKAĆ (FEMALE), MUŽ (SCAMP) and ANATEMA (ANATHEMA), feature citations that strongly state a preference for male children or include explicit references to the undesirability of female children and to the despair of their parents. The noun AŽDAJA (DRAGON) has a derogatory meaning of "an evil, horrible woman", defined in 2.б., and the basic meaning of "a young female horse" for the noun AJGIRUŠA is complemented by a figurative and pejorative one of "a lustful, voluptuous woman". The verb ŠEFOVATI (TO BOSS, TO BOSS AROUND) is accompanied by the following citation: "Women work–men boss around". The wry humour of the lexicographer is clear in some of these examples, but the choice of the citations that express a preference for male children is harder to interpret. The insight into historical attitudes through the prism of the lexicographer that this dictionary provides makes it a valuable source of social commentary and invites a comparison between past and present sensibilities.

References

Baker, Paul. *Using corpora to analyze gender*. Bloomsbury Academic, 2014.

Caldas-Coulthard, Carmen Rosa and Rosamund Moon. "'Curvy, hunky, kinky': Using corpora as tools for critical analysis". *Discourse & Society* Vol. 21, no. 2 (2010): 99–133.

Docs, IBM Bluemix. *Personality models*. IBM Bluemix, 2017. Accessed December 23, 2016, Available online.

[31] Compiled into an untagged corpus by Cvetana Krstev. (Accessed 13 August 2017).

Fournier, Hannah S and Delbert W Russell. "A study of sex-role stereotyping in the Oxford English Dictionary 2E". *Computers and the Humanities* Vol. 26, no. 1 (2004): 13–20.

Gucul-Milojević, Sandra, Vanja Radulović and Cvetana Krstev. "A view on the representation of women in Serbian newspaper texts". In *Applications of Finite-State Language Processing: Selected Papers from the 2008 International NooJ Conference*, Váradi, Tamás, Judit Kuti and Max Silberztein, 160–170. Cambridge Scholars Publishing, 2010. Accessed July 5, 2016, Available online.

Hoey, Michael. *Lexical priming: A new theory of words and language*. Routledge, 2004.

Kilgarriff, Adam, Pavel Rychly, Pavel Smrz and David Tugwell. "The Sketch Engine". In *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*, 105–115. 2004. Accessed November 17, 2016, Available online.

Krstev, Cvetana and Duško Vitas. "An effective method for developing a comprehensive morphological e-dictionary of compounds"/ In *Arena Romanistica — The 28th Conference on Lexis and Grammar, Bergen, 29th September–3rd October 2009*, 204–212. University of Bergen, Department of Foreign Languages, 2009.

Moon, Rosamund. "From gorgeous to grumpy: adjectives, age and gender.". *Gender & Language* Vol. 8, no. 1 (2014).

Pearce, Michael. "Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine". *Corpora* Vol. 3, no. 1 (2008): 1–29. Accessed November 15, 2016, Available online.

Utvić, Miloš. "Izgradnja referentnog korpusa savremenog srpskog jezika". Doktorska disertacija, Univerzitet u Beogradu, Filološki fakultet, 2014. Accessed October 17, 2016, Available online.

Utvić, Miloš. "Annotating the Corpus of Contemporary Serbian". *INFOtheca* Vol. 12, no. 2 (2011): 36a–47a.

Vitas, Duško. "Prikaz jednog sistema za automatsku obradu teksta". In *ROLP – Računarska obrada lingvisitčkih podataka, Bled, 7-9. X 1982.*, Tancig, Peter and Duško Vitas, Institut "Jožef Stefan", 457–461. 1982.

Крстев, Цветана, Милош Утвић и Јелена Јаћимовић. "Ако коза лаже, рог не лаже: Где су и ко су жене у српској дневној штампи". *Књиженство* Vol. 5, (2015). Accessed July 3, 2016, Available online.

РМС. *Речник српскохрватскога књижевног језика (I–III Матица српска – Матица хрватска, IV–VI Матица српска)*. Нови Сад: Матица српска, 1967-1976.

Станојчић, Живојин и Љубомир Поповић. *Граматика српскога језика*. Завод за уџбенике и наставна средства, 2000.

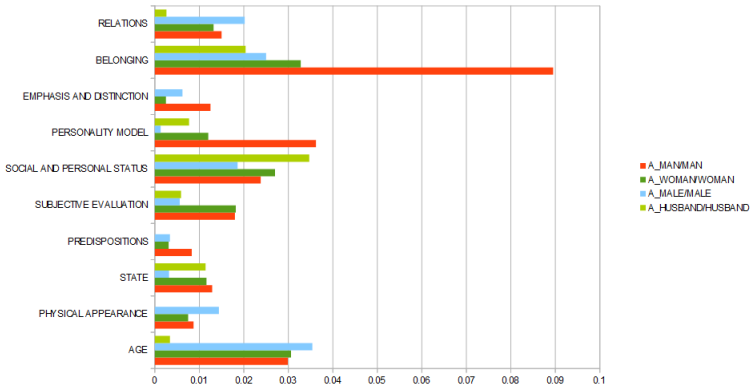

Figure 1. Overview of the groups

.

Appendix

Charts

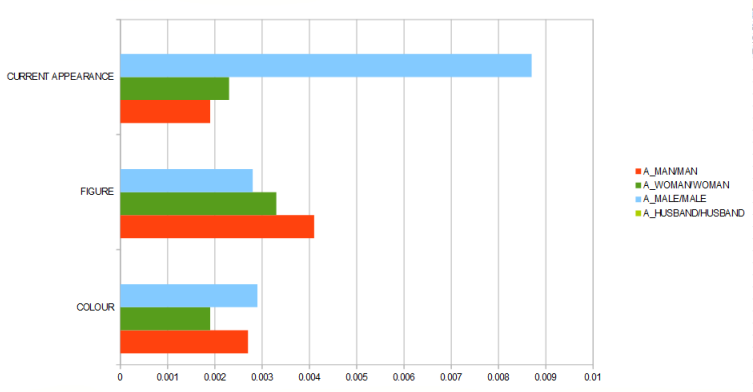

Figure 2. Physical appearance

.

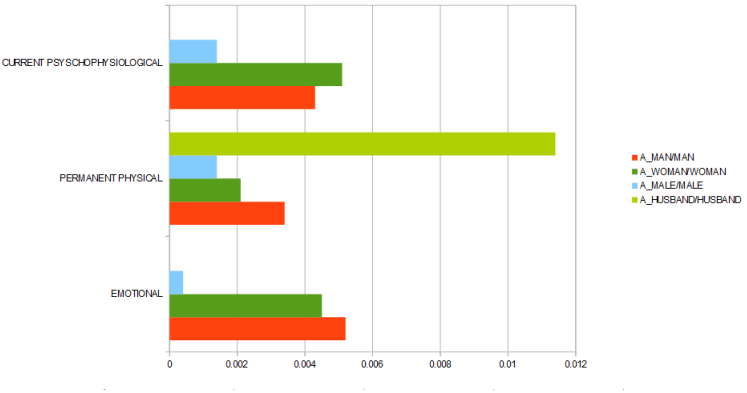

Figure 3. State

.

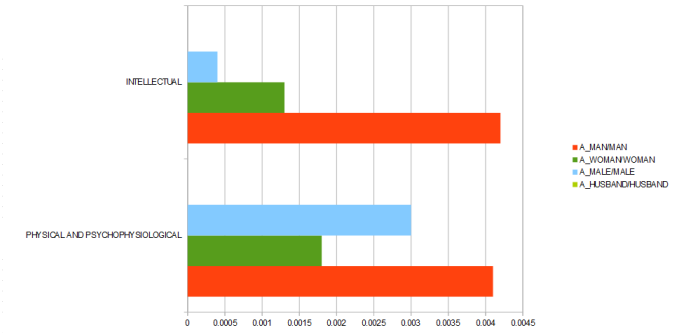

Figure 4. Predispositions

.

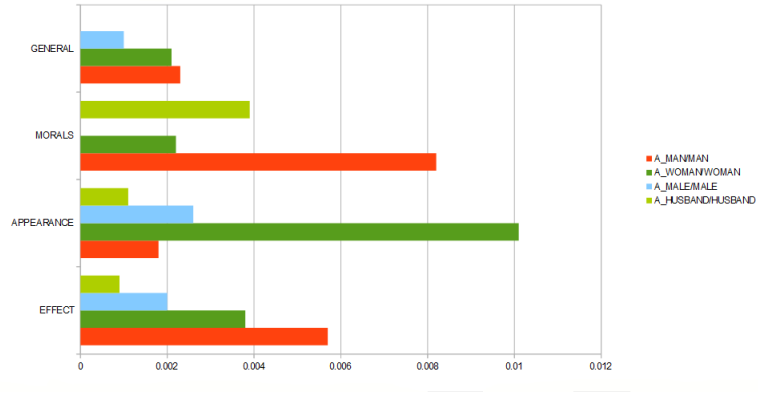

Figure 5. Subjective evaluation

.

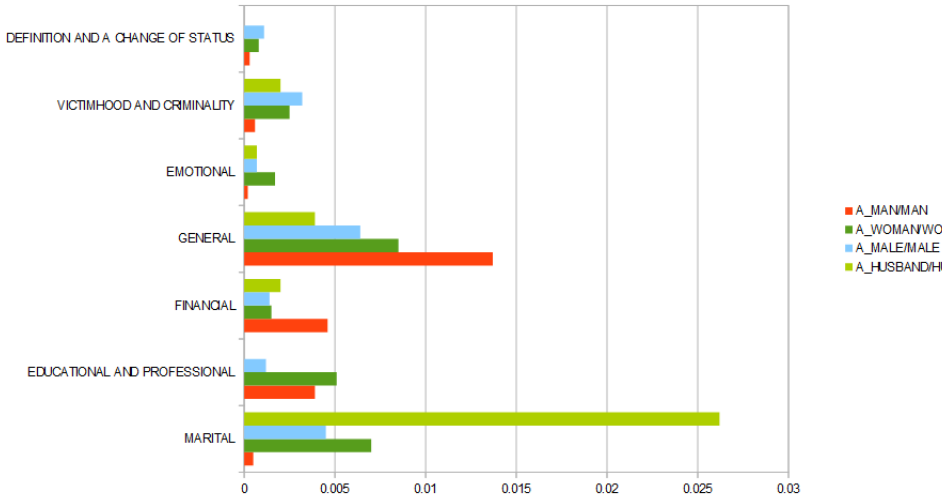

Figure 6. Social and personal status

.

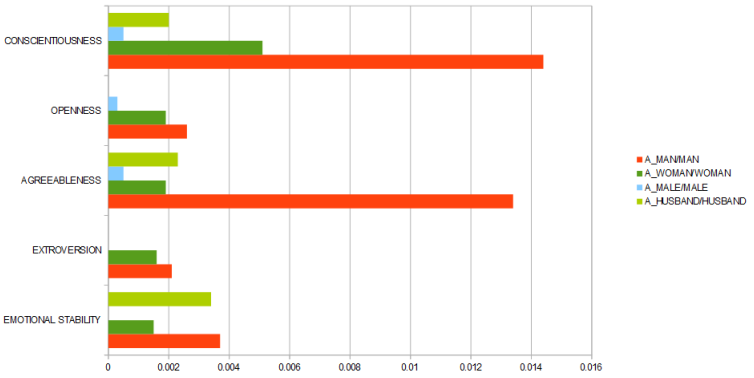

Figure 7. Personality model

.

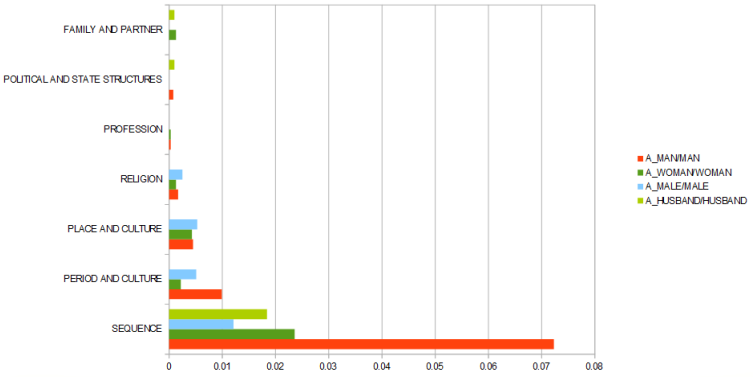

Figure 8. Belonging

.

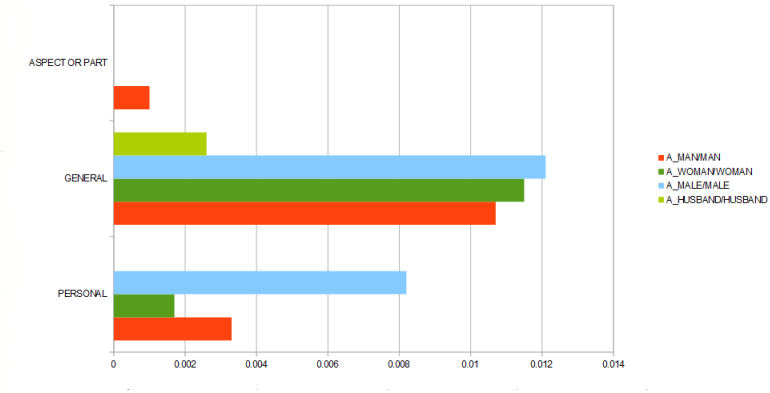

Figure 9. Relations

.

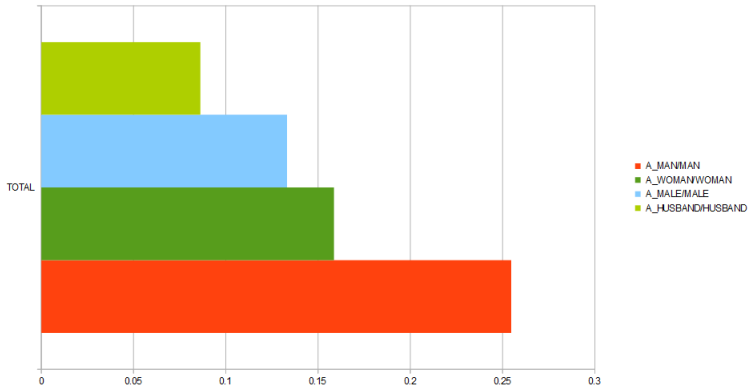

Figure 10. Overview of the overall distribution

.