

Elena Spaziani

Nizy, koški, peregovory and vybory: Lexicalization Profiles

1. Introduction

The phrase *pluralia tantum* refers to “nouns, like *oats* or *trousers*, which appear only in a plural form” (Matthews 2007, *pluralia tantum*). The formalist and structuralist traditions, especially in Russian studies, have approached nouns exhibiting a defective paradigm as arbitrary anomalies. However, their cross-linguistic diffusion and productivity (Corbett 2000; Gardelle, Acquaviva 2023; Koptjevskaja-Tamm, Wälchli 2001) prove that further investigation is necessary. Within the Canonical Typology framework, Corbett (2019) offers an extensive cross-linguistic typology of *pluralia tantum*. This is predicated on the synchronic state of different types of mismatches between the morphosemantic and morphosyntactic properties of nouns. As a result, the canonical *plurale tantum* must denote more than one object, while having a plural-only morphological inflection and syntactic agreement (e.g. German *Eltern* ‘parents’, Corbett 2019: 97). From a semantic perspective, the singular-plural idiosyncrasy finds little synchronic motivation. Conversely, diachronic studies offer a more adequate explanation, interpreting these cases as the result of lexicalization, i.e. a historical process leading to semantic opacity and morphological fusion among its components (Lipka *et al.* 2004: 7).

Accordingly, a distinction is drawn between a compositional, non-lexicalized plural and a non-compositional, lexicalized plural¹. The latter does not have a grammatical interpretation, as it contributes to the lexical meaning of the noun (see, e.g., Alexiadou 2019). Lexical plurality encompasses several defective conditions which extend beyond plurals like *oats* “whose only distinctive property is that they have no singular” (Acquaviva 2008: 16). Another relevant condition is exemplified by lexicalized plurals that are semantically isolated from their formal singular correspondent (e.g. English *fund* – *funds*, Corbett 2000: 176). The first type will be addressed in this work as morphological *plurale tantum*, while the second type as semantic *plurale tantum*. Several “polysemy analyses” (Corbett 2019: 62) focus on the defective expansion of Russian nouns, uncovering its dynamicity and underlying semantic mechanisms. Among these, a philological reconstruction of the

¹ “If the meaning of a plural noun cannot be completely inferred from the meaning of the base and of the plural morpheme, the plurality of that noun is lexical in a semantic sense: *books* is compositional and inflectional, but *goods* is lexical, because its sense of ‘merchandise’ remains inaccessible if one only knows the meaning of *good* and that of *-s*.” (Acquaviva 2008: 43).

FIGURE 1
Conceptual illustration of the lexicalization process through metonymy

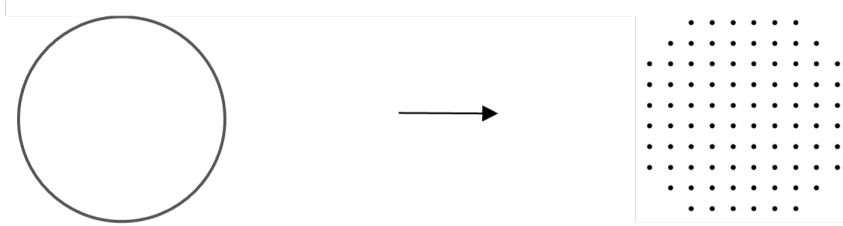
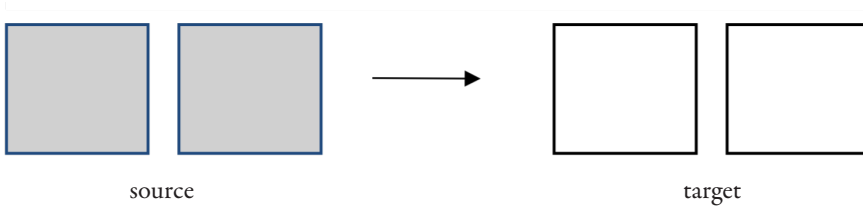


FIGURE 2
Conceptual illustration of the lexicalization process through metaphor



semantic and morphological *pluralia tantum* in the Slavic sub-family is offered by Degtjar-ev (2014). According to Braun (1930), the defective expansion undergoes an organic formation, whereby the semantic *pluralia tantum* are regarded as the antecedent stage of the morphological *pluralia tantum*. Finally, a cognitive perspective by Šemaeva (2014; 2016) and Besedina *et al.* (2015) unveils the mechanisms of the lexicalization, which involves two phases. The grammatical plurality is defocused and a subsequent semantic shift, through either metonymy or metaphor, occurs. Metonymy and its mereological implications have been shown to be associated with grammatical number (Brdar 2007) and lexical plurality, particularly in nouns denoting spatial or temporal extensions (Aquaviva 2016). In his overview on metonymy, Koch (2001) defines shifts within the categories of number as contiguity-based relations involving a figure / ground effect, which, according to Blank (1997), do not involve semantic change. The same shift is referred to as a neutralization of the grammatical opposition by Zaliznjak and Padučeva (1974). This occurs when the singular form expresses a plural meaning, leading to a state of functional synonymy, most prominently in specialized contexts. Percov (2001) observes that, even in the case of a neutralization, the grammatical forms maintain their semantic invariants, as the singular refers to a single class of items, while the plural refers to multiple items in the class, as **FIGURE 1** shows.

On the other hand, in a metaphorical model, the conceptual and semantic derivation draws on features related to plurality, more often to duality, and a concept from another domain; thereby, a plural source concept such as *eyes* motivates a new plural meaning in the target concept, as in *bags* ‘dark circles under the eyes’ (see Šemaeva 2014; 2016). This model, exemplified by **FIGURE 2**, accounts for analogical defective expansions, which quite

productively involve the class of nouns denoting bipartites and induments, as Šaljan (2015) and Ljaševskaja (2004) note (e.g., *адидасы* ‘Adidas sneakers’).

Despite the valuable results offered by these works, the semantic evolution within the paradigms leading to different types of *pluralia tantum* remains unsystematized, thereby precluding the formulation of a predictive model for defective expansion. Therefore, the aim of the present corpus-driven investigation is to explore modes of defectiveness in standard Russian morphology. This is achieved by identifying recurrent patterns through a quantitative diachronic analysis of the lexicalization of four nouns: *низ* ‘bottom’; *кошка* ‘cat’; *непезовор* ‘discussion’ and *выбор* ‘choice’. Specifically, relying on a multi-layered annotation of concordance lines involving both the singular and the plural forms of the selected nouns, the type of defective expansion is assessed, focusing on the frequency of individual senses and on the evolving distributional proximity between the two forms. The time span considered ranges from the 18th century, for which the singular form of many morphological *pluralia tantum* is attested (Zacharova 2009), to the 21st century. Within this timeframe, lexicographical sources indicate that the nouns selected for this study display a regular paradigm in the earlier period, but progressively evolve toward a plural lexicalization in the modern stage: *непезоворы* is lemmatized as a morphological *plurale tantum*, together with *выборы*, despite the existence of a singular lemma *выбор*; while some senses for the plural of *низ* and *кошка* are registered as semantic *pluralia tantum*.

The paper is structured as follows: the second section summarizes the defective evolution of the nouns based on two dictionaries, corresponding respectively to the 18th and 21st centuries. Section 3 sets forth the method used to annotate and analyze the development of the nouns – the *Behavioral profiles* procedure (Gries 2010). The number of concordance lines retrieved from the corpus is outlined in Section 4. The linguistic categories selected for the annotation are listed and motivated in Section 5. As the annotation constitutes the core of the study, relevant aspects pertaining to it are discussed in Section 6, mostly concerning the rationale behind certain judgments and observations related to specific contexts. In the sections dedicated to the results, the sense-level frequency of the forms and the annotated profiles are shown and statistically analyzed, leading to concluding remarks in the final paragraph.

2. Defective Evolution Across Two Dictionaries

A first overview of the defective evolution of the nouns selected was obtained comparing their senses in two representative dictionaries: *Slovar’ russkogo jazyka XVIII veka* (henceforward SRJa) and *Bolšoj Akademičeskij Slovar’* (henceforward BAS). The former is a historical dictionary that depicts the linguistic state of the 18th century only. The latter is regarded as the most comprehensive dictionary of the contemporary Russian language, from Puškin to modern days. The lexicographical portrait of the four paradigms between the two time periods indicates that a significant number of senses underwent a process of lexicalization in the plural form. This evolution is evident in the shift of lexicographical labels

TABLE I
Defective evolution across SRJa and BAS

Sense	SRJa	BAS
<i>низы</i> 'low-lying areas'	no label	<i>только мн.</i> 'plural only'
<i>низы</i> 'low notes'	absent	<i>обычно мн.</i> 'usually in the plural'
<i>низы</i> 'lower classes'	absent	<i>только мн.</i> 'plural only'
<i>кошки</i> 'multi-tailed whip'	<i>чаще мн.</i> 'more frequent in the plural'	<i>только мн.</i> 'plural only'
<i>кошки</i> 'crampons'	absent	<i>только мн.</i> 'plural only'
<i>кошки</i> 'Felidae'	absent	<i>только мн.</i> 'plural only'
<i>переговоры</i> 'negotiations'	no label	plural lemma
<i>выборы</i> 'elections'	no label	plural lemma

across the two dictionaries. In some cases, the entry evolves from the absence of any label indicating defectiveness to the introduction of a label marking predominant or exclusive plural usage (*обычно мн.* 'usually in the plural' or *только мн.* 'only in the plural'); in others, a label marking predominant plural use (*обычно мн.* or *чаще мн.* 'usually / more frequent in the plural') is replaced by a label or lemma indicating exclusive plural usage (*только мн.* 'only in the plural'). The senses labeled as *только мн.* 'only in the plural' in the modern dictionary will be regarded here as semantic *pluralia tantum*; while the lemmas registered only in the plural will be considered morphological *pluralia tantum*². TABLE I provides a summary of the functional shifts observed in the change of lexicographical labels, while the following is a more detailed account for each paradigm.

Низ – *низы* 'bottom – bottoms' was originally a functional word, attested as an adverb in the *Staroslavjanskij slovar'*, a historical dictionary of Old Slavonic based on manuscript evidence from the 10th-11th centuries (Cejtlin *et al.* 1994: 379). The SRJa registers its lexical evolution as a noun with predominantly singular usage and little to no indication of specialized plural senses. Only at a later stage does the paradigm further develop and lexicalize distinct senses associated with the plural. For example, the sense 'low-lying areas' is recorded in the SRJa without remarks on its plural use (Sorokin 1998: 136-137), whereas the BAS explicitly registers it as used exclusively in

² The selection of *pluralia tantum* based on lexicographical labels serves as a useful starting point, providing an initial overview of their defective evolution. However, it should be noted that this selection relies on specific lexicographical sources, whose criteria may differ from those adopted in others, as shown by Külmoja *et al.* (2020).

the plural (Gerd 2009: 407-408). Likewise, the paradigm develops two additional senses confined to the plural form, namely 'low notes' and 'lower classes,' both absent in the SRJA.

Кошка – кошки 'cat – cats' presents a similar evolution (see Sorokin 2005: 215-216; Gerd 2011: 561-563). The sense 'multi-tailed whip,' shifts from a predominant to an exclusive use in the plural form³, while the senses 'crampons' and 'Felidae,' both absent in the SRJA, emerge in the BAS as semantic *pluralia tantum*.

Переговор – переговоры 'discussion – negotiations' boasts a regular paradigm in the SRJA (Sorokin 2011: 49-50) as its entry is registered in the singular form and the sense 'negotiations' has no indication about its use in the plural. On the other hand, in the BAS (Gorbačević 2007: 60-61) it is lemmatized in the plural form as a morphological *plurale tantum*, while admitting an occasional use for the singular form, labeled as *устар.* 'historical' in the sense 'discussion'.

Выбор – выборы 'choice – elections' has a unique evolution as it displays a detachment of the plural form for the derived sense 'elections' from the original paradigm. Similarly to *переговор* 'negotiations,' in the SRJA (Sorokin 1988: 189-190) this sense is registered under a lemma in the singular form with no specific indication of its use in the plural; while in the BAS it is lemmatized as a morphological *plurale tantum*, separately from the primary sense 'choice' under the singular lemma (see Gorbačević 2005: 348-349 for *выбор* 'choice' and Gorbačević 2005: 350 for *выборы* 'elections').

3. *Lexicalization Profiles*

As Zaliznjak (2004) notes, the numerous factors underlying semantic derivation are often examined separately and only with respect to individual lexical items. For this reason, a more accurate account of plural lexicalization should rely on a multidimensional analysis. In the current study, a corpus-driven approach was adopted to address this requirement and, specifically, the *Behavioral profiles* procedure developed by Gries (2010) was chosen. This method was developed within the framework of Corpus Linguistics, motivated by the paradigmatic principle of distributional semantics by Firth (1957: 11) "you shall know a word by the company it keeps". It involves both qualitative and statistical analysis of a large collection of data and it is based on a comprehensive annotation of a given lexical item within concordance lines extracted from a corpus. Given its flexibility, i.e. the possibility to include multiple linguistic categories and features in the annotation (syntax, morphology, semantics, etc.), the procedure has been applied to several theoretical questions, mainly from a synchronic perspective, e.g. quasi-synonymy, polysemy (Divjak, Gries 2006; Gries

³ Vasmer (1950-1958: 360) and Berneker (1924: 589) link the sense 'multi-tailed whip' to its German equivalent *Katze* 'cat', in the meaning *Knute* 'knout', which, by contrast, shows no indication of defectiveness.

2019). Recently, it has also been extended to a diachronic perspective in the analysis of the Spanish verb *sentir* (Jansegers, Gries 2020).

Drawing on the latter work, this study develops and applies a tailored version of the profiling method to lexicalized plurals, termed *Lexicalization Profiles*. The main difference from the original profiling lies in the choice of the linguistic categories used for annotation: beyond the basic syntagmatic and paradigmatic information, the sense relations involved in a polysemous chain were included to monitor the process of lexicalization. As a result, *Lexicalization Profiles* combines the qualitative aspect of a traditional polysemy analysis and the quantitative aspect of a proper corpus-driven profiling. As for the procedure, its phases strictly follow the original workflow. Firstly, the data were manually retrieved extracting concordance lines for a given form from a balanced corpus. For this purpose, The Main Section of the Russian National Corpus (henceforward RNC) was selected, as it encompasses various text themes and genres that reflect the Russian standard written language from the 18th century to modern days. Secondly, the retrieved contexts were annotated according to a sense inventory and linguistic categories, which yield a set of features. In the final stage, the annotated data were converted into a co-occurrence table, containing the profile of each sense, i.e. the relative frequency of co-occurrence between each sense and each feature. Finally, the profiles of each sense were explored statistically.

4. Data Retrieval

In the data retrieval, four time periods were selected, ranging from the 18th century to the 21st century: 1700-1799 (T1); 1800-1899 (T2); 1900-1999 (T3); 2000-2025 (T4)⁴. Contexts were retrieved from the Main Section of the RNC using a systematic-random sampling procedure: after querying the corpus for both the singular and plural forms of the four nouns, one context was randomly sampled with an interval of five years⁵. Altogether, 401 contexts were considered, as detailed in TABLE 2. The table also presents the total number of instances for each noun form in the Main Section of the RNC, revealing a higher frequency of the plural in *переговор* and *выбор* compared with *низ* and *кошка*, which aligns with their established lexicographical status.

⁴ It must be clarified that this type of periodization is not intended to reflect stages in the development of the Russian standard language norm, but rather to ensure a statistically robust sample. Century-based segmentation favors data comparability across languages and case studies, aligning with established practices in quantitative, corpus-driven diachronic research (cfr. Hilpert, Gries 2016: 49). Owing to the smaller dataset and the narrower temporal scope of the present study compared to Jansegers and Gries 2020, the original periodization procedure, consisting of five 200-year intervals, was adapted accordingly.

⁵ The lexicographical independence of the plural *выборы* is reflected in its dual lemmatization in the RNC. Consequently, the contexts for the plural form were retrieved alternately from both the singular and the plural lemmas.

TABLE 2
Contexts sampled for the two forms of the four nouns in each time period (T)
and total occurrences from the main section of the RNC

Time Period	<i>низ</i>	<i>низы</i>	<i>кошка</i>	<i>кошки</i>	<i>переговор</i>	<i>переговоры</i>	<i>выбор</i>	<i>выборы</i>
T1	16	–	9	9	8	8	16	8
T2	21	8	20	19	16	20	20	19
T3	20	20	20	20	11	20	20	20
T4	5	5	4	5	1	5	5	5
Total	62	32	53	52	36	53	61	52
Total RNC (Main)	3,939	1,429	11,291	5,865	66	19,277	29,874	22,301

5. *Multi-layered Annotation Scheme*

The retrieved concordance lines were annotated using a multi-layered scheme. A first layer was dedicated to sense disambiguation, involving the selection of the appropriate sense for each form in its contextual occurrence. A sense inventory was adopted for this purpose, and it was obtained by factorizing the two dictionaries considered (SRJa and BAS), with additional sources being added during the annotation process. A fine-grained sense distinction was adopted, wherein the semantic extensions delineated by the dictionaries (usually through symbols like // or ||) were regarded as separate senses. The result of this factorization is a radial-chain model of semantic derivation in which the senses are ordered both radially from the primary sense and sequentially from derived senses (Apresjan 1995: 182). This model is encoded in the structure of an identifying numerical label (see FIGURE 3). The numerical component preceding the decimal point marks the order of radial derivation from the primary sense. The component following the decimal point indicates the order of chain derivation, specifying the sequential development of senses that arise from an already derived meaning. For instance, the primary sense of *низ* ‘lower part’ is

FIGURE 3
Radial-chain sense label structure

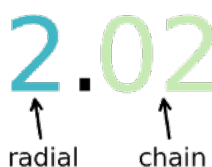


TABLE 3
Overview of the feature set for each linguistic layer

Layer	Category	Feature
1	primary; metonymic; metaphoric; taxonomic	X; X – X _i ; X = Y; X <> Y
2	abstractness; concreteness; animacy	abstract; concrete; inanimate; animate
3	grammatical case	nominative; genitive; dative; accusative; instrumental; prepositional
4	colligation L ₁ ; colligation R ₁	noun; verb; adjective; preposition; adverb; conjunction; pronoun; numeral; particle
	diaphasic preference	neutral; specialized
	text theme	daily life; politics; economy; pedagogy; education; law; religion; administration; war; biology; physics; arts; history; medicine; philosophy; travel; geography; nature; technology; sports; music; cuisine

assigned the label 1.01, while a directly derived sense such as *нуж* ‘lower floor’ is labeled 1.02. During annotation, each form was semantically disambiguated by assigning a numerical label that identifies its corresponding sense in the predefined sense inventory.

Four additional layers encoding linguistic features along the paradigmatic and syntagmatic axes were selected. Altogether, a set of 57 features was considered (see the comprehensive overview in TABLE 3) according to which 84 senses were annotated.

The first of these layers was intended not only to characterize the form within its contextual framework, but also to provide further metalinguistic interpretation, aligned with extant literature concerning the lexicalization of plurals (see Section 1). These features make explicit the semantic derivation encoded in the sense labels by specifying the type of relation among the senses and identifying potentially lexicalized plural. In particular, metonymic, metaphoric, and taxonomic relations were selected, following Blank’s theoretical framework (1997; 2003). The second and third layers aim to describe the semantic and morphological properties of a given sense in a specific context. With respect to semantic properties, the categories of abstractness / concreteness and animacy / inanimacy were selected, given their close involvement in the process of sense derivation. As noted by Birich (1995), such shifts are frequently observed, for instance from concrete to abstract meanings, particularly in cases of metaphorization (see Skljarevskaja 1993). Among morphological properties, grammatical case was annotated for each form in context, due to its interaction with grammatical number (see Plungjan 2011). Finally, three distributional categories were considered, in order to capture the syntagmatic properties of each form:

TABLE 4
Annotation of the singular form *низ*

Context	Sense	Sense Relations	Semantic Categories	Morphological Categories	Colligation LI/RI	Diaphasic Preference	Text Theme
тогда <i>низ</i> церкви, даже в том виде, как он есть, был бы очень хорош. (Н.А. Бестужев. <i>Письмо А.С. Связевой</i> , 1843)	1.02	X<>Y	concrete; inanimate	nominative	adverb; noun	neutral	daily life

- a) The first one pertains to the colligational pattern, that is the grammatical structure with which the form co-occurs. This construction, and in general the combinatorial profile, is considered a relevant factor in semantic change, as it can contribute to a sense modulation with respect to the individual context, provoking an effect of semantic reversal (Sinclair 1998). Specifically, many (see Blank 1997; Dobrovolskij 2004) note how in occasional or unusual combinations new senses may arise and potentially enter the standard language over time. In this case, the context window immediately next to the target form was annotated, consisting of a word on the left (LI) and a word on the right (RI).
- b) The second category concerns what was here defined as ‘diaphasic preference’, which draws from Sinclair’s (1998) ‘semantic preference’ and from Biber’s (1994) ‘level of discussion’. This level refers to the co-selection of the register in each concordance line, which was annotated as either neutral, in the event of surrounding high-frequency lexical items, or specialized, in the event of surrounding specialized terms. The latter is particularly relevant in the emergence of new senses, especially in the plural form (see Prochorova 1996), which eventually penetrate the standard language, involving a ‘diasystemic change’ (Blank 1997).
- c) The textual category focuses on the specific topic of a text. As Biber (1994) notes, text themes are important to discern different sub-registers, and they are usually presented as an open class of features. In the present framework, the selection of the individual features was mostly based on the meta-data of the Main Section in the RNC and a specific text theme was annotated for each context.

The annotation scheme is outlined in **TABLE 4**, where a sample context is annotated to demonstrate the procedure⁶.

⁶ Hereinafter the examples and their sources are taken from the RNC. Unless otherwise indicated, all translations are provided by the author.

6. Annotation of the Lexicalized Plurals: A Distributional Principle

The main difficulty related to the annotation pertains to the stages of lexicalization and to the diachronic evolution of the sense relations. A metonymic relation was annotated for senses undergoing a semantic shift from the singular to the plural form, without altering the meaning; whereas a metaphoric relation was annotated for senses lexicalized in the plural by analogy to nouns characterized by a condition of plural dominance (see Section 1). In both cases, the lexicalization involves a loss in the morpho-semantic compositionality, whereby the plural form evolves into a conventionalized lexical item with an autonomous semantic status. As this is a rather complex phenomenon to evaluate from a diachronic perspective, a distributional principle was established to distinguish between a grammatical (compositional) and a lexicalized plural, whereby:

- a) A plural is considered grammatical if it is surrounded by other nouns, belonging to the same lexical field and displaying a regular paradigm, which are also used in the plural, with a grammatical-distributive function:

(1) A.S. Šiškov, *Zapiski* (1780-1814)

Между тьмь какъ пересылки и
 between that-N.INS.SG how correspondence-F.NOM.PL and
переговоры сии съ графомъ
 negotiation-M.NOM.PL this-M.NOM.PL with count-M.INS.SG
Аракчеевымъ продолжались
 Arakčeev- M.INS.SG continue-PST-IPFV-REFL-3.PL

‘Meanwhile, the correspondences and negotiations with Count Arakčeev continued.’

(2) Petr III, *Manifest o darovanii vol'nosti i svobody vsemu rossijskomu dvorjanstvu* (1762)

в губерниях и провинциях меж собой
 in gubernija-F.PRP.PL and province-F.PRP.PL between self-N.INS.SG
выборы чинить, объявляя
 election- M.ACC.PL carry.out-INF announce-PTCP- ADV-PRS
только, кто выбран будет
 only who-NOM.SG PTCP-PASS-M.NOM.SG be-FUT-3.PL

‘In gubernijas and provinces, to conduct the elections among themselves, announcing only who will be elected.’

(3) A.I. Bogdanov, *Opisanie Sankpeterburga* (1751)

кошками и линьками гнаты
 multi.tailed.whip-F.INS.PL and whip-M.INS.PL chase-PTCP-PASS-M.NOM.PL
были
 be-PST-3.PL

‘They were chased with multi-tailed whips and flogging whips.’

- b) A plural is considered lexicalized if it is surrounded by: other nouns, belonging to the same lexical field and displaying a regular paradigm, which are used in the singular form with a denominative function; other nouns displaying a defective paradigm, used in the plural form with a denominative function:

(4) N. M. Murav'ev, *Konstitucija* (1825)

<i>Он</i>	<i>имеет</i>	<i>право</i>	<i>созывать</i>	<i>обе</i>
he-3.SG	have-PRS.3.SG	right-N.ACC.SG	summon-INF	both-F.ACC.PL
<i>Палаты</i>	<i>и</i>	<i>Верховную</i>	<i>Думу</i>	<i>в</i>
Chamber-F.ACC.PL	and	Supreme-F.ACC.SG	Duma-F.ACC.SG	in
<i>случае</i>	<i>переговоров</i>	<i>или</i>	<i>суда</i>	
case-M.PR.P.SG	negotiation- M.GEN.PL	or	trial-M.GEN.SG	

'He has the right to summon both Chambers and the Supreme Duma in case of negotiations or trial.'

(5) Ekaterina II, *Gramota na prava i vygody gorodam Rossijskoj Imperii* (1785)

<i>какъ</i>	<i>для</i>	<i>дозволенныхъ</i>	<i>городовымъ</i>
as	for	permit-PTCP-PASS-M.GEN.PL	urban-M.DAT.PL
<i>обывателямъ</i>	<i>выборовъ,</i>	<i>такъ и</i>	<i>для</i>
inhabitant-M.DAT.PL	election-M.GEN.PL	as and for	
<i>выслушанія</i>	<i>предложеній</i>	<i>Генераль-Губернатора</i>	
listening-N.GEN.SG	suggestion-N.GEN.PL	Governor-General-M.GEN.SG	

'Both for the elections permitted by the inhabitants of the city and for listening to the suggestions of the Governor General.'

(6) A.T. Bolotov, *Žizn' i priklučenija Andreja Bolotova, opisannye samim im dlja svoich potomkov. Pisma 91-96* (1800)

<i>Прежде</i>	<i>бывшее</i>	<i>наказание</i>	<i>солдат</i>
before	former- N.NOM.SG	punishment-N.NOM.SG	Soldier-M.GEN.PL
<i>и</i>	<i>всех</i>	<i>военных</i>	<i>батожемъ,</i>
and	all-M.GEN.PL	military-M.GEN.PL	baton-N.INS.SG
			<i>кошками</i>
			whip-F.INS.PL
<i>и</i>	<i>кнутом</i>	<i>отменено</i>	
and	knout-M.INS.SG	abolish-PTCP-PASS-N.NOM.SG	

'The former punishment of soldiers and all militaries by horsewhip, multi-tailed whip and knut has been abolished.'

(7) S. Koljubakin, *Zemlja dalëkaja i blizkaja* (1975)

<i>снимают</i>	<i>лыжи,</i>	<i>надевают</i>	<i>кошки,</i>
take.off-PRS-3.PL	ski-F.ACC.PL	put.on-PRS-3.PL	crampon-F.ACC.PL
<i>разматывают</i>	<i>веревки</i>		
unwind-PRS-3.PL	rope-F.ACC.PL		

'(they) are taking off their skis, putting on crampons, unwinding the ropes.'

As evidenced by the contexts, semantic and syntactic coordination serves as an indicator of such a distinction, as it establishes a relation of functional equivalence between the plural form and its co-occurring noun. Accordingly, a difference between the two stages of plurals can be noted in (1) and (4). In (1) the plural form *переговоры* ‘negotiations’ is co-occurring with the plural form of *пересылка* ‘correspondences’. The latter is a regular, singular-dominant noun which, in this context, conveys grammatical plurality. As indicated by the coordination with the conjunction *и* ‘and’, *переговоры* is aligned with the same grammatical function, yielding an iterative, plural interpretation. By contrast, in (4) the plural form *переговоры* co-occurs with the regular, singular-dominant noun *суд* ‘trial’. Their coordination within the hypothetical construction *в случае чего* ‘in case of’ suggests a neutralization of the singular-plural opposition (see Section 1), with both forms being used equivalently with a denominative function. Indeed, the legislative text refers to both *суд* ‘trial’ and *переговоры* ‘negotiations’, not as definite multiple events, rather as a hypothetical complex event, constituted by multiple phases. Likewise, in (5) *выборы* refers to a single event, on par with *выслушание* ‘hearing’⁷. In the third example, *кошки* ‘multi-tailed whip’ appears with *линьки* ‘whips’, used here as a grammatical plural, as the reference is to multiple objects. Whereas in (6) *кошки* is co-occurring with two singular nouns, *бато́жье* ‘baton’ and *кну́том* ‘knout’. Here, the context is not informing about a banishment against multiple instruments, but against the type of corporal punishment, denoted by the plural in a denominative function. As for *кошки* in (7), the sense ‘crampons’ may be developed by analogy with such nouns as *лыжи* ‘skis’ or *ботинки* ‘boots’ whose semantics alludes to items needed and used in pairs, as outlined in Section 1. Prochorova (1996) classifies both *кошки* ‘crampons’ and *лыжи* ‘skis’ as lexicalized plurals denoting bipartite items, mostly derived through metaphorization. Likewise, Skljarevskaja (1993: 43) cites *кошки* ‘crampons’ as an example of a genetic or lexicalized metaphor. The metaphoric model is also observable at the syntagmatic level, under the assumption that nouns following this pattern tend to occur in similar contextual environments. In (7), *кошки* ‘crampons’ is co-occurrent with *лыжи* ‘skis’, with whom it shares the same colligational pattern: both *лыжи* and *кошки* can be the object of the verb *надеть* ‘to put on’ and *снимать* ‘to take off’.

7. Results: Co-occurrence Tables and Sense-Level Frequency

The annotation was followed by the creation of a co-occurrence table showing the relative frequency of co-occurrence between the 57 features and each of the 84 senses annotated, which corresponds to their profiles. **TABLE 5** illustrates a fragment of the resulting co-occurrences in the plural form of *переговоры*, which quantifies the evolution of the features. The lexicalization in the sense ‘negotiations’ is visible through the change in the morphological and colligational patterns, which correlates with metonymy: as the latter increases in

⁷ Within the contexts retrieved, both lexicalized plurals appear for the first time in specialized-legislative texts, respectively a draft constitution by the Decembrist movement and a legislative act issued by Catherine the Great.

TABLE 5

Relative co-occurrence frequency between senses and features of the plural form in *переговор* across the four time periods (1.01: 'discussion'; 2.01: 'negotiations'; 3.01: 'condemnation')

Category	Feature		переговоры T1			переговоры T2		переговоры T3		переговоры T4	
sense	1.01; 1.02; 2.01; 3.01	1.01	2.01	3.01	1.01	2.01	1.01	2.01	1.01	2.01	
metonymic	X-X _i	0.00	0.60	1.00	0.29	1.69	0.50	2.00	0.00	2.00	
grammatical case	prepositional	0.00	0.00	0.00	0.00	0.31	0.10	0.40	0.00	0.75	
colligation LI	preposition	0.50	0.00	0.00	0.14	0.38	0.20	0.10	0.00	0.50	
colligation RI	preposition	0.00	0.20	0.00	0.71	0.31	0.70	0.80	0.00	0.75	

frequency towards the last periods, there is also an increase in the use of prepositions as LI and RI and in prepositional case. This colligation corresponds to a prepositional noun phrase typically denoting events like *на переговорах в Лондоне* 'during the negotiations in London'.

From the co-occurrence tables, the frequency on the sense level was extracted to monitor the semantic evolution of each form in each period. Overall, the plural form *низы* displays the highest number of lexicalized plurals, entailing two more senses which were not registered in the two dictionaries selected but found in the contexts: 'surroundings of the Volga' and 'low (minimum) engine speed'. The plural form *кошки* is second in number of senses, followed by *выборы* and *переговоры*. In terms of frequency, it was first noticed that the set of senses presents a different distribution in the two grammatical forms both diachronically and synchronically. A detailed account of each paradigm is provided below.

As FIGURE 4 shows, the plural form *низы* corresponds to its singular form only in three senses across the entire timespan, which are also the most frequent ones: 'lower course of the river', 'lower part' and 'lower classes'. While the singular form maintains as primary the sense 'lower part', the plural form maintains as predominant the sense 'lower classes', appeared in the early 20th century⁸. Other less frequent lexicalized senses ('low notes', 'low (minimum) engine speed', 'surroundings of the Volga') show an erratic tendency through the centuries.

⁸ The meaning 'lower classes' is already attested in the late nineteenth century's publicistic discourse, anticipating corpus evidence by several decades. Early uses of the plural noun *низы* as in 'lower classes' are documented, for example, in Šelgunov's *Očerki ruskoj žizni* (1888), where the

FIGURE 4
Sense distribution over time: *низ* (sg) – *низы* (pl)

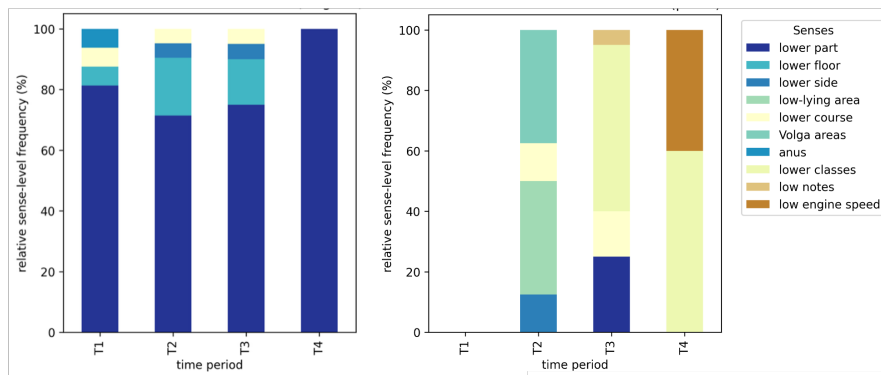


FIGURE 5
Sense distribution over time: *кошка* (sg) – *кошки* (pl)

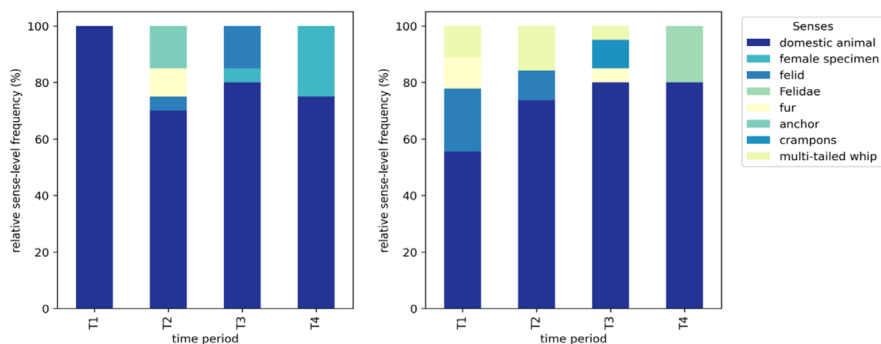


FIGURE 6
Sense distribution over time: *выбор* (sg) – *выборы* (pl)

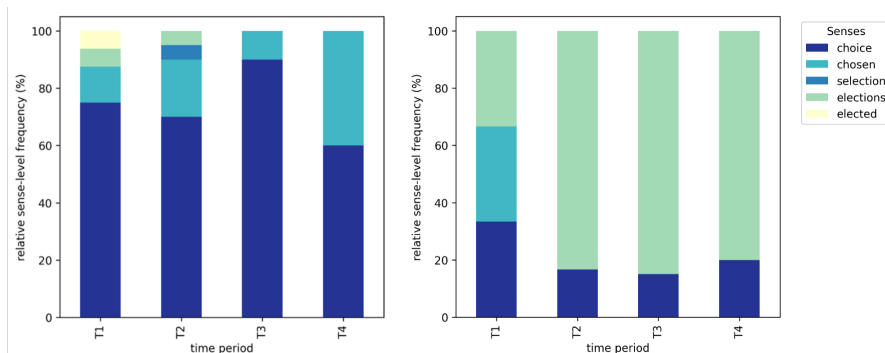
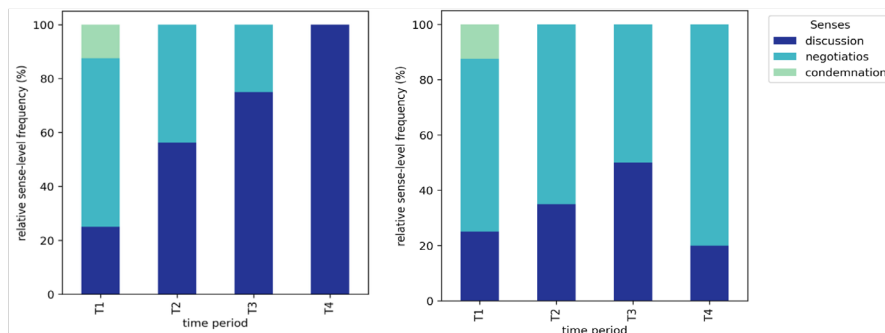


FIGURE 7
Sense distribution over time: *переговор* (sg) – *переговоры* (pl)



The plural form *кошки* mainly corresponds to the singular form in the primary sense ‘cat’ as ‘domestic animal’. Whereas the derived senses, ‘multi-tailed whip’, ‘crampons’ and ‘Felidae’, tend not to stabilize and to stay enclosed in specialized discourse, as attested by their lower frequency of occurrence in **FIGURE 5**.

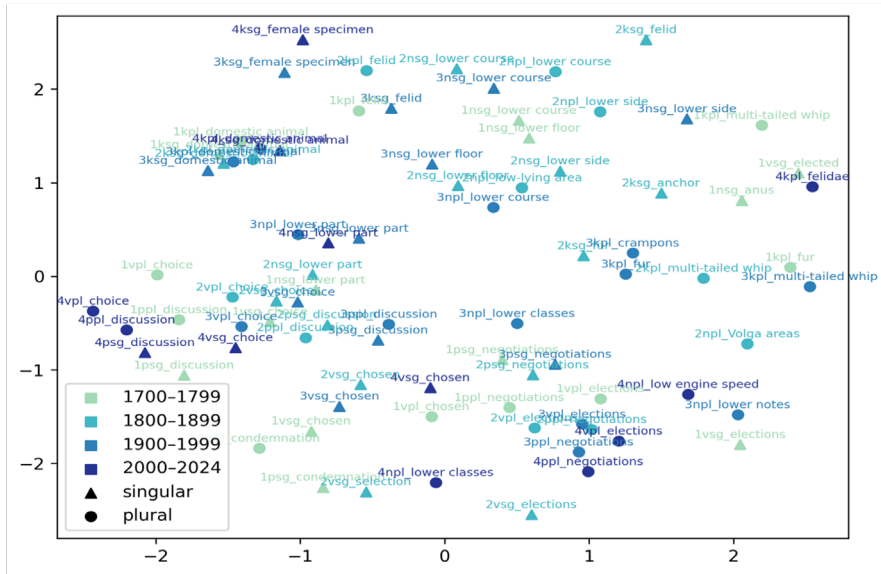
Across the four periods, *выбор* (**FIGURE 6**) maintains in the singular form the primary sense ‘choice’, from which the sense ‘elections’ derives. As demonstrated by the frequencies, the semantic detachment of the latter in the plural form is situated between the second and the third period, marking its disappearance from the sense set in the singular.

Переговор (**FIGURE 7**) displays a similar tendency. The plural form emerges several decades after the singular form, yet its usage for the sense ‘negotiations’ increases in frequency during the same period, particularly towards the first half of the 19th century. Around the second period, a stronger reversal of frequencies in the sense ‘negotiations’ is assessed, representing the temporal point of the lexicalization in the plural. The reversal reaches its extreme in the fourth period, where the singular form is solely associated with the sense ‘discussion’.

phrase *читающие низы* ‘the reading masses’ occurs. Corpus data document the subsequent stabilization of this sense in standard discourse. The retrieved contexts from the first half of the twentieth century suggest a growing degree of lexical autonomy, although the sense remains only partially stabilized, as evidenced by the presence of periphrases such as *в так называемых низах* ‘among the so-called lower classes’. By the twenty-first century the sense appears to be fully lexicalized in the plural form, e.g. from RNC (Anatolij Saluckij’s *Nemoj Nabat*): “Низы, мидлы, верхи – все гражданское общество пришло в движение” ‘The lower classes, the middle classes, and the upper classes – the whole civil society has been set in motion’.

FIGURE 8

MDS низ – низы, кошка – кошки, переговоры – переговоры, выбор – выборы



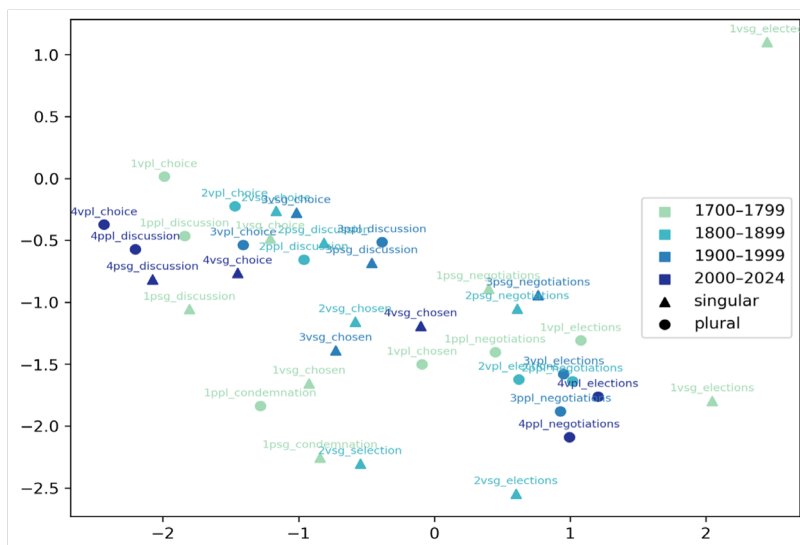
7.1. Results: MDS

To explore and visualize their distributional similarities, the profiles of the 84 annotated senses were processed as vectors as in Jansegers and Gries (2020). This process involves two steps: I. computing the Euclidean distance between each pair of profile vectors in a distance matrix; II. reducing the dimensions of the features to a lower-dimensional space, through the MDS (Multidimensional scaling) technique⁹. The results are shown in a two-dimension plot, where the spatial distance between the data points is assumed to reflect their degree of similarity. **FIGURE 8** represents the 84 senses across the four time periods, from which the main trends can be identified:

1. The degree of proximity to the primary sense appears to be contingent on the nature of the sense relation. In this regard, senses having a taxonomic relation are found to be more proximate to the primary sense, while metonymic and metaphoric relations are found to be more distant.
2. Lexicalized plurals exhibit a common condition especially in the fourth period, as they are outliers placed at the periphery of the space.

⁹ The distance matrix, i.e. the pairwise vector distances, was computed in Python using the *SciPy* package. The MDS dimensionality-reduction technique was chosen by Jansegers and Gries (2020) over the Hierarchical clustering method, given the diachronic nature of the data. MDS was performed in Python using the *scikit-learn* package..

FIGURE 9
MDS *выбор* – *выборы* and *переговор* – *переговоры* (XVIII-XXI)



3. The upper half of the figure seems to cluster concrete and mostly animate senses, while the lower half shows a prevalence of abstract and inanimate senses. For instance, the semantic concretization of the primary sense ‘lower part of’ is evident in the evolution of its corresponding data points. In the first period it is closer to the abstract sense of *выбор* ‘choice’, in the last periods it approaches *кошка* as ‘domestic animal’, which is indeed concrete.
4. The periphery in the right half of the plot gathers more specialized senses (*низ* as medical ‘anus’, nautical ‘hold’, musical ‘low notes’ and technical ‘low engine speed’; *кошка* in commercial contexts ‘fur’, technical ‘crampons’ and ‘multi-tailed whip’). On the other hand, the central and left quadrants encompass primary senses, which could be considered as high-frequency lexis (‘choice’, ‘discussion’, ‘domestic animal’, ‘lower part’, etc.).
5. The senses that have a correspondence between the singular and the plural are more distant in case of abstract denotation than in concrete denotation. A clear example is given by the comparison between *выбор* ‘choice’ - *выборы* ‘choices’ and what can be considered as the baseline, that is *кошка* ‘domestic animal’ - *кошки* ‘domestic animals’.

Specifically, comparing the evolution of the two deverbal nouns lemmatized in the plural form by the second dictionary, *выборы* and *переговоры*, similar trends can be seen. As **FIGURE 9** illustrates, the data points representing their semantic derivation cluster in the same space the primary deverbal senses, ‘choice’ – ‘discussion’, and the secondary eventive senses ‘elections’ and ‘negotiations’. The latter show a correspondence between

the singular and the plural form, with a wider distance in *переговоры*, until the formal disappearance of the singular, which occurs earlier for *выборы*, as already noted in FIGURE 6. Overall, the data points of *выборы* and *переговоры* do not show a particular reason to account for the lexicographical detachment. The latter could probably be motivated by the frequency of occurrence of the primary sense in *выбор* ‘choice’, which is indeed more frequent than ‘discussion’ in *переговоры*. Similarly, the lexicalized plurals of *низ* and *кошка* stay in the condition of outliers, showing more distance from the primary sense, which stays predominant in the polysemic chain.

8. Conclusions

This study contributed to the existing research on the evolution of *pluralia tantum*, offering a first corpus-driven diachronic analysis of four Russian nouns from the 18th to the 21st century. Considering the multi-layered annotation and the resulting profiles, some final remarks can be presented. A first distributional principle to discern different stages between a grammatical and a lexicalized plural was established. This led to the observation that most of the plurals that were here considered underwent a process of lexicalization between the end of the 18th century and the beginning of the 20th century. The mechanisms of lexicalization resulted to be mostly based on metonymy and, on one occasion, on metaphor. The interaction of the features in the co-occurrence tables shows that the semantic derivation is often accompanied by a change in the morphological and colligational pattern. Nonetheless, a proper correlation analysis among the different features and sense evolution should be conducted in future studies.

With regards to the defective paradigm’s expansion of the four nouns, the frequency of the senses and the distributional proximity between the two forms over time serve as useful indicators. Specifically, two types of defective expansion may be hypothesized. The first one is a semantic expansion, leading to a possible detachment from the original paradigm and to the creation of a secondary paradigm. This implies: 1. the dominance of the singular form in the predominant sense; 2. a clear distance between the derived and the primary sense and between the singular and the plural, with the plural being an outlier. For instance, the evolution of *низы* as ‘lower classes’ or *кошки* as ‘crampons’, could lead to a possible detachment from the original paradigm, following the model: *выбор*₁ ‘choice’ – *выборы*₁ ‘choices’; *выборы*₂ ‘elections’. The second type is a paradigmatic expansion, which concerns the evolution of the original paradigm. This involves: 1. a sharp decrease of frequency for the singular form in the predominant sense; 2. a clear distance between the singular and the plural, with the plural being an outlier. As one limitation of this analysis is the number of nouns and contexts considered, this hypothesis should be further investigated in a large-scale study.

Resources

- RNC: *Russian National Corpus*, <<https://ruscorpora.ru/>> (latest access: 15.04.25)

Literature

- Acquaviva 2008: P. Acquaviva, *Lexical Plurals: A Morphosemantic Approach*, Oxford 2008.
- Acquaviva 2016: P. Acquaviva, *Conceptualization of Part Structure*, in: M. Carrara, A. Arapinis, F. Moltmann (eds.), *Unity and Plurality: Logic, Philosophy, and Linguistics*, Oxford 2016, pp. 195-218.
- Alexiadou 2019: A. Alexiadou, *Morphological and Semantic Markedness Revisited: The Realization of Plurality Across Languages*, "Zeitschrift für Sprachwissenschaft", XXXVIII, 2019, 1, pp. 123-154, DOI: <https://doi.org/10.1515/zfs-2019-0004>.
- Apresjan 1995: Ju.D. Apresjan, *Izbrannye trudy*, 1. *Leksičeskaja semantika*, Moskva 1995.
- Besedina et al. 2015: N.A. Besedina, Y.V. Shemaeva, I.V. Borisovskaya, N.V. Zimovets, *Grammar and Interpretation*, "Journal of Language and Literature", xv, 2015, 2, pp. 165-167, DOI: <https://doi.org/10.7813/jll.2015/6-2/33>.
- Biber 1994: D. Biber, *An Analytical Framework for Register Studies*, in: D. Biber, E. Finegan (eds.), *Sociolinguistic Perspectives on Register*, New York-Oxford 1994, pp. 31-56.
- Blank 1997: A. Blank, *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*, Tübingen 1997.
- Blank 2003: A. Blank, *Polysemy in the Lexicon and in Discourse*, in: B. Nerlich, Z. Todd, V. Herman, D.D. Clarke (eds.), *Polysemy: Flexible Patterns of Sense in Mind and Language*, Berlin-New York 2003, pp. 267-294.
- Birich 1995: A. Birich, *Metonimija v sovremennom russkom jazyke (semantičeskij i grammatičeskij aspekt)*, München 1995.
- Braun 1930: M. Braun, *Das Kollektivum und das Plurale tantum im Russischen: Ein bedeutungsgeschichtlicher Versuch*, Leipzig 1930.
- Brdar 2007: M. Brdar, *Metonymy in Grammar: Towards Motivating Extensions of Grammatical Categories and Constructions*, Osijek 2007.
- Corbett 2000: G.G. Corbett, *Number*, Cambridge 2000.

- Corbett 2019: G.G. Corbett, *Pluralia Tantum Nouns and the Theory of Features: A Typology of Nouns with Non-canonical Number Properties*, "Morphology", XXIX, 2019, pp. 51-108, DOI: <https://doi.org/10.1007/s11525-018-9336-0>.
- Degtjarev 2014: V.I. Degtjarev, *Kategorija čisla v slavjanskih jazykach (istoriko-semantičeskoe issledovanie)*, Rostov-na-Donu 2014.
- Divjak, Gries 2006: D. Divjak, S.T. Gries, *Ways of Trying in Russian: Clustering Behavioral Profiles*, "Corpus Linguistics and Linguistic Theory", II, 2006, 1, pp. 23-60, DOI: <https://doi.org/10.1515/CLLT.2006.002>.
- Dobrovol'skij 2004: D.O. Dobrovol'skij, *Lexical Semantics and Combinatorial Profile: A Corpus-based Approach*, in: G. Williams, S. Vessier (eds.), *Proceedings of the 11th EURALEX International Congress*, Lorient 2004, pp. 797-796.
- Firth 1957: J. Firth, *A Synopsis of Linguistic Theory, 1930-55*, in: *Studies in Linguistic Analysis*, Oxford 1957, pp. 1-31.
- Gardelle, Acquaviva 2023: L. Gardelle, P. Acquaviva, *Pluralia tantum and singularia tantum*, in: P. Ackema, S. Bendjaballah, E. Bonet, A. Fábregas (eds.), *The Wiley Blackwell Companion to Morphology*, Oxford 2023, pp. 2043-2071, DOI: <https://doi.org/10.1002/9781119693604.morphcom065>.
- Gries 2010: S.T. Gries, *Behavioral Profiles: A fine-grained and Quantitative Approach in Corpus-based Lexical Semantics*, "The Mental Lexicon", V, 2010, 2, pp. 323-346, DOI: <https://doi.org/10.1075/ml.5.3.04gri>.
- Gries 2019: S.T. Gries, *Polysemy*, in: E. Dąbrowska, D. Divjak (eds.) *Cognitive Linguistics. Key Topics*, Berlin-Boston 2019, pp. 23-43, DOI: doi.org/10.1515/9783110626438-002.
- Hilpert, Gries 2016: M. Hilpert, S.Th. Gries, *Quantitative Approaches to Diachronic Corpus Linguistics*, in: M. Kytö, P. Pahta, *The Cambridge Handbook of English Historical Linguistics*, Cambridge 2016, pp. 36-53, DOI: <https://doi.org/10.1017/CBO9781139600231.003>.
- Langacker 2008: R.W. Langacker, *Cognitive Grammar: A Basic Introduction*, New York 2008.
- Ljaševskaja 2004: O.N. Ljaševskaja, *Semantika ruskogo čisla*, Moskva 2004.
- Lipka et al. 2004: L. Lipka, S. Handl, W. Falkner, *Lexicalization and Institutionalization*, "Skase Journal of Theoretical Linguistics", I, 2004, 1, pp. 2-19.
- Koch 2001: P. Koch, *Metonymy. Unity in Diversity*, "Journal of Historical Pragmatics", II, 2001, 2, pp. 201-244, DOI: <https://doi.org/10.1075/jhp.2.2.03koc>.

- Koptjevskaja-Tamm, Wälchli 2001: M. Koptjevskaja-Tamm, B. Wälchli, *The Circum-Baltic languages. An Areal-typological approach*, in: Ö. Dahl, M. Koptjevskaja-Tamm (eds), *The Circum-Baltic Languages*, II. *Grammar and Typology*, Amsterdam 2001, pp. 615-750.
- Külmoja, Mukovskaya 2020: I. Külmoja, L. Mukovskaya, *On the Representation of the Category of Number in Nouns in Estonian and Russian Dictionaries*, "Eesti Rakenduslingvistika Ühingu Aastaraamat", XVI, 2020, pp. 141-55, DOI: <https://doi.org/10.5128/ERYa16.09>.
- Jansegers, Gries 2020: M. Jansegers, S.Th. Gries, *Towards a Dynamic Behavioral profile: A Diachronic Study of Polysemous sentir in Spanish*, "Corpus Linguistics and Linguistic Theory", XVI, 2020, 1, pp. 145-187, DOI: <https://doi.org/10.1515/cllt-2016-0080>.
- Percov 2001: N.V. Percov, *Invarianty v russkom slovoizmenenii*, Moskva 2001.
- Plungjan 2011: V. Plungjan, *Vvedenie v grammatičeskiju semantiku: grammatičeskie značenija i sistemy jazykov mira*, Moskva 2011.
- Prochorova 1996: V.N. Prochorova, *Russkaja terminologija (leksiko-semantičeskoe obrazovanie)*, Moskva 1996.
- Šaljan 2015: A.H. Šaljan, *Pluralia Tantum Nouns and Nouns, Mostly Used in the Plural Form, in Modern Russian*, "Alustath Journal for Human and Social Sciences", 2015, 215 (1), pp. 159-176, DOI: <https://doi.org/10.36473/ujhss.voic215.612>.
- Šemaeva 2014: E.V. Šemaeva, *Kognitivnye osnovy leksikalizacii form množestvennogo čisla imeni suščestvitel'nogo: na materiale anglijskogo i russkogo jazykov. Avtoreferat dissertacii na soiskanie učenoj stepeni kandidata filologičeskich nauk*, Belgorod 2014.
- Šemaeva 2016: E.V. Šemaeva, *Identificirujuščaja metaforičeskaja model' leksikalizacii form množestvennogo čisla anglijskich i russkich imen suščestvitel'nych*, "Kognitivnye issledovanija jazyka", XXVI, 2016, pp. 188-189.
- Sinclair 1998: J. Sinclair, *The Lexical Item*, in: E. Weigand (ed.) *Contrastive Lexical Semantics*, Amsterdam 1998, pp. 1-24, DOI: <https://doi.org/10.1075/cilt.171.02sin>.
- Skljarevskaja 1993: G.N. Skljarevskaja, *Metafora v sisteme jazyka*, Sankt-Peterburg 1993.
- Zacharova 2009: E.V. Zacharova, *Osobnosti funkcionirovanija kategorii čisla iměn suščestvitel'nych s potencial'no polnoj paradigmoj (na materiale russkogo čisla rubeža XX-XXI vekov)*, Ul'janovsk 2009.
- Zaliznjak 2004: A.A. Zaliznjak, *Fenomen mnogoznačnosti i sposoby ego opisanija*, "Voprosy jazykoznanija", 2004, 2, pp. 20-45.

- Zaliznjak, Padučeva 1974: A.A. Zaliznjak, E.V. Padučeva, *O kontekstnoj sinonimii edinstvennogo i množestvennogo čisla sušestvitel'nych v russkom jazyke*, "Informacionnye voprosy semiotiki, lingvistiki i avtomatičeskogo perevoda", 1974, 4, pp. 29-37.

Dictionaries

- Berneker 1924: E. Berneker, *Slavisches etymologisches Wörterbuch*, Heidelberg 1924.
- Blagova et al. 1994: R.M. Cejtin, R. Večerka, Ė.S. Blagova (red.), *Staroslavjanskij slovar' (po rukopisjam X-XI vekov)*, Moskva 1994.
- Gerd 2009: A.S. Gerd (red.), *Bolšoj akademičeskij slovar'*, XII, Moskva 2009.
- Gerd 2011: A.S. Gerd (red.), *Bolšoj akademičeskij slovar'*, XVI, Moskva–Sankt-Peterburg 2011.
- Gorbačevič 2005: K.S. Gorbačevič (red.), *Bolšoj akademičeskij slovar'*, III, Moskva–Sankt-Peterburg 2005.
- Gorbačevič 2007: K.S. Gorbačevič (red.), *Bolšoj akademičeskij slovar'*, VIII, Moskva–Sankt-Peterburg 2007.
- Matthews 2007: P. H. Matthews, *The Concise Oxford Dictionary of Linguistics*, Oxford 2007.
- Šmelev 1986: D.N. Šmelev (red.), *Slovar' russkogo jazyka XI-XVII vv.*, XI, Moskva 1986.
- Sorokin 1988: Ju.S. Sorokin (red.), *Slovar' russkogo jazyka XVIII veka*, IV, Sankt-Peterburg 1988.
- Sorokin 1998: Ju.S. Sorokin (red.), *Slovar' russkogo jazyka XVIII veka*, X, Sankt-Peterburg 1998.
- Sorokin 2005: Ju.S. Sorokin (red.), *Slovar' russkogo jazyka XVIII veka*, XV, Sankt-Peterburg 2005.
- Sorokin 2011: Ju.S. Sorokin (red.), *Slovar' russkogo jazyka XVIII veka*, XIX, Sankt-Peterburg 2011.
- Vasmer 1950-1958: M. Vasmer, *Russisches Etymologisches Wörterbuch*, Heidelberg 1950-1958 (transl. by N. Trubačev, *Ėtimologičeskij slovar' russkogo jazyka*, Moskva 1986).

Abstract

Elena Spaziani

Nizy, koški, peregovory and vybory: *Lexicalization Profiles*

The paper provides a systematic diachronic analysis of the lexicalization in the plurals of four nouns: *niz* 'bottom'; *koška* 'cat'; *peregovor* 'discussion' and *vybor* 'choice'. The corpus-driven study was conducted adopting the procedure of behavioral profiles, which consisted of three phases: retrieving 401 concordance lines from the Russian National Corpus; annotating 84 senses according to a set of 57 linguistic features; visualizing and analyzing the co-occurrence information through a dimensionality reduction technique (MDS) in a two-dimensional space, where the distance between data points is indicative of their semantic similarity. During the annotation a distributional principle was established to distinguish the grammatical plural from the lexicalized plural. The semantic derivation underlying lexicalization was annotated as metonymic for most plurals. Considering both the distance between the original sense and the lexicalized sense and the sense level frequency, a defective paradigm seems to expand: paradigmatically, as in *peregovory* 'negotiations', or semantically, as in *vybory* 'elections', *nizy* 'lower classes', and *koški* 'crampons'.

Keywords

pluralia tantum; Russian; Lexicalization; Corpus-Driven Analysis; Behavioral Profiles.