Russian feminitives: what can corpus data tell us?

Anonymous authors

Abstract

Recent years have seen considerable debate concerning Russian feminitives, i.e. derived formations that designate female professionals, such as *advokatka*, *advokatša*, *advokatessa*, *ženščina-advokat* or *advokat-ženščina* that all refer to female lawyers. In this article, we investigate the use of feminitives based on data from the Araneum Russicum Maximum corpus and the Russian National Corpus. It is shown that the choice of feminitive to some extent depends on the morphophonological properties of the base word. It is furthermore argued that suffixed feminitives are more frequent than compounds like *ženščina-advokat* and *advokat-ženščina,* and that the distribution has changed over time. Suffixed feminitives reveal a stronger tendency to combine with gender-related epithets (e.g., *obajatel'naja agentka* 'charming agent'), while the type *ženščina-X* is frequently used with the epithet *pervyj* 'first'. Our article is an empirical study of the actual use of feminitives in corpus data, which we hope will inform future metalinguistic discussion and prescriptivist thinking about feminitives in Russian.

В последние годы ведется активная дискуссия о русских феминитивах, т. е. о производных единицах, обозначающих женщин по роду занятий, например *адвокатка*, *адвокатша*, *адокатесса*, *женщина-адвокат* и *адвокат-женщина* для указания на женщину, которая занимается адвокатской деятельностью. В этой статье использование феминитивов изучается на основе данных корпуса Araneum Russicum Maximum и из Национального корпуса русского языка. Мы показываем, что выбор феминитива в определённой степени зависит от морфонологических свойств производящей основы. Кроме того, мы демонстрируем, что суффиксальные феминитивы имеют более высокую частотность, чем композиты типа *женщина-адвокат* и *адвокат-женщина*, и что их распределение со временем меняется. Суффиксальные феминитивы склонны сочетаться с определениями, так или иначе связанными с гендером (напр., *обаятельная агентка*), в то время как тип *женщина-X* часто используется со словом *первая*. Наша статья представляет собой эмпирическое исследование феминитивов в корпусах, и можно надеяться, что полученные выводы будут полезны для дальнейших металингвистических дискуссий и для формулирования прескриптивных указаний.

1. Introduction

"Russian words denoting females have had a difficult fate", says Fufaeva (2020, p. 12). In Soviet times, Russian used masculine words like *advokat* 'lawyer', *inžener* 'engineer' and *èkskursovod* 'guide' to designate the majority of professions, even when the professional in question is female. In formal contexts, this happened almost without exceptions. However, in recent years it has become more widespread to use separate words to refer to female professionals, so-called feminitives (see Arkhangelskaya, 2014; Fufaeva, 2018; Georgievskaja, 2018; Morozova, 2018; Piperski, 2018, 2019; Krongauz & Severin, 2019; Bobylëva, 2021). Should a female lawyer be referred to as an *advokatka*, *advokatša*, *advokatessa*, *ženščina-*

1

advokat or *advokat-ženščina*? Russian employs a number of suffixes that denote female persons (Townsend, 1968, p. 180-185; Švedova et al., 1980, §380–393), and in addition compounds like *ženščina-advokat* or *advokat-ženščina* are attested in contemporary Russian discourse. Feminitives formed by means of these and other word-formation patterns[1] have become a hotly debated topic in metalinguistic discussions (see the selection of articles on feminitives on the webpage of the Russian feminist association *ONA* 'she': https://ona.org.ru/search/феминитивы). Linguists have also become interested in empirical studies of the use of feminitives and their interaction with other forms of expressing or avoiding gender in present-day Russian and earlier periods (Badanina, 2017; Fufaeva, 2020; Kirey-Sitnikova, 2021; Magomedova & Slioussar, 2021). Our aim with the present study is to contribute to this field. Without going into the metalinguistic and prescriptivist issues at stake (Rozental' et al., 1998), our research question is: What *linguistic* factors motivate the distribution of feminitives? We carry out a quantitative analysis of a large sample of data and establish the frequencies for various feminitives. We hope that our findings will inform future metalinguistic discussion and prescriptivist thinking about feminitives.

Feminitive formation in Russian is also interesting from the point of view of morphological theory. This is a case of suffix rivalry, a topic that has received considerable attention in recent years. A classic example is the competition between *-ity* and *-ness* in English (see, e.g., Arndt-Lappe, 2014). Lindsay and Aronoff (2013) analyze rivalry between a number of English suffixes, such as *-ity*, *-ment* and *-ation*. Works about morphological rivalry in Russian include Nesset and Janda's studies of "suffix shift" in Russian verbs (Nesset & Janda, 2010), Naccarato's (2019) analysis of rival word-formation constructions for Russian compounds, and Bobkova and Mortermini's (2020) investigation of the rivalry between relational adjectives in *-sk-*, *-n-*, and *-ov-.* While all these authors focus on productivity and mostly consider formal factors (the shape of the stem), we will not discuss productivity and we furthermore analyze both formal and semantic factors relevant for the choice of feminitive.

In order to shed light on the distribution of feminitives, we present two empirical studies. First, we undertake a large-scale quantitative study using the Araneum Russicum Maximum corpus (Benko, 2014), a web-corpus with approximately 20 billion tokens. Second, we present the results of a narrower case study of data from the Russian National Corpus (RNC, both the main and the newspaper subcorpora, which taken together contain about 600 million tokens).[2] The two studies complement each other. On the one hand, the Araneum corpus provides a large amount of data that facilitate the study of morphophonological factors that influence the distribution of feminitives, but it does not provide any relevant metadata. This part of the analysis sheds light on the behavior of the relevant word-formation patterns in the language at large. On the other hand, the study of the smaller, but richly annotated RNC enables us to zoom in on a more detailed analysis of ten frequent professions with several feminitive word-formation patterns and track their development over time. Since this case study is limited to ten words, it is not clear whether the observed tendencies generalize to the Russian language as a whole. However, since the ten words combine with several word-formation patterns, these words enable us to draw conclusions about competition between the relevant patterns. Admittedly, corpus-data offer limited opportunities to study the socio-

---

[1] Throughout the article, we refer to feminitive formation *patterns* as they include both suffixation and compounding. Formations as *ženščina-X* are sometimes analyzed as appositive constructions, but their exact status is beyond the scope of our study.

[2] The Araneum Russicum Maximum is available at http://unesco.uniba.sk. The Russian National Corpus can be accessed at www.ruscorpora.ru.

linguistic factors related to the use of feminitives (age, region, social status, register etc.), but the data and analysis presented in our study creates a basis for further investigations of the socio-linguistics of feminitives in present-day Russian.

Our findings can be summarized as follows. First, data from the Araneum corpus indicate that morphophonological factors, in particular stem-final consonant and stress, are important for the choice of feminitive form. We present a flow-chart that accommodates the relevant statistical tendencies.

Second, in addition to confirming the results of the quantitative analysis, the case study of the RNC data suggests that suffixed feminitives are more frequent than compounds of the types *ženščina-X* (e.g., *ženščina-advokat*) and, especially, *X-ženščina* (e.g., *advokat-ženščina*). This conclusion holds for both the main and the newspaper subcorpora.

Third, it is shown that the distribution of feminitives has changed over time. The model *ženščina-X* came into use in the first part of the 20th century, followed by the suffix *-ka* (e.g., *advokatka*) in the middle of the century. The suffix *-essa* has become widely used for lawyers (*advokatessa*) in the 21st century, but it is still marginal for the remaining professions under scrutiny. The suffix *-ša* is in the process of changing from denoting the wife of a professional to denoting a female professional, as illustrated by *kapitanša* from *kapitan* 'captain'. While in the 19th century *kapitanša* was used about a captain's wife, this word now frequently refers to a female captain.

A fourth finding is that adjectival modifiers combining with feminitives reveal interesting behaviors. While the model *ženščina-X* is frequently used with the epithet *pervyj* 'first' (e.g., *pervaja ženščina-advokat* 'first female lawyer'), suffixed models more often attract gender-related epithets (e.g., *xrupen'kaja advokatessa* 'delicate lawyer' and *obajatel'naja agentka* 'charming agent').

Our article is structured as follows. Section 2 discusses the quantitative study of the Araneum corpus, before we turn to data from the RNC in section 3. Subsection 3.1 provides an overview of the data, while subsection 3.2 is devoted to the historical development and subsection 3.3 explores epithets that combine with feminitives. Our findings are summed up in section 4.

2. Quantitative analysis of data from the Araneum Maximum corpus

In order to perform a corpus study of feminitives, we need a comprehensive list of professions. There is at least one large official list of this kind, namely *Obščerossijskij klassifikator professij rabočix, dolžnostej služaščix i tarifnyx razrjadov* 'All-Russian Classification of Workers' Professions, Employee Positions and Wage Grades', which contains a total of 5,556 entries, ranging from common terms such as *bortprovodnik* 'flight attendant' and *artist dramy* 'drama actor' to very specific terms such as *redaktor rukovodstv dlja plavanija* 'editor of navigation guides' and *naladčik avtomatov èlementnogo proizvodstva* 'adjuster of automatic element production machines'. From these 5,556 entries, we extracted a total of 1,158 types of head nouns in nominative case denoting professions: for example, the four professions mentioned above would contribute *bortprovodnik* 'flight attendant', *artist* 'actor', *redaktor* 'editor', and *naladčik* 'adjuster' to the final list, while all other words were not included (such as *drama* and *rukovodstvo*), since they do not denote professions that could potentially form feminitives. The number of nominative nouns is five times smaller than the number of entries because many entries share a common headword: e.g., *agent* 'agent' is the headword of 12 entries.

Obviously, using *Obščerossijskij klassifikator …* as a source has certain drawbacks. It does not include umbrella terms (e.g., *muzykant* 'musician') and recent loanwords (e.g., *dizajner* 'designer'). However, it is the most complete list available, which justifies its use.

Out of these 1,158 names of professions, only 15 belong to the feminine grammatical gender, namely:

akušerka 'midwife', *vyšival'ščica* 'embroiderer', *vjazal'ščica* 'knitter', *gorničnaja* 'maid', *kasteljanša* 'laundrywoman', *kovrovščica* 'carpet weaver', *kruževnica* 'lacemaker', *manikjurša* 'manicurist', *mašinistka* 'typist', *modistka* 'milliner', *mojščica* 'washer', *pedikjurša* 'pedicurist', *sanitarka* 'nurse', *sestra* 'sister, nurse', *šveja* 'seamstress'[3]. However, the object of our study were not these 15 nouns, but rather the remaining 1,143 nouns and feminitives formed from them.

In order to find out what feminitive forms occur naturally, we performed a corpus study using Araneum Russicum Maximum (Benko, 2014), a corpus of Russian texts collected from the Web. This corpus suits our purposes quite well, because it is very large, containing 20 billion tokens, and mostly includes texts written in the last two decades. Obviously, this corpus has some drawbacks: it lacks sociolinguistic markup; some automatically generated texts, duplicate texts, and texts in other languages such as Bulgarian or Ukrainian are not always filtered out; and part-of-speech tagging and lemmatization are not perfect. However, these shortcomings do not make the corpus unsuitable for our purposes.

Using 1,143 nouns in our sample[4], we formed feminitives using all the most common suffixes, namely *-ka*, *-ica/-nica*, *-ša*, and *-inja* (Švedova et al., 1980, §380–393; Fufaeva, 2020, p. 281–295)[5]; we also included compounds of the types *ženšcina-X* and *X-ženščina*[6]. For instance, for *direktor* 'director' the following feminitives were formed: *direktorka*, *direktorica*, *direktorša*, *direktorinja*, *ženščina-direktor*, and *direktor-ženščina*. Where possible, idiosyncratic forms (in this case, *direktrisa*) were also added to the list. All of these forms were automatically inflected for case and number, and the resulting list was fed to Araneum. Summing up the counts for the forms of each lemma, we obtained a list of frequencies for all possible feminitives. A sample extract from our data can be seen in Table 1.

Masculine form	-ka	-ica / -nica	-ša	-inja	ženšcina-X	X-ženšcina
fel'dšer 'paramedic'	8	1,282	1	0	83	3
konditer 'confectioner'	1,712	0	72	0	0	1

Table 1: Distribution of feminitives for some nouns in Araneum Russicum Maximum.

As expected, the resulting dataset presented some challenges. First, many names of professions in the All-Russian Classification are very rare. A typical sample of professions from our list is as follows: *groxotovščik* 'grater operator', *groxotčik* 'grater operator', *gruzčik* 'loader', *gruntoval'ščik* 'grubber', *gruntovščik* 'primer', *gummirovščik* 'rubber gum operator',

---

[3] Masculine nouns corresponding to some of these nouns can also be formed, e.g. *akušer*, *vyšival'ščik*, etc., but they are not listed in the All-Russian Classification.

[4] We emphasize that this is a sample of bases rather than a sample of suffixed units as in (Bobkova & Montermini, 2020).

[5] More accurately, one should distinguish between the suffix proper and the inflectional ending linked to this suffix, e.g. *-k-a*. However, for the sake of brevity we will use the notation *-ka*, etc. throughout this paper.

[6] Other less frequent and non-neutral compounds such as *devuška-direktor* 'young.woman-director', *staruška-direktor* 'old.woman-director', etc. were not included. This also applies to rare suffixes like *-essa* as in *poètessa* 'female poet' and *advokatessa* 'female lawyer' and *-isa* as in *direktrisa* 'female director', although we will return to *advokatessa* in Section 3 below. We also do not consider suffixes that have more specific meanings like *-ička* 'female teacher of a certain subject', as in *geografička* 'female geography teacher', *biologička* 'female biology teacher', etc.

gurtovščik 'cattle drover', *davil'ščik* 'crush operator', *daktilolog* 'fingerprint specialist', *dvornik* 'janitor', containing only two words that are frequent and understandable to anyone except a specialist, namely *gruzčik* and *dvornik*. Second, the results had to be filtered manually for some words, since not all derivatives with *-ka* and *-ica/-nica* are actually feminitives. For instance, *mel'nica* means 'mill' rather than being a feminitive from *mel'nik* 'miller', and many words in *-ka* are univerbations based on adjective + noun phrases, e.g. *kur'erka* is not only a female *kur'er* 'courier', but also a colloquial shorthand for *kur'erskaja služba* or *kur'erskaja dostavka* 'delivery service'; *konditerka* may denote a female *konditer* 'confectioner', but is more likely to mean *konditerskie izdelija* 'confectionery', etc.

As we are focusing not on morphological productivity, but rather on the formation and use of stable lexical items, we restricted our analysis to feminitives formed from 329 nouns where the masculine form occurred at least 1,000 times in the Araneum corpus (the threshold was chosen manually to ensure an amount of data that seems sufficient for each word). In cases of unwanted homonymy, we performed manual filtering of the concordances to get a more realistic estimate for the counts of a feminitive; in cases where the number of hits was too large, we had to extrapolate the true number of hits based on 100 randomly selected concordance lines. Thus, for instance the actual count for *konditerka* 'female confectioner' in Table 1 dropped from 1,712 to 0, since there was not a single example of this word used as a feminitive among 100 randomly selected contexts.

After filtering the data, we can establish the most frequent suffixed feminitive form for each noun.

> *-ka* (59 words): *iskusstvovedka* 'art historian', *èkspertka* 'expert', *konsul'tantka* 'consultant', …

> *-ica/-nica* (119 words): *sledovatel'nica* 'investigator', *učitel'nica* 'teacher', *načal'nica* 'head', …

> *-ša* (78 words): *dizajnerša* 'designer', *ekskursovodša* 'tourist guide', *agronomša* 'agronomist'…

> *-inja* (9 words): *biologinja* 'biologist', *šefinja* 'chef', *psixologinja* 'psychologist', *geologinja* 'geologist', *mikrobiologinja* 'microbiologist', *narkologinja* 'expert in narcology', *bibliografinja* 'bibliographer', *kartografinja* 'cartographer', *ornitologinja* 'ornithologist'.

For other words, the counts for all suffixed forms were too small (not exceeding two), or the *ženščina-X* form was preferred over the suffixed feminitives.

The lists presented above allow us to draw some conclusions about the distribution of the suffixes. Namely, *-nica* is preferred with bases in *-tel'*, such as *učitel'* — *učitel'nica* 'teacher'[7]; *-ica* is used with bases in *-nik*, e.g. *načal'nik* — *načal'nica* 'head' (these two observations can already be found in Švedova et al., 1980, §382–383); bases ending in *f* and *g* take the feminitive suffix *-inja*, e.g. *šef* — *šefinja* 'chef'. These rules are categorical, i.e. they are applicable to all nouns with the corresponding stems. However, the choice between *-ka* and *-ša* is much less straightforward and requires a statistical analysis.

Out of 59 words having *-ka* as the dominant suffix, 41 never occurred with *-ša*; most notably, these were the stems ending in *-ist* and *-k* (*parašjutistka* rather than *parašjutistša* 'parachutist', *gornjačka* rather than *gornjakša* 'miner'). Out of 78 words having *-ša* as the dominant suffix, 29 never occurred with *-ka*, most of them having a stem in *-r* (*auditorka*

---

[7] Thus, a combination *-tel'* + *-nica* is a morphologically complex affix consisting of a dependent and its carrier as defined by Stump (2020).

rather than *auditorša* 'auditor'). The remaining 18 + 49 = 67 words exhibit a competition between *-ka* and *-ša*. We focus on these 67 words together with 11 other words exhibiting this competition, where both *-ka* and *-ša* exist but are not the most common way of forming a feminitive (e.g., *advokatka/advokatša* 'lawyer', the most common feminitive in Araneum being *ženšcina-advokat*).

In order to find out how to choose between these two suffixes, we constructed a decision tree using two phonological variables as predictors and a binary outcome variable (*-ka / ša* being the most frequent variant)[8]. The two phonological variables in question are stem-final consonant and stress position in the base. These variables have already been hinted at by Švedova et al. (1980, §385), and Uhlik (2019). 78 words are obviously too few for a definitive statistical analysis, but some observations can still be made. A baseline accuracy of 78% (61 out of 78) can be achieved by stating that all the nouns in question prefer *-ša*. However, the best possible accuracy with these two variables is 85% (66 out of 78), which is reached by separating dental stops (*d* or *t*) from other stem-final consonants at the root level of the decision tree. 14 out of 23 stems (61%) ending in a dental stop prefer *-ka* with stress playing no role, whereas if a stem ends in a sonorant, there is an overwhelming preference for *-ša*; it is the preferred form for 95% of the stems (52 out of 55), while there are only 3 words with a preference for *-ka*, all of them being stressed on the last syllable (*kollekcioner* 'collector', *stažër* 'trainee', and *kočegar* 'stoker'[9]).

Our findings can be summed up in the form of a flowchart (Figure 1) that shows the preferred feminitive suffixed form for any stem. The reader should bear in mind that it only presents statistical tendencies rather than exceptionless rules.

---

[8] The analysis was performed in R using `rpart` package (R Core Team, 2021; https://rdocumentation.org/packages/rpart/versions/4.1-15).

[9] Note that *kočegarka* 'steamshop' was filtered out manually at the preceding stages of analysis.
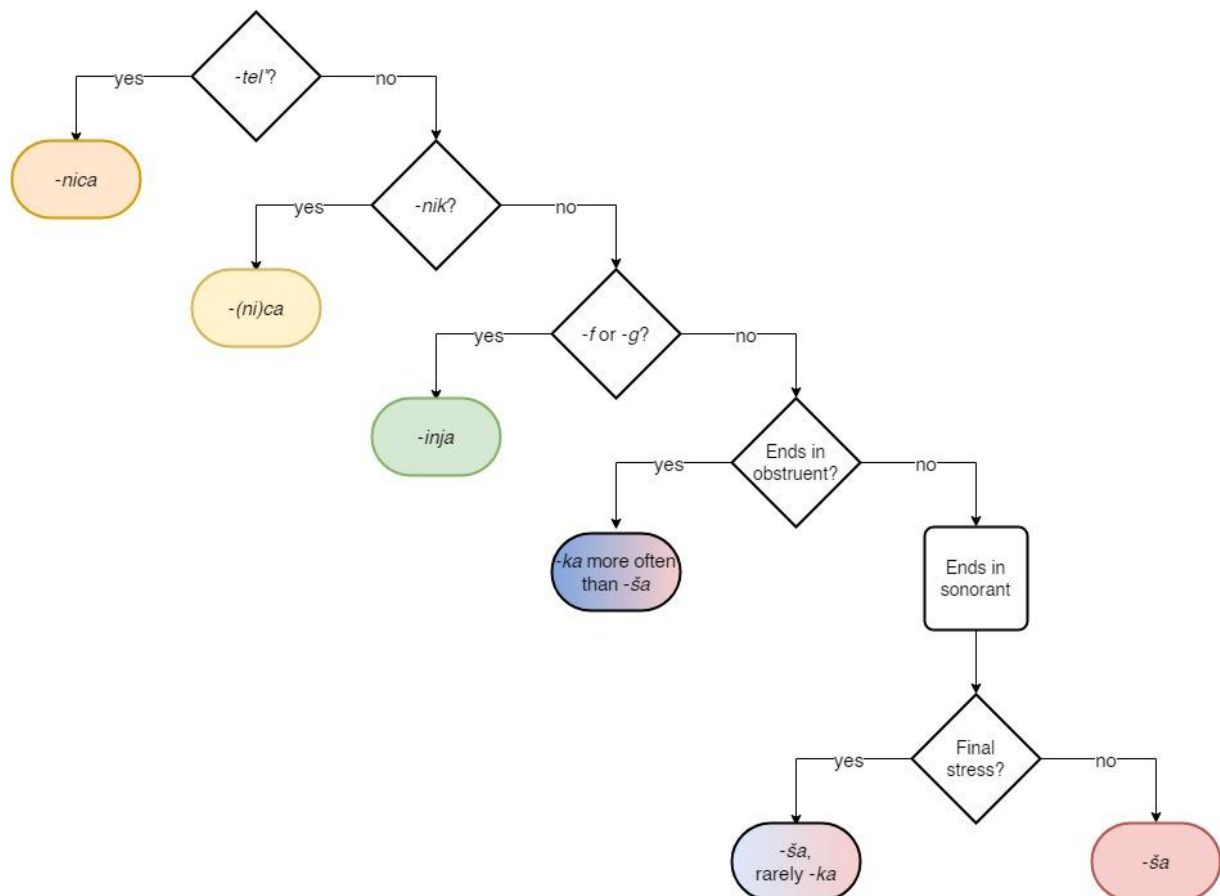
Figure 1: Choice of feminitive according to the Araneum corpus

In summary, the data from the Araneum corpus shows that the choice of feminitive depends on word-formation (derivational suffixes like *-tel'* and *-nik*), but also on morphophonology (the stem-final consonant). In the following, we will zoom in on ten widely used words which combine with several feminitives. These words will be analyzed on the basis of data from the RNC.

3. Qualitative analysis of data from the Russian National Corpus

3.1 Overview

For all nouns in the cleaned sample from the Araneum corpus, we calculated the Simpson diversity index (i.e., the probability that two randomly selected feminitive forms of a word belong to the same word-formation pattern, see Simpson, 1949; also known as the Herfindahl–Hirschman index) and chose the 10 most diverse words with at least two competing suffixed feminitives. This resulted in the following list of professions, which forms the basis of our analysis of data from the RNC:

(1) *advokat* 'lawyer', *agent* 'agent', *inžener* 'engineer', *trener* 'coach', *dizajner* 'designer', *konsul'tant* 'consultant', *režisser* 'film director', *kapitan* 'captain', *instruktor* 'instructor', and *èkskursovod* 'guide'.

We then searched for all attestations of possible feminitive forms of these nouns in the main and newspaper subcorpora of the RNC. The newspaper corpus was included in the study, because language use in mass media is of particular interest with regard to feminitives, which

represent a hotly debated topic in metalinguistic discussions. We searched for feminitives involving the following word-formation patterns:

(2) Word-formation patterns included in the analysis of RNC data:
 a. *-ka* (e.g., *advokatka*)
 b. *-ša* (e.g., *advokatša*)
 c. *-essa* (e.g., *advokatessa*)
 d. *ženščina-X* (e.g., *ženščina-advokat*)
 e. *X-ženščina* (e.g., *advokat-ženščina*)

The result was a database of 586 attestations spanning from 1800 to 2017. The general distribution of the relevant word-formation patterns is given in Table 2 and Figure 2.[10] We provide numbers for both the main subcorpus and the newspaper subcorpus. Although there are some differences, the distribution is relatively similar in both subcorpora. It is worth pointing out that the newspaper subcorpus mainly contains "traditional" newspaper articles. We speculate that the distribution may be different in less formal genres such as blogs and debate forums, but our database does not allow us to test this hypothesis.

The following observations can be made on the basis of our dataset. First of all, if we compare suffixed feminitives to compounds, it is clear that the suffixed types are more widespread. In both subcorpora, suffixed feminitives represent about 80% of the data. In particular, the *X-ženščina* pattern is marginal in our dataset. Among the suffixes, *-ša* is more frequent than *-essa*, which in turn is more frequent than *-ka*. On the face of it, this finding may be surprising, since in section 2 it was shown that *-ka* is widely used. However, upon closer inspection the high frequency of *-ša* in our dataset is not at variance with the results from section 2, where our analysis of data from the Araneum corpus showed that stems in sonorants prefer *-ša*. Since half the nouns in (2) end in the sonorant /r/ and these nouns constitute 64% of the examples (tokens) in our database, the data from the RNC lend further support to the findings from the Araneum corpus.

Why is *ženščina-X* more frequent than *X-ženščina*? Our data do not provide a clear answer, but we note that in both word-formation patterns the leftmost component tends to control agreement.[11] Thus, *ženščina-advokat* typically combines with adjectival modifiers in feminine grammatical gender, e.g. *pervaja* 'first (nominative feminine)':[12]

(3) Segodnja v obščem sobranii "ligi ženskogo ravnopravija" pervaja russkaja ženščina-advokat g-ža Flejšic vystupaet s dokladom "ženščina i pravo". ("Russkoe Slovo" 1910)
 'Today at the general assembly of the "League of women's equal rights" the first Russian female lawyer Mrs. Flejšic will give a lecture on the topic "woman and law".'

Advokat-ženščina, on the other hand, frequently takes adjectival modifiers with masculine grammatical gender, e.g. *veduščym* 'leading (instrumental masculine)':

---

[10] Seven examples in our database involve both compounding and suffixation at the same time. For instance, in *babuška-inženerša* '(lit.) grandmother-female engineer' both the compounding of *babuška* and *inžener* and the addition of the suffix *-ša* is used to clarify the gender of the referent. In Table 2 and Figure 2, these examples are counted as suffixed models.

[11] Notice that we have only investigated preposed agreement targets.

[12] Unless otherwise indicated, numbered examples are from the RNC. For each example a year is given in addition to the name of the author (for fiction) or the periodical (for non-fiction).

(4) Ona javljalas' veduščim advokatom-ženščinoj po ugolovnym delam v Irlandii.[13]
'She is the leading female lawyer in the field of criminal law in Ireland.'

Thus, the agreement pattern of *ženščina-X* gives the language user an opportunity to emphasize feminine gender through agreement with *ženščina*, while the feminine gender is not emphasized in the typical agreement pattern of *X-ženščina*. We speculate that the feminine gender of the referent is important in many contexts, and that this may explain why *ženščina-X* dominates over *X-ženščina* in our dataset. However, testing this hypothesis is beyond the scope of the present study.

	Main corpus		News corpus		Total
	#	%	#	%	
-ša	296	62.9%	73	62.9%	369
-ka	57	12.1%	7	6.0%	64
-essa	24	5.5%	18	15.5%	42
ženščina-X	83	17.6%	17	14.7%	100
X-ženščina	9	1.9%	1	0.9%	10
Total	470	100%	116	100%	586

Table 2: Distribution of feminitives in the main subcorpus (left) and the newspaper subcorpus (right) of the Russian National Corpus. Raw numbers are given in columns marked #, percentages in columns marked %.
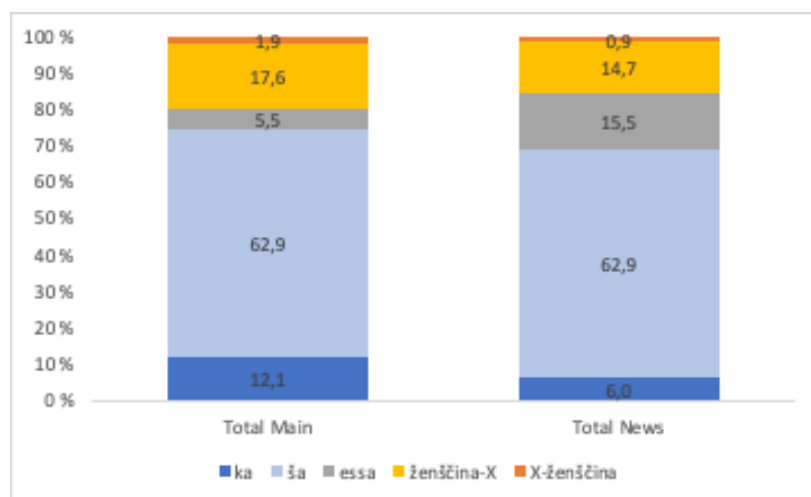

Figure 2: Distribution of feminitives in the main subcorpus (left) and the newspaper subcorpus (right) of the RNC.

Let us now turn to the individual lexical items under scrutiny. Based on the frequencies displayed in Table 3 and Figure 3, a number of observations can be made. First of all, it is striking that each lexical item shows a different behavior and many items show variation among different word-formation patterns. Thus, although the aggregate data from the Araneum corpus and the RNC explored above show that the choice of suffix depends on the morphophonological properties of the stem, the relevant generalizations are no more than statistical tendencies, which apply in different degrees to different lexical items.

A second observation concerns the distribution of the suffixes. While *-essa* is attested only for *advokat*, *-ka* and especially *-ša* are widely distributed. The relationship between these two suffixes are asymmetric. While *-ka* is almost exclusively attested with stems in obstruents, *-ša*

---

[13] Our database does not contain examples with relevant agreement patterns; this example is from the internet: https://glosbe.com/en/ru/criminal%20lawyer.

occurs after both obstruents and sonorants. There is furthermore only one lexical item, *konsul'tant*, where -*ka* is the suffix with the highest relative frequency. The -*ša* suffix, on the other hand, is the most frequent word-formation pattern for eight out of the ten lexical items in Table 3 and Figure 2. This testifies to the general robustness of -*ša* vis-à-vis its "competitors" -*ka* and -*essa* in our dataset. In our analysis of the Araneum corpus in section 2, it was shown that nouns ending in sonorants have an affinity for -*ša*, and the data from the RNC confirms this, since five out of ten words under scrutiny in this section end in /r/. However, we hasten to add that our study of data from the RNC concerns a small set of lexical items, so we cannot draw conclusions about the mophophonological shape of stems in the language at large.[14]

As a third point, notice that the *ženščina-X* pattern has a somewhat "flatter" distribution than the suffixes, in the sense that it is attested for all lexical items under scrutiny. Apart from -*ša*, *ženščina-X* is the only word-formation pattern that combines with all lexical items in Table 3 and Figure 3. This makes sense; while suffixes are closely connected to the stem and therefore often sensitive to the stem's morphophonological properties, compounds of the *ženščina-X* type are not expected to be sensitive to morphophonology, and should therefore be compatible with all types of stems. As we will see in section 3.3, however, the *ženščina-X* word-formation pattern shows sensitivity to its *syntactic* environment, but before we turn to syntax, we will explore the historical development of the ten lexical items under scrutiny.

	ka	ša	essa	ženščina-X	X-ženščina	Total
agent	33	43	0	8	2	86
advokat	6	20	42	19	0	87
dizajner	0	13	0	1	0	14
inžener	5	59	0	26	0	90
instruktor	0	16	0	4	1	21
kapitan	0	145	0	20	2	167
konsul'tant	11	2	0	1	0	14
režissër	2	31	0	12	5	50
trener	0	28	0	4	0	32
èkskursovod	7	12	0	5	0	24
Total	64	369	42	100	10	

Table 3: Distribution of feminitives for individual lexical items in the RNC (main and newspaper subcorpora combined).

---

[14] We would like to emphasize that we do not make any claims about productivity of the suffixes under scrutiny. Our sample of ten words from the RNC is too small to facilitate claims about productivity, and we focus on high frequency lexemes that are not well suited for an investigation of productivity.
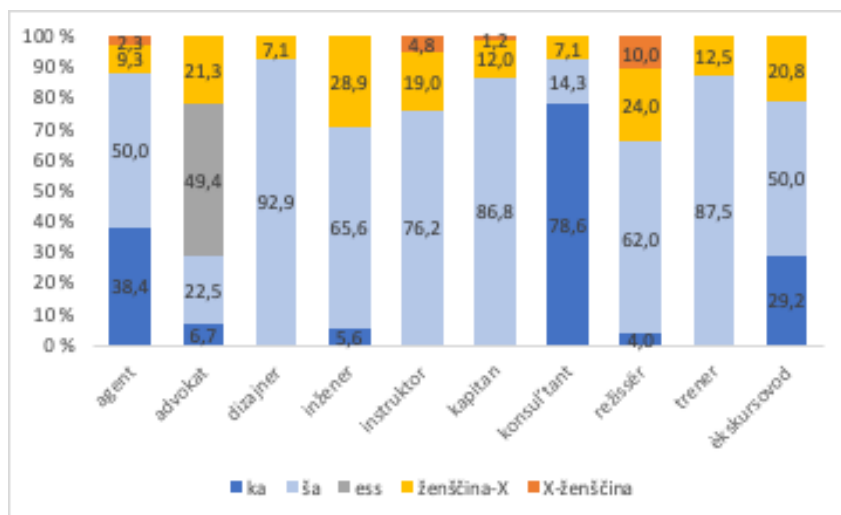
Figure 3: Distribution of feminitives for individual lexical items in the RNC (main and newspaper subcorpora combined).

3.2 Historical change

The changes in the distribution over time are shown in Tables 4-5 and Figures 4-5. We have divided the data into periods of fifty years. The first period is labeled "before 1850", because it includes thirty one example from the 18th century. The last period spans the first fifteen years of the 21st century.

In Table 4 and Figure 4, we report items per million. We see that -*ša* is the only pattern that was widely used in the 19th century for the words under scrutiny. Measured in items per million, its frequency decreases in the 19th century, and then increases again in the 20th century. The *ženščina-X* type becomes widespread in the beginning of the 20th century, while -*ka* is widely used from the middle of the 20th century. The -*essa* suffix is frequently used about female lawyers in the 21st century. The *X-ženščina* type is marginal in our database, but it is attested from the first half of the 20th century, in the same way as *ženščina-X.* These linguistic changes reflect changes in society. For instance, some of the words we analyze represent relatively recent professions (e.g., *dizajner*) and other professions were not available for women in earlier times. As pointed out by Fufaeva (2020, p. 105), for instance, women were only admitted to law schools in 1911, and the first female lawyers appeared later.

	Before 1850	1851-1900	1901-1950	1951-2000	2001-2015
ša	1,36	1,08	0,82	0,95	1,38
ženščina-X	0,00	0,00	0,39	0,32	0,44
ka	0,00	0,07	0,02	0,44	0,19
ess	0,00	0,00	0,07	0,06	0,39
X-ženščina	0,00	0,00	0,01	0,03	0,08
Total words in corpus	22 817 261	59 525 713	88 612 642	96 389 525	77 488 101

Table 4: Changes in the distribution of the relevant word-formation patterns over time based on data from the RNC (main and newspaper subcorpora combined). Items per million word.
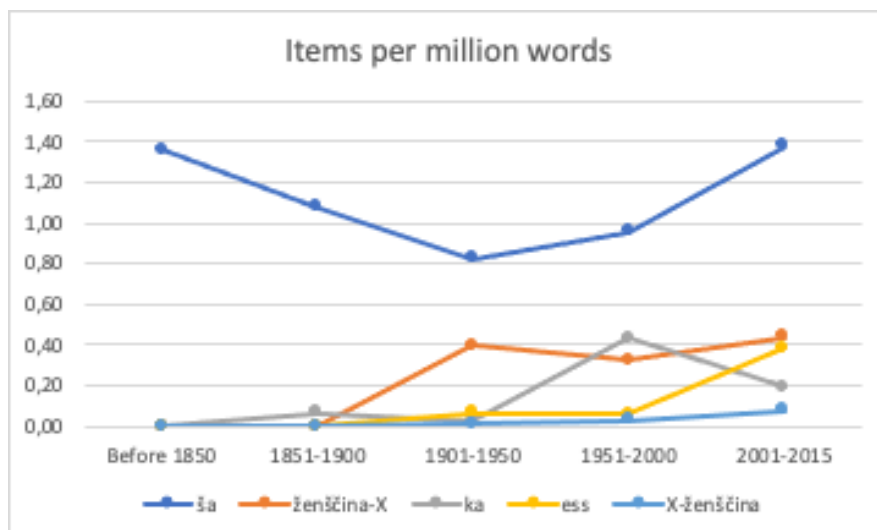
Figure 4: Changes in the distribution of the relevant word-formation patterns over time based on data from the RNC (main and newspaper subcorpora combined).

Table 5 and Figure 5 show the relative distribution of the relevant patterns in different periods. Although, as shown in Figure 4, the absolute frequency of *-ša* increases in the 20th century, the *proportion* of *-ša* is lower in the 20th and 21st centuries, as competing patterns emerge. It has been argued in the literature that the use of *-ša* is restricted (Mozdzierz, 1999, p. 165). Rozental' et al. (1998, p. 200-201) claim that a reduced use of feminitives in *-ša* is observed for two reasons: to leave out the cases of polysemy with the second meaning of *-ša,* referring to the 'wife of a professional', and to avoid negative connotations. Some scholars, like Doleschal (2015, p. 1168) and Lopatin and Uluxanov (2016, p. 777), list the Russian suffix *-ša* as productive only in colloquial Russian. Our data in Tables 4 and 5 do not indicate a reduced use of *-ša* for the nouns under scrutiny. Although our data suggest that the *proportion* of *-ša* has decreased, the frequency measured in items per million is on approximately the same level in the 21st century as it was in the first part of the 19th century. We hasten to add that our results only concern ten nouns, which are frequent and display considerable variation between the patterns under scrutiny. Whether our results can be generalized to other groups of nouns is a question we will leave open for future research.

Dmitrieva (see Fufaeva, 2020, p. 146, for discussion) found that for female authors the use of feminitives decreased in the middle of the 20th century (1930-1960). While our focus is not on female authors, our data lends some support to this observation. As shown in Figure 4, the use of *-ša* dipped in the beginning of the 20th century, and this is also the period where the *ženščina-X* type becomes widely used.

	Before 1850		1851-1900		1901-1950		1951-2000		2001-2015	
	#	%	#	%	#	%	#	%	#	%
ša	31	100%	64	94%	73	62%	92	53%	107	56%
ka	0	—	4	6%	2	2%	42	24%	15	8%
essa	0	—	0	—	6	5%	6	3%	30	15%
ženščina-X	0	—	0	—	35	30%	31	18%	34	18%
X-ženščina	0	—	0	—	1	1%	3	2%	6	3%
Total	31	100%	68	100%	117	100%	174	100%	192	100%

Table 5: Changes in the relative distribution of feminitives over time based on data from the RNC (main and newspaper subcorpora combined). Raw numbers are given in columns marked #, percentages in columns marked %.
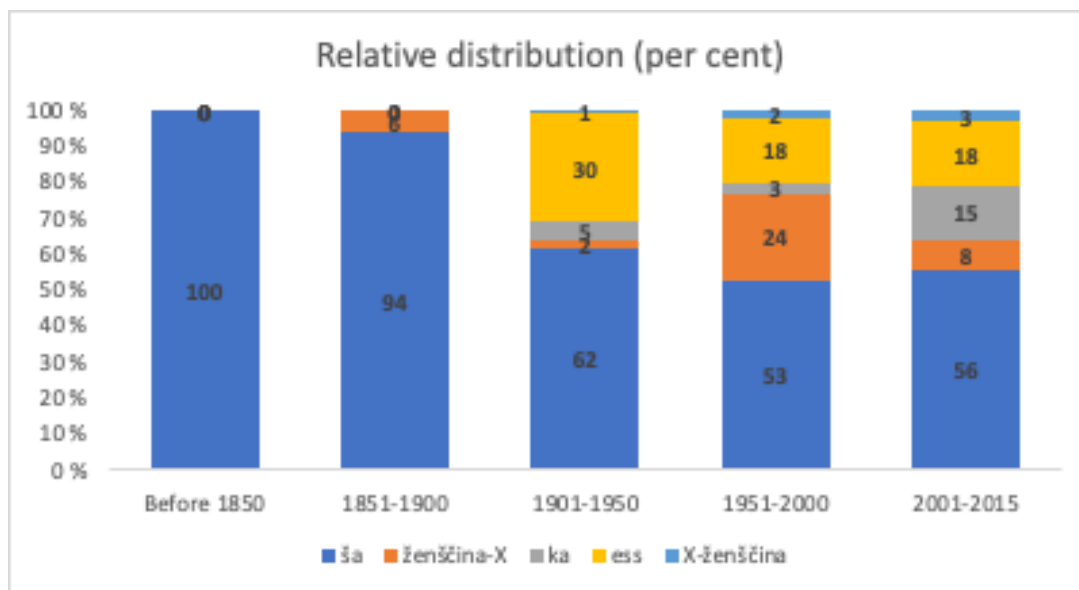
Figure 5: Changes in the relative distribution of feminitives over time based on data from the RNC (main and newspaper subcorpora combined).

As mentioned, the proportion of *-ša* displays a gradual decrease in the 19th century and then becomes larger again in the 20th century, as shown in Table 4 and Figure 4. This may be related to a semantic shift. Traditionally, this suffix was used to designate the wife of a professional, as in the following example where *kapitanša* refers to a 'captain's wife':[15]

(5) Bog mne posobil prijutit'sja v dolžnosť èkonomki u staruški vdovy, kapitanši 1 ranga, syn kotoryj byl odnim iz učitelej, gde naxodilis' moi rebjatiški. (Skobelev 1838-1844)
'With God's help I was able to get a job as a housekeeper at an old widow's house. She had been the wife of a captain of first rank, whose son had been one of the teachers where my children had gone to school.'

However, with the emergence of women in professions that were earlier reserved for men, the *-ša* suffix came to be used about female professionals:

(6) – Vam razrešili vyezd, – skazala kapitanša KGB. (Goljaxovskij 1984-2001)
'"You have permission to leave", the female KGB captain said.'

In order to find out more about the change from 'professional's wife' to 'female professional', we went through all examples with *kapitanša* in our dataset. The data are summarized in Table 5, which covers the same time periods as Table 4 above. As can be seen from the table, *kapitanša* is attested in the meaning 'captain's wife' in all periods, but examples from the end of the 20th century and the beginning of the 21st century tend to describe realia from earlier times, as in the following example, which is about a story by A.K. Tolstoj (1814-1875):

(7) Imenno zdes' na pamjať čitatelju dolžny prijti pervye stranicy znamenitoj povesti grafa A.K. Tolstogo "Upyr'": mnogoljudnyj bal u kapitanši Sugrobinoj, strannoe povedenie

---

[15] Strictly speaking, examples like (5) are not "feminitives" since they denote a wife of a professional rather than a female professional. We have included examples like (5) in our database, since this enables us to investigate empirically the semantic shift from 'wife' to 'female professional'.

odnogo iz gostej, vzdumavšego vozmuščat'sja ni bol'še, ni men'še, kak prisutstviem upyrej! ("*Znanie – sila*" 2008)

'Right here the first pages of the famous short novel "The Vampire" by count A.K. Tolstoj might come to mind: A ball with many guests at the house of the captain's wife Sugrobina and the strange behavior of one of the guests who got upset over the presence of no less than vampires.'

In the meaning 'female captain', the first attestation in our dataset is from 1912. The example describes a female soccer team:

(8) Simpatičnye kapitanši, xavbekši i gol'kiperši promenjali svoi jubočki na inye mužepodobnye tualety. (*Vesti* 1912)
'The sweet captains, halfbacks and goalkeepers replaced their skirts with other masculine-looking garments.'

We have only 21 examples involving female captains, so it is impossible to draw firm conclusions. However, the data suggest that the meaning 'female captain' became more frequently used in the second half of the 20th century, and that it is still in use in the 21st century. Of the 21 examples in our dataset, 16 involve female captains in the military and the police, while 3 examples refer to captains of ships and 2 examples are about soccer teams. All the examples involving the military and the police are from the 1970s and later.

	Before 1850	1851-1900	1901-1950	1951-2000	2001-2015	Total
wife	29	52	31	9	2	123
female professional	0	0	5	13	3	21
Total	29	52	36	22	5	144

Table 6: Changes in the distribution of kapitanša *as 'captain's wife' vs. 'female captain' over time. Data from the RNC (main and newspaper subcorpora combined).*

3.3 Epithets

We now turn to the syntactic environments where feminitives occur. In order to investigate the stylistic and socio-linguistic aspects of feminitives, it would be necessary to read each example in a larger context and discuss its connotations. However, this is not feasible in a corpus-based analysis of the type we present in this article. We will therefore limit ourselves to the immediate context of the feminitives and focus on adjectives and other preposed modifiers, which we for the convenience of the reader will refer to as "epithets". Epithets are of particular interest since earlier studies have found that morphological patterns with strong connotations are often accompanied by other evaluative elements in the sentence (Beliaeva & Knoblock, 2019). As we will see, suffixed feminitives frequently combine with gender-related epithets that often characterize the appearance of the referent, while *pervyj* 'first' is the dominant epithet for *ženščina-X*.

We divide the epithets into three broad classes: "gender-related", "professional-related" and "other". The first group involves descriptions of the person as a woman, while the second focuses on the profession of the person in question. In the third group, we find those epithets that do not fit into either group one or two.

Examples of gender-related epithets include:

(9) a. *malen'kaja, legon'kaja advokatessa* 'small and ethereal female lawyer'
    b. *xrupen'kaja advokatessa* 'fragile female lawyer'

14

    c.   *obajatel'naja agentka* 'charming female agent'
    d.   *xorošen'kaja konsul'tantka* 'cute female consultant'
    e.   *paročka figuristyx instruktorš* 'a couple of curvy female instructors'
    f.   *s dlinnoj, suxoj, očen' nekrasivoj inženeršej* 'with a tall, severe and very unattractive female engineer'
    g.   *taraxtjaščaja, slovno pogremuška, advokatša* 'a female lawyer babbling like a rattlebox'
    h.   *xudye zlye inženerši* 'skinny nasty female engineers'

As the examples show, these epithets often characterize the appearance of the woman. Some epithets also describe the person's behavior or personality in a way that is associated with traditional stereotypes for women. In our, admittedly limited, dataset these epithets carry negative connotations, e.g. *taraxtjaščij* 'babbling' and *zloj* 'angry, nasty' in (9g-h). Notice that it is not always straightforward to draw a line between gender-related and professional-related epithets. In (9h), we have classified *zloj* as "gender-related", since it occurs together with *xudoj* 'skinny'. Taken together, the two epithets arguably create a negative stereotype of a skinny and nasty woman. Quite often in our dataset the gender-related stereotypes are invoked through several epithets that occur together, rather than by one single adjective, as in (9a, f and h).

Here are some epithets that focus on the profession, rather than the gender of the referent:

(10)   a.   *predpriimčivaja advokatša* 'industrious female lawyer'
    b.   *izvestnaja advokatessa* 'well-known female lawyer'
    c.   *otvažnaja agentša* 'brave female agent'
    d.   *aktivnaja trenerša* 'active female coach'
    e.   *talantlivejšaja režisserša* 'very talented female film director'
    f.   *tolkovaja instruktorša* 'sensible female instructor'
    i.   *provincial'naja inženerša* 'provincial female engineer'
    j.   *idiotka-režisserka* 'idiotic female film director'

As shown in (10), we include epithets that describe and evaluate the professional activity or reputation of the person as a professional, such as *predpriimčivyj* 'enterprising', *aktivnyj* 'active', *talantlivejšij* 'very talented' and *izvestnyj* 'well-known'. *Provincial'nyj* 'provincial' represents a borderline case; we have included it in the professional-related group since the adjective arguably suggests that the engineer in question is evaluated negatively as a professional since she comes from an unsophisticated part of the country. Finally, we included *idiotka* 'idiot' here, since this epithet involves a negative evaluation of the relevant person's professional achievements as a film director.

Here are some examples from the third group of epithets, which are neither gender-related, nor focus on the profession of the referent. This is a small, but heterogeneous class:

(11)   a.   *ustalaja advokatša* 'tired female lawyer'
       b.   *zamučennaja inženerša* 'exhausted female engineer'
       c.   *pervaja v mire kapitanša* 'first female captain in the world'

The epithets *ustalyj* 'tired' and *zamučennyj* 'exhausted' describe the state the relevant person is in without focusing on gender or profession. Notice that these epithets may occur in examples where the referent is evaluated as an efficient professional:

(12)   Ustalaja advokatša tolkovo ob''jasnila: esli prjamo sejčas zanesti sud'e dvadcatku, delo vernut na dosledovanie […]. (*RIA Novosti* 2005)
       'The tired female lawyer explained sensibly: if we give a twenty to the judge right now, the case will be returned for further investigation.'

Here, the adverb *tolkovo* 'sensibly' involves a positive evaluation of the person as a professional, but the epithet *ustalyj* is nevertheless not directly relevant for the profession of the person, and is therefore placed in the "other" category.

The epithet *pervyj* 'first' occupies a special position in our dataset. We have included it in the "other" category, since it does not directly describe the professional skills of the person in question. At the same time, sentences like the following emphasize the achievement of being the first female captain in the world:

(13)   Ustanovlen pamjatnik pervoj v mire ženščine-kapitanu. (*Izvestija* 2001)
       'A monument has been erected to commemorate the first female captain in the world.'

Table 7 shows the distribution of the three types of epithets for the four most frequent word-formation patterns under scrutiny, and the results are visualized in Figure 5. As shown, the distribution is quite similar for the three suffixes *-ša, -ka and -essa,* which often co-occur with gender-related epithets. *Ženščina-X* also has more examples with gender-related than professional-related epithets, although numbers are too small to facilitate strong conclusions. *Ženščina-X* differs from the suffixed feminitives, in that it displays a larger proportion of epithets that are neither gender-related, nor professional-related. Notably, the dominant epithet for *ženščina-X* is *pervyj* 'first': out of the 18 examples with *ženščina-X* and an epithet in the "other" group, 16 involve *pervyj*, as in (13) above. In sentences of this type, the focus is that the person in question is the first female in some profession. We speculate that *ženščina-X* is favored in such examples, since the word *ženščina* conveys that the relevant person is a woman in a more explicit way than feminitive suffixes.

As can be seen from Table 7, *-ša* is more often attested with epithets than the other word-formation patterns. This is expected, since *-ša* is by far the most frequent word-formation pattern in our dataset, as shown in section 3.1. It is also possible that *-ša* has a strong affinity to epithets, because *-ša* has been claimed to involve evaluative connotations (Rozental' et al., 1998, p. 200-201; Doleschal, 2015, p. 1168). However, Fufaeva (2020, p. 264) has shown that for *-ša* such connotations display considerable variation across native speakers of Russian.

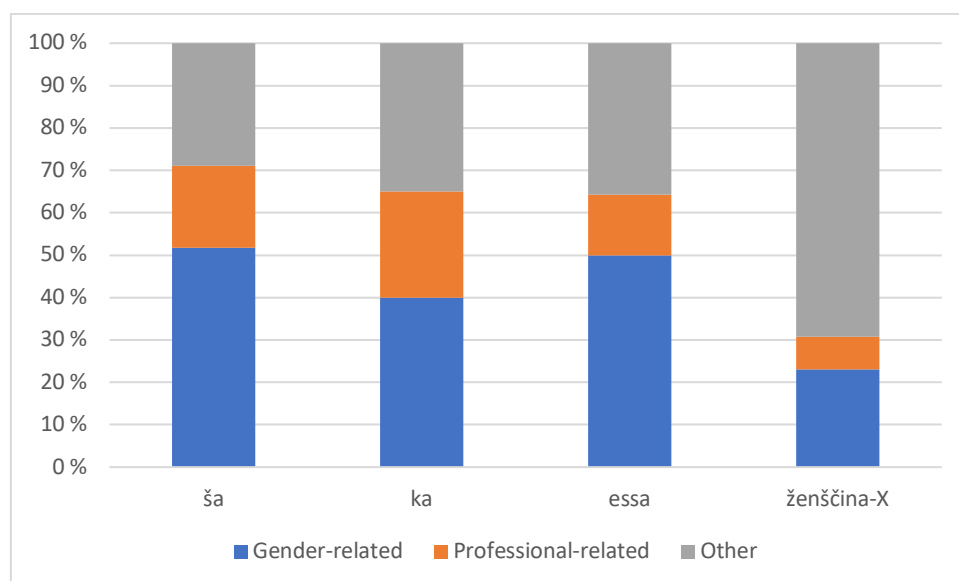|	-ša	-ka	-essa	ženščina-X	Total
Gender-related	42	8	7	6	63
Professional-related	16	5	2	1	24
Other	24	7	5	18	53
Total	82	19	14	25	140

Figure 5: Distribution of epithets for four word-formation patterns

Before we leave the epithets, a remark is in order. It would be helpful to have a baseline, against which we could compare the results for epithets reported above. For instance, it would be interesting to compare with epithets occurring with non-feminitives like *agent, advokat,* etc. when such words denote women. However, there is no straightforward way to extract such cases from the RNC, since these words very often refer to males. An anonymous referee suggested limiting the searches to cases with an adjective in the feminine gender followed by the relevant nouns. However, such searches returned relatively few examples and considerable "noise" (irrelevant examples). Instead, we chose to search for collocations in CoCoCo ("Collocation, Colligations, Corpora", Kopotev et al., 2015).[16] Here, it is possible to identify collocations of adjectives followed by nouns such as *agent, advokat,* etc. For all the ten nouns we investigate in the present study, CoCoCo returned a total of 97 epithets. Two observations can be made. First, the majority were professional, such as *glavnyj inžener* 'chief engineer' and *finansovyj konsul'tant* 'financial consultant', while no epithets were gender-related. Second, almost all epithets were neutral adjectives that did not involve any evaluation of the referent. Rare exceptions include *kruglen'kij kapitan* 'rotund captain' and *ozabočennyj kapitan* 'anxious captain'. While a more detailed investigation of epithets for non-feminitives must be left for future research, our findings suggest that feminitives show an affinity to gender-related and evaluative epithets.

4. Concluding remarks

Our empirical analysis of data from the Araneum Corpus and the RNC has enabled us to draw some conclusions about the use of feminitives in Russian. The investigation of the Araneum corpus showed that the choice of suffix to some extent depends on the morphophonological properties of the stem. In particular, the suffixes are sensitive to whether the stem ends in an obstruent or sonorant consonant, and to the stress pattern of the word. The feminitive suffix *-ka* is preferred after obstruents, while *-ša* dominates after sonorants, especially in words with non-final stress. There are also morphological generalizations involved, insofar as professions with the suffixes *-tel'* and *-nik* prefer feminitives in *–(n)ica.* We summarized the generalizations

---

[16] CoCoCo is available at https://cococo.cosyco.ru.

in a flowchart that shows the interaction of the morphophonological and morphological factors under scrutiny. While the corpus data show clear patterns, we emphasize that our generalizations capture statistical tendencies, not categorical rules.

Our analysis of the RNC focused on ten professions which are frequent and combine with several types of feminitives. This analysis lends further support to the conclusions above. In addition, we showed that in our dataset the suffixed word-formation patterns are more frequent than the compounds of the type *ženščina-X* and *X-ženščina*. In particular, the *X-ženščina* type is marginal in our dataset. We scrutinized data from the newspaper and main subcorpora of the RNC, and found that the distributions in the two subcorpora were very similar, suggesting that with regard to the use of feminitives, journalistic texts of the kinds represented in the newspaper subcorpus do not stand out as different from other texts in the RNC.

A diachronic analysis of the data from the RNC showed that the distribution of the relevant word-formation patterns has changed over time. While *-ša* has been widely used for the nouns under scrutiny since the 19th century, the other types appear on the scene somewhat later. The *ženščina-X* and *X-ženščina* types are in use from the first half of the 20th century, while *-ka* is widely used from the middle of the 20th century. The *-essa* suffix is frequently used about female lawyers in the 21st century.

A case study of *kapitanša* indicates a gradual change from the meaning 'captain's wife' to 'female captain'. Early attestations of the latter meaning come from the beginning of the 20th century, but with the appearance of more female captains in society, the *-ša* suffix has come to denote female professionals. However, in texts describing historical realia, *kapitanša* is still used about captain's wives even in the 21st century.

We explored the use of epithets for the word-formation patterns under scrutiny. In our dataset, all patterns combine with gender-related epithets more frequently than with professional-related epithets. The *ženščina-X* type is also frequently attested with the epithet *pervyj* 'first' (e.g., *pervaja ženščina-advokat* 'first female lawyer').

In general, our study of data from the two corpora has shown that the choice of word-formation pattern for feminitives is sensitive to linguistic factors, such as the morphological and morphophonological properties of the stem. We acknowledge that the linguistic factors we have explored in the present study may be overridden by native speakers' conscious choices, which may be based on ideology and other non-linguistic factors. The study of the interaction of linguistic and non-linguistic factors deserves further investigation. However, this interesting problem will have to be left open for future research.

Statement about conflict of interest
We are not aware of any conflicts of interest.

References

Arkhangelskaya, T. A. (2014). *Simmetrija i asimmetrija grammatičeskoj kategorii roda i semantičeskoj kategorii pola v slavjanskix jazykax*. PhD dissertation, Olomouc.

Badanina, I. V. (2017). Funkcionirovanie feminativov v jazyke interneta. In A. V. Dolžnikova & V. V. Barabaš (eds.), *Russkij jazyk v internete: ličnost', obščestvo, kommunikacija, kul'tura: Sbornik statej i mežduna-rodnoj naučno-praktičeskoj konferencii*. Moscow: Peoples' Friendship University of Russia.

Beliaeva, N. & Knoblock, N. (2019). Conceptual Blending in the Construction of Morphological Blends. Paper presented at the *International Cognitive Linguistics Conference* (ICLC-15), Kwansei Gakuin University, Nishinomiya, Japan, August 6-11, 2019.

Bobylëva, M. (2021). *My tak govorim. Obidnye slova i kak ix izbežat'*. Moscow: AST.

Doleschal, U. (2015). Gender marking. In: *Word-Formation: An International Handbook of the Languages of Europe*, 1159-1171.

Fufaeva, I. (2018). Tkaxa, blogerša i novaja glava goroda. Počemu feminitivy byli, est' i budut ograničennoj čast'ju russkogo jazyka. *Nož*, December 27, 2018: https://knife.media/word-formation-difficulties/

Fufaeva, I. (2020). *Kak nazyvajutsja žensčiny. Feminitivy: istorija, ustrojstvo, konkurencija.* Moscow: Korpus.

Georgievskaja, E.N. (2018). Komu ugrožajut feminitivy? Russian feminist association *ONA*, February 8, 2018: https://ona.org.ru/post/170655867278/linguistic-menace

Kirey-Sitnikova, Y. (2021). Prospects and challenges of gender neutralization in Russian. *Russian Linguistics*: https://link.springer.com/article/10.1007/s11185-021-09241-6

Kopotev, M. Escoter, L., Kormacheva, D., Pierce, M., Pivovarova, L., & Yangarber, R. (2015). CoCoCo: Online Extraction of Russian Multiword Expressions. *The 5th Workshop on Balto-Slavic Natural Language Processing* (10–11 September 2015, Hissar, Bulgaria). Sofia: INCOMA Ltd, 2015. pp. 43-45.

Krongauz M. & Severin A. (2019). Feminitivy ukorenjatsja, esli izmenitsja sama žizn'. *Kul'tura*, March 1, 2019: https://portal-kultura.ru/articles/kulturnaya-politika/236355-maksim-krongauz-feminitivy-ukorenyatsya-esli-izmenitsya-sama-nasha-zhizn/

Lopatin, V. V. & Uluxanov, I. S. (2016). *Slovar' slovoobrazovatel'nyx affiksov sovremennogo russkogo jazyka*. Moscow: Azbukovnik.

Magomedova, V. D. & Slioussar, N. A. (2021). Gender and case in Russian nouns denoting professions and social roles. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2021" Moscow, June 16–19, 2021*.

Morozova, K. (2018). Slovo *avtorka* protivoestestvenno: filologi o tom, počemu feminitivy vsex besjat. *Sobaka.ru*, April 12, 2018: https://www.sobaka.ru/city/science/71329

Mozdzierz, B. M. (1999). The rule of feminization in Russian. In Margaret Mills (ed.), *Slavic Gender Linguistics*, 165-182. Amsterdam/Philadelphia: Benjamins.

Nesset, T. & Janda, L. A. (2010). Paradigm structure: Evidence from Russian suffix shift. *Cognitive Linguistics* 21(4), 699-725.

Piperski, A. Ch. (2018). Juristka i juristec: Lingvist o buduščem feminitivov i izmenenijax v jazyke. *Wonderzine*, March 29, 2018: http://www.wonderzine.com/wonderzine/life/life-interview/234057-alexander-piperski

Piperski, A. Ch. (2019). Kovarnye suffiksy. Kak sovremennye feminitivy menjajut russkij jazyk. https://nplus1.ru/material/2019/11/28/russian-feminitives

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Rozental', D.È., Džandžakova, E.V. & Kabanova, N.P. (1998). *Spravočnik po russkomu jazyku*. Moscow: CheRo.

Simpson, E. H. (1949). Measurement of diversity. *Nature* 163, 688.

Švedova, N. Ju et al. (eds.) (1980). *Russkaja grammatika*, vol. I. Moscow: Akademija nauk SSSR.

Townsend, C. E. (1968). *Russian Word-Formation*, New York: McGraw-Hill.

Uhlik, M. (2019). Izražanje spolske korelacije pri položajnih poimenovanjih v ruščini in slovenščini. *Slavisticna revija* 67 (2), 313–324.