



NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Olga Lyashevskaya

**CORPUS-BASED PROFILES OF
RUSSIAN NOUNS: FROM
GRAMMATICAL NUMBER TO
LEXICAL SEMANTICS**

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: LINGUISTICS

WP BRP 33/LNG/2015

CORPUS-BASED PROFILES OF RUSSIAN NOUNS: FROM GRAMMATICAL NUMBER TO LEXICAL SEMANTICS²

A grammatical profile indicating the relative frequency distribution of the inflected forms of a word in a corpus is a tool for exploring lexical semantics. However, previous attempts to infer semantically relevant hierarchies of nouns from frequency biases within their grammatical forms seem to have failed. In this paper we explore the distinctive power of grammatical profiles of Russian nouns using the ratio of plural forms as observed in the Russian National Corpus (cf. *roditel'* 'parent' having 95% plural forms and *mama* 'mom' having just 2% plural forms). We claim that since frequent nouns for the most part are semantically ambiguous, their profiles cannot reveal any straightforward effects for large lexical classes. Instead of working on the macro-lexical level we focus on micro-effects within specific taxonomic groups, studying grammatical profiles of body part names, kinship terms, names of vehicles and emotions. The analysis involves the notion of functional frames which represent how objects/events are typically used and typically observed. Our case studies show that grammatical profiles help to structure each group and correlate with certain properties of functional frames associated with nouns.

JEL Classification: Z

Key words: grammatical profiles, grammatical number, nouns, lexical semantics, Russian language

1 National Research University Higher School of Economics, Faculty of Humanities, School of Linguistics, professor; Vinogradov Institute of the Russian Language RAS (Moscow, Russia), senior researcher. E-mail: olesar@yandex.ru.

2 This study (research grant No 14-01-0069) was supported by The National Research University – Higher School of Economics' Academic Fund Program in 2014/2015.

1. Introduction

Russian nouns have a wide range of patterns if we measure the distribution of their singular and plural forms in a corpus. In this paper we offer a particular implementation of the Grammatical Profile method (Janda & Lyashevskaya 2011a, 2011b; cf. also inflectional profiles in Newman 2008) as applied to the grammatical category of Russian substantive number. A grammatical profile here is the distribution of singular and plural forms of a given noun in a corpus analyzed in comparison with the mean distributions of larger lexical and semantic classes. According to the Russian Standard subcorpus of the Russian National Corpus (RNC)³, substantives in general occur 74% in singular and 26% in plural forms. However, the mean ratio of singular and plural forms within the lexical class of days of week (e.g. *ponedel'nik* 'Monday', *vtornik* 'Tuesday', etc.) varies from 89% to 11%. Moreover, if we look at the individual cases across this lexical class, they behave quite differently: the subclass of weekday names (Monday to Friday) have less than 10% plural forms, weekend days have 15–16% of plural forms, their sortal name *vykhodnoj* 'day off' has 65% plural forms, and the sortal name for weekdays *budni* has 100% plural forms (i.e. being pluralia tantum), cf. Fig. 1. We can speculate that the differences in grammatical profiling correlate with some relevant properties of the meaning such as taxonomic category (e. g. common, proper, animate, concrete inanimate, abstract, etc.), inherent and functional qualities of the denotatum, lexical construal (Lyashevskaya 2004) as well as with certain syntagmatic preferences such as constructional choice.

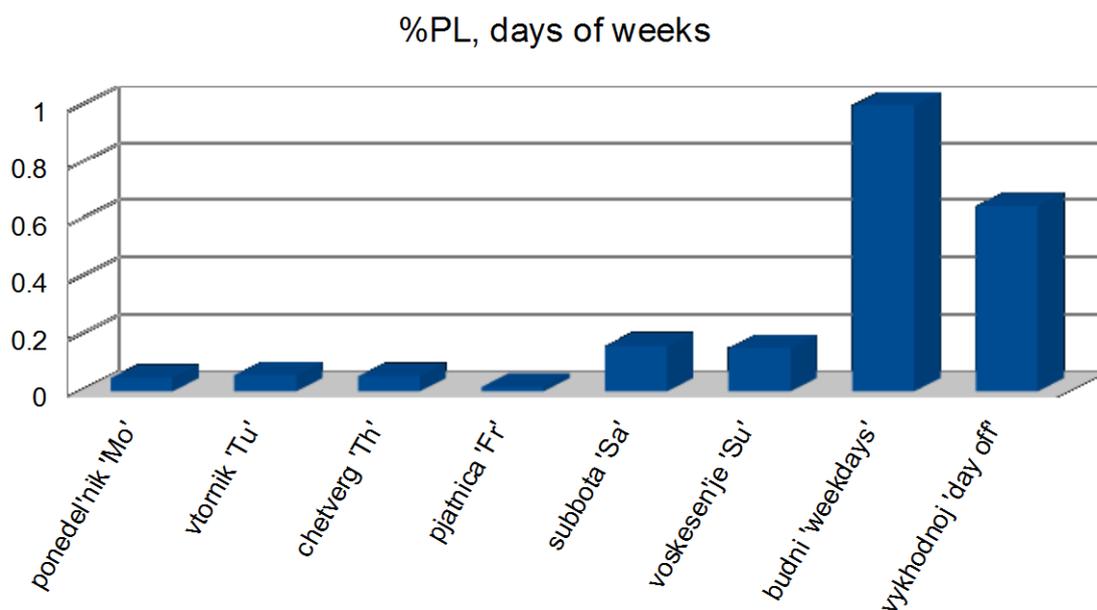


Fig 1. Ratio of plural forms: names of days of week⁴.

Substantive number is traditionally listed among the categories whose grammatical behavior is by and large explained by semantic factors (on the hypothesis of semantic motivatedness see Wierzbicka 1988; Lyashevskaya 2004). So far the behavior patterns under investigation are countability/uncountability and the absence of plural or singular forms in production, which defines the boundaries of the classes of singularia tantum and pluralia tantum. Grammatical profiles give us new behavioral data on what happens between the poles, so we believe that they can serve as a tool to explore more gradual, otherwise overlooked effects in lexical semantics.

³ A 6 MW corpus of texts created in 1760-2010 with manually disambiguated grammatical homonyms, available online at <http://ruscorpora.ru/mycorpora-main.html>.

⁴ Data for the noun *sreda* is not provided since its meanings 'Wednesday' and 'environment' are not distinguished in the corpus.

The rest of the paper is structured as follows. Section two gives an overview of the previous research. Section three presents the corpus data which are then analyzed in section 4; we report the behavior of nouns in four lexical groups: body part names (4.1), names of vehicles (4.2), kinship terms (4.3), and emotions (4.4). Section five concludes.

2. Background

The idea of using corpus data on grammatical frequencies in order to structure the lexicon is not new. Greenberg (1974/1990) suggested that different semantic groups might have different distributions of grammatical cases (both with and without prepositions). According to his hypothesis, the mean proportion of case forms within the group of abstract quality names (or for example, the names of body parts or measurements, etc.) differs from the mean proportion of case forms measured over the the entire set of vocabulary. Greenberg checked his hypothesis with the Russian frequency data he had at hand, which was not accidental: at that moment Russian was one of the few languages for which there was a dictionary with the frequency lists of cases and preposition-case combinations for each noun (Šteinfeldt 1963). Greenberg was looking for the “magic” ratio, which would allow him to group words in relevant semantic classes; unsurprisingly, he did not succeed. From today's perspective, his observations can be reinterpreted as a semantically motivated shift in the frequencies of grammatical forms. For example, the word which refer to forest tends to be used to a large extent in locative constructions (cf. *v lesu* ‘in the forest’; hence the large proportion of Locative case forms), while the name that means ‘path’ is used more often in the Dative case (cf. *po tropinke* ‘along the path’). However, the overall overlap of individual effects is too complex to make clear-cut lexical groupings possible.

Some 25 years later the first attempt to account for the singular and plural form frequencies of Russian nouns was made (Corbett et al. 2001). Since we are interested in the interaction of number forms and lexical classes, Brown et al. (2013) is more relevant for the purpose of our study. They make use of the frequency data from the Uppsala corpus of contemporary Russian to study the relationship between number use and the Smith-Stark cross-linguistic hierarchy of number availability:

Speaker > Addressee > Kin > Non-human rational > Human rational > Human non-rational > Animate > Concrete inanimate > Abstract inanimate

Brown et al. suggest that in morphologically rich languages like Russian, the animacy hierarchy predicts not only the availability of the category of number but that the higher or lower proportion of plural forms depends on the position of the name on the hierarchy. Even though Brown et al. had to reject their main hypothesis (cf. Fig. 2 which does not confirm that the plural proportions are decreasing as we move rightward along the hierarchy), they identify two interesting effects. Firstly, the middle part of the hierarchy (names of humans and animals) shows the highest plural proportions, with a strong decrease in the left part (2nd person pronouns, kinship terms and names of gods and creatures) and a stepwise decrease in the right part of the plot (concrete inanimate and abstract inanimate). Secondly, some nouns tend to be ‘more plural’ regardless of their lexical class; such items are plotted as outliers in kinship and abstract classes. Brown et al. define cases like *bliznec* ‘twin’ as ‘locally unmarked for plural’ (following Tiersma 1982), since their use in singular contexts is unusual and they are expected to occur mostly in the plural. In our view, grouping the main body of Russian inanimate nouns into two classes is too general to see the structure of the lexicon. In our research we focus on micro-effects within quite compact taxonomic groups.

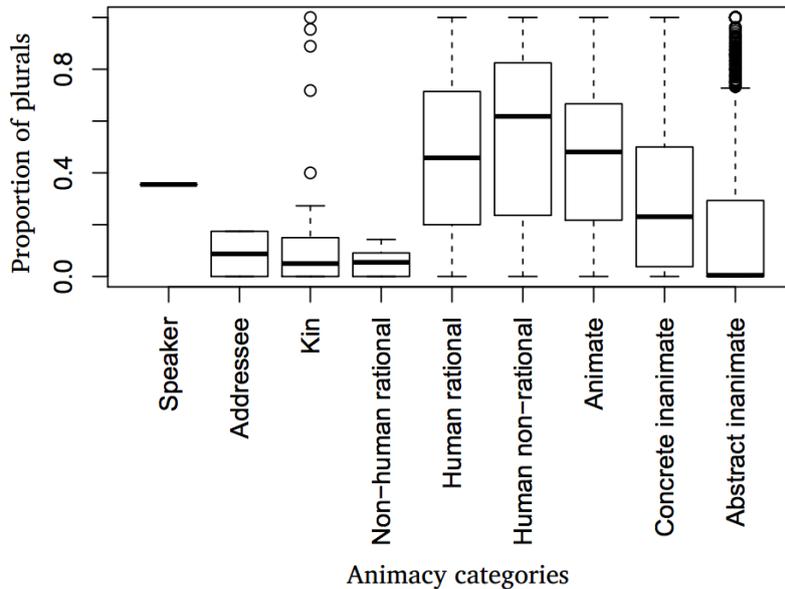


Fig. 2. The proportion of plural forms in lexical groups which correspond to the Smith-Stark hierarchy (Brown et al. 2013: 235).

3. Data

For the purpose of this study, we compiled a database of 2900 Russian nouns which occur 100 and more times in the Russian Standard subcorpus. The same threshold was used in the study of TAM-profiles⁵ of Russian verbs (Janda and Lyashevskaya 2011): we assume that the threshold of 5 occurrences used in Brown et al. (2013) is too low and can report false trends. Given these constraints, the database includes 413 common animate, 1006 common concrete inanimate, 1087 abstract, 319 proper animate, 41 proper inanimate nouns, and some borderline cases. Each noun was provided with the text frequencies of singular forms and plural forms (case distinctions were not taken into account), and the ratio of the plural to the total number of occurrences was calculated as the main indicator of number behavior, NumGP. We collected the 10 most frequent bigrams for each noun from the corpus in order to see in which patterns the word typically occurs. In addition, the nouns were tagged according to their lexical category (common *vs.* proper; animate *vs.* inanimate concrete *vs.* inanimate abstract), taxonomic class (e.g. human, animal, transport, emotions, etc., cf. the RNC classification⁶) and other known attributes of their behavior such as (a) singularia tantum *VS* pluralia tantum; b) countable *VS* uncountable; c) individual *VS* sort / mass / class *vs.* cumulative reference (cf. *bereza* ‘birch tree’, *rastenie* ‘plant’ and *rastitel’nost’* ‘vegetation’). Animate and inanimate nouns are treated separately. Items that fall into more than one taxonomic category (as attested in the corpus) are classified according to their prototypical meaning (e.g. *krylo* ‘wing’ is assigned to body parts and not to the parts of devices, cf. the wings of an airplane). Singular uses with paucal numerals (e. g. *dve ruki* ‘plant’) are excluded from analysis.

4. Case studies

We report four case studies for the lexical classes of body parts, vehicles, kinship terms, and emotions. For each group, we provide a list of nouns sorted by NumGP. Fig. 3 shows the distributions of NumGPs in these groups by presenting the median and interquartile range and allows us to compare these values with the distributions in the main lexical categories (common animate, inanimate concrete, abstract; proper animate, proper inanimate) and the overall distribution

⁵ Tense, aspect, and mood.

⁶ <http://ruscorpora.ru/en/corpora-sem.html>.

in our database. Our data in general agree with those provided by Brown et al. (2013) but demonstrate that specific lexical classes within larger lexical categories behave differently. Half of the body part terms are plural-oriented (NumGP > 50%), which contradicts the ‘usual’ behavior of concrete nouns. Two-thirds of emotional terms have NumGP < 10% while the median NumGP for the total body of abstract nouns is 13%. In the following subsections we look at the level of individual lexemes and explain their position on the NumGP scale. Our analysis involves the notion of functional frames which represent how objects/events are typically used and typically observed (cf. the similar concept of ‘functional structure’ in the lexical meaning in Pustejovsky 1991; also ‘functional predicate’ in Rakhilina 2000).

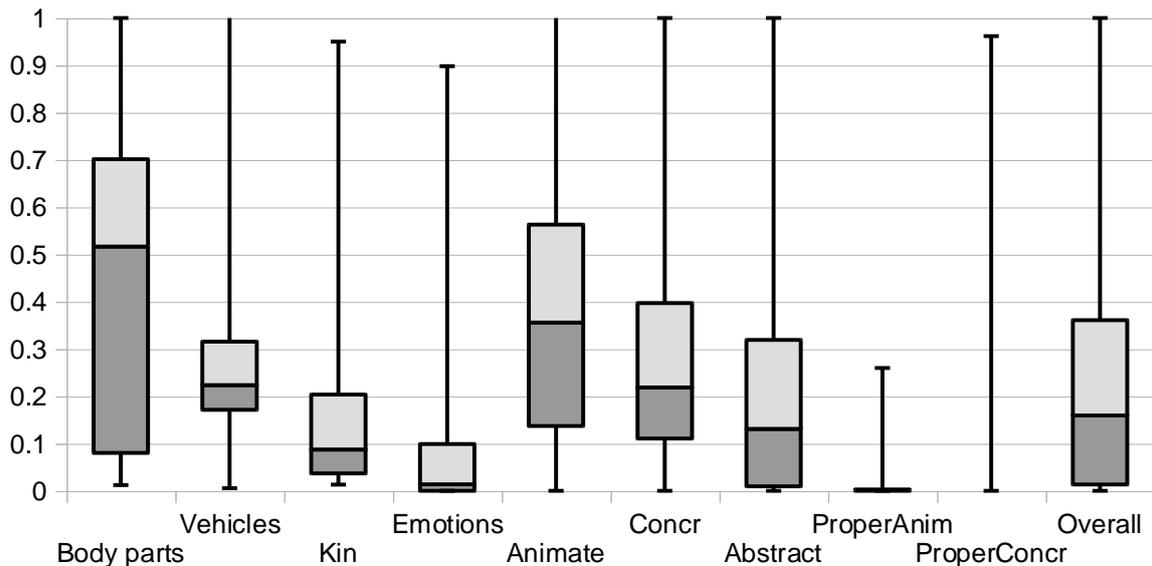


Fig. 3. Distributions of NumGP within target lexical classes and main lexical categories.

4.1. Body parts

The distribution of number GPs across the names of body parts (as well as other kinds of parts) is governed by the following hierarchy:

singular parts > parts that form pairs and sets

Moreover, the distinction between active (mobile) and passive (fixed) body parts may be of importance.

Noun	PL, %	PL, abs.f.	Noun	PL, %	PL, abs.f.	Noun	PL, %	PL, abs.f.
<i>usta</i> 'mouth'	100%	128	<i>palec</i> 'finger'	58%	897	<i>lico</i> 'face'	13%	612
<i>veko</i> 'eyelid'	90%	99	<i>lapa</i> 'paw'	58%	201	<i>duša</i> 'soul'	9%	224
<i>glaz</i> 'eye'	90%	5860	<i>plečo</i> 'shoulder'	57%	1004	<i>serdce</i> 'heart'	8%	172
<i>guba</i> 'lip'	87%	887	<i>visok</i> 'temple'	55%	97	<i>spina</i> 'back'	8%	87
<i>brov'</i> 'eyebrow'	87%	291	<i>ščeke</i> 'cheek'	55%	302	<i>golova</i> 'head'	7%	387
<i>zub</i> 'tooth'	82%	751	<i>ukho</i> 'ear'	53%	510	<i>rot</i> 'mouth'	6%	57
<i>ljogkoe</i> 'lung'	80%	86	<i>ruka</i> 'arm/hand'	52%	4588	<i>život</i> 'stomach'	6%	21
<i>nozdrja</i> 'nostril'	79%	85	<i>bedro</i> 'hip'	51%	62	<i>nos</i> 'nose'	5%	51
<i>koleno</i> 'knee'	74%	530	<i>lokot'</i> 'elbow'	34%	93	<i>lob</i> 'forehead'	5%	29
<i>rog</i> 'horn'	71%	93	<i>ladon'</i> 'palm'	29%	157	<i>šeja</i> 'neck'	5%	33
<i>krylo</i> 'wing'	70%	337	<i>kulak</i> 'fist'	28%	116	<i>grud'</i> 'breast'	4%	41
<i>noga</i> 'foot/leg'	70%	2356	<i>čerep</i> 'skull'	23%	56	<i>želudok</i> 'stomach'	2%	3
<i>nogot'</i> 'nail'	67%	136	<i>khvost</i> 'tail'	14%	54	<i>gorlo</i> 'throat'	1%	4
<i>oko</i> 'eye'	62%	72	<i>morda</i> 'snout'	13%	28	<i>podborodok</i> 'chin'	1%	2

Table 1. Number profiles of body part names.

Table 1 shows that the names of pairs and sets have 28% to 90% plural forms, while the names of single parts have less than 25% plural forms. However, within the subcategory of pairs and sets there are still some nouns the NumGPs of which are still significantly lower than expected, as they often refer to the one (selected from the pair / set) element which is in the focus of an observer (speaker). These cases include either an actively moving part of the body (cf. *stuknut' kulakom* 'to bang one's fist', *maknut' rukoj* 'to wave one's hand', *pogrozit' pal'cem* 'to wag one's finger', *čertit' nogtem* 'draw by one's fingernail') or an actively used location (cf. *povesit' na plečo* 'to hang on the shoulder', *skazat' na ukho* 'to whisper in one's ear', *teč' po ščeke* 'to flow frown one's cheek'). Thus, in accordance with the hierarchy of activity, *ruka* 'hand/arm' have more plural forms than *noga* 'foot/leg', and *palec* 'finger/toe' have more plural forms than *zub* 'tooth'). *Brov'* 'eyebrow' and *guba* 'lip' refer to the parts which are less 'active' than hands and fingers and are usually described as pairs.

4.2. Vehicles

The nouns in this group describe not only means of travel but also (and often) the location where a scene is developing, see Table 2. Therefore, it is important where the observer of the scene is situated, either s/he see the scene from outside (e.g. the observer looks at the road where a vehicle/vehicles move(s), cf. *mčatsja avtomobili*, *gruzoviki*, *avtobusy*, *poezda* 'rushing cars, trucks, buses, trains') or from inside where the vehicle is unique since the observer can not be situated in more than one closed space (cf. *v avtobuse bylo žarko* 'it was hot in the bus').

Noun	PL, %	PL, abs.f.	Noun	PL, %	PL, abs.f.	Noun	PL, %	PL, abs.f.
<i>sani</i> 'sledge'	100 %	136	<i>samoljot</i> 'plane'	22%	57	<i>paroxod</i> 'steamship'	15%	16
<i>tank</i> 'tank'	75%	170	<i>lokomotiv</i> 'locomotive'	20%	24	<i>velosiped</i> 'bicycle'	12%	12
<i>gruzovik</i> 'truck'	36%	77	<i>tramvaj</i> 'tram'	20%	156	<i>teležka</i> 'trolley'	12%	12
<i>avtomobil</i> 'car'	36%	246	<i>motocikl</i> 'motorcycle'	19%	66	<i>kareta</i> 'carriage'	10%	19
<i>traktor</i> 'tractor'	32%	45	<i>poezd</i> 'train'	18%	63	<i>taksi</i> 'cab'	1%	1
<i>korabl</i> 'ship'	29%	154	<i>avtobus</i> 'bus'	17%	25	<i>metro</i> 'subway'	1%	1
<i>vagon</i> 'coach'	28%	33	<i>lodka</i> 'boat'	16%	39			

Table 2. Number profiles of vehicle names.

The two names which have more than 50% plural forms are *sani* 'sledge' and *tank* 'tank'. The former is countable plurale tantum (named for the profiled pair of parts, see Lashevskaja 2002); the latter, according to corpus data, is described externally, usually as a set of vehicles used in military operations, cf.:

— *Вперёд, на Запад!.. Танки_{PL} идут ромбом!..* [Сергей Довлатов. Заповедник (1983)]
'Forward, to the West!.. **Tanks_{PL}** in rhombus formation!..'

The perspective of an inner observer sitting in the tank is attested in 3 of 58 singular uses, cf.:

Семнадцатилетним он сгорел в танке_{SG}, защищая от фашистов Украину. [Марина Палей. Поминование (1987)]
'He was seventeen when he burned in the **tank_{SG}**, protecting Ukraine from the fascists.'

Railway coaches which form sets are designated by the singular-oriented noun (NumGP = 28%): our data shows that it frequently refers to deictically unique locations, in the same way as *buses*, *boats*, *planes*, etc. which have NumGP ranging from 10 to 36%. Taxis, bicycles, motorcycles and carts are also, for the most part, deictically unique as they are mentioned as individual means/instruments in various motion frames. *Metro* 'subway' denotes not only the means of transport, but also the urban system (by default unique in the city), a space under ground, and a landmark on the ground (cf. *vstrečаемsja u metro* 'we are meeting at the subway').

4.3. Kinship terms

The deictical uniqueness is very characteristic of kinship terms as they refer only to the role of a person in the speech situation, for example, in family talks, cf. dad, mother, husband, wife, mother-in-law, stepfather and even babysitter, see Table 3. Dads, mothers, husbands and the like do not form any "natural" pairs or sets. What is more, the terms for sons, grandmothers, granddaughters, sisters, nephews and other family members which in principle could refer to more than one member of the family, are usually used as singular pointers to the only relative in the situation, for example, a grandmother who lives in the family or the only child.

"Ну, мне мама и бабушка_{SG} рассказывали, и я кое-что читал, да и в церкви слышал". [митрополит Антоний (Блум). О христианстве (1995)]

'Well, my mother and grandmother told me (about that), and I read a bit, and heard in the Church'.

Я, например, для внучки_{SG} настегла своими руками лоскутное одеяло, зная, что оно будет её оберегать, давать ей энергию. [Народный костюм: архаика или современность? // «Народное творчество», 2004]

'For example, I sewed a patchwork quilt for my **granddaughter_{GL}** with my own knowing that it will protect her and give her energy.'

Even though it is typical to have two grandfathers and two grandmothers, neither *deduški* nor *babuški* are “natural” pairs, because they do not live together and there is no common household (cf. the collocation *babuška i deduška*, which does not have a one-word equivalent).

Noun	PL, %	PL, abs.f.	Noun	PL, %	PL, abs.f.	Noun	PL, %	PL, abs.f.
<i>roditel'</i> ‘parent’	95%	933	<i>syn</i> ‘son’	11%	227	<i>djadja</i> ‘uncle’	4%	39
<i>predok</i> ‘ancestor’	84%	133	<i>nevesta</i> ‘bride’	10%	38	<i>ded</i> ‘grandfather’	4%	33
<i>potomok</i> ‘descendant’	77%	86	<i>žena</i> ‘wife’	9%	251	<i>mat'</i> ‘mother’	3%	94
<i>rodstvennik</i> ‘relative’	74%	301	<i>tjotka</i> ‘aunt’	9%	30	<i>mamen'ka</i> ‘mommy’	2%	2
<i>rebjonok</i> ‘child’	65%	2763	<i>dočka</i> ‘daughter’	9%	37	<i>mama</i> ‘mom’	2%	37
<i>vnuk</i> ‘grandson’	48%	127	<i>babka</i> ‘grandmother’	8%	12	<i>supruga</i> ‘spouse (f)’	2%	3
<i>brat</i> ‘brother’	26%	509	<i>batjuška</i> ‘father’	8%	27	<i>matuška</i> ‘mother’	2%	4
<i>sestra</i> ‘sister’	26%	243	<i>muž</i> ‘husband’	7%	118	<i>deduška</i> ‘grandfather’	2%	12
<i>ženikh</i> ‘fiancé (m)’	15%	41	<i>zjat'</i> ‘son in law’	6%	6	<i>tjotja</i> ‘aunt’	1%	6
<i>doč</i> ‘daughter’	12%	132	<i>djadjuška</i> ‘uncle’	6%	7	<i>papa</i> ‘dad’	1%	11
<i>vnučka</i> ‘granddaughter’	12%	15	<i>babuška</i> ‘grandmother’	4%	43			
<i>plemjannik</i> ‘nephew’	11%	13	<i>otec</i> ‘father’	4%	149			

Table 3. Number profiles of kinship terms.

The tradition of naming characters by their kin role in fiction and magazine texts as well as in everyday spoken discourse adds up to the low ratio of singular forms, cf.:

Новый год обернулся двойным праздником: Катерина — родная сестра_{SG} хозяина — приехала погостить из далёкой Сибири, правда ненадолго, проездом. Кoryтин сам ездил на станцию, к поезду, её встречать и привез прямо к накрытому столу. Сестра_{SG} на родине не гостила давно. Было о чём поговорить. Вот и просидели у ёлки далеко за полночь, пели и даже танцевали под музыку. Но по привычке и обычаю людей немолодых сестра_{SG} хозяина всё равно проснулась довольно рано. [Б. Екимов. Пиночет (1999)]

‘New Year turned into a double celebration: Katerina—the **sister**_{SG} of the owner—came to visit him from remote Siberia, however briefly, only passing through. Korytin himself went to the station, to meet her and bring directly to the served table. The **sister**_{SG} had not been at her parents home since long. They had a lot to talk about. They sat by New Year tree long past midnight, sang and even danced to the music. But nevertheless, out of the habit of elderly people, the owner's **sister**_{SG} woke up quite early.’

Вот, например, он очень любил маленького сына_{SG}, но: "Наркотик был дай Бог! Вернее, не дай Бог. Потому и ломка оказалась страшной". Любовь к сыну_{SG} оказалась наркотиком, а значит, её нужно вырезать из сердца под корень. Отцовская привязанность принесла только зло: сын_{SG} превратился в инфантильного жирного борова, который не может расстаться с беззаботным детством и не в силах взять на себя взрослую ответственность за свои поступки. [И. Новикова. Преодоление иллюзий (о романе Александра Мелихова «Любовь к отеческим гробам») // «Октябрь», 2003]

‘Well, for example, he loved his little **son**_{SG} but ‘The drug was a heaven! Or rather, heaven forbid. Therefore the drug withdrawal was terrible’. Love for his **son**_{SG} was a drug, which means it needed to be cut out of the heart at the root. Paternal affection brought only evil; his **son**_{SG} turned into an infantile fat hog, who could not part with his carefree childhood and was not able to take on adult responsibilities for his actions.’

As a result, a number of kin terms are used in plural only occasionally (less than 5 occurrences in our data set) and thus can be considered as “potential singularia tantum” (Čelcova 1976), cf. *test'* ‘father-in-law’, *tjošča* ‘mother-in-law’, *mačekha* ‘stepmother’, *kum* ‘godfather’, *prababka* ‘great-grandmother’, *batja* ‘dad’, *papen'ka* ‘daddy’, *matuška* ‘mother’, *papaša* ‘pop’.

In contrast, there is a group of kin names that have more than 50% plural uses, in which they refer to the class of relatives (cf. *rodstvenniki* ‘relatives’, *potomki* ‘descendants’, *predki* ‘ancestors’) or pairs (cf. *roditeli* ‘parents’ and a ‘reciprocal’ pair *suprugi* ‘spouses’). The singular forms of these words are often substituted in the lexical system by other, more frequent, nominations, cf. *roditel'* ‘parent’ such as *otec i mat'* ‘father and mother’; *suprugi* ‘spouses’ and *muž i žena* ‘husband and wife’; so they receive mostly stylistically marked use, cf. (high) *čado* ‘child’, (high or highly formal) *roditel'* ‘a parent’, *suprug* ‘spouse’, cf:

*Целеустремлённость родителей, решивших вывести своё **чaдо**_{SG} в люди, поистине безгранична.* [М. Давыдова. Кто в доме хозяин? (2003) // «100% здоровья», 2003.01.15]
 ‘The tenacity of parents who decided to bring their **child**_{SG} into the world is truly boundless.’

Супру_{SG}, которого дотолe баловали со всем пылом неистраченной материнской любви и нежности, вдруг оказывается в собственных глазах "третьим лишним". [М. Давыдова. Кто в доме хозяин? (2003) // «100% здоровья», 2003.01.15]
 ‘A **spouse**_{SG}, who hitherto had been pampered with all the fervor of unspent maternal love and tenderness, suddenly appears in his own eyes "the third wheel"’ [both sentences are an ironical stylization of the high style]

In addition, if a kinship name figuratively refers to the class of people of the same generation or social group (e.g. *brat'ja i sjostry* ‘brothers and sisters’, *vnuki* ‘grandchildren as descendants’, *dedy* ‘grandfathers as ancestors’), it receives higher proportion of plural forms, cf.:

*Настроим мы дач, и наши **внуки**_{PL} и **правнуки**_{PL} увидят тут новую жизнь...* [А. П. Чехов. Вишневы сад (1904)]
 ‘We’ll build the dachas and our **grandchildren**_{PL} and **great-grandchildren**_{PL} will see a new life here ...’
 — *Братья*_{PL} и *сёстры*_{PL} — *проникновенно* сказал он, — *у меня только что... от нежности содрогнулась душа.* [Василий Шукшин. Калина красная (1973)]
 ‘—**Brothers**_{PL} and **sisters**_{PL}—he said with feeling—I just ... my soul shuddered of tenderness.’

Thus, the semantic hierarchy of kinship terms which corresponds to NumGP is as follows:

deictically unique > pairs and sets > classes

Another functional-semantic factor is the proportion of appellative and hypocoristic uses. About half of the names with NumGP close to zero are primarily appellatives, cf. *mama* ‘mom’, *papa* ‘dad’. Hypocoristic uses (e.g. *babuška* ‘granny’, *matuška* ‘mother’) can be also associated with greater individualization.

We have already mentioned some disproportions of number use forced out by the structure of the lexical system. The same factor can explain the low ratio of plural forms in nomina feminina such as *supruga* ‘spouse (fem.)’. The plural form *suprugi* is a form of the noun *suprug* which refers to a masculine spouse in singular and mainly a couple in plural. Given that, the use of the female name in plural is limited to rather exotic examples, cf.:

*Дело в том, что со всеми своими тремя **супругами**_{PL} (в хронологическом порядке) я познакомился в одном*

4.4. Emotions

If we ignore words that occur less than 5 times in the plural (occasional uses), we see that this group can be divided into four cases: 1) singularia tantum, 2) names with a small NumGP (2% to 14%); 3) names with larger ratio of plural uses (22% to 67%), and 4) the class name *emocija* 'emotion' which is mostly plural (cf. also *čuvstvo* 'feeling' which is excluded because of homonymy), see Table 4.

Noun	PL, %	PL, abs.f.	Noun	PL, %	PL, abs.f.	Noun	PL, %	PL, abs.f.
<i>emocija</i> 'emotion'	90%	97	<i>udovol'stvie</i> 'pleasure'	5%	41	<i>toska</i> 'yearning'	0%	0
<i>pereživanie</i> 'feelings'	67%	78	<i>vesel'je</i> 'fun'	3%	3	<i>udivlenie</i> 'surprise'	0%	0
<i>stradanie</i> 'suffering'	50%	166	<i>dosada</i> 'vexation'	2%	4	<i>otčajanie</i> 'despair'	0%	0
<i>strast'</i> 'passion'	40%	183	<i>sožalenie</i> 'regret'	2%	9	<i>prezrenie</i> 'contempt'	0%	0
<i>obida</i> 'offense'	22%	64	<i>grust'</i> 'sorrow'	2%	2	<i>zavist'</i> 'envy'	0%	0
<i>trevoga</i> 'anxiety'	14%	43	<i>revnost'</i> 'jealousy'	1%	1	<i>skuka</i> 'boredom'	0%	0
<i>naslaždenie</i> 'pleasure'	11%	20	<i>zloba</i> 'malice'	1%	1	<i>styd</i> 'shame'	0%	0
<i>pečal'</i> 'sadness'	11%	21	<i>otvraščenie</i> 'aversion'	1%	1	<i>izumlenie</i> 'amazement'	0%	0
<i>radost'</i> 'joy'	10%	83	<i>voskhiščenie</i> 'admiration'	1%	1	<i>ispug</i> 'fright'	0%	0
<i>vostorg</i> 'delight'	8%	30	<i>ljubov'</i> 'love'	0%	6	<i>jarost'</i> 'rage'	0%	0
<i>užas</i> 'horror'	7%	47	<i>gore</i> 'grief'	0%	1	<i>ravnodušie</i> 'indifference'	0%	0
<i>strakh</i> 'fear'	6%	54	<i>sčast'je</i> 'happiness'	0%	0			
<i>sočuvstvie</i> 'sympathy'	6%	8 ⁷	<i>žalost'</i> 'pity'	0%	0			

Table 4. Number profiles of emotion names.

The plural forms can bear the following meanings: several speech acts expressing an emotional state or one longer act consisting of multiple sub-units (*vyslušivat' sožalenija* 'listen to one's regrets'); heterogeneous emotion, mixed feelings (*strakhi* 'fears'); several objects of emotional attitude (*izo vsekh ljubovej* 'among all persons I loved'); several events of the social interaction (*meždu pervoj i vtoroj ljubovjami* 'between first and second love'); 'sorts' of emotional states (*raznoobraznye naslaždenija* 'various types of pleasure'); several objects or situations which cause an emotion (*udovol'stvija* 'pleasures') and some other readings which relate to an emotion as if it consists of multiple units, cf.:

Я шёл так, словно мне шестнадцать лет, всё апрельское волнение_{SG} и юношеские страхи_{PL} воскресли во мне. [Василий Аксенов. Пора, мой друг, пора (1963)]

'I was walking as if I was sixteen, and all the excitement_{SG} of April and all the youth fears_{PL} rose within me'.
Любопытным получился образ бабки Зубра. Писал её учитель под Салтычиху, как нас учили про крепостников. Ругалась по-черному. Обливала девок кипятком в своих отчаянных злобах_{PL}. [Даниил

7 Five out of eight occurrences of *sočuvstvie* in plural in our database belong to Sergej Bočarov who discusses Apollon Grigoriev's essay on Puškin.

Гранин. Зубр (1987)]

'An image of Zubr's grandmother occurred to be interesting. The teacher posed her as Saltychikha, just as we were taught about the feudals. She swore her head off and poured boiling water over her serf girls in her desperate **malices_{PL}**'.

Some emotions tend to accumulate inside the subject, cf. *obidy* 'insults'. Emotions like *stradanija* 'sufferings' (in plural) are mostly heterogeneous; the singular form of the noun usually has either generic or physical reading (cf. *telesnoe stradanie* 'physical body suffering'). *Strasti* 'passions' (in plural) describes external manifestations of any strong feelings while *strast'* (in singular) is a particular (internal) emotion, cf.:

— Наталия Юрьевна, некоторое время назад в вашем театре разгорелся скандал, коллектив раскололся, об этом рассказывала и «ОГ». **Страсти_{PL}** поутихли? [Ида Рыбкина. Страна чудес открыла двери // «Общая газета», 1995]

'- Natalia Yurievna, some time ago there was a scandal in your theater, the team split up, OG newspaper also wrote about it. Have the **passions_{PL}** calmed down?'

The singular-oriented nouns (*singularia tantum* and those that occur in plural only occasionally, e. g. *udivlenie* 'surprise', *ispug* 'fright', *gore* 'grief', *revnost'* 'jealousy') are hardly consistent with the idea of heterogeneity and external manifestations and refer mostly to a current internal emotional state.

The most frequent n-grams in the RNC which include emotional terms show the difference in singular-oriented nouns and plural-oriented nouns, too, cf. *s trevogoj* 'with anxiety', *ne v obidu* 'no offense (intended), lit. not in offense', *strast' k* 'passion for', on the one hand, and *stradanija i X* 'sufferings and smth.', *pereživaniya i X* 'feelings and smth.', *emocii i X* 'emotions and smth.', on the other hand.

5. Conclusions

The lexical classes we have analyzed exhibit a variety of grammatical behavior including nouns with full paradigms, *singularia tantum*, *pluralia tantum*, paradigms with certain biases of singular and plural forms, with occasional uses of either singular or plural forms. Grammatical profiles help to highlight certain aspects of meaning which explains how their denotata typically function, how they are conceptualized (cf. countable, uncountable, and deictically unique nouns), and how the nouns are used in speech (cf. appellatives and idiomatic expressions).

References

- Brown D., Corbett G.G., Fedden S., Hippisley A., Marriott P. (2013). Grammatical typology and frequency analysis: number availability and number use, *Journal of Language Modelling*, Vol. 1 (2), pp. 227–241.
- Čelcova L.K. (1976). *Forma množstvenného čísla ako objekt leksikografie* [Plural number forms as an object of lexicography], Cand. sc. dissertation, Moscow.
- Corbett G.G. (2000). *Number*, Cambridge University Press, Cambridge.
- Corbett G.G., Hippisley A., Brown D., Marriott P. (2001). Frequency, regularity and the paradigm: a perspective from Russian on a complex relation, in Bybee J., Hopper P. (eds.), *Frequency and the Emergence of Linguistic Structure*, John Benjamins, Amsterdam, pp. 201–226.
- Greenberg J.H. (1974/1990). The relation of frequency to semantic feature in a case language (Russian), in Denning K., Kemmer S. (eds.), *On Language, Selected Writings of Joseph H. Greenberg*, Stanford, pp. 207–226.

- Janda L.A., Lyashevskaya O. (2011a), Aspectual pairs in the Russian National Corpus, *Scando-Slavica*, Vol. 57 (2), pp. 201-215.
- Janda L.A., Lyashevskaya O. (2011b). Grammatical profiles and the interaction of the lexicon with aspect, tense and mood in Russian, *Cognitive Linguistics*, Vol. 22 (4), pp. 719-763.
- Lashevskaja O.N. (2001). Slozhnyje predmety i ikh chasti [Complex objects and their parts], *Russian linguistics*, Vol. 25 (3), pp. 285-313.
- Lyashevskaya O. (2004). *Semantika russkogo chisla* [Semantics of Russian number], Moscow.
- Lyashevskaya O.N., Paducheva E.V. (2011). Ontologicheskie kategorii imen emocij [Names of emotions and their ontological categories], *Nauchnaja i tekhnicheskaja informacija. Ser. 2. Informacionnyje processy i sistemy* [Scientific and Technical Information, Ser. II, Information Processes and Systems], No. 5.
- Newman J. (2008), Aiming low in linguistics: Low-level generalizations in corpus-based research, in *Proceedings of the 11th International Symposium on Chinese Languages and Linguistics*, National Chiao Tung University, Hsinchu, Taiwan, May 24, 2008.
<http://www.ualberta.ca/~johnnewm/Aiming%20Low.pdf>.
- Pustejovsky J. (1991), The generative lexicon, *Computational Linguistics*, Vol. 17 (4), pp. 409-441.
- Rakhilina E.V. (2000). Kognitivnyj analiz predmetnykh imen: semantika i sochetaemost' [Cognitive Analysis of Concrete Nouns: Semantics and Collocation], Moscow, *Russkie slovari*.
- Šteinfeldt E. (1963). *Russian Word Count*, Moscow.
- Tiersma P.M. (1982). Local and general markedness, *Language*, Vol. 58, pp. 832–849.
- Wierzbicka A. (1988). *The Semantics of Grammar*, Amsterdam: John Benjamins.

Olga Lyashevskaya

National Research University Higher School of Economics, Faculty of Humanities, School of Linguistics, professor; Vinogradov Institute of the Russian Language RAS (Moscow, Russia), senior researcher. E-mail: olesar@yandex.ru.

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

© Lyashevskaya, 2015