

**MORPHOLEXICAL PROPERTIES OF ENGLISH AND
SLOVENIAN MASCULINE/FEMININE NOUNS
FOR PROFESSIONS**

Gašper Ilc

Andrej Stopar

Faculty of Arts

University of Ljubljana

Ljubljana, Slovenia

Key words: associative meaning, collocators, corpus frequency, feminatives, masculinatives, word-formation, semantic preference.

Summary: The article presents an analysis of masculine and feminine nouns denoting professions in English and Slovenian. The researched expressions are first discussed from the point of view of word-formation (derivation and compounding), then corpora are employed to examine the frequency of the forms in both languages. The corpus data are complemented with an analysis of collocators with the purpose of identifying the semantic preferences and associative meanings of gender-marked expressions for professions. The results reveal that some feminine nouns for professions are stylistically unmarked (especially in Slovenian), whereas others co-occur (in English and Slovenian) with words that reflect the societal attitudes to gender roles, appearance and character.

**MORFO-LEKSIKALNE ZNAČILNOSTI ANGLEŠKIH IN
SLOVENSKIH MOŠKIH/ŽENSKIH SAMOSTALNIKOV
ZA POKLICE**

Gašper Ilc

Andrej Stopar

Filozofska fakulteta

Univerze v Ljubljani

Ljubljana, Slovenija

Ključne besede: asociativni pomen, feminativi, kolokatorji, maskulinativi, oblikotvorje, pogostnost v korpusu, semantična preferenca.

Povzetek: Članek predstavi kontrastivno analizo moških in ženskih samostalnikov za poklice na jezikovnem paru angleščina-slovenščina. Obravnavani izrazi so najprej predstavljeni z besedotvornega vidika, prek izpeljave oziroma zlaganja. Prispevek nato s pomočjo korpusov preverja pogostnost opazovanih oblik v obeh jezikih. Zbrane korpusne podatke razčleni tudi s pregledom njihovih kolokatorjev, s čimer opredeli semantične preference in pomenske asociacije obravnavanih feminativov in maskulinativov. Rezultati pokažejo, da so feminativi za poklice lahko nezaznamovani (sploh v slovenščini) ali pa se pojavljajo z izrazi, ki (v angleščini in v slovenščini) odsevajo družbeni odnos do spolnih vlog, videza in osebnostnih lastnosti.

1 Introduction

It is well-known that several distinct levels of (un)markedness can be observed with nouns specifically denoting male or/and female referents. Starting with morphology, morphologically marked nouns for gender involve (un)markedness in terms of the (non-)presence of special gender-related morphology, be it inflectional or derivational. In English, for example, the noun *actor* is morphologically unmarked as it lacks any gender-related morphological markings, whereas the gender-related morpheme *-ess* in *actress* makes the female-referring noun marked. Exactly the same can be observed for their Slovenian counterparts *igralec* (masculine noun, male or female referents, morphologically unmarked) and *igralka* (feminine noun, female referents, morphologically marked)¹. At the level of semantics, it has traditionally been assumed that in a pair of masculine/feminine nouns (e.g., *actor/actress*), the masculine noun is unmarked as it can denote both female and male referents, whereas the feminine noun is marked, allowing only the denotation of female referents (cf. Huddleston in Pullum, 2002: 1680ff; Toporišič, 2000: 266; Vidovič-Muha, 2018). The current usage, however, shows that the unmarkedness of masculine nouns for both genders is disputable (Kranjc in Ožbot, 2013: 234). The reason for this development can lie in the fact that it is often the

¹ This is a slight oversimplification. Both nouns are derived from the verbal base *igral-*, and then the derivational morpheme *-ec* is added for the masculine noun, and *-ka* for the feminine noun (cf. Vidovič-Muha, 2018: 410). The morpheme *-ec* carries no overt masculine marker, but this is the case of masculine nouns of the 1st declension, whereas the morpheme *-ka* carries the overt feminine marker *a*, typical of feminine nouns of the 1st declension (cf. Vidovič-Muha, 2018: 415).

case that the feminine noun has acquired some negative connotations through usage, so the users prefer the non-connotative masculine noun instead (e.g., *poet/poetess*, *master/mistress*). This observation leads us to the third type of markedness, stylistic, which is the focus of the present paper. In particular, the paper addresses the question of the stylistic (un)markedness of English and Slovenian gender-marked expressions for professions (henceforth: masculine/feminine nouns), analysing the data from different referential corpora: BNC, English web corpus 2013 (enTenTen13), and the Brown Family of English, and FidaPLUS v2 for Slovenian.

The corpus-based approach adopted herein examines gender-marked expressions for professions in their typical (co-)texts. The data on the co-occurrence of lexical units in corpora are used to analyse various aspects of their use. In line with Sinclair (1996), our choice of units of meaning depends on their linguistic co-text – his term ‘extended units of meaning’ thus takes into account the attractions between the word and its typical environments, which are based on collocation, colligation, semantic preference and semantic prosody.

Collocations are frequent co-occurrences of lexical units; Stubbs (2001) exemplifies such combinations with the phrase *rancid butter*. Gabrovšek (2005: 210-12) points out that collocations can be either semantically or lexically motivated – a semantically acceptable combination may still be unacceptable with regard to the frequency of its use (compare **glacial age* and *ice age*). More recent definitions of collocations consider several properties of collocations to define them as “lexically and pragmatically constrained recurrent cooccurrences of at least two lexical items which are in a direct syntactic relation with each other” (Bartsch, 2004: 76). A colligation, on the other hand, refers to the frequent co-occurrences of words based on the syntactic patterns that include, for instance, negation, mood, person, number, tense, and quantifiers. For example, the verb *deny* requires a gerund, whereas the verb *decline* requires an infinitive: *He denied visiting him.* vs. *He declined to visit him.*

Corpora enable us to determine the discursive (or lexical/semantic) fields typical for a unit of meaning (Stubbs, 2007: 178; McIntyre, 2018: 87). The noun *saucer*, for example, is frequently found next to *teapot*, *teacup*, *plate*, and *drop*, but also next to *flying* (BNC via Sketch Engine, 2019). *Saucer* thus often appears in texts that describe tea drinking, setting the table, and similar; however, the same word is also frequent in texts related to unknown flying objects. In some cases, these word combinations can be classified as proper collocations (*matching saucers*), while in others they do not correspond to our initial definition of collocations. Such an example can

be seen in the clause *to stack the dirty cups and saucers on the tray*: here the nouns *saucers* and *tray* do not represent a collocation, but rather exhibit a tendency to appear in the same text. Stubbs (2001) presents a similar case by discussing the co-occurrence of *large* and various words denoting quantity (*number, amount, scale*). Gabrovšek (2005, 211) explains that such combinations have been classified by Crowther, Dignen and Lea (2002) as examples of categorial collocations – frequent co-occurrences within the discursive field. The latter are also closely related to the term semantic preference (Stubbs, 2001 and 2007), which is the tendency of the word to co-occur with words that belong to the same semantic set. Philip (2010) claims that semantic preference is shared by all speakers of a language.

Closely related to semantic preference is Hoey's (2005) associative meaning, which focuses on the wider context and on the process of learning a word that includes acquiring its referential meaning, syntactic environments, connotations, and associations. Speakers expect the lexical unit to behave in a certain way and associate it with a set of meanings that (may) entail its collocates (Hoey, 2005: 26). The difference between Hoey's associative meaning and Sinclair's semantic preference is in that the former also entails the secondary meanings of associations, evaluations and connotations. Philip (2010) describes this as a psychological aspect of understanding the lexical unit which is not necessarily shared by all speakers. For instance, the frequent use of adjectives describing appearance (*beautiful*) when discussing women, and the use of adjectives expressing importance (*leading*) when discussing men may reflect value judgments with potential societal ramifications (Caldas-Coulthard and Moon, 2010).

When the semantic aspects of lexical units are assigned a pragmatic function, semantic prosody comes into play (Philip, 2010). McIntyre (2018: 87-88) explains that, according to Sinclair, "semantic prosody describes the discourse function of the unit of meaning" and emphasizes that semantic prosodies "belong to units of meaning and not to individual words". He discusses the nouns *job* and *career* and shows that they both convey a similar semantic preference, yet corpus hits show that the typical collocates of *career* more often express a positive semantic prosody. In this respect semantic prosody refers to the positive and negative associations acquired by neutral words when they co-occur with other words (Šorli, 2012 and 2013; Gabrovšek, 2014; Jurko, 2015).

The sections that follow use the extended units of meanings – especially semantic preference and associative meaning – to analyse the corpus data for masculine/feminine nouns for professions.

2 Gender-marked nouns for professions: basic theoretical tenets

With a few exceptions (see below), English allows only so-called notional or semantic gender, meaning that the noun acquires its gender through reference with a real person (cf. Quirk et al., 1999: 314; Huddleston and Pullum, 2002: 484). For instance, the noun *teacher* is morphologically unmarked for gender, and displays semantic gender agreement with its referent (1a-b). When there is no referent (i.e., generic reference), the noun can refer to either male or female referents or both (1c).

(1) a) *She was a very popular **teacher** because of **her** personal interest in them.*

b) *I'm not sure what **my dad** does. I think **he** is a **teacher** in the Air Force.*

c) *This is a first-level administrative post within an education department, usually for an experienced **teacher**. **He** or **she** works under an assistant education officer².*

In some cases, English has special forms to distinguish masculine/feminine nouns. These can be morphologically unmarked pairs (e.g., *boy/girl*, *fox/vixen*), or morphologically marked pairs³. The feminine nouns are derived either by the gender-marking morphemes *-ess*, *-ette*, *-ine*, *-trix*, or by compounding with the component elements *she*, *woman*, whereas the masculine nouns by the gender-marking morpheme *-er*, or the gender-marking component elements *he* and *man* (cf. Quirk et al., 1999: 1546f.; Huddleston in Pullum, 2002: 1680ff.; Bauer, 1996: 220f.).

For everyday usage, it is advisable (cf. Huddleston in Pullum, 2002: 1680) to avoid using morphologically marked feminine nouns unless their use is semantically or pragmatically motivated. For instance, the difference between *a count* and *a countess* or *an abbot* and *an abbess* is not just gender-related but also semantic, as they refer to different societal and legal positions. In a similar fashion, the use of the feminine nouns *actress* and *waitress* is justified, as they also denote different theatre roles and job specifics respectively. Another reason for using a morphologically marked feminine noun is contextual: *he married a rich heiress* is not the same as *he married a rich heir* (Huddleston in Pullum, *ibid.*).

The main reason for avoiding some feminine nouns may lie in the fact that diachronically these nouns acquired negative connotative meanings and

² All the examples listed in the paper are taken from the following corpora: BNC and English Web 2013 for English and FidaPLUS v2 for Slovenian.

³ There is also a third group of nouns, whereby the gender is fully grammaticalized and involves the personification of inanimate nouns (e.g., states, vehicles, emotions, etc.).

are consequently felt as pejorative in present-day usage (cf. Holmes, 2001: 126). Thus, derivatives such as *poetess*, *authoress* and *editorette*, are of rare occurrences (see Table 1).

Table 1. *poetess*, *authoress* and *editorette* vs. *poet*, *author* and *editor*

query	BNC		enTenTen13			BNC		enTenTen13	
	No.	per M ⁴	No.	per M		No.	per M	No.	per M
poetess	25	0.22	1 762	0.08	poet	2 917	25.96	334 032	14.7
authoress	23	0.20	1 089	0.05	author	6 809	60.61	2 726 767	120.0
editorette	0	0	1	0	editor	5 440	48.42	1 316 870	57.9

In contrast to English, in Slovenian, gender is fully grammaticalized, meaning that the gender of a noun can either be the same of different from the gender of the referent. As a rule, the noun preserves its gender regardless of its referent (2a). There are, however, some exceptions. Some nouns such as *starešina* ('a chief') and *vodja* ('a leader')⁵ can be either feminine or masculine depending on the referent (2b,c) due to their morphological make-up (cf.: Uhlik, 2019).

(2) a) *Ona je dober/*dobra pedagog.*

she_{FEM} is good_{MASC}/*good_{FEM} educator_{MASC}
 'She is a good educator.'

b) *Kot skrben vodja je priljubljen pri delavcih.*

as attentive_{MASC} leader_{MASC} is popular_{MASC} at workers
 'Being attentive, he is very popular among the workers.'

c) *Vodja je prisluhnila našim željam.*

leader_{FEM} is listened_{FEM} our wishes.
 'The leader listened to our wishes.'

When analysing feminine derivatives, Vidovič-Muha (2018: 408) identifies the suffixes *-a*, *-(ar)ka*, *-esa*, *-ica*, *-ina*, *-inja*, *-isa*, *-ja*, *-ka*, *-(ov)ka*, *-na* and *-ulja*, with the following productive suffixes: *-(ar)ka*, *-ica*, and *-inja*. With regard to the suffixes used to derive nouns denoting male professions, the author (Vidovič-Muha, 2018: 410) discusses the suffix *-ec*. In her paper on designators for female holders of professions in Slovenian and Russian, Derganc (2017: 23f) argues that those in Slovenian are

⁴ Per million words.

⁵ *Vodja* can denote both male and female referents; however, its derivatives *računovodja* ('an accountant') and *poslovodja* ('a manager') have feminine derivatives ending in *-inja*: *računovodkinja* and *poslovodkinja*, respectively.

stylistically unmarked. It is noteworthy that the corpus data show that in most cases the masculine nouns are much more numerous than their feminine counterparts (cf. Štumberger 2015). As already pointed out by Derganc (2017: 126), this is an expected observation, since, first, in the past the masculine nouns were used generically, and, second, the most prestigious positions and jobs were occupied by men. In addition to these two factors, special attention also has to be paid to the social context (cf. Vidovič-Muha, 2018: 413). For example, in Table 2, high values for the masculine noun *zdravnik* ('a doctor') can be attributed to the generic use, whereas the high values for the feminine nouns *sobarica* ('a chambermaid'), *vzgojiteljica* ('a kindergarten teacher'), and the masculine noun *varilec* ('a welder') result from the fact that these are the professions typically practised by females or males.

Table 2. *Frequency of selected masculine/feminine nouns for professions in FidaPLUS v2*

	No.		No.
zdravnik	34 775	zdravnica	4 811
učitelj	17 842	učiteljica	8 782
sobar	111	sobarica	414
vzgojitelj	4 965	vzgojiteljica	6 535
varilec	418	varilka	9

3 Methodology

For the purposes of the analysis, we first extracted all English feminine nouns for professions from the British National Corpus, BNC, (100 million words, from 1960-1993, written and spoken discourse) in line with the morphological specifications presented in section 2. The final list after the extraction contains 39 feminine nouns for professions derived by the suffixes *-ess*, *-trix* and the component element *woman*. To determine the frequency of feminine nouns in English, the data from the BNC were compared with the data obtained from the English web corpus 2013, enTenTen13, (19 billion words, written language from the .uk web sites from 2013). The data from the Brown Family corpora (American and British English, divided into four subcorpora at 30-year intervals – 1931, 1961, 1991 and 2006) were used to examine the frequency and the usage tendencies from a diachronic perspective.

The basis for the analysis of Slovenian feminine nouns was the database by the Employment Service of Slovenia, available online at: <https://www.ess.gov.si/ncips/opisi-poklicev>. The database contains

masculine nouns only, so the next step was to check the existence of corresponding feminine nouns in the *Dictionary of Standard Slovenian* (SSKJ² available at: fran.si), and in the FidaPLUS v2 corpus. The final list after the extraction contains 289 feminine nouns for professions derived by the suffixes *-esa*, *-ica*, *-inja* and *-(ar)ka*.

Stylistic (un)markedness was ascertained with the help of the collocation candidate tool provided by the Sketch Engine interface. For each investigated noun, collocation candidates were sought in the span from -5 to +5 words, and then the logDice measurement was examined. According to Rychlý (2008), the logDice highest value is 14, with the value of 7 and above indicative of collocational association (cf. Gablasova et al., 2017: 168f).

The analysis of extended units of meaning is based on Word Sketch, a feature of Sketch Engine that provides a summary of a word's grammatical and collocational behaviour (Kilgarriff et al., 2014). The data collected in this way enabled the study of stylistic (un)markedness based on the concepts of semantic preference and associative meaning (Sinclair, 1996; Hoey, 2005; Philip, 2010).

4 Results

4.1 Corpus frequencies

In English, the highest proportion of feminine nouns for professions is derived by the suffix *-ess* (21), followed by the compounds with *woman* (15). Only one noun ending in *-ette* was identified (*usherette*), and two nouns ending in *-atrix* (*aviatrix* and *prosecutrix*). In Table 3, we present those feminine nouns that were found at least 200 times in the enTenTen13 corpus. It should be noted that due to the different sizes of the enTenTen13 and BNC corpora, only a value per million words is comparable. The data from the Brown Family subcorpora is comparable in terms of raw frequency, since all subcorpora contain the same number of words.

Table 3. *Frequency of feminine nouns in English*

feminine noun	enTenTen13		BNC		Brown Family			
	No.	per M	No.	per M	1931	1961	1991	2006
actress	310 864	13.70	1 259	11.21	18	25	61	38
mistress	132 414	5.80	1 179	3.45	25	32	22	39
waitress	73 030	3.20	361	3.21	3	8	14	13
hostess	59 524	2.60	422	3.76	11	31	20	10

stewardess	13 665	0.47	97	0.86	4	3	9	7
headmistress	6 243	0.27	255	2.26	1	6	0	0
huntress	6 218	0.27	7	0.06	0	1	0	0
saleswoman	3 842	0.17	22	0.20	0	0	0	3
policewoman	3 834	0.17	228	2.03	0	3	0	2
laundress	1 818	0.08	13	0.12	0	0	1	0
washerwoman	958	0.04	22	0.20	1	0	0	0
anchorwoman	897	0.04	2	0.02	0	0	0	0
manageress	803	0.04	104	0.93	0	1	1	0
schoolmistress	789	0.03	26	0.24	2	0	0	0
sculptress	726	0.03	11	0.10	0	0	0	0
usherette	629	0.03	28	0.25	0	0	0	0
aviatrix	535	0.02	1	0.01	0	0	0	0
directress	529	0.02	1	0.01	0	0	0	0
fisherwoman	509	0.02	2	0.02	0	0	0	0
instructress	347	0.02	6	0.05	1	0	0	0
ambassadress	297	0.01	1	0.01	0	0	0	0
conductress	202	0.01	12	0.11	0	0	0	0

The corpus data in Table 3 shows that the most commonly used feminine nouns for professions in English are the nouns *actress* and *waitress*, but the data from the Brown Family reveal that there was a decrease between 1991 and 2006 in the use of the noun *actress*. With all other feminine nouns ending in *-ess*, there is a decrease in usage. Special attention must be paid to the nouns *mistress* and *huntress*. The data in Table 3 suggest more frequent values in the enTenTen13 corpus than in others; however, we believe that the increase should not be attributed to the more frequent use of the two nouns when denoting female professionals. The noun *mistress* is polysemous as it may also denote a female lover. A better insight into the actual use of the noun *mistress* when denoting female professionals can thus be observed in the case of the compounds *headmistress* and *schoolmistress*: in both cases there is a sharp decrease in usage. The noun *huntress*, on the other hand, can also denote characters in computer games and films, so for this reason the occurrence of *huntress* is relatively high in the enTenTen13 corpus. There is a slight increase in the

use of some compounds with *woman* (e.g., *anchorwoman*, *saleswoman*); however, in some other cases, the usage is on the decrease (e.g., *policewoman*).

In Slovenian, the majority of feminine nouns denoting professions are derived by the suffix *-(ar)ka* (190), to be followed by the suffixes *-ica* and *-inja* (each around 50). Only two nouns with the suffix *-esa* were identified: *klovnesa* ('a clown') and *stewardesa* ('a stewardess'). In some cases, it may seem that the feminine nouns for the same profession may be derived by two different suffixes, for example, *mesarica* and *mesarka*. However, *mesarica* ('the butcher's wife'), according to the SSKJ², does not denote a profession but a social position of a woman (cf. Vidovič-Muha, 2018: 408), whereas the noun *mesarka* ('a butcher') pertains to the profession, and has as such also been standardised in the National Qualifications Framework (SOK). The SSKJ² defines the noun *kovačica* in a similar fashion as *mesarica* ('the blacksmith's wife'), but the SOK uses the same word for a female profession (i.e., 'a female blacksmith'). The masculine noun *geograf* ('a geographer') has two feminine counterparts *geografka* and *geografinja*, with only the latter being standardised in the SOK. In addition to female nouns, our analysis has identified the morpheme *-ar*, which is used to derive masculine nouns from feminine nouns (e.g., *babičar* 'a male midwife').

Table 4 shows those feminine nouns that have at least 200 occurrences in the FidaPLUS v2. The data are provided separately for the 1979-1990 and 200-2006 subcorpora. Due to different sizes of the subcorpora only the value per million words is comparable⁶.

Table 4. *Frequency of feminine nouns in Slovenian*

feminine noun	subcorpus 1979-1990		subcorpus 2000-2006		total	feminine noun	subcorpus 1979-1990		subcorpus 2000-2006		total
	No.	Per M	No.	per M			No.	per M	No.	per M	
igralka	9,091	45.00	37,246	72.00	46,337	sobarica	248	1.23	774	1.49	1,022
direktorica	6,724	33.00	25,340	49.00	32,064	gasilka	279	1.00	724	1.00	1,003
babica	4,916	24.00	15,374	30.00	20,290	izvajalka	203	1.00	796	2.00	999
učiteljica	4,628	23.00	11,924	22.90	16,552	fotografinja	152	1.00	817	2.00	969
delavka	3,931	20.00	10,066	19.00	13,997	ekonomistka	264	1.00	695	1.00	959
novinarka	3,242	16.00	9,458	18.00	12,700	varuhinja	203	1.01	740	1.42	943
sodnica	2,488	12.40	9,385	18.02	11,873	kozmetičarka	149	1.00	775	1.00	924

⁶ Table 4 also lists nouns *babica*, *detektivka*, *kuharica*, *lovka* and *spremljevalka*, even though these are polysemous items: *babica* ('a midwife' or 'a grandmother'), *detektivka* ('a female detective' or 'a detective book/film'), *kuharica* ('a female cook', or 'a cookery book'), *lovka* ('a female hunter' or 'a tentacle'), and *spremljevalka* ('a companion' or 'a runner-up in a beauty competition').

sekretarka	3,644	18.10	8,183	15.71	11,827	stilistka	25	0.12	874	1.68	899
tajnica	3,448	17.20	7,424	14.30	10,872	menedžerka	132	1.00	742	1.00	874
ravnateljica	2,647	13.17	6,692	12.85	9,339	negovalka	170	1.00	681	1.00	851
svetovalka	2,200	10.94	5,892	11.30	8,092	predavateljica	167	0.83	674	1.29	841
poslanka	3,077	45.31	4,726	9.07	7,803	zastopnica	247	1.23	572	1.10	819
zdravnica	1,818	9.04	5,922	11.37	7,740	opazovalka	258	1.28	554	1.06	812
manekenka	1,392	7.00	5,986	11.00	7,378	vodnica	178	0.89	630	1.21	808
plesalka	1,412	7.02	5,929	11.40	7,341	referentka	217	1.08	589	1.13	806
prodajalka	1,735	8.63	5,562	10.68	7,297	lektorica	211	1.00	583	1.00	794
vzgojiteljica	2,013	10.01	4,488	8.62	6,501	napovedovalka	219	1.00	569	1.00	788
voznica	1,215	6.04	4,789	9.19	6,004	uradnica	230	1.14	530	1.02	760
pomočnica	1,204	5.99	3,956	7.59	5,160	računovodkinja	207	1.03	517	0.99	724
natakarica	1,417	7.00	3,383	6.00	4,800	etnologinja	243	1.00	456	1.00	699
spremljevalka	1,038	5.16	3,663	7.03	4,701	vojakinja	121	0.60	578	1.11	699
režiserka	804	4.00	3,613	6.94	4,417	dirigentka	117	1.00	536	1.00	653
športnica	936	4.66	3,211	6.16	4,147	terapevtka	117	0.58	530	1.02	647
kuharica	999	5.00	2,866	6.00	3,865	sociologinja	188	0.94	440	0.84	628
tehničarica	240	1.19	3,430	6.58	3,670	fizioterapevtka	148	1.00	476	1.00	624
odvetnica	677	3.37	2,334	4.48	3,011	bolničarka	186	1.00	428	1.00	614
oblikovalka	526	2.62	2,454	4.71	2,980	inženirka	177	1.00	435	1.00	612
šivilja	714	3.55	1,928	3.70	2,642	pediatrinja	109	0.54	484	0.93	593
psihologinja	590	2.93	1,872	3.59	2,462	upravnica	152	0.76	391	0.75	543
izvedenka	529	3.00	1,875	4.00	2,404	scenaristka	87	0.43	438	0.84	525
prometnica	661	3.29	1,595	3.06	2,256	maserka	103	1.00	404	1.00	507
strokovnjakinja	440	2.19	1,790	3.44	2,230	biologinja	114	1.00	384	1.00	498
prevajalka	502	2.50	1,680	3.23	2,182	posrednica	133	0.66	361	0.69	494
inšpektorica	522	3.00	1,616	3.00	2,138	receptorka	155	0.77	334	0.64	489
pedagoginja	546	2.72	1,461	2.80	2,007	oskrbnica	113	0.56	363	0.70	476
asistentka	326	2.00	1,612	3.00	1,938	poslovodkinja	134	0.67	319	0.61	453
čistilka	472	2.00	1,461	3.00	1,933	vizažistka	8	0.04	434	0.83	442
mojstrica	359	2.00	1,396	3.00	1,755	strežnica	154	0.77	276	0.53	430
glasbenica	326	2.00	1,393	3.00	1,719	zobozdravnica	140	0.70	286	0.55	426
trenerka	361	1.80	1,338	2.57	1,699	kostumografinja	92	0.01	309	1.00	401
stewardesa	287	1.43	1,384	2.66	1,671	cvetličarka	83	0.01	304	1.00	387
veleposlanica	640	3.18	1,015	1.95	1,655	arheologinja	96	0.01	287	1.00	383
lovka	424	2.00	1,211	2.00	1,635	točajka	133	0.66	231	0.44	364

kustosinja	345	2.00	1,272	2.00	1,617	izdelovalka	91	0.01	266	1.00	357
frizerka	287	1.00	1,327	3.00	1,614	antropologinja	61	0.01	293	1.00	354
policistka	336	1.67	1,253	2.41	1,589	klovnesa	43	0.01	310	1.00	353
organizatorka	306	1.52	1,212	2.33	1,518	veterinarka	78	0.39	269	0.52	347
arhitektka	290	1.00	1,225	2.00	1,515	administratorica	112	1.00	233	0.10	345
pravica	392	1.95	1,037	1.99	1,429	notarka	62	0.01	270	1.00	332
koreografinja	274	1.00	1,125	2.00	1,399	animatorka	61	0.01	259	0.01	320
zgodovinarica	347	1.73	1,028	1.97	1,375	revizorka	53	0.26	242	0.46	295
solistka	330	1.64	1,027	1.97	1,357	telefonistka	103	0.51	181	0.35	284
blagajničarka	305	2.00	1,042	2.00	1,347	farmaceutka	42	0.01	211	0.01	253
detektivka	204	1.00	1,049	2.00	1,253	mlinarica	108	1.00	125	0.01	233
knjižničarka	321	2.00	923	2.00	1,244	oskrbovalka	6	0.03	223	0.43	229
dramaturginja	283	1.00	886	2.00	1,169	založnica	78	0.39	147	0.28	225
upraviteljica	275	1.37	816	1.57	1,091	matičarka	31	0.01	185	0.01	216
kiparka	328	2.00	758	1.00	1,086	hišnica	72	0.01	142	0.01	214

Comparing the data from both subcorpora, we can observe that the number of feminine nouns for professionals is either on the increase (e.g., *igralka* ('an actress'), *novinarica* ('a journalist'), *režiserka* ('a film director')) or it maintains its value over time (e.g., *delavka* ('a worker'), *zgodovinarica* ('a historian'), *solistka* ('a soloist')). There are only few examples where a decrease has been observed (e.g., *sekretarka* ('a secretary'), *telefonistka* ('a telephonist')). The reason for such declines could be that these professions are slowly disappearing.

4.2 Collocators and semantic preferences

Corpus data reveal the attractions between the word and its typical environments. The masculine/feminine pairs of nouns below present the most frequent combinations, with special reference to semantic fields and semantic preferences.

The English data are as follows.

I Actress/actor (1,259/3,600 hits in BNC): Using the BNC corpus and the Word Sketch function in Sketch Engine, it can be observed that some of the most frequent premodifiers of the feminine form are *best*, *actor*, *bit-part*, *out-of-work*, *sought-after* and *singing*. The word also co-occurs with nouns such as *actor*, *singer*, and *dancer*, and is often accompanied by names of well-known individuals (*Hepburn*). Among the verbs appearing in

its proximity are also *star, marry, support, win, lead, watch* and *become*. The masculine form *actor*, on the other hand, is frequently premodified by words such as *character, best, Hollywood, talented, comedy* and *unemployed*; furthermore, it can be followed by, for instance, *actress, singer, musician, and award*; and it also co-occurs with the names of individuals (*Branagh*). The verbs that co-occur with the masculine form include *vote, lead, dress, become, name, concern, play* and *portray*. A comparison of the most frequent combinations in EnTenTen13 (via Sketch Engine, with a span of +/-5 words) shows that the feminine form is noticeably more frequently combined with the modifiers *porn* and *supporting*.

Considering the semantic preferences, both forms combine with expressions describing related professions; place names related to the profession; names of individuals; expressions describing the type/quality of work; activities related to the profession; and expressions related to awards. Some specific connotations of the feminine form are brought to the fore by the collocators such as *aspiring, supporting, porn, affair* and *steamy*; some of these may also appear with the masculine form, but less frequently.

II Policewoman/policeman (228/3,193 hits in BNC): The nouns *policewoman/policeman* have been largely replaced by the gender-neutral *police officer*; nevertheless, the data from BNC reveal some interesting differences between them. The feminine form appears after the premodifiers like *zoggian, plain-clothed, unlucky, probationer, off-duty* and *daredevil*. The word is often followed by nouns such as *baillet, teleanou, field-worker* and *firewoman*; it also appears in the proximity of words like *radio, rape, gun, fancy, staff* and *retreat*. The most frequent collocators (based on the logDice value) of the feminine form are also *unluckiest, smartly* and *brave*, which do not appear next to the masculine form. The masculine form *policeman* often co-occurs with *off-duty, neighbourhood, plain-clothes, sleeping, secret* and *ordinary*; it is used with verbs such as *injure, assault, kill, patrol, arrest* and *kick*; and it co-occurs with nouns that frequently include *civilian, soldier, woman, fireman, detective* and *policewoman*.

Regarding semantic preferences, both forms appear with expressions that refer to police activities or describe individuals connected to police work. The feminine noun is more frequently found next to personal names, in sci-fi contexts, and with adjectives that point to (positive) personal traits (*smartly, brave*). The observed word *rape* refers to an attempted rape of a female police officer.

III Waitress/waiter (361/932 hits in BNC): The feminine form *waitress* is often premodified by *cocktail*, *17-year-old*, *assistant*, *Russian*, *head*, *young* and *good*. It co-occurs with nouns such as *job*, *service*, *waiter*, *chambermaid* and *cleaner*, and verbs like *call*, *ask*, *bring*, *be*, *serve* and *arrive*. Among the frequent collocators are also the words *Hooters*, *topless* and *flirt* (logDice: 6.6, 5.8 and 5.3 respectively). The masculine noun appears with the premodifiers *tide*, *obsequious*, *white-jacketed*, *head*, *dumb* and *wine*. *Waiter* also co-occurs with *waitress*, *attendant*, *chef*, *barman*, *maid* and *menu*. It can also be found next to *beckon*, *hover*, *summon*, *bow*, *serve*, *bring*, *pour* and similar verbs.

With respect to semantic preferences, the pair *waitress/waiter* attracts collocators referring to related professions, typical activities, workplace objects, and expressions related to the activities of customers. Both also frequently co-occur with expressions related to appearance and character in which waitresses are more frequently positively perceived (*attentive*), and waiters negatively (*dumb*). The feminine noun also co-occurs with expressions that sexualize the profession (*topless*).

Moving to the Slovenian data, the feminine nouns can be divided into two main groups: Group A consists of those feminine nouns that do not differ from the masculine ones and are stylistically unmarked, whereas Group B includes those whose meaning is specialized or at least partly stylistically marked. The authors see these groups as extreme ends of a continuum, which is supported by examples that exhibit the features of both groups.

Representing Group A are the masculine/feminine pairs: *zdravnica/zdravnik* ('a female/male doctor'), *sodnica/sodnik* ('a female/male judge'), and *vzgojiteljica/vzgojitelj* ('a female/male kindergarden teacher'). Group B includes the pairs *administratorka/administrator* ('a female/male administrator'), *tajnica/tajnik* ('a female/male secretary'), *kuharica/kuhar* ('a female/male cook'), *maserka/maser* ('a female/male masseuse/masseur'), and *čistilka/čistilec* ('a female/male cleaner'). The selection of examples aims to present both prototypical and borderline examples in each group.

Group A (semantically equivalent, stylistically unmarked forms):

A-I Zdravnica/zdravnik ('female/male doctor'; 7,857/127,248 hits in FidaPLUS v2):⁷ The most frequent collocators of the feminine noun (logDice above 7, span +/-5 words) are *dežurna* ('on-call'), *specialistka*

⁷ The number of hits stated here differs slightly from the one in Table 4; this is due to the sub-corpora not containing all the data from FidaPLUS v2.

(‘specialist’), *pregledala* (‘examined’), *medicine* and *Quinn* – the results are slightly skewed here because of the series *Dr. Quinn, Medicine Woman* appearing in the numerous TV listings included in the corpus. The function Word Sketch, on the other hand, lists collocators that include the adjectives *dežurna* (‘on-call’), *oddelčna* (‘departmental’), *cenjena* (‘highly valued’), *zavodska* (‘institutional’); nouns such as *znanstvenica* (‘a scientist’), *pravnica* (‘a jurist’), *upraviteljica* (‘a manager’), *pedagoginja* (‘a pedagogue’); and verbs like *oskrbeti* (‘treat’), *predpisati* (‘prescribe’), *prihiteti* (‘hurry’), *odrediti* (‘order’), and *pregledati* (‘examine’). The masculine form co-occurs with the adjectives *dežurni* (‘on-call’), *splošni* (‘general’), *izbrani* (‘chosen’), *hribovski* (‘mountain’), *družinski* (‘family’); nouns like *zobozdravnik* (‘a dentist’), *bolnik* (‘a patient’), *sestra* (‘a nurse’), *zdravnica* (‘a doctor’); and verbs that include *predpisati* (‘prescribe’), *svetovati* (‘advise’), *priporočati* (‘recommend’), *pregledati* (‘examine’), and *oskrbeti* (‘treat’). Some of the most frequent co-occurring expressions (logDice above 7, span +/-5 words) are *pojdit* (‘go’), *predpiše* (‘prescribe’), *dežurni* (‘on-call’), *zaupajo* (‘trust’), *splošni* (‘general’), *bolniki* (‘patients’), *sestre* (‘nurses’).

The semantic preference for both forms is as follows: adjectives for the type of work done by doctors; verbs describing decisions related to medical treatment; and related professions.

A-II *Sodnica/sodnik* (‘female/male judge’; 11,946/115,838 hits in FidaPLUS v2): The feminine noun frequently (logDice above 7, span +/-5 words) appears in TV listings, which is evidenced by the frequent collocator *raztresena* (‘absent-minded’), a word used in the Slovenian title of the series *Ally McBeal*. Similarly troublesome for the analysis are those hits referring to the series *Judge Judy* and *Judging Amy*, since the Slovenian translations of their titles include the word *sodnica* (‘female judge’). More insight can be gained by examining the Word Sketch data. The frequent modifiers of the feminine noun include *preiskovalna* (‘investigative’), *okrajna* (‘county’), *razpravljajoča* (‘hearing’), and *dežurna* (‘on-call’). Female judges preside (*predsedovati*), order (*odrediti*), judge (*razsoditi*), grant (*ugoditi*), and interrogate (*zaslišati*). The feminine noun also appears next to nouns such as *odločitev* (‘decision’), *vprašanje* (‘question’), *nepristranskost* (‘impartiality’), *tožnica* (‘claimant’), *odvetnica* (‘lawyer’), and *pravnica* (‘jurist’). Some of the frequent collocators also include words related to sports: *strelska* (‘shooting’), *trenerka* (‘coach’), *turnir* (‘tournament’). The findings are very similar when it comes to the masculine form: *ustavni* (‘constitutional’), *predsedovati* (‘preside’),

izvolitev ('election'), *porota* ('jury'), *nogometni* ('football'), *tekma* ('match'), etc.

Both masculine and feminine forms appear in contexts related to law and sports. The semantic preferences are as follows: the position within the profession; activities related to decision-making; and related professions. A comparison of adjectives premodifying the two forms reveals a significant difference: the adjectives denoting important roles within the profession are more frequently found next to the masculine form: *ustavni* ('[masculine] constitutional') – 11,042 hits / *ustavna* ('[feminine] constitutional') – 442 hits; *vrhovni* ('[masculine] supreme') – 2781 hits / *vrhovna* ('[female] supreme') – 253 hits; *okrajni* ('[male] county') – 491 hits / *okrajna* ('[feminine] county') – 654 hits. The finding is relevant from the point of view of associative meaning and extra-linguistic circumstances, which may also explain the attraction of female judges (and doctors) for TV produces/audiences.

A-III Vzgojiteljica/vzgojitelj ('female/male kindergarten teacher'; 6,535/4,965 hits in FidaPLUS v2): The word *vzgojiteljica* often co-occurs (logDice above 7, span +/-5 words) with the following collocators: *vrtec* ('kindergarten'), *učiteljice* ('teachers'), *pomočnice* ('assistants'), *predšolski* ('pre-school'), *mentorica* ('mentor'), *otroci* ('children'), *opravili* ('done'), *poskrbele* ('made sure'), etc. Word Sketch lists the typical adjectival promodifiers such as *predšolska* ('pre-school'), *prizadevna* ('dedicated'), *diplomirana* ('graduate'), *usposobljena* ('qualified'), *zaposlena* ('employed'), *upokojena* ('retired'); some of the verbal collocators are *poučevati* ('teach'), *uresničiti* ('realize'), *razložiti* ('explain'), *spodbujati* ('encourage'), *skrbeti* ('care'), *zaposliti* ('employ'), *ustvarjati* ('create'), *načrtovati* ('plan'); while the nominal collocators include *učiteljica* ('teacher'), *vzgojitelj* ('kindergarten teacher'), *pomočnica* ('assistant'), *varuhinja* ('guardian'), *ravnateljica* ('principal'), *starš* ('parent') and others.

The masculine noun frequently appears with *učitelji* ('teachers'), *predšolskih* ('pre-school'), *otrok* ('children'), *vzgojiteljice* ('kindergarten teachers'), *starši* ('parents'), *vrtci* ('kindergartens'), *pedagogi* ('pedagogues'), etc. Other collocators include *dijaški dom* ('student dormitory'), *skrbnik* ('guardian'), *gojenci* ('boarders'). The hits thus reveal a strong link between the masculine noun and secondary-school dormitories and similar institutions, which is confirmed in Word Sketch that lists adjectives like *usposobljeni* ('qualified'), *pozorni* ('attentive'), *dežurni* ('on-call'), *strog* ('strict'), *zasebni* ('private'); verbs such as *pretepsti* ('beat'), *vzgajati* ('raise'), *spodbujati* ('encourage'), *poučevati* ('teach'), *skrbeti* ('care'), *spraševati* ('ask'); and nouns like *vzgojiteljica* ('kindergarten

teacher'), *učitelj* ('teacher'), *katehet* ('catechist'), *stars* ('parent'), *gojenec* ('boarder'), *skrbnik* ('guardian'), *šolnik* ('educator'), etc.

The described corpus data do not reveal any substantial differences between the masculine and feminine nouns: both are similarly frequent and appear with similar collocators. The semantic preferences can be described as: relating to work environment; people in related professions; and activities typical for the profession. Both forms also co-occur with similar expressions that describe an individual's character and their professional qualifications. Nevertheless, the data suggests that some elements of gender stereotyping may be present: for instance, female kindergarten teachers are *dedicated*, while male ones are *strict*. A similar finding can be associated with the working environment of student dormitories, where the masculine form is more common.

Group B (narrower semantic meaning, partial stylistic marking):

B-I Administrator/administrator ('female/male administrator'; 345/2,876 hits in FidaPLUS v2): The SSKJ² defines the feminine form as "a lower office worker" and as "a woman who administers webpages". The masculine form is defined as "one who manages, a manager"; the same definition includes the use of the word in the field of information technology ("one who manages a computer system or an online forum") and, as the last option, the meaning of "a lower office worker". The state-run Employment Service of Slovenia only lists office-related positions in connection with this job title.

The corpus data (via Word Sketch) reveal that the feminine noun appears next to adjectives such as *notarjeva* ('notary's'), *nezaposlena* ('unemployed'), *policistova* ('policeman's'), *radodarna* ('generous'), *ljubezniva* ('kind'), *vestna* ('conscientious'), and *postavna* ('shapely'); verbs like *pridirjati* ('come running'), *ožigosati* ('stamp'), and *prepisovati* ('copy'); and nouns that include *pisanje* ('writing'), *strežnica* ('hospital attendant'), *računovodkinja* ('accountant') and *menadžerka* ('manager'). The masculine form is found next to adjectives such as *apostolski* ('apostolic'), *sistemski* ('systemic'), *mrežni* ('network'), *dekanijski* ('deanery's'), *kolonialni* ('colonial') and *sposobni* ('competent'); verbs like *pregledovati* ('check'), *dodeljevati* ('assign'), *upravljati* ('manage') and *vzdrževati* ('maintain'); and nouns such as *forum* ('forum'), *špica* ('peak'), *strežnik* ('computer server'), *povezava* ('connection'), *baza* ('database'), *moderator* ('moderator'), *teoretik* ('theoretician') and *programmer* ('programmer'). The typical collocators (logDice above 7, span +/-5 words) show that the most frequent ones involve expressions related to IT, for instance names, nicknames and other expressions common for online

content: *moderator* ('moderator'), *pridružen* ('joined'), *timi666* (a nickname), *sistemiški* ('system'), *sporočilo* ('message'), etc.

The semantic preference of the feminine form can be identified in its co-occurrence with professions denoting superior positions; (gentle) character traits; appearance; office-related activities and locations. The masculine form collocators confirm its semantic preference for management positions, expressions describing high professional qualifications, and administration in the field of IT.

B-II *Tajnicaltajnik* ('female/male secretary'; 11,016/12,931 in FidaPLUS v2): The SSKJ² defines the feminine form as "one who conducts administrative, organizational and technical business". The definition of the masculine form is identical, but it also includes the additional meaning of "a high official of some organization, body". The corpus-based analysis reveals, however, that the feminine form often combines with *glavna* ('main/head') in the sense of "high official". Despite this, the classification of occupations provided by the Employment Service of Slovenia lists the feminine form only in connection with its narrower meaning, and even states that the position is more commonly filled by women.

The frequent collocators for *tajnica* (Word Sketch) are *telefonska/avtomatska* ('phone/automatic') – referring to an answering machine – *direktorjeva* ('director's'), *ministrove* ('minister's'), *županova* ('mayor's'), *Clintonova* ('Clinton's') and similar. The word combines with verbs such as *prevezati* ('transfer calls'), *se oglašati* ('answer'), *povedati* ('tell'), *vstopiti* ('enter'), *obveščati* ('inform'), *poklicati* ('call'), and *z njimi varati* ('cheat with them'). It stands in proximity to nouns like *izpopolnjevanje* ('training'), *klub* ('club'), *kongres* ('congress'), *stol* ('chair'), *pisarna* ('office'), *fikus* ('fig'), *računovodkinja* ('accountant'), *blagajničarka* ('treasurer'), *receptorka* ('receptionist'), *snažilka* ('cleaner'), *ljubica* ('lover'), *razmerje* ('affair'), etc. Corpus hits include sentences describing a diplomat who obtained an office, a fig plant and a secretary, or bankers who are afraid to lose their armchairs, substantial salaries, secretaries and fig plants.

The most frequent collocators of the masculine form (logDice above 7, span +/-5 words) are prominent functions in organizations or names of individuals occupying these positions: *glavni* ('head/main'), *sindikata* ('union's'), *sinode* ('synod's') and *demokratov* ('Democrats'). According to Word Sketch, the form *tajnik* is often close to adjectives such as *glavni* ('main'), *vatikanski* ('Vatican'), *generalni* ('general'), and *papeški* ('papal'). The typical verbs in the proximity of the masculine form are *službovati* ('serve'), *povedati* ('tell'), *pojasniti* ('explain'), *zatrđiti*

(‘assure’), and *nagovoriti* (‘address’). The masculine form also appears near nouns like *blagajnik* (‘treasurer’), *podtajnik* (‘deputy secretary’), *podžupan* (‘deputy mayor’), *računovodja* (‘accountant’), *knjižničar* (‘librarian’), as well as near names of well-known personalities.

Regarding semantic preference, we can conclude that the feminine form is frequently linked to the business space led by an influential male (director) and surrounded by objects typical for such a space (office plants). In addition, the activities that *tajnica* is involved in are related to the flow of information and involve other professions below the managerial level. The examples clearly reflect the associative meaning of the feminine noun (a secretary is objectified by being equated with furniture and office plants). Some additional connotations can be identified in contexts related to socially less desirable interactions (affairs). On the other hand, the masculine form is mostly connected with the position of a high official. It co-occurs with important institutions, activities that involve addressing larger groups of people, and related professions.

B-III *Kuharica/kuhar* (‘female/male cook’; 3,907/14,892 hits in FidaPLUS v2): The form *kuharica* is polysemous – besides the frequent meaning of “a woman who is (professionally) occupied with cooking”, the meaning of “cookbook” is also common (see the SSKJ²). The masculine form is defined as “one who is (professionally) occupied with cooking”.

The list of collocators for *kuharica* (logDice above 7, span +/-5 words) includes words such as *Kalinšek* (a well-known cook-book author), *zaposlimo* (‘[we] employ’), *Felicita* (a well-known cook-book author), *kuharji* (‘cooks’), *natakarica* (‘waitress’), *izvrstna* (‘excellent’), *Vendelina* (a well-known cook-book author), *čistilka* (‘a cleaning lady’), *odlična* (‘excellent’), *recepti* (‘recipes’), and *sobarica* (‘a chambermaid’). The function Word Sketch lists adjectives like *farovška* (‘parochial’), *ravnateljeva* (‘principal’s’), *izvrstna* (‘excellent’), *žlehtna* (‘wicked’), *izučena* (‘educated’), *župnikova* (‘priest’s’), *vedoželjna* (‘eager to learn’), and *pozabljiva* (‘forgetful’). The feminine noun is often coordinated with the nouns *čistilka* (‘a cleaning lady’), *kuhar* (‘a cook’), *natakarica* (‘a waitress’), *sobarica* (‘a chambermaid’), and *hišnik* (‘a caretaker’). Some examples of the most frequent words that co-occur with the feminine noun are *menza* (‘canteen’), *gostilna* (‘inn’), *kuhinja* (‘kitchen’), *vrtec* (‘kindergarten’), *dom* (‘home’), *recept* (‘recipe’), *praksa* (‘practice’), *izkušnja* (‘an experience’), and *družina* (‘family’). Among the verbal collocators the following can be observed: *peči* (‘bake’), *kuhati* (‘cook’), *zaposliti* (‘employ’), *obleči* (‘wear’), and *upokojiti* (‘retire’).

The masculine noun co-occurs with *vaja* ('an exercise'), *natakar* ('a waiter'), *gospa* ('Ms.'), *hiša* ('house'), *nedoločen* ('permanent'), *najboljši* ('the best'), *dragocenosti* ('valuables'), and *poročila* ('reports'). Word Sketch lists the following premodifying collocators: *pojoč* ('singing'), *ljubiteljski* ('amateur'), *ladijski* ('ship's'), *izvrsten* ('excellent'), *dvoren* ('court'), *izučen* ('trained'), *pozabljiv* ('forgetful'), and *vešč* ('skilful'). *Kuhar* often stands close to the nouns like *natakar* ('a waiter'), *kuharica* ('a cook'), *slaščičar* ('a confectioner'), *jedec* ('an eater'), *gurman* ('a gourmet'), *menza* ('a canteen'), *restavracija* ('a restaurant'), *kuhinja* ('a kitchen'), *hotel* ('a hotel'), *gostilna* ('an inn'), *pomočnik* ('a helper'), *kapa* ('a hat'), *sekcija* ('a section'), *ekipa* ('a team'), *tekmovanje* ('a competition'), and *tečaj* ('a course'). The verbs frequently found next to *kuhar* are *servirati* ('serve'), *izmišljati* ('make up'), *speči* ('bake'), *skuhati* ('cook'), *odrezati* ('cut off'), *izobraževati* ('educate'), *postreči* ('serve'), *razjariti* ('enrage'), *najeti* ('hire'), *pripravljati* ('prepare'), *pokvariti* ('spoil'), *zaposliti* ('employ'), and *pohvaliti* ('praise').

The semantic preferences for both nouns include expressions for related professions; character traits; professional qualifications; types of work; work environments; and activities related to the occupation. The analysis also reveals several socially constructed differences. The feminine noun is present in some specific work environments reserved for female cooks (parish house), and it appears with professions that typically require lower qualifications (cleaner). The masculine form, on the other hand, can refer to a ship's cook; it is more frequently described positively (excellent, trained); and is often associated with competitions (section, team, competition).

B-IV Maserka/maser (masseur/masseuse; 512/1,216 hits in FidaPLUS v2): The SSKJ² uses the same definition for both forms: "a woman/one who professionally massages". Word Sketch lists some frequent premodifiers such as *tajska* ('Thai'), *spretna* ('skilful'), *izurjena* ('trained'), *prikupna* ('cute'), *diplomirana* ('graduate'); verbs *zmasirati* ('massage'), *obdelovati* ('do'); and nouns *maser* ('a masseur'), *aromaterapevtka* ('an aromatherapist'), *fizioterapevtka* ('a physiotherapist'), *psihologinja* ('a psychologist'), *serviser* ('a repairman'), *sestra* ('a nurse'), *roka* ('a hand'), etc. The corpus also includes the names of well-known sports masseuses. Some less frequent collocators (logDice under 7, span +/-5 words) of the feminine noun include *erotična* ('erotic'), *nežna* ('gentle'), *čarobna* ('magic'), *oblečena* ('dressed'), and *ljubezen* ('love'). The data for the masculine form are very different. According to Word Sketch, a masseur is *tajski* ('Thai'), *izšolan* ('educated'), *izurjen* ('skilled'), *izkušen* ('experienced'), *usposobljen* ('trained'), *ukrajinski* ('Ukrainian'),

profesionalen ('professional'), and *priznan* ('renowned'), but not *prikupen* ('cute'). The most frequent verbs next to *maser* include *pregnete* ('knead'), *gladi* ('stroke'), *masira* ('massage'); it co-occurs with the nouns like *pediker* ('a pedicurist'), *fizioterapevt* ('a physiotherapist'), *mehanik* ('a mechanic'), *kozmetičarka* ('a beautician'), *serviser* ('a repairman'), *psiholog* ('a psychologist'), *zdravnik* ('a doctor'), and *trener* ('a coach'). The most frequent hits do not include the adjective *erotični* ('erotic'). In addition, the masculine noun is more prevalent in the fields of sport and medicine.

The semantic preferences shared by both the masculine and feminine nouns can be identified in expressions related to professional qualifications; typical activities; and similar professions. The differences can be detected in the words describing personal traits and appearance (cute), as well as in corpus hits involving erotic massages – such collocators co-occur with the feminine form and testify to the often perceived sexualization of the profession.

B-V Čistilka/čistilec ('female/male cleaner'; 1,950/1,767 hits in FidaPLUS v2): Word Sketch lists the feminine form with adjectives such as *optimistična* ('optimistic'), *pridna* ('hard-working'), *bosanska* ('Bosnian'), *priljubljena* ('popular'), *mlada* ('young'); verbs like *pospravljati* ('tidy up'), *zboleti* ('fall ill'), *preživljati* ('support'), *pobirati* ('pick up'), *čistiti* ('clean'), *zaposeliti* ('employ'); nouns as *čistilec* ('a cleaner'), *kuharica* ('a cook'), *vrtnarka* ('a gardener'), *zaposlitev* (employment), *plača* ('a salary'), and *čiščenje* (cleaning). The most frequent collocators (logDice above 7, span +/-5 words) include words like *Fata* (a female Bosnian name), *sobarica* ('a chambermaid'), and *hišnik* ('a caretaker'). The masculine noun most frequently stands next to words that suggest that it refers to cleaning appliances: *visokotlačni* ('high-pressure'), *kemični* ('chemical'), *parni* ('steam'), *vakuumski* ('vacuum'), etc. Only a few collocators refer to people, e.g. a sentence in the corpus states that a shop is "meant for professional and amateur [male] cleaners, as well as housewives". The co-occurring verbs are *izobraževati* ('educate'), *čistiti* ('clean'), *zaposeliti* ('employ'), while the nouns include *smetar* ('a garbage collector'), *vzdrževalec* ('a maintenance worker'), *čistilnica* ('a laundry shop'), *kuhar* ('a cook'), *natakar* ('a waiter'), *traktor* ('a tractor'), and *prah* (dust).

Considering the semantic preference, we can conclude that it encompasses words referring to character traits; typical activities related to cleaning; and similar professions that require lower qualifications. The associative meaning of the feminine noun can be observed in the hits that involve the collocators *bosanska* (Bosnian) and *Fata* (a stereotypical female

Bosnian name). The masculine form mostly refers to appliances, signalling that the profession of a cleaner is mostly perceived as a female one.

5 Discussion

The first part of the paper is focused on the derivation of feminine and masculine nouns for professions in English and Slovenian. The analysis shows that the dominant form of feminine nouns in both languages involves derivation via affixation, while English also features compounding. Masculinatives are rare in both languages; they are mostly formed in relation to professions which have been traditionally considered part of the female sphere. Besides the masculine suffix *-ec*, which was previously discussed by Vidovič-Muha (2018:410), the analysis also identifies the Slovenian masculine suffix *-ar*. The suffix is not productive but can nevertheless be observed in the word *babičar* ('male midwife'). The most notable difference between the two languages is that contemporary English seems to be abandoning the use of gender-marked forms (except for the rare feminatives that remain), whereas in Slovenian, as evidenced by corpus data, the use of feminatives is on the rise. The reason for this asymmetry is rooted in the fact that English has an almost non-existent grammatical gender system, and in most cases uses the same words to describe a man or woman working in a certain profession. In contrast, Slovenian grammatical gender is fully grammaticalized and may or may not overlap with biological sex. Slovenian nouns denoting professions have clear referents with regard to gender, and their contemporary use – evidenced in corpora – shows a clear tendency towards the agreement between grammatical and biological genders. Thus, in the case of female referents the use of feminine nouns is now preferred, while the use of masculine generic forms is on the decline.

Our corpus analysis shows that the differences between masculine and feminine nouns for professions are best revealed by examining their collocators, semantic preferences, and semantic associations (Sinclair 1996 and Philip 2010).

The collocators of English masculine and feminine nouns for the examined professions (*actress/actor*, *policewoman/policeman* in *waitress/waiter*) reveals that they share some semantic environments: related professions; places and surroundings associated with the profession; personal names; words describing the quality of their work; and the typical actions associated with the profession. Some specific associative meanings are also revealed if we examine the collocators describing human traits and appearance. The feminatives often co-occur with expressions for stereotypically female character traits and personality (including sexual

connotations). This is in line with Caldas-Coulthard and Moon (2010), who explain such asymmetries with stereotypes of gender, sexualization of women, as well as with age and behaviour.

Slovenian masculine/feminine pairs for professions reveal similar findings. Regarding semantic preference, we can observe expressions related to typical actions; typical places and surroundings; related professions; character traits; and appearance. Yet the corpus data also show some significant differences. The pairs of Slovenian masculine and feminine nouns referring to professions can be divided into two groups. The first one comprises pairs that are semantically equivalent and stylistically unmarked (see Derganc, 2017), while the pairs in the second group are subject to (a degree of) semantic narrowing and stylistic marking.

A prototype for the first group is the pair *zdravnica/zdravnik* ('female/male doctor'), both of which can co-occur with expressions referring to the type of work, typical actions and related professions. The feminine nouns in such pairs are stylistically unmarked. A similar conclusion can be reached for the pairs *sodnica/sodnik* ('female/male judge') and *vzgojiteljica/vzgojitelj* ('female/male kindergarten teacher'), but with a caveat. The form of the pair *sodnica/sodnik* overlap in the fields of law and sport; however, the data also reflect the societal status of the profession – the most important positions (e.g., constitutional judges) are still mostly reserved for men. The pair *vzgojiteljica/vzgojitelj* can be described as a borderline case between marked and unmarked meaning. The corpus data suggest a degree of semantic narrowing and stylistic marking: the meaning of "teacher in a student dormitory" refers to men, while the use of adjectives (diligent women and strict males) points to gender stereotyping.

Partially marked forms that also involve semantic narrowing have been presented using the pairs *administatorka/administrator* ('female/male administrator'), *tajnica/tajnik* ('female/male secretary'), *kuharica/kuhar* ('female/male cook'), *maserka/maser* ('masseuse/masseur'), and *čistilka/čistilec* ('female/male cleaner'). The masculine nouns *administrator* and *tajnik* are associated with socially more prominent professions, which is confirmed by the definitions in the SSKJ². *Administatorka* tends to be in a subordinate position (notary), and is often described as shapely (appearance) or kind (character). A similar conclusion can be reached for *tajnica*, which co-occurs with expressions signalling her objectification (figus plant) and sexualization (affair). Socially-constructed differences have been identified for the pair *kuharica/kuhar* as well. The feminine noun appears in specific work environments (parsonage) and with professions that require lower qualifications, while the masculine noun appears in

contexts related to competitions. Still, in this case, the meaning is not narrow enough to justify, for instance, separate dictionary definitions. The same can be concluded for the pair *maserka/maser*, which demonstrates, again, that feminatives are more likely to appear with expressions describing appearance (cute) and sexual activities (erotic). The final pair – *čistilka/čistilec* – shows that the profession is often associated with a specific ethnic group (Bosnian) and with stereotypes about this group (the name Fata). It can be also observed that this is a profession dominated by women, since the masculine form is mostly used to describe appliances.

6 Conclusion

The contrastive corpus analysis of collocators in English and Slovenian presented in this paper has shown that in both languages there are cases in which the feminine nouns can be described as stylistically unmarked. However, the study has identified numerous cases of masculine and feminine nouns for professions occurring in contexts that reflect the status of their referents in society. Due to these extra-linguistic factors, feminine nouns are more frequently accompanied by stereotypical views of gender roles, appearance and character: thus, their female referents are often sexualized and objectified. The masculine nouns, on the other hand, typically highlight the important position of their referents.

In the end, it should be pointed out that the main focus of the paper has been a synchronic analysis of the subject matter, relying on corpora that mostly contain examples from written language. Considering these limitations, we hope that our findings will motivate future research on the use of masculine and feminine nouns also from a diachronic perspective as well as from the point of view of different text-types.

Corpora and Tools

SSKJ2: Dictionary of Standard Slovenian, www.fran.si.

BNC: <https://old.sketchengine.co.uk/>.

EnTenTen13: <https://old.sketchengine.co.uk/>.

FidaPLUS v2: <https://old.sketchengine.co.uk/>.

SketchEngine, 2018: <https://www.sketchengine.eu/>.

References

- Bartsch, Sabine. (2004). *Structural and functional properties of collocations in English: a corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: Gunter Narr.
- Bauer, Laurie. (1996). *English Word-formation*. Cambridge: Cambridge University Press.
- Caldas-Coulthard, Carmen Rosa, Moon, Rosamund. (2010). Curvy, hunky, kinky: Using corpora as tools for critical analysis. In *Discourse & Society* No 21/2. 99-133. <https://doi.org/10.1177/0957926509353843>.
- Crowther, Jonathan, Dignen, Sheila, Lea, Diana (Ed.). (2002). *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.
- Derganc, Aleksandra. (2017). Poimenovanje za ženske nosilke poklicev in položajev v slovenščini in ruščini [Naming for women holders of professions and positions in Slovenian and Russian]. In *Jezikoslovni zapiski* [Linguistic notes] No 23/2. 125-130. <https://doi.org/10.3986/JZ.23.2.6906>.
- Gablasova, Dana, Brezina, Vaclav, McEnery, Tony. (2017). Collocations in corpus-based language learning research: identifying, comparing, and interpreting the evidence. In *Language Learning* No 67/S1. 155-179. <https://doi.org/10.1111/lang.12225>.
- Gabrovšek, Dušan. (2014). Extending binary collocations: (lexicographical) implications of going beyond the prototypical a - b. In *ELOPE: English language overseas perspectives and enquiries* No 11/2. 8-20. <https://doi.org/10.4312/elope.11.2.7-20>.
- Gabrovšek, Dušan. (2005). *Words Galore: Aspects of general and Slovenian-English contrastive lexicology*. Ljubljana: FF UL, Oddelek za anglistiko in amerikanistiko.
- Hoey, Michael. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Holmes, Janet. (2001). *A corpus-based view of gender in New Zealand*. In *Gender Across Languages: The linguistic representation of women and men*. Marlis Hellinger and Hadumod Bussman, ed. Amsterdam: John Benjamins. 115-136.
- Huddleston, Rodney, Pullum, Geoffrey K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Jurko, Primož. (2015). Evaluative Meaning in Translation. In *ELOPE: English language overseas perspectives and enquiries* No 12/2. 193-209. <https://doi.org/10.4312/elope.12.2.193-209>.
- Kilgarriff, Adam. Baisa, Vít. Bušta, Jan. Jakubiček, Miloš. Kovář, Vojtěch. Michelfeit, Jan. Rychlý, Pavel. Suchomel. Vít. (2014). The Sketch Engine: ten years on. In *Lexicography* No 1/1. 7-36. <https://doi.org/10.1007/s40607014-0009-9>.
- Kranjc, Simona, Ožbot, Martina. (2013). *Vloga spolno občutljivega jezika v slovenščini, angleščini in italijanščini* [The role of gender sensitive language in Slovenian, English and Italian]. In *Družbena funkcijskost jezika: vidiki, merila, opredelilive* (Obdobja 32) [Social Functionality of Language: Aspects, Criteria,

- Definitions]. Andreja Žele (ed.). Ljubljana: Znanstvena založba Filozofske fakultete. 233-239.
- McIntyre, Dan. (2018). *Irony and semantic prosody revisited*. In *The Pragmatics of Irony and Banter*. Manuel Jobert and Sandrine Sorlin (eds.). Amsterdam: John Benjamins. 81-99.
- Philip, Gill E. (2010). *Why prosodies aren't always present: Insights into the idiom principle*. In *Proceedings of the Copus Linguistics Conference CL2009*. Michaela Mahlberg, Victorina González-Díaz and Catherine Smith. Liverpool: University of Liverpool. URL: <http://ucrel.lancs.ac.uk/publications/cl2009/>.
- Quirk, Randolph et al. (1999). *A Comprehensive Grammar of English*. Harlow: Longman.
- Rychlý, Pavel. (2008). *A lexicographer-friendly association score*. In *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN*. Petr Sojka and Aleš Horák, ed. Brno: Masaryk University. 6-9.
- Sinclair, John. (1996). The search for units of meaning. In *Textus: English Studies in Italy* No 9. 75-106.
- Stubbs, Michael. (2001). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Stubbs, Michael. (2007). *Quantitative data on multi-word sequences in English: The case of the word world*. In *Text, Discourse and Corpora: Theory and Analysis*. Michael Hoey, Michaela Mahlberg, Wolfgang Teubert and Michael Stubbs, eds. London: Continuum. 163-189.
- Šorli, Mojca. (2013). Corpus-based lexicographical descriptions with a special focus on pragmatics: the case of the Slovene lexical database. In *ELOPE: English language overseas perspectives and enquiries* No 10/2. 45-65. <https://doi.org/10.4312/elope.10.2.45-65>.
- Šorli, Mojca. (2012). *Semantična prozodija v teoriji in praksi – korpusni pristop k proučevanju pragmatičnega pomena: primer slovenščine in angleščine* [Semantic prosody in theory and practice – a corpus approach to the study of pragmatic meaning: Slovene and English]. In *Dvojezična korpusna leksikografija: slovenščina v kontrastu: novi izzivi, novi obeti* [Bilingual corpus lexicography: Contrasting Slovenian: New challenges, new outlook]. Mojca Šorli (ed.). Ljubljana: Trojina, zavod za uporabno slovenistiko. 90-116.
- Štumberger, Saška. (2015). *Veränderungen in der gesellschaftlichen Stellung von Frauen in Slowenien seit 1991: Personenbezeichnungen im Wörterbuch Slovar novejšega besedja slovenskega jezika (2012)*. In *New Approaches to Gender and Queer Research in Slavonic Studies*. Dennis Scheller-Boltz, ed. Wiesbaden: Harrasowitz. 101-109.
- Toporišič, Jože. (2000). *Slovenska slovnica* [Slovenian grammar]. Maribor: Obzorja.
- Uhlik, Mladen. (2019). Izražanje spolske korelacije pri položajnih poimenovanjih v ruščini in slovenščini [Expressing gender correlation in positional denominations in Russian and Slovenian]. In *Slavistična revija* [Slavistic revue] 67/2. (In press).
- Vidovič-Muha, Ada. (2018). *Slovensko skladijsko besedotvorje* [Slovenian syntactic word formation]. Ljubljana: Filozofska fakulteta.