Gender variation and gender markedness in Russian nouns

Varvara Magomedova
HSE University, Moscow, Russia; varya.magomedova@gmail.com

Natalia Slioussar
HSE University, Moscow, Russia; Saint Petersburg State University, St. Petersburg, Russia; slioussar@gmail.com

Abstract: In this paper, we analyze different factors affecting gender assignment in Russian nouns. We conducted two studies combining experimental, dictionary, and corpus data (relying both on traditional corpora and on the web-as-a-corpus approach). The first study is dedicated to gender assignment to expressive nouns in a situation when the gender of the base noun and the expressive suffix give contradictory cues. We demonstrated extensive gender variation in these cases and estimated the role of different factors affecting it. The gender of the base noun and the inflection associated with the expressive suffix appeared to be the most important. Animacy is also relevant (certain options are possible only for nouns denoting people) but statistically insignificant. The second study is on gender assignment to nouns that have a zero inflection in nominative singular and a palatalized, alveolo-palatal, or post-alveolar stem-final consonant. In the previous studies, it was noted that the phonological properties of this consonant affect the gender distribution in these nouns. Our data confirmed this generalization and provided several new insights. We also demonstrated that native Russian speakers are sensitive to it. We discuss the implications of our results for different approaches in theoretical morphology and focus on gender markedness analyzing a special status of masculine gender.

Keywords: augmentative, diminutive, experiment, expressivity, gender, Russian, variation
Acknowledgements: The work on this project was carried out within the Basic Research Program at the HSE University, Russia.
For citation: Magomedova V., Slioussar N. Gender variation and gender markedness in Russian nouns. *Voprosy Jazykoznanija*, 2023, 2: 7–28.
DOI: 10.31857/0373-658X.2023.2.7-28

Вариативность и маркированность грамматического рода у русских существительных

Варвара Дмитриевна Магомедова
Национальный исследовательский университет «Высшая школа экономики», Москва, Россия; varya.magomedova@gmail.com

Наталия Анатольевна Слюсарь
Национальный исследовательский университет «Высшая школа экономики», Москва, Россия; Санкт-Петербургский государственный университет, Санкт-Петербург, Россия; slioussar@gmail.com

Аннотация: В этой статье мы исследуем различные факторы, влияющие на род существительных в русском языке, и предлагаем новый теоретический анализ маркированности рода. Статья опирается на экспериментальные, словарные и корпусные данные. Первое исследование

посвящено определению рода у существительных субъективной оценки в ситуации, когда род исходного существительного и экспрессивный суффикс дают противоречивые подсказки. Мы показали, что в этой ситуации наблюдается значительная вариация по роду, а также оценили роль различных факторов, влияющих на распределение. Наиболее важными оказались род исходного существительного и тип склонения, связанный с суффиксом субъективной оценки. Одушевленность также важна (некоторые варианты возможны только для существительных, обозначающих людей), но статистически незначима. Второе исследование посвящено определению рода у существительных, имеющих нулевое окончание в именительном падеже единственного числа и оканчивающихся на мягкую или шипящую согласную. В предыдущих исследованиях было отмечено, что конечная согласная основы влияет на распределение родов у этих существительных. Наши данные подтвердили это заключение и позволили сделать несколько новых выводов. Мы также показали, что носители русского языка чувствительны к этим признакам. Мы обсуждаем значение полученных результатов для различных подходов в теоретической морфологии, уделяя особое внимание проблеме маркированности рода и особому статусу мужского рода в русском языке.

Ключевые слова: аугментатив, вариативность, диминутив, род, русский язык, эксперимент, экспрессивность

Благодарности: В данной работе использованы результаты проекта, выполненного в рамках Программы фундаментальных исследований НИУ ВШЭ.

Для цитирования: Magomedova V., Slioussar N. Gender variation and gender markedness in Russian nouns. *Voprosy Jazykoznanija*, 2023, 2: 7–28.

DOI: 10.31857/0373-658X.2023.2.7-28

1. Introduction

1.1. The structure of the paper

Russian has three grammatical genders: masculine, feminine, and neuter. Gender in Russian has been at the center of several theoretical debates. Firstly, feature markedness is a matter of controversy. Masculine is the most frequent, while neuter is the least frequent, but it is used in impersonal constructions, where the unmarked form is expected. Some experimental studies point to neuter as the unmarked option, while many others suggest that masculine is unmarked. Theoretical accounts are similarly split. Secondly, many nouns exhibit systematic gender variation, challenging different approaches to gender assignment. In particular, it is intuitively clear that complex interactions can be found between the gender of the noun and its inflectional and derivational affixes and even its phonological form. But it is not clear how to account for them.

In this paper, we review different aspects of these interactions and try to account for the prima facie contradictory observations mentioned above. We conducted two studies combining experimental, dictionary, and corpus data (relying both on traditional corpora and the web-as-a-corpus approach). The first study is dedicated to gender assignment to expressive nouns in a situation when the gender of the base noun and the expressive suffix give contradictory cues, and the second one to gender assignment to nonce nouns depending on the final consonant of the stem. Only nominative forms are discussed — we plan to analyze case-related differences in our further studies.

The paper is organized as follows. Firstly, we outline the system of nominal inflectional classes in Russian and their connections to grammatical gender, focusing in more detail on expressive nouns. Secondly, we present different approaches to feature markedness. After that, we turn to the two studies we conducted and finally discuss their implications.

1.2. Gender and declensions in Russian

Russian nouns are distributed between three genders: masculine (M), feminine (F), and neuter (N), and are inflected for six cases and two numbers. In most cases, the gender of the noun cannot be unambiguously determined from its inflectional affix, only from agreeing adjectives, participles, and verb forms. However, as we show below, there is a conspicuous connection between gender and inflectional affixes, which was analyzed in different theoretic frameworks, e.g. [Asarina 2009; Caha 2021; Corbett, Fraser 2000; Kramer 2015; Rice 2005].

Nouns have different sets of inflections depending on the inflectional class, or declension, they belong to.[1] Traditional reference grammars (e.g., [Russian grammar 1980]), as well as many other studies (e.g., [Aronoff 1994; Halle 1994]), identify three declensions with several subclasses and various exceptions. Alternative approaches either divide the first declension into two classes with masculine and neuter nouns respectively and therefore identify four declensions [Corbett, Fraser 1993; Müller 2004], or make a primary distinction between the 1st and 2nd ('core') declensions on the one hand and the less frequent 3rd declension on the other hand [Zaliznyak 1987; Wiese 2008].

Our findings do not allow teasing apart these approaches. Therefore, we will rely on the model with four declensions outlined in Table 1 for the sake of convenience. Classes I and IV have different inflections only in the nominative and accusative singular (the reason why many researchers unite them).

Table 1

A system of four inflectional classes for Russian nouns

Classes	Descriptions	Examples	Percentage of nouns in the RNC[2]
class I	M nouns ending in a consonant in nominative singular (with a zero inflection)	*zakon* 'law', *kon'* 'horse' (C/C'-final nouns[3])	46%
class II	F nouns ending in *-(j)a* in nominative singular	*komnata* 'room', *zemlja* 'earth'	29%
	M nouns (only denoting people) ending in *-(j)a* in nominative singular	*papa* 'dad', *djadja* 'uncle'	1%
class III	F nouns ending in a consonant in nominative singular (with a zero inflection)	*kost'* 'bone' (C'-final nouns)	5%
class IV	N nouns ending in *-o* or *-e* in nominative singular	*okno* 'window', *more* 'sea'	18%
indeclinable	nouns of different genders	*kivi* 'kiwi', *pal'to* 'coat'	1%

Table 1 shows that declensions are closely related to gender. It goes from declension to gender, but it is also illustrative to go from gender to declension. 97% of masculine nouns have zero

[1] All declensions have the same inflections in plural, so the discussion of declension-related differences below only concerns singular.

[2] Percentages of nouns in the Russian National Corpus, or RNC, are taken from [Slioussar, Samoilova 2015]. Their counts were based on the grammatically disambiguated subcorpus and did not take substantivized adjectives into account. There is also a very small number of exceptional cases with irregular inflection.

[3] Here and below we use the term 'C'-final nouns' for the nouns that end in a palatalized, alveolo-palatal, or postalveolar consonant, and the term 'C-final nouns' for the nouns that end in other consonants.

inflection in nominative singular, 85% of feminine nouns end in *-(j)a*, and almost 100% of neuter nouns end in *-o*/*e* (like in Table 1, these counts are based on [Slioussar, Samoilova 2015]).

The connection between gender and declension is so strong that it leads to indeclinability of some nouns, gives rise to defective paradigms and gender variation. For example, all declinable C-final nouns belong to the declension class I and are masculine. Russian has many feminine proper names (first and last names, toponyms etc.) that are C-final, but they are all indeclinable as a result.

A different pattern is found in the nouns denoting professions and social roles. Historically, most of them are masculine and belong to class I. When women gained access to these roles and professions, new feminine words were derived from a number of these nouns, but by far not from all of them. At the same time, it became possible to use masculine nouns for women, either with masculine or with feminine agreement. Notably, using feminine agreement is grammatical only in the nominative case [Graudina et al. 1976; Zaliznyak 2002], although some exceptions are found in real usage [Sitchinava 2011]. Thus, these words did not become proper common gender nouns like *plaksa* 'crybaby$_{M/F}$', for which both masculine and feminine agreement is always possible — apparently, because class I contains both feminine and masculine nouns. As a result, Zaliznyak [2002] suggests treating these cases as two separate nouns, a masculine one and a feminine one with a severely defective paradigm.

There is a vast literature on these nouns: e.g., [Panov 1968; Muchnik 1971; Graudina et al. 1976; Corbett 1982; 1991; Asarina 2009; Pesetsky 2013; Steriopolo, Wiltschko 2010; Lyutikova 2015; Steriopolo 2019; Matushansky 2013; Caha 2019; Privizentseva, to appear]. In [Slioussar, Magomedova, to appear], we also look at these nouns in more detail using corpus and experimental data. We argue that the availability of feminine agreement in the nominative case relies on analogy: Russian has other feminine nouns with a zero inflection in the nominative singular (namely, class III nouns), but all oblique case singular inflections do not coincide in these classes.

Another problematic group are nouns with diminutive and augmentative suffixes. These suffixes are not supposed to change the gender of the base noun (see [Vinogradov 1947]), but as a result, we get many combinations that do not fit the picture outlined in Table 1. This leads to extensive variation. Studying this variation is the goal of this paper, so we discuss expressive nouns in a separate subsection.

Finally, indeclinable nouns are interesting for the following reasons. Firstly, neuter gender is chosen by default for them, although otherwise masculine is the most frequent. Secondly, their gender may be affected both by semantic analogy (the gender of a declinable noun with a similar meaning) and by their phonology: e.g., nouns ending in *-(j)a* are more likely to be feminine, although in this case, *-(j)a* is not an inflectional affix, but a part of the root. This leads to an even greater gender variation than in expressive nouns, and we study it in a separate paper [Chuprinko et al., to appear, a; b].

Apart from expressive nouns, our paper focuses on the following problem. As Table 1 shows, C'-final nouns can be masculine or feminine (the same is true for nouns ending in *-(j)a*, but they are feminine almost 30 times more often than masculine). In our second study, we show that the distribution of C'-final nouns with different stem-final consonants in two genders is not equal and that native speakers of Russian are sensitive to that.

1.3. Nouns with diminutive and augmentative affixes

Russian has a great variety of diminutive and augmentative affixes. Normally, nominal derivational affixes are associated with a particular set of inflectional affixes, which define the gender of the derived word:

(1) *grib-*∅ 'mushroom$_M$' → *grib-nic-a* 'mycelium$_F$', *grib-nik-*∅ 'mushroom collector$_M$'

The picture with expressive affixes is more complex. A number of examples are given in Table 2. Firstly, some affixes are associated with two sets of inflections. Secondly, diminutive and augmentative

nouns may keep the gender of the base noun (henceforth, we will call it the 'base gender').[4] Since we study gender variation in these nouns, we do not indicate their gender in Table 2 and elsewhere — only when corpus or experimental examples with particular gender agreement are analyzed.

Table 2

Some diminutive and augmentative affixes, inflections associated with them and declensions associated with these inflections

Expressive suffixes	Declension	Examples
Diminutive: *-ok-∅*, *-ek-∅*, *-ik-∅*, *-čik-∅*	class I (M)	*stul'čik* 'chair$_{DIM}$' from *stul* 'chair$_{M}$' *vinčik* 'wine$_{DIM}$' from *vino* 'wine$_{N}$'
Diminutive: *-k-a*, *-onk-a*, *-en'k-a* Augmentative: *-in-a*, *-(j)ug-a*, *-(j)ar-a*	class II (predominantly F)	*knižonka* 'book$_{DIM}$' from *kniga* 'book$_{F}$' *zverjuga* 'animal$_{AUG}$' from *zver'* 'animal$_{M}$'
Diminutive: *-išk-a*/*o*, *-yšk-a*/*o*, *-(j)ušk-a*/*o* Augmentative: *-išč-a*/*e*	class II (predominantly F) or IV (N)	*rybišča*/*e* 'fish$_{AUG}$' from *ryba* 'fish$_{F}$' *kotiška*/*o* 'cat$_{DIM}$' from *kot* 'cat$_{M}$'

Many researchers claimed that expressive suffixes do not change the base gender in Russian [Vinogradov 1947; Corbett 1982; Hippisley 1996; Rice 2005]. Several more recent papers noted that in real usage, variation is observed [Doleschal 2000; Savchuk 2011; Sitchinava 2011], but did not explore how extensive this variation is or which factors are relevant. Our paper aims to fill this gap.[5]

Expressive suffixes were also analyzed in many works by Olga Steriopolo [Steriopolo 2008; 2015; 2017; Steriopolo et al. 2021]. Initially, she argued that some suffixes always change the base gender, while the others never do, and developed an account in the Distributed Morphology framework. In [Steriopolo et al. 2021], she argues that suffixes with emotive meaning always change the base gender, while suffixes with size meaning never do. In [Steriopolo 2017], she admits the existence of some variation, but only for animate nouns, and attempts to explain it with underspecification of a gender feature for common gender nouns. Our data do not confirm these hypotheses: extensive gender variation was found both for animate and for inanimate nouns, and it did not correlate with the meaning of the suffix.

1.4. Markedness

Asymmetries between different gender or number features are usually explained in terms of feature markedness. However, the very definition of markedness depends on the theoretical framework (see, e.g., [Haspelmath 2006] for an overview). Structural approaches

[4] Expressive suffixes are often considered to be between inflectional and derivational, as they do not create a word with a new meaning, only add a nuance [Booij 2000]. We will call them derivational in this paper, but our data do not bear on this question.

[5] Notably, some researchers claim that in case of affixes like *-išč-a*/*e*, the *-a* variant is used with feminine base nouns, and the *-e* version with masculine and neuter base nouns — see [Russian grammar 1980, I: 213; A reference grammar of Russian 2004: 146; Privizentseva, to appear]. This would be logical, but, paradoxically, language defies this simple logic. Naturally occurring data we collected show that this is definitely not the case: e.g., *mašinišče* 'car$_{AUG}$' from *mašina* 'car$_{F}$' or *domiška* 'house$_{DIM}$' from *dom* 'house$_{M}$' can be easily found, more examples are provided in Table 2 and below. However, we did not analyze how frequently different variants are selected depending on the base noun — we plan to come back to this problem in our further research, also taking case into acount. In this study, we took various expressive nouns as a starting point and analyzed their gender agreement options.

rely on the so-called representational markedness: a [+a] feature value is more marked than a [–a] value, and absence of a feature is usually assumed to be the least marked option. As a result, the least marked (or 'unmarked') forms are expected to appear in structures where no gender feature is available (e.g., in impersonal constructions such as *Stemnelo* 'It got dark'). In case of derived nouns, structural approaches usually predict that some affixes will always change the gender of the base noun, while the others never will, are not well suited to describe variation and definitely do not predict that the unmarked option should be the most frequent.

In competition-based theories, like Optimality Theory (OT), the unmarked gender is the default, which is used when there are no specific requirements for gender assignment. It is expected to be the most frequent and the most productive. A similar approach to markedness is adopted in many functional and typological studies. The classic version of OT does not describe variation, but OT-based probabilistic theories, like Harmonic Grammars and Maximum Entropy models, are specifically geared to account for it. In any case, the unmarked option is expected to have higher chances to be selected — also in case of gender of derived nouns.

Very often, these approaches point to the same feature value as unmarked: for example, to the singular value of the number feature. However, this is not the case for Russian gender. Masculine is the most frequent in Russian, while neuter is the least frequent (see Table 1), but neuter is used in impersonal constructions.

Structural approaches consider neuter to be unmarked in Russian (e.g., [Kramer 2015; Matushansky 2015; Nevins 2011]) and encode this intuition in the representation of gender features. For example, for Nevins, feminine is [+FEM], [−MASC], masculine is [−FEM], [+MASC] and neuter is [−FEM], [−MASC]. Kramer encodes feminine as [+FEM], masculine as [−FEM] and neuter corresponds to no gender features. Optimality-theoretic approaches consider masculine to be unmarked (e.g., [Rice 2005]), while neuter is considered to be the most marked. This view is shared by many other authors from different frameworks, e.g., [Jakobson 1960]. Finally, some authors consider masculine to be unmarked on the word level and neuter to be unmarked on the sentence level [Corbett, Fraser 2000].

Several experimental studies looking at agreement processing in Russian found that masculine behaves differently from the two other genders [Akhutina et al. 1999; 2001; Romanova, Gor 2017; Slioussar 2018]. Slioussar and Malko [2016], who studied gender agreement attraction[6] in production and in comprehension, found that neuter behaves as unmarked in the former case and masculine in the latter. A series of earlier similar production experiments on Slovak [Badecker, Kuminiak 2007], in which the gender system is similar to Russian, revealed the same picture: neuter behaved as unmarked.

We think that there is no 'correct' approach to the problem of gender markedness. Masculine is special in one sense, and neuter in some other. We will formulate our view on this problem in the discussion section.

2. Study 1: Gender assignment in expressive nouns

In our study of expressive nouns, we could not use traditional corpora because the variation we are interested in is usually beyond the literary norm, and corpora mostly contain edited

[6] Attraction errors (such as the number error in the example *The key to the cabinets are rusty*, in which the dependent noun, an attractor, disrupts subject-predicate agreement) were demonstrated to occur more frequently in production and to be missed more easily in comprehension in various languages. It was also noted that the effect depends on the feature values of the attractor and the head: for example, when the head is plural and the attractor is singular, no attraction is observed. This made attraction important for the study of feature markedness.

texts. Therefore, we used the web-as-a-corpus approach and searched the internet for different examples for which the base gender and inflectional affixes give conflicting cues, to see which gender agreement is possible for them. After that, we conducted an experiment with a subset of such nouns to find out which derivate gender is preferred by native Russian speakers. This subset was balanced according to several factors that we identified as relevant in the web-as-a-corpus study — the base gender, the inflection of the derivate, and its animacy, — so we could estimate their influence numerically.

2.1. Web-as-a-corpus data

Compiling a complete list of Russian expressive nouns is not feasible, so providing numerical data (such as the share of nouns with a given combination of gender assignment cues that show masculine agreement) is difficult. Therefore, the goal of this part of the study was mostly to establish what is possible and frequent and what is not: which combinations of the base gender, inflections, and derivate gender are easily found on the web and which can hardly be found. We used the Google search engine (www.google.com).

We took the animacy scale into account, dividing our examples into three groups: denoting people, denoting animals, and inanimate. The gender feature is semantically loaded for the first group and semantically empty for the third one, while the situation in the second group is complicated, so this factor may be relevant for derivate gender assignment. We did not observe any correlations between gender change and the meaning of the suffix (diminutive vs. augmentative, emotive vs. size-related).

Our first general observation was that for masculine and feminine base gender, all combinations of inflections and derivate gender can easily be found — examples are provided below. However, nouns with neuter base gender almost never select expressive affixes associated with class I or class II inflections. We could find only several examples like *jajco* 'egg$_{N}$' — *jajčik* 'egg$_{DIM}$', *pivo* 'beer$_{N}$' — *pivčik* 'beer$_{DIM}$' or *odejalo* 'blanket$_{N}$' — *odejalka* 'blanket$_{DIM}$'. These examples showed masculine or feminine agreement, respectively, which can be explained by their class I and class II inflections, but, since they are very few and very infrequent, it is hard to make any conclusions based on them.

Nouns with masculine base gender usually keep their base gender in case of conflicting cues, but many examples of gender change can also be found. Below, examples with different inflections and occupying different positions on the animacy scale are provided.

(2) Masculine base gender + inflection *-a* (class II) = masculine agreement

a. inanimate

Da	*i*	***pidžačonka***	***takoj***	***byl…***	*nu*	*takoj*	*prjam :)*
and	also	jacket$_{DIM.NOM.SG}$	such$_{M.NOM.SG}$	was$_{M.SG}$	well	such$_{M.NOM.SG}$	just

'And the jacket$_{DIM}$ was also like… well, just like, you know.'

b. animal

Vidal?	***Dikij***	***zverjuga!***
saw$_{M.SG}$	wild$_{M.NOM.SG}$	animal$_{AUG.NOM.SG}$

'[You] saw [it]? A wild beast$_{AUG}$!'

c. human

Ne	*kapitan*	*s*	*gvozdem,*	***praporiška***	***nesčastnyj***.
not	captain$_{NOM.SG}$	with	nail$_{INS.SG}$	ensign$_{DIM.NOM.SG}$	pitiful$_{M.NOM.SG}$

'Not a captain with a nail, but a pitiful ensign$_{DIM}$.'

(3) Masculine base gender + inflection *-a* (class II) = feminine agreement

a. inanimate

Mal'čik *milyj,* *tol'ko* ***pidžačonka*** ***kvelaja***.
boy$_{NOM.SG}$ nice$_{M.NOM.SG}$ just jacket$_{DIM.NOM.SG}$ tacky$_{F.NOM.SG}$
'The boy is nice, just his jacket$_{DIM}$ is tacky.'

b. animal

Moj *pës —* ***zverjuga*** *na redkost'* ***spokojnaja***.
my$_{M.NOM.SG}$ dog$_{NOM.SG}$ animal$_{AUG.NOM.SG}$ rarely calm$_{F.NOM.SG}$
'My dog is a surprisingly calm animal$_{AUG}$.'

c. human

Praporiška *ty* ***naša,*** *ty* *prosto* *njaška.*
ensign$_{DIM.NOM.SG}$ you our$_{F.NOM.SG}$ you simply sweetie$_{NOM.SG}$
'You [little] ensign$_{DIM}$, you are such a sweetie.'

(4) Masculine base gender + inflection *-o*/*e* (class IV) = masculine agreement

a. inanimate

Umišče-to ***kakoj*** *u* *našego* *brata* *mužika!*
mind$_{AUG.NOM.SG}$-PART what$_{M.NOM.SG}$ at our$_{M.GEN.SG}$ brother$_{GEN.SG}$ man$_{GEN.SG}$
'How clever we are, simple men! Lit.: What a mind$_{AUG}$ our brother simple man has!'

b. animal

Bezumno ***krasivyj*** *i* ***obajatel'nyj*** ***kot'iško!***
insanely beautiful$_{M.NOM.SG}$ and charming$_{M.NOM.SG}$ cat$_{DIM.NOM.SG}$
'An incredibly beautiful and charming kitty$_{DIM}$!'

c. human

Voobšče *etot* *Lučnikov* ***naivnyj*** ***idiotiško*** *okazalsja.*
generally this$_{M.NOM.SG}$ Luchnikov$_{NOM.SG}$ naïve$_{M.NOM.SG}$ idiot$_{DIM.NOM.SG}$ turned_out$_{M.SG}$
'Generally, that Luchnikov turned out to be a naïve idiot$_{DIM}$.'

(5) Masculine base gender + inflection *-o*/*e* (class IV) = neuter agreement

a. inanimate

Eto *že* ***umišče*** ***kakoe*** *nado* *imet'* ***fašistskoe****…*
this PART mind$_{DIM.ACC.SG}$ what$_{N.ACC.SG}$ must have$_{INF}$ fascist$_{ADJ.N.ACC.SG}$
'What a fascist mind one must have (to come to this)…'

b. animal

Vot ***moë*** ***kotiško*** *)))*
here my$_{N.NOM.SG}$ cat$_{DIM.NOM.SG}$
'Here is my kitty$_{DIM}$.'

c. human

Tupoe ***švedskoe*** ***idiotiško****!*
stupid$_{N.NOM.SG}$ Swedish$_{N.NOM.SG}$ idiot$_{DIM.NOM.SG}$
'A stupid Swedish idiot$_{DIM}$!'

Nouns with feminine base gender change it much more often than the ones with masculine base gender, but, as (6)–(9) show, examples of both types (with changed and unchanged gender) can still be found. As far as we can judge, the only exceptions are C-final derivates. They always change their gender to masculine if they are inanimate or denote animals, feminine agreement can be found only with nouns denoting humans. A similar picture can be found in the very few examples with neuter base gender and class I inflections (e.g., *jajco* 'egg$_{N}$' — *jajčik* 'egg$_{DIM}$').

(6) Feminine base gender + inflection -∅ (class I) = feminine agreement
a. inanimate: no examples found

b. animal: no examples found

c. human
Mamčik ***pozvonila*** *v* *školu..*
mom$_{DIM.NOM.SG}$ called$_{F.SG}$ to school$_{ACC.SG}$
'Mommy$_{DIM}$ called the school…'

(7) Feminine base gender + inflection -*o*/*e* (class IV) = feminine agreement
a. inanimate
Vooruženny *ljudi* *v* *forme* *i* ***ogromnaja*** ***mašinišče***.
armed$_{NOM.PL}$ people$_{NOM.PL}$ in uniform$_{LOC.SG}$ and giant$_{F.NOM.SG}$ car$_{AUG.NOM.SG}$
'Armed people in uniform and a giant car$_{AUG}$.'

b. animal
Mgnov'enno *rybki* *isčezli* *i* ***pokazalas'*** ***rybišče***.
immediately fishes$_{DIM.NOM.PL}$ disappeared$_{PL}$ and appeared$_{F.SG}$ fish$_{AUG.NOM.SG}$
'Immediately the [little] fishes$_{DIM}$ disappeared, and a [huge] fish$_{AUG}$ appeared.'

c. human
Tam ***moja*** ***babuško*** *i* *deduško.*
there my$_{F.NOM.SG}$ grandma$_{DIM.NOM.SG}$ and grandpa$_{DIM.NOM.SG}$
'My grandma$_{DIM}$ and grandpa$_{DIM}$ are there.'

(8) Feminine base gender + inflection -∅ (class I) = masculine agreement
a. inanimate
Vkusnyj ***bulčik***
yummy$_{M.NOM.SG}$ bun$_{DIM.NOM.SG}$
'A yummy bun$_{DIM}$ (the title of a YouTube video).'

b. animal
Segodnja ***prixodil*** ***milen'kij*** ***sobačok***.
today came$_{M.SG}$ cute$_{M.NOM.SG}$ dog$_{DIM.NOM.SG}$
'A cute doggy$_{DIM}$ came today.'

c. human
Mamčik ***prišël*** — *obradovalas'* *dočka.*
mom$_{DIM.NOM.SG}$ came$_{M.SG}$ cheered$_{F.SG}$ daughter$_{NOM.SG}$
' "Mommy$_{DIM}$ came," — cheered up the daughter.'

(9) Feminine base gender + inflection -*o*/*e* (class IV) = neuter agreement
a. inanimate
Krutoe ***mašinišče,*** ***blestjavoe***!
cool$_{N.NOM.SG}$ car$_{AUG.NOM.SG}$ shiny$_{N.NOM.SG}$
'It's a cool, shiny car$_{AUG}$!'

b. animal
A *čto* *eto* *za* ***rybišče*** ***takoe*** ***interesnoe***?
and what this for fish$_{AUG.NOM.SG}$ such$_{N.NOM.SG}$ intresting$_{N.NOM.SG}$
'And what an interesting fish$_{AUG}$ is this?'

c. human
Roždestvenskaja *istorija* *pro* ***omskoe*** ***babuško***.
Christmas$_{ADJ.F.NOM.SG}$ story$_{NOM.SG}$ about Omsk$_{ADJ.N.ACC.SG}$ grandma$_{DIM.ACC.SG}$
'A Christmas story about a grandma$_{DIM}$ from Omsk.'

We have also found several examples with feminine and neuter mixed cues showing masculine agreement:

(10) Feminine base gender + inflection *-o*/*e* (class IV) = masculine agreement
Očen' *mn'e* ***ponravils'a —*** ***delovityj,*** ***uverennyj*** *v* *sebe* ***sobačiško.***
very me_{DAT} $appealed_{M.SG}$ $busy_{M.NOM.SG}$ $confident_{M.NOM.SG}$ in self $dog_{DIM.NOM.SG}$
'I liked him very much — such a busy, confident $doggy_{DIM}$.'

To summarize, we found extensive gender variation in expressive nouns. The following factors may have influence: the base gender, the inflectional class, and animacy. We also observed different patterns for the three genders. Masculine base gender is preserved more often than feminine, while neuter nouns just do not attach expressive affixes associated with other genders. Inflections associated with masculine gender (class I) trigger gender change more often than those associated with feminine (class II) or neuter (class IV). To estimate the role of these factors numerically, we conducted an experiment.

2.2. Experimental data

2.2.1. Experiment 1: Method

Thirty native speakers of Russian (17 female and 13 male) volunteered to take part in the experiment. Their age ranged from 18 to 45, the mean age being 23.5. They all provided informed consent.

The materials were 40 nouns with diminutive and augmentative affixes (*-ok-*∅, *-ek-*∅, *-ik-*∅, *-čik-*∅, *-išk-a*/*o*, *-(j)ušk-a*/*o*, *-išč-a*/*e*, *-in-a*, *-(j)ug-a*, *-(j)ar-a*). We checked on the internet that they all occur naturally. Their properties are listed in Table 3. We did not include neuter base gender because, as we noted in the previous section, such nouns almost never attach expressive affixes that could trigger gender change. To distract participants' attention from the goal of the study, we also included 40 filler nouns and pseudonouns without expressive affixes (animate and inanimate, from different declensions, masculine, feminine, or neuter).

Table 3

Fourty nouns with expressive affixes used as stimuli

Base gender	Inflection	Animacy
M	*-a:* class II (F)	5 nouns denoting people and 5 inanimate nouns
	-o/*e:* class IV (N)	5 nouns denoting people and 5 inanimate nouns
F	*-*∅*:* class I (M)	5 nouns denoting people and 5 inanimate nouns
	-o/*e:* class IV (N)	5 nouns denoting people and 5 inanimate nouns

To avoid asking participants explicitly about the gender of stimulus nouns, we instructed them to select an adjective from the list for every stimulus noun. During the experiment, nouns were shown one by one on the computer screen, and participants were asked to say aloud the combination of the selected adjective and the noun. The list contained seven adjectives with stressed inflections (to avoid problems due to unstressed vowel reduction): *krutoj* 'cool', *rodnoj* 'native, dear', *ploxoj* 'bad', *zloj* 'evil', *bol'šoj* 'big', *nebol'šoj* 'not big, small', and *takoj* 'such' (we instructed to select the last option if the others are not suitable). As a result, participants thought that the experiment was about semantic connotations of different nouns, while in fact we were interested in gender agreement.

In the instruction, the adjectives from the list were provided in the three gender forms (in nominative singular) not to prime any gender. Remembering these adjectives during the experiment could be difficult, so we printed out a reminder that we gave to the participants while they were reading the instruction. The reminder contained only the beginnings of the adjectives: e.g., *krut…* for *krutoj* 'cool', *bol'š…* for *bol'šoj* 'big', etc. Participants believed that this was done so that they could remember the adjectives faster, but the true reason was again not to prime any gender with the inflections.

Participants passed the experiment one by one in a quiet room. We used the *PsyScope* platform (http://psy.ck.sissa.it/) to demonstrate the instruction and stimuli and recorded participants' answers. First the experimenter pressed a key and a fixation point appeared in the center of the screen (1000 ms), followed by a stimulus noun (500 ms). Then the blank screen was demonstrated until the participant pronounced the selected adjective and the noun. After that, the experimenter pressed a key again, and the next trial started.

The order of stimulus presentation was random. We selected the interval of 500 ms so that participants could read stimuli, but could not start thinking about them deeply. The main experiment was preceded by a training session.

2.2.2. Experiment 1: Results and discussion

We collected 1200 answers. In 614 out of them (51%), expressive derivates did not preserve the base gender. The feminine base gender was lost in 461 cases (77%), while the masculine base gender was lost only in 153 cases (26%). Gender changes were not random and were predominantly triggered by the inflectional affix, as expected (92% cases). But in 47 examples, masculine gender was assigned to nouns with feminine base gender and suffixes ending in *-o* or *-e* (class IV). We will come back to these cases in the discussion section. Table 4 presents the results in more detail.

Table 4

The results of the Experiment 1

Base gender	Inflection	Observed base gender change
M	*-a:* class II (F)	in 90 out of 300 answers (30%) changed to F
	-o/*e:* class IV (N)	in 63 out of 300 answers (21%) changed to N
F	*-∅:* class I (M)	in 282 out of 300 answers (94%) changed to M
	-o/*e:* class IV (N)	in 132 out of 300 answers (44%) changed to N in 47 out of 300 answers (16%) changed to M

The statistical analysis was done in the R programming environment (www.r-project.org). We built mixed effects logistic regressions with random slopes and random intercepts by participants and by items. We used the *glmer* function from the *lme4* package [Bates et al. 2015]. For post hoc analyses, Tukey's tests were conducted using the *glht* function from the *multcomp* package [Bretz et al. 2010]. We used binary factors coded as 1 vs. 0 as fixed effects, namely: masculine base gender, feminine base gender, final *-o*/*e*, final *-a*, final C (hard consonant), stress position (on the suffix vs. on the stem), animacy, diminutive meaning, augmentative meaning, emotive meaning of the adjective, size-related meaning of the adjective. We have built a separate model for each factor.

Both the base gender and the inflection appear to be highly relevant: the former is preserved in 49% answers, while the latter triggers changes in 47% answers (the remaining 4% cases changed to masculine despite feminine and neuter mixed gender cues). We found that masculine base gender is significantly more likely to be preserved than feminine ($\beta = -3.84$, SE = 0.68,

$p < 0.001$). The zero inflection (final C, associated with masculine) is significantly more likely to trigger gender change than *-a* (associated with feminine) or *-o*/*e* (associated with neuter) ($\beta = -5.39$, SE = 1.16, $p < 0.001$; $\beta = -4.69$, SE = 0.82, $p < 0.001$, respectively). The difference between *-a* and *-o*/*e* did not reach significance.

49% animate nouns and 52% inanimate nouns changed their base gender, and this difference was not significant. Combining web-as-a-corpus and experimental data, we can conclude that animacy might play an important role in some cases, but on average, its influence is weak. For example, C-final expressive nouns may preserve their feminine base gender only when they denote people, but, since they change it in the absolute majority of cases, this categorical difference (possible vs. impossible) gives very small numerical effects.

We also estimated the influence of several other factors, but none of them were significant. We checked stress position (on the suffix vs. on the stem), expressive suffix meaning (diminutive vs. augmentative), age and sex of the participants. We compared every suffix to all the others and found that none of them showed idiosyncratic behavior.

As we noted in the introduction, Steriopolo et al. [2021] claimed that suffixes with emotive meaning trigger gender change, while suffixes with size meaning do not. However, we found no correlations between adjective choice and gender change, not even a tendency. We also saw no evidence of such correlation in our web-as-a-corpus data.

Finally, some suffixes we used are associated with two different inflectional classes, e.g., *-išk-a*/*o* with class II (predominantly feminine) and class IV (neuter) (see also Table 2). Somewhat surprisingly, this does not preclude Russian speakers from deriving words like *stjuardessiško* ‘female flight attendant$_{DIM}$’ from *stjuardessa*, although the *-išk-a* variant could be used, avoiding a gender conflict. Words like *stjuardessiško* were among our stimuli and showed variable agreement.

We can conclude that two factors are predominantly responsible for gender variation in expressive nouns: the base gender and the inflection of the derivate. For both factors, masculine gender has a special status. It is preserved significantly more often and triggered by the inflection significantly more often than feminine or neuter. Moreover, all (relatively few) examples that cannot be explained by these two factors have masculine gender.

3. Study 2: Gender assignment depending on the stem-final consonant

As we noted in the Introduction, the correlation between gender and declensions is very strong in Russian and shows up in different cases, gender assignment in expressive nouns being only one of them. The goal of our first study was not to establish this fact as such, but to explore the details. In our second study, we focus on a correlation that is much less obvious.

Table 1 shows that while C-final nouns can only be masculine (with the exception of indeclinable nouns and nouns with defective paradigms), C’-final nouns can be both masculine and feminine. C’-final nouns ending in alveolo-palatal and postalveolar consonants have different spelling rules depending on their gender. Feminine nouns have the soft sign ‹ь› at the end (it does not affect pronunciation), while masculine nouns do not (cf. *rož’* ‹рожь› ‘rye$_F$’ and *muž* ‹муж› ‘husband$_M$’). No obvious orthographic or phonological cues are available for C’-final nouns ending in a palatalized consonant: their gender cannot be determined from their nominative singular form (cf. *len’* ‹лень› ‘laziness$_F$’ and *den’* ‹день› ‘day$_M$’).

However, several authors noted that the distribution of C’-final nouns between the two genders depends on the particular stem-final segment [Vrabie 1992; Nesset 2003]. They did not test their observations statistically. We provide a statistical analysis of dictionary data and test whether native speakers are sensitive to these correlations in an experiment with nonce words. Our results show that gender assignment depends not only on inflections, but also on the phonological properties of the stem.

3.1. Dictionary data

The distribution of genders in C'-final nouns was investigated by Emil Vrabie [1992] based on four Russian dictionaries [BAS 1948–1965; MAS 1981–1985; Ozhegov 1960; Ushakov 1935–1940] and later by Tore Nesset [2003] based on the *Grammatical Dictionary of the Russian Language* [Zaliznyak 1987]. Both authors found that the phonological properties of the stem-final segment are important, but did not provide a statistical analysis, so we do so below. In Table 5, we summarize some dictionary and corpus data illustrating that the gender distribution is indeed not random. For the statistical analysis, we used the dictionary compiled by Hagen [2018], which is based on several popular dictionaries of Russian and extended with novel words from web sources.

Table 5

C'-final nouns in the grammatically disambiguated subcorpus of the Russian National Corpus (RNC)[7], in the *Grammatical Dictionary of the Russian Language* (GDRL) [Zaliznyak 1987] and in the Hagen dictionary [Hagen 2018]

Final consonant	RNC (Nom.Sg forms)		GDRL (lemmas)		Hagen (lemmas)	
	M	F	M	F	M	F
/b'/	34 (24%)	110	1	11	—	13
/p'/	3 (2%)[8]	169	—	19	—	22
/v'/	13 (1%)	1448	1	20	1	21
/f'/	0	2	—	2	—	2
/m'/	0	16	—	3	—	5
/d'/	748 (51%)	707	10	55	5	59
/t'/	713 (5%)	13184	17	3414	8	5508
/z'/	319 (49%)	327	7	34	3	37
/s'/	80 (14%)	491	5	57	1	60
/n'/	2354 (45%)	2842	126	112	102	115
/r'/	2160 (76%)	677	177	34	63	28
/l'/	6648 (71%)	2653	1083	215	681	224

The Hagen dictionary contains 8552 C'-final nouns. For the statistical analysis, we excluded all nouns with two highly frequent suffixes: *-tel'* associated with the masculine (940 nouns) and *-ost'* associated with the feminine (5333 nouns). This left us with 2272 nouns, 1750 inanimate and 522 animate. The influence of animacy on gender is not always straightforward, therefore we calculated statistics on two sets: with animate nouns and without them. The results were very similar, so below, statistics on the full set are provided.

The statistical analysis was done in the R programming environment (www.r-project.org). We used general logistic regressions.

We found that the place of articulation and sonority of the stem-final segment significantly affect the gender distribution. There are only two masculine nouns with labial or labiodental stem-final segments: *golub'* 'pigeon$_M$' and *červ'* 'worm$_M$', so in this group, conducting a statistical analysis does not make any sense. Nouns ending in coronal obstruents are significantly more often feminine than masculine ($\beta = -3.64$, SD = 0.25, $p < 0.001$), while nouns ending in sonorants

[7] The counts are based on [Slioussar, Samoilova 2015].

[8] These three forms are *rup'* (a reduced form of the noun *rubl'* 'ruble').

are significantly more often masculine ($\beta = 0.53$, SD = 0.09, $p < 0.001$). Nouns ending in alveolo-palatal and postalveolar consonants, in which the gender is signaled orthographically, by the presence or absence of the soft sign, are significantly more often masculine ($\beta = 2.08$, SD = 0.14, $p < 0{,}001$). In general, feminine nouns occur in all groups, while masculine nouns tend to concentrate in groups with strident and sonorant stem-final segments.

3.2. Experimental data

We conducted two experiments to test whether native speakers of Russian are sensitive to the generalizations concerning the stem-final segment. The first one was a pilot study that was included in the Experiment 1 on gender assignment to expressive nouns described above. The second one (Experiment 2) was an independent study in which we used more stimuli and a different method to facilitate data collection.

3.2.1. Pilot study

When describing Experiment 1, we mentioned that apart from expressive nouns, we used 40 real and nonce nouns without diminutive or augmentative suffixes as fillers. Among them, there were 12 real and 12 nonce nouns ending in palatalized consonants, on which we focus in this section. The design of the experiment was presented above, so here we will only discuss the stimuli and the results.

In the real group, six nouns were masculine and six feminine. They had the following stem-final consonants: /b'/, /d'/, /n'/, /l'/. We collected 360 answers and found only seven errors in gender assignment: three errors with *žen'šen'* 'ginseng$_M$', two errors with *stupen'* 'step$_F$', one error with *kisel'* 'starch drink$_M$' and with *prorub'* 'ice-hole$_F$'. Obviously, with such a low number of errors no interesting conclusions can be made — only that the participants paid enough attention to the task. The high accuracy is not surprising because native Russian speakers in general do not have problems with determining the gender of a noun unless there is variation — see, e.g., [Rusakova 2013].[9]

As for the 12 nonce nouns, we had two examples with each of the following stem-final segments: /b'/, /d'/, /s'/, /n'/, /l'/, /r'/. The results for them are presented in Table 6 (p. 21). The share of masculine answers changes depending on the final consonant.

For the statistical analysis, we used mixed effects logistic regression with random slopes and random intercepts by participants and by items, like in Experiment 1. We found that nonce nouns ending in *-l'* and *-r'* were assigned masculine significantly more often than nonce nouns in the other groups ($\beta = 2.48$, SE = 1.15, $p = 0.03$; $\beta = 4.00$, SE = 1.54, $p < 0.01$, respectively), while other differences were not significant. This indicates that native Russian speakers are indeed sensitive to the properties of the stem-final consonant because, unlike all other palatalized consonants, /r'/ and /l'/ are more characteristic for masculine nouns than for feminine nouns.

Let us also note that in total, participants selected masculine gender in 59% answers, while real C'-final Russian nouns are more often feminine than masculine. As many other findings in our paper, this points to a special status of masculine gender. The validity of conclusions from this pilot study may be compromised by the low number of stimuli, so we conducted Experiment 2 to check them on a larger stimulus set.

[9] Some C'-final nouns show gender variation, usually because of a diachronic gender and declension change from class III feminine to class I masculine [Savchuk 2011], but we did not include such words in the experiment.

Table 6

The results of the pilot experiment and Experiment 2: The number and percentage of masculine answers

Final consonant	Pilot experiment	Experiment 2
/b'/	22 (37%)	69 (70%)
/p'/	—	68 (69%)
/v'/	—	68 (69%)
/f'/	—	68 (69%)
/m'/	—	70 (71%)
/d'/	26 (43%)	65 (66%)
/t'/	—	71 (72%)
/z'/	—	76 (77%)
/s'/	38 (63%)	70 (71%)
/n'/	30 (50%)	66 (67%)
/l'/	51 (85%)	83 (84%)
/r'/	44 (73%)	93 (94%)

3.2.2. Experiment 2: Method

Thirty-three native speakers of Russian (15 female and 18 male) volunteered to take part in the experiment. Their age ranged from 18 to 55, the mean age being 25.1. They all provided informed consent.

The pilot study showed that it made sense to focus on nonce nouns. Therefore, we created 36 nonce nouns with the following stem-final consonants: /b'/, /p'/, /v'/, /f'/, /m'/, /d'/, /t'/, /z'/, /s'/, /n'/, /l'/, and /r'/ (three stimuli for each consonant). We also created 36 filler nonce nouns, 18 ending in a non-palatalized consonant and 18 ending in *-a*, to distract participants' attention from the main goal of the study.

In Experiment 1, we opted for oral production because we wanted our participants to answer as fast as possible, to avoid any metalinguistic considerations on their side (what is normative and what is not, etc.). In case of nonce nouns, the prescriptive norm and other similar considerations cannot affect the answers, so we opted for a written version of the task. It is less immediate, but the experiment was easier to conduct and to analyze.

We presented nonce nouns on the screen as a list and asked participants to come up with an adjective that seemed associated with this word and to type it in the blank provided. The order of stimulus presentation was random. Like in Experiment 1, participants believed that we were interested in semantic connotations of nonce nouns, while we analyzed the gender of the adjectives. We used Google Forms to implement this design.

3.2.3. Experiment 2: Results and discussion

We collected 1188 answers to target nonce nouns. The share of masculine adjectives depending on the stem-final consonant is shown in Table 6. The statistical analysis was the same as in Experiment 1 and in the pilot study: we used mixed effects logistic regression with random slopes and random intercepts by participants and by items.

The results of the pilot study were confirmed on a more representative stimulus set. Nonce nouns ending in *-l'* and *-r'* were again considered masculine significantly more often than other nonce nouns ($\beta = 1.92$, SE = 0.84, $p = 0.03$; $\beta = 2.95$, SE = 0.62, $p < 0.001$, respectively), while other differences were not significant. The overall percentage of masculine gender was even higher in Experiment 2 than in the pilot study.[10] Although Russian has only one or no masculine nouns in every labial group (ending in /b'/, /p'/, /v'/, /f'/, /m'/), most answers to such nonce words were masculine. I.e., masculine gender is readily overgeneralized — if the stimulus has the right inflection (the majority of nonce nouns ending in *-a* were classified as feminine).

4. Discussion

Our data may be interesting not only by themselves, but also because they are more easily compatible with some theoretical approaches than with the others. We demonstrate this briefly in this section. Then, we turn to the problem of gender markedness.

4.1. Theoretical approaches

Distributed Morphology, or DM, is the most widespread formal framework in theoretical morphology. It offers a syntax-based approach to morphological structure with gender being a feature on the n-head. This framework also argues for late vocabulary insertion: the syntactic tree is built from universal heads and features, while actual morphemes are inserted only at a very late stage, at the interface with the phonological component. In case of gender agreement, a gender feature of the highest n-head of the tree is retrieved. Therefore, expressive suffixes may be expected to assign gender in case they are represented as n-heads and carry a gender feature [Kramer 2015]. Accordingly, Steriopolo, who argued that certain expressive suffixes or suffixes with certain meanings always change the base gender in Russian, while the others never do, represented the former as syntactic heads and the latter as modifiers [Steriopolo 2008; 2015; 2017; Steriopolo et al. 2021].

What is not expected in this framework is the variation that we observed in our first study — expressive suffixes sometimes change the gender of the base noun and sometimes do not, and this is not defined by the animacy of the noun or the meaning of the suffix. This would require two sets of identical expressive affixes that would differ only in one respect: one suffix in a pair must be a head and the otherwise identical twin must be a modifier. Moreover, even this inelegant solution would not account for the asymmetries between the three genders. It could be compatible with random variation, while we observed that masculine base gender is preserved more often and affixes trigger gender change to masculine more often than to feminine or to neuter.

Another question in formal approaches is whether inflectional class defines gender or vice versa (see, e.g., [Asarina 2009; Kramer 2015; Privizentseva, to appear]). Our first study shows that none of the two categorical answers may be correct. Matushansky [2015] suggests building a more complex system in which lexical gender and inflectional class are among other factors that contribute to the gender assignment, but it has not been developed so far.

Our second study of nouns with palatalized stem-final segments was based on nonce nouns, for which no gender information is stored. Therefore, it is not surprising that native speakers may rely on various analogies with real nouns when they decide on their gender. Still, analogies based on the phonological properties of the final segment are more difficult to encode in DM because

[10] This might have happened because the pilot study had a more heterogeneous set of stimuli, both real and nonce.

in this framework, phonological properties of the root are not available at the stage of gender assignment.[11] It is even more difficult to encode how these properties interact with the special status of masculine gender, which is chosen more often in nonce nouns than we would expect based on the distribution of real nouns.

Distributed Morphology and most structural theories assume a separate storage of elements and an abstract structure where these elements are inserted. Relational and constructional approaches to morphology suggest that there is no separate storage: all words are stored in the form of constructions or relational schemes — see, e.g., [Booij 2010; Masini, Audring 2019; Jackendoff, Audring 2019]. Although these theories may have problems accounting for purely syntactic processes, our data on morphological variation are more readily compatible with these approaches.

According to this group of theories, the variation in gender assignment is due to different schemas available in the lexicon. The fact that in Experiment 1, masculine gender was assigned to almost all C-final expressive nouns ignoring their non-masculine lexical gender may be explained by the distribution of C-final nouns in Russian. This group is very numerous and contains only masculine nouns (apart from some indeclinable nouns, nouns with defective paradigms and expressive nouns under discussion). Therefore, schemas with a C-final stem and feminine or neuter gender are either unavailable or extremely rare. The correlation between the stem-final segment and gender in C'-final nouns may be explained in the same way: e.g., schemas with a stem-final sonorant C' and masculine gender are more frequent than analogous schemas with feminine gender, and this influences gender assignment to nonce nouns.

A model of Russian gender developed by Doleschal [2000] is similar in the spirit to Relational Morphology. Doleshal even takes variation into account. She analyzes some expressive examples making important observations, but does not aim to draw a wider picture, identifying the relative importance of different factors that affect variation in this group.

Finally, competition-based approaches, like Optimality Theory (OT), are especially well-suited to account for our data. In OT, different types of information (phonological, syntactic, semantic) may be available at once and different cues may compete during gender assignment. In the classic version of OT, one variant is selected as a result, but OT-based probabilistic theories, like Harmonic Grammars and Maximum Entropy models (e.g., [Boersma, Pater 2016; Goldwater, Johnson 2003; Hayes, Wilson 2008]) were developed to account for variation. If we focus on gender assignment, there is an influential Optimal Gender Assignment Theory built by Rice [2006] on German, French, and Russian data and substantially modified by Corteen [2019] to improve the predictions for German nouns, the model suggested by Galbreath [2010] for Russian, and some others. Importantly, none of these models included gender variation. In [Chuprinko et al. to appear, b], we offer an optimality-theoretic account of gender variation in Russian indeclinable nouns, where an interplay of different factors can also be detected.

At the same time, as we show in the next section, we also need some insights from structural approaches, namely, the concept of representational markedness.

4.2. Gender markedness

Now let us come back to the problem of gender markedness. As we noted in the Introduction, structural approaches assume that neuter is unmarked in Russian, while competition-based and functional approaches view masculine as unmarked. The question is not trivial because both genders may demonstrate some special behavior in different tasks and different contexts.

In the two studies reported in this paper, the special status of masculine could be observed. Masculine base gender was preserved more often, masculine was assigned more often in case

[11] Unlike most other DM models, Borer [2003; 2013] suggests that the phonological form of roots and at least some other morphemes must be available for morphosyntactic processes.

of conflicting gender cues. Special properties of the neuter gender did not have a chance to reveal themselves in these studies. However, we know independently that neuter is used in impersonal sentences and is the most frequent gender in indeclinable nouns [Corbett 1982; Muchnik 1971; Murphy 2000; Chuprinko et al., to appear, a; b].

How to reconcile these prima facie conflicting observations? We suggest the following solution. We assume that neuter is representationally unmarked, as it is claimed in structural approaches, and therefore tends to be chosen when no gender cues are present. This is the case in impersonal sentences and in indeclinable nouns. As for the special properties of masculine, we believe that they can be explained by its high frequency and productivity. Masculine is also the most diverse gender in terms of the final segments.[12] As a result, if any cues point to masculine, it is more likely to be chosen than any other gender. We would call it a 'greedy gender', using an analogy to greedy operators in regular expressions in programming.[13]

A group of examples does not readily fit this picture: in Experiment 1, there were 47 answers in which masculine gender was chosen for the nouns with feminine base gender and suffixes ending in *-o* or *-e* (class IV, associated with neuter). We also found such examples on the internet — one of them was given in (10). One might assume that masculine not only has the strongest cues, but may also be chosen as a last resort option when the system is stuck in case of conflicting cues. However, this solution would be difficult to reconcile with the choice of neuter in impersonal sentences or with its prevalence in indeclinable nouns.

The 47 examples in question ended in *-išče* and *-iško*. We analyzed all nouns with these affixes in the *Grammatical Dictionary of the Russian Language* [Zaliznyak 1987] and found that nouns classified as masculine were more numerous than those classified as neuter.[14] Of course, the dictionary mostly ignores variation observed in real usage, but this still means that many *-išče* and *-iško* nouns with masculine agreement can be found. We argue that these affixes become associated not only with neuter, but also with masculine as a result. This can explain why masculine was used in the 47 answers under discussion without assuming that masculine is chosen as a last resort option in some cases and neuter in the others.

Our approach may also be used to account for discrepancies in the previous experimental findings. As we discussed in the introduction, Slioussar and Malko [2016], who studied gender

[12] Many patterns found in the nominative singular in feminine and neuter can also be found in masculine: there are masculine nouns ending in *-a* like *papa* 'dad'; masculine nouns ending in *-o* and *-e* like *domiško* 'house$_{DIM}$' and many other expressive derivates that preserved their base gender; masculine nouns with zero inflections ending in all possible consonants. Feminine nouns with zero inflections can only end in alveolo-palatal, postalveolar and palatalized consonants, while neuter nouns always end in vowels. There are also indeclinable masculine nouns with endings not typical for nominative singular forms, like *-i* in *viski* 'whiskey' or *-u* in *kenguru* 'kangaroo'.

[13] Most structural approaches assume that the absence of any features is the least marked option: e.g., for Kramer [2015], feminine is [+FEM], masculine is [−FEM], and neuter corresponds to no gender feature in Russian. Adamson and Šereikaitė [2019] develop a different account for Lithuanian. Lithuanian has three genders, like Russian, but nouns may be only masculine or feminine. Neuter agreement can be found only in impersonal constructions, with pronouns meaning 'nothing' or 'nobody', etc. For this reason, Adamson and Šereikaitė encode neuter as the absence of gender features, but do not see this option as a part of the markedness hierarchy, as it is not realized on any noun. In their system, masculine (encoded as [−FEM]) is regarded as the least marked option. Adamson and Šereikaitė's approach cannot be applied to our data because Russian does have neuter nouns, while our approach has a potential to be extended to their data, providing an alternative explanation why both neuter and masculine show special properties in Lithuanian and getting rid of the uncommon assumption that the absence of any features is not the least marked option. We leave this question for further research.

[14] In general, masculine nouns are dramatically more frequent than neuter nouns in Russian (especially if we do not take into account abstract neuter nouns formed with several productive suffixes), so, even though they attach a great variety of expressive affixes, while neuter nouns strongly prefer class IV affixes, masculine nouns turn out to be more numerous even in the class IV affix group.

agreement attraction in Russian, found that in production, neuter showed special properties usually associated with the unmarked feature value, while in comprehension, masculine was different from the two other genders. We believe that neuter attractors are the least disruptive for gender agreement in production because the neuter feature is the least marked representationally and thus the least noticeable. At the same time, gender agreement errors with masculine subjects are the most noticeable in comprehension — probably, because masculine as the most frequent gender with the strongest cues projects the strongest expectations about the gender of the upcoming predicate (see also [Slioussar 2018] and [Akhutina et al. 1999, 2001; Romanova, Gor 2017] on attributive agreement).

5. Conclusions

In this paper, we demonstrated that Russian nouns with diminutive and augmentative affixes exhibit gender variation and estimated the role of different factors affecting this variation using web-as-a-corpus and experimental data. Two factors were identified as crucial: the gender of the base noun and the inflection associated with the expressive suffix. Animacy is also relevant because it excludes certain options (expressive nouns ending in a non-palatalized consonant with non-masculine base gender may preserve their base gender only when they denote people), but numerically, its influence is small because the options it excludes are not frequent anyway. The meaning of the suffixes (diminutive vs. augmentative, emotive vs. size-related) did not play any role, although the latter distinction was argued to be crucial by Steriopolo et al. [2021].

We also focused on the observation made by [Vrabie 1992; Nesset 2003] that the gender of nouns with a zero inflection in nominative singular and a palatalized, alveolo-palatal or post-alveolar consonant at the end of the stem depends on the phonological properties of this consonant. We supported this observation with more data and a statistical analysis. In an experimental study, we demonstrated that native Russian speakers are sensitive to this generalization.

We explored the implications of our findings for different approaches in theoretical morphology and for the problem of gender markedness. Our data are not readily compatible with structural frameworks like Distributed Morphology, but can be modelled in optimality-theoretic, relational, or constructional approaches, in which different types of information (phonological, syntactic, semantic) may be available at once and different cues may compete during gender assignment. At the same time, our data support the validity of the concept of representational markedness that is introduced in structural approaches.

Different authors disagreed upon which gender is unmarked in Russian: masculine or neuter. We argue that neuter is the least marked representationally, so it is used in impersonal sentences and is the most frequent gender in indeclinable nouns — when gender assignment cues are absent or very weak (see [Chuprinko et al., to appear, a; b] for a detailed analysis). Masculine is a greedy class due to its highest productivity, frequency, and diversity of surface forms. Therefore, it has the strongest cues and tends to be chosen more often than other genders in case of conflict. This approach explains the observations we made in this study, as well as other reported findings.

ABBREVIATIONS

ACC — accusative
ADJ — adjective
AUG — augmentative
DAT — dative
DIM — diminutive
DM — Distributed Morphology
F — feminine
GEN — genitive
INF — infinitive
INS — instrumental
LOC — locative (“prepositional case”)
M — masculine
N — neuter
NOM — nominative
OT — Optimality Theory
PL — plural
SG — singular

REFERENCES

A reference grammar of Russian 2004 — Timberlake A. (ed.). *A reference grammar of Russian*. Cambridge: Cambridge Univ. Press, 2004.

Adamson, Šereikaitė 2019 — Adamson L., Šereikaitė M. Gender representation and defaults in Lithuanian. *Glossa: A Journal of General Linguistics*, 2019, 4(1): Article 121.

Akhutina et al. 1999 — Akhutina T., Kurgansky A., Polinsky M., Bates E. Processing of grammatical gender in a three-gender system: Experimental evidence from Russian. *Journal of Psycholinguistic Research*, 1999, 28: 695–713.

Akhutina et al. 2001 — Akhutina T., Kurgansky A., Kurganskaya M., Polinsky M., Polonskaya N., Larina O., Bates E., Appelbaum M. Processing of grammatical gender in normal and aphasic speakers of Russian. *Cortex*, 2001, 37: 295–326.

Aronoff 1994 — Aronoff M. *Morphology by itself: Stems and inflectional classes*. Cambridge (MA): MIT Press, 1994.

Asarina 2009 — Asarina A. Gender and adjective agreement in Russian. Handout presented at the 4th Annual Meeting of the Slavic Linguistics Society, Univ. of Zadar, Croatia, 2009.

Badecker, Kuminiak 2007 — Badecker W., Kuminiak F. Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, 2007, 56: 65–85.

BAS 1948–1965 — Chernyshev V. I., Barkhudarov S. G., Vinogradov V. V., Filin F. P. (eds.). *Slovar' sovremennogo russkogo literaturnogo yazyka* [Dictionary of modern Russian literary language]. Moscow; Leningrad: Academy of Sciences of the USSR Press, 1948–1965.

Bates et al. 2015 — Bates D., Maechler M., Bolker B., Walker S. *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1–8. http://CRAN.R592project.org/package=lme4. 2015.

Boersma, Pater 2016 — Boersma P., Pater J. Convergence properties of a gradual learning algorithm for Harmonic Grammar. *Harmonic Grammar and Harmonic Serialism*. McCarthy J. J., Pater J. Sheffield: Equinox, 2016: 389–434.

Booij 2000 — Booij G. Inflection and derivation. *Morphology. An international handbook on inflection and word-formation.* Vol. 1. Booij G., Lehmann C., Mugdan J., Skopeteas S., Kesselheim W. (eds.). Berlin: Walter de Gruyter, 2000, 654–661.

Booij 2010 — Booij G. Construction morphology. *Language and Linguistics Compass*, 2010, 4: 543–555.

Borer 2003 — Borer H. Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. *The nature of explanation in linguistic theory*. Moore J., Polinsky M. (eds.). Stanford: CSLI Publications, 2003, 31–67.

Borer 2013 — Borer H. *Structuring sense.* Vol. III: *Taking form*. Oxford: Oxford Univ. Press, 2013.

Bretz et al. 2010 — Bretz F., Hothorn T., Westfall P. *Multiple comparisons using R*. Boca Raton: CRC Press, 2010.

Caha 2019 — Caha P. *Case competition in Nanosyntax: A study of numerals in Ossetic and Russian*. Berlin: Language Science Press, 2019.

Caha 2021 — Caha P. Modeling declensions without declension features. The case of Russian. *Acta Linguistica Academica*, 2021, 68(4): 385–425.

Chuprinko et al., to appear, a — Chuprinko K., Magomedova V., Slioussar N. Gender variation in indeclinable inanimate nouns and gender markedness in modern Russian. Submitted manuscript, 2023.

Chuprinko et al., to appear, b — Chuprinko K., Magomedova V., Slioussar N. Modelling gender variation in Russian indeclinable nouns: optimality over structuralism, hierarchical MaxEnt and degrees of idiosyncrasy. Submitted manuscript, 2023.

Corbett 1982 — Corbett G. G. Gender in Russian: An account of gender specification and its relationship to declension. *Russian Linguistics*, 1982, 6: 197–232.

Corbett 1991 — Corbett G. G. *Gender*. Cambridge: Cambridge Univ. Press, 1991.

Corbett, Fraser 1993 — Corbett G. G., Fraser N. M. Network Morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics*, 1993, 29: 113–142.

Corbett, Fraser 2000 — Corbett G. G., Fraser N. M. Gender assignment: A typology and a model. *Systems of nominal classification*. Senft G. (ed.). Cambridge: Cambridge Univ. Press, 2000, 293–325.

Corteen 2019 — Corteen E. C. *The assignment of grammatical gender in German: Testing Optimal Gender Assignment Theory*. Ph.D. diss. Univ. of Cambridge, 2019. https://doi.org/10.17863/CAM.37638

Doleschal 2000 — Doleschal U. Gender assignment revisited. *Gender in grammar and cognition.* Part I: *Approaches to gender*. Dahl Ö., Unterbeck B., Rissanen M. (eds.). Berlin: Mouton de Gruyter, 2000, 117–166.

Galbreath 2010 — Galbreath B. L. E. *Gender assignment in contemporary standard Russian: A comprehensive analysis in Optimality Theory.* Ph.D. diss. Univ. of Virginia, 2010.

Goldwater, Johnson 2003 — Goldwater Sh., Johnson M. Learning OT constraint rankings using a maximum entropy model. *Proc. of the Stockholm Workshop on Variation within Optimality Theory.* Spenader J., Eriksson A., Dahl Ö. (eds.). Stockholm: Stockholm Univ., 2003, 111–120.

Graudina et al. 1976 — Graudina L. K., Itskovich V. A., Katlinskaya L. P. *Grammaticheskaya pravil'nost' russkoi rechi* [Grammatical correctness of the Russian speech]. Moscow: Nauka, 1976.

Hagen 2018 — Hagen M. *Polnaya paradigma. Morfologiya* [Full paradigm. Morphology]. 2018. RAR archive. Retrieved on January 10, 2019 from http://www.speakrus.ru/dict/.

Halle 1994 — Halle M. The Russian declension: An illustration of the theory of Distributed Morphology. *Perspectives in phonology*. Cole J., Kisseberth C. (eds.). Stanford: CSLI Publications, 1994, 29–60.

Haspelmath 2006 — Haspelmath M. Against markedness (and what to replace it with). *Journal of Linguistics*, 2006, 42: 25–70.

Hayes, Wilson 2008 — Hayes B., Wilson C. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 2008, 39(3): 379–440.

Hippisley 1996 — Hippisley A. Russian expressive derivation: A Network Morphology account. *The Slavonic and East European Review*, 1996, 74: 201–222.

Jackendoff, Audring 2019 — Jackendoff R., Audring J. Relational morphology in the parallel architecture. *The Oxford handbook of morphological theory*. Oxford: Oxford Univ. Press, 2019, 390–408.

Jakobson 1960 — Jakobson R. O. The gender pattern of Russian. *Studii și Cercetări Lingvistice*, 1960, 11: 541–543.

Kramer 2015 — Kramer R. T. *The morphosyntax of gender*. Oxford: Oxford Univ. Press, 2015.

Lyutikova 2015 — Lyutikova E. A. Soglasovanie, priznaki i struktura imennoi gruppy v russkom yazyke [Features, agreement, and structure of the Russian noun phrase]. *Russkij jazyk v nauchnom osveshchenii*, 2015, 30: 44–74.

MAS 1981–1985 — *Slovar' russkogo yazyka: v chetyrekh tomakh* [Dictionary of the Russian Language: In 4 vols.]. Evgen'eva A. P. (ed.). Moscow: Russkii Yazyk, 1981–1985.

Masini, Audring 2019 — Masini F., Audring J. Construction morphology. *The Oxford handbook of morphological theory.* Audring J., Masini F. (eds.). Oxford: Oxford Univ. Press, 2019, 365–389.

Matushansky 2013 — Matushansky O. Gender confusion. *Diagnosing syntax.* Cheng L. L.-S., Corver N. (eds.). Oxford: Oxford Univ. Press, 2013, 271–294.

Matushansky 2015 — Matushansky O. *N is for "not there".* A talk presented at the conf. "Formal Approaches to Slavic Linguistics 24", New York, 2015.

Muchnik 1971 — Muchnik I. *Grammaticheskie kategorii glagola i imeni v sovremennom russkom literaturnom yazyke* [Grammatical categories of verb and nominal in contemporary standard Russian]. Moscow: Nauka, 1971.

Müller 2004 — Müller G. A Distributed Morphology approach to syncretism in Russian noun inflection. *Formal Approaches to Slavic Linguistics 12*. Browne W., Rivero M.-L. (eds.). Ann Arbor (MI): Michigan Slavic Publications, 2004, 353–373.

Murphy 2000 — Murphy D. L. *The gender of inanimate indeclinable common nouns in modern Russian.* Ph.D. diss. Univ. of Ohio, 2000.

Nesset 2003 — Nesset T. The assignment of gender and declension to Russian nouns in soft consonants: Predictability and rule interaction. *Journal of Slavic Linguistics*, 2003, 11: 287–322.

Nevins 2011 — Nevins A. Marked targets versus marked triggers and impoverishment of the dual. *Linguistic Inquiry*, 2011, 42(3): 413–444.

Ozhegov 1960 — Ozhegov S. I. *Slovar' russkogo yazyka* [Dictionary of the Russian language]. 4th edn. Moscow: Gosudarstvennoe Izdatel'stvo Inostrannykh i Russkikh Slovarei, 1960.

Panov 1968 — Panov M. *Russkii yazyk i sovetskoe obshchestvo: Morfologiya i sintaksis sovremennogo russkogo literaturnogo yazyka* [Russian language and Soviet society: Morphology and syntax of the modern Russian literary language]. Moscow: Nauka, 1968.

Pesetsky 2013 — Pesetsky D. *Russian case morphology and the syntactic categories*. Cambridge (MA): MIT Press, 2013.

Privizentseva, to appear — Privizentseva M. Mixed agreement in Russian: Gender, declension, and morphological ineffability. To appear in *Natural Language and Linguistic Theory*.

Rice 2005 — Rice C. Optimizing Russian gender: A preliminary analysis. *Formal Approaches to Slavic Linguistics 13*. Franks S., Gladney F., Tasseva-Kurktchieva M. (eds.). Ann Arbor (MI): Michigan Slavic Publications, 2005, 265–275.

Rice 2006 — Rice C. Optimizing gender. *Lingua*, 2006, 116: 1394–1417.

Romanova, Gor 2017 — Romanova N., Gor K. Processing of gender and number agreement in Russian as a second language. *Studies in Second Language Acquisition*, 2017, 39: 97–128.

Rusakova 2013 — Rusakova M. V. *Elementy antropotsentricheskoi grammatiki russkogo yazyka* [Elements of anthropocentric grammar of Russian]. Moscow: Yazyki Slavyanskikh Kul'tur, 2013.

Russian grammar 1980 — Shvedova N. Yu. (ed.). *Russkaya grammatika* [Russian grammar]. In 2 vols. Moscow: Nauka, 1980.

Savchuk 2011 — Savchuk S. Korpusnoe issledovanie variantov rodovoi prinadlezhnosti imen sushchestvitel'nykh v russkom yazyke [A corpus study of gender variation in Russian nouns]. *Computational linguistics and intellectual technologies. Papers from the Annual International Conf. "Dialogue"*, 2011, 10: 562–579.

Sitchinava 2011 — Sitchinava D. V. Rod [Gender]. *Materialy dlya proekta korpusnogo opisaniya russkoi grammatiki* [Materials for the project of the corpus description of Russian grammar]. As a manuscript. Moscow, 2011. http://rusgram.ru/%D0%A0%D0%BE%D0%B4.

Slioussar 2018 — Slioussar N. Gender, declension and stem-final consonants: An experimental study of gender agreement in Russian. *Computational linguistics and intellectual technologies. Papers from the Annual International Conf. "Dialogue"*, 2018, 17: 688–700.

Slioussar, Magomedova, to appear — Slioussar N., Magomedova V. *Gender stereotypes in agreement processing in Russian*. Submitted manuscript.

Slioussar, Malko 2016 — Slioussar N., Malko A. Gender agreement attraction in Russian: Production and comprehension evidence. *Frontiers in Psychology*, 2016, 7: 613–632.

Slioussar, Samoilova 2015 — Slioussar N., Samoilova M. Chastotnosti razlichnykh grammaticheskikh kharakteristik i okonchanii u sushchestvitel'nykh russkogo yazyka [Frequencies of different grammatical features and inflectional affixes in Russian nouns]. *Proc. of the conf. "Dialogue"*, 2015 (online articles). http://www.dialog-21.ru/digests/dialog2015/materials/pdf/SlioussarNASamoilovaMV.pdf.

Steriopolo 2008 — Steriopolo O. *Form and function of expressive morphology: A case study of Russian*. Ph.D. diss. Univ. of British Columbia, 2008.

Steriopolo 2015 — Steriopolo O. Syntactic variation in expressive size suffixes: A comparison of Russian, German, and Spanish. *SKASE Journal of Theoretical Linguistics*, 2015, 12: 2–21.

Steriopolo 2017 — Steriopolo O. Nominalizing evaluative suffixes in Russian: The interaction of declension class, gender, and animacy. *Poljarnyj vestnik: Norwegian Journal of Slavic Studies*, 2017, 20: 18–44.

Steriopolo 2019 — Steriopolo O. Mixed gender agreement in the case of Russian hybrid nouns. *Questions and Answers in Linguistics*, 2019, 5(2): 91–105.

Steriopolo et al. 2021 — Steriopolo O., Markopoulos G., Spyropoulos V. A morphosyntactic analysis of nominal expressive suffixes in Russian and Greek. *The Linguistic Review*, 2021, 38: 1–42.

Steriopolo, Wiltschko 2010 — Steriopolo O., Wiltschko M. Distributed gender hypothesis. *Studies in Slavic linguistics: Proc. of the Formal Description of Slavic Languages 7.5*. Zybatow G., Dudchuk P., Minor S., Pshekhotskaya E. (eds.). Bern: Peter Lang, 2010, 155–172.

Ushakov 1935–1940 — Ushakov D. N. (ed.). *Tolkovyi slovar' russkogo yazyka* [Explanatory dictionary of Russian]. Moscow: Sovetskie Entsiklopedii, 1935–1940.

Vinogradov 1947 — Vinogradov V. V. *Russkii yazyk: grammaticheskoe uchenie o slove* [Russian language: A grammatical theory of word]. Moscow: Prosveshchenie, 1947.

Vrabie 1992 — Vrabie E. Gender distribution of Russian soft-sign nouns with word frequency highlights. *Russian Language Journal*, 1992, 46: 57–83.

Wiese 2008 — Wiese B. Categories and paradigms. On underspecification in Russian declension. *Explorations in nominal inflection.* Müller G., Gunkel L., Zifonum G. Berlin: De Gruyter Mouton, 2008, 321–372.

Zaliznyak 1987 — Zaliznyak A. A. *Grammaticheskii slovar' russkogo yazyka. Slovoizmenenie* [Grammatical dictionary of Russian. Inflection]. 2nd edn. Moscow: Russkii Yazyk, 1987.

Zaliznyak 2002 — Zaliznyak A. A. *"Russkoe imennoe slovoizmenenie" s prilozheniem izbrannykh rabot po sovremennomu russkomu yazyku i obshchemu yazykoznaniyu* ["Russian nominal inflection" and selected works on the modern Russian language and general linguistics]. Moscow: Yazyki Slavyanskikh Kul'tur, 2002.

Получено / received 08.11.2022 *Принято / accepted 22.11.2022*