

A long birth: The development of gender-specific paucal constructions in Russian

Tore Nessel, UiT The Arctic University of Norway

Abstract in English

This article investigates the diachronic development of Russian numeral constructions consisting of a paucal numeral (*dva* 'two', *tri* 'three', *četyre* 'four') followed by an adjective and a noun. Based on statistical analysis of more than 6,000 corpus examples, it is shown that a split took place in the second half of the twentieth century when feminine nouns developed a different agreement pattern from that of masculine and neuter nouns. This split is argued to represent the final step in a long "birth process" of gender-specific paucal constructions that started with the loss of the dual in the Middle Ages. It is suggested that we are witnessing a cascading effect, whereby the feminine pattern develops when the pattern for masculine and neuter nouns are approaching stabilization. The article furthermore includes a discussion of the hypothesis that "S-curves" represent a template for language change. While the documented changes resemble S-curves, the proposed analysis also addresses some general problems with testing the S-curve hypothesis empirically.

Abstract in French

Cet article examine le développement diachronique des constructions numériques du russe qui se composent d'un paucal (*dva* 'deux', *tri* 'trois', *četyre* 'quatre') suivi d'un adjectif et un nom. Sur la base d'une analyse statistique de plus de 6000 exemples tirés de corpus, nous montrons qu'un changement a eu lieu dans la deuxième moitié du vingtième siècle lorsque les noms féminins ont développé un modèle d'accord différent de celui des noms masculins et neutres. Il a été argumenté que ce changement représente la dernière étape d'un long « processus de naissance » des constructions paucales spécifiques au genre qui a commencé avec la disparition du duel dans le Moyen Age. Il est proposé que nous assistons à un effet en cascade par lequel le modèle féminin continue à se développer tandis que le modèle des noms masculins et neutres s'approche la stabilisation. En outre, l'article inclut une discussion sur l'hypothèse que les « courbes en S » représentent un gabarit de changement linguistique. Si les changements documentés ressemblent à des courbes en S, l'analyse proposée aborde également quelques problèmes généraux liés au test empirique de l'hypothèse des courbes en S.

Abstract in German

Dieser Artikel untersucht die diachrone Entwicklung russischer Numeralkonstruktionen, die aus einem paucalen Numeral (*dva* 'zwei', *tri* 'drei', *četyre* 'vier') gefolgt von einem Adjektiv und einem Nomen bestehen. Basiert auf einer statistischen Analyse von über 6,000 Korpusbeispielen wird gezeigt, dass in der zweiten Hälfte des zwanzigsten Jahrhunderts eine Spaltung stattfand, in der feminine Nomen ein eigenes Beugungsmuster entwickelten, das sich von dem für maskuline und neutrale Nomen unterscheidet. Es wird argumentiert, dass diese Spaltung der letzte Schritt eines langen „Geburtsprozesses“ genus-spezifischer Paucalkonstruktionen ist, der im Mittelalter mit dem Verlust des Duals begann. Es wird vorgeschlagen, dass wir Zeugen eines Kaskadeneffektes sind, in dem das feminine Muster sich erst bildet als sich das der maskulinen und neutralen Nomen einem stabilen Zustand nähert.

Der Artikel enthält des Weiteren eine Diskussion der Hypothese, dass „S-Kurven“ eine Schablone für den Sprachwandel repräsentieren. Während die dokumentierten Veränderungen S-Kurven ähneln, spricht die vorgeschlagene Analyse auch einige generelle Probleme mit dem empirischen Testen solcher S-Kurven an.

Keywords

Russian, numerals, paucal, gender, S-curves

1. Introduction

Russian numeral constructions are notorious for their complexity. Numerals combine with nouns in the genitive, but while most numerals take nouns in the genitive *plural* as illustrated in (1a), the so-called paucal numerals (*dva* ‘two’, *tri* ‘three’, *četyre* ‘four’) govern nouns in (what looks like) the genitive *singular*, as shown in (1b):¹

- (1) a. V moem kabinete stoit **pjat’ stolov**_{GEN.PL.}
‘In my office there are five tables.’ (Jurskij, 1997-1998)
b. V kabinete **dva stola**_{GEN.SG.}
‘In the office there are two tables.’ (Vološina and Kul’kov, 2009)

The examples with paucal numbers have inspired a number of different analyses. Some scholars identify the noun as the genitive singular (e.g. Andersen 2006, Pesetsky 2013), while others analyze the nouns as a separate “numerative” case (Isačenko 1962, Zaliznjak 1967, Mel’čuk 1985, Rappaport 2002) or a paucal (or dual) number (Corbett 1993, Pereltsvaig 2010, Bailyn and Nevins 2008). Since this interesting theoretical question is tangential to the present study, I will not provide further discussion. Remaining agnostic I will adopt a conservative approach and classify the relevant nouns as genitive singular, since this does not involve the postulation of categories that are not attested outside the paucal constructions.

The focus of the present study is on a further complication concerning the form of the adjective in paucal constructions. The following examples show that even if the noun appears to have a singular form, the preceding adjective is in the plural. The adjective may have (what looks like) the nominative/accusative plural ending *-ye*, as in (2a), but we also find the genitive plural ending *-yx* on the adjective, as shown in (2b):

- (2) a. Preslovutaja zagadočnost’ pozvoljaet Putinu uspešno rešat’ **dve važnye**_{NOM/ACC} **zadači** [...].
‘His notorious inscrutability allows Putin to successfully solve **two important tasks** [...].’ (*Eženedel’nyj žurnal*, 2003)
b. Ix vysokoe bystrodejstvie možet rešit’ **dve važnyx**_{GEN} **zadači** [...].
‘Their high speed can solve **two important tasks** [...].’ (*Gornaja promyšlennost’*, 2004)

This rivalry in terms of adjectival inflection is attested in constructions with a paucal numeral (*dva* ‘two’, *tri* ‘three’, *četyre* ‘four’) in the nominative or nominative-like accusative followed by an adjective and a noun, and these constructions are therefore the topic of the present

¹ Throughout this article, numbered examples are from the Russian National Corpus, freely available at www.ruscorpora.ru. For each example, a year and a source are provided. The sources are either the name of a periodical (in italics) or, for examples not taken from periodicals, the name of the author (not italicized). The numeral constructions are boldfaced and subscripts indicate relevant morphosyntactic features. All examples are given in transliterated orthography.

study.² The question is what factors determine the distribution of the adjective forms, and how this distribution has changed over time. The purpose of the present investigation is to shed light on these questions. Throughout the study, I will refer to the adjective forms in (2) as “nominative/accusative” and “genitive”, although I am aware that other analyses are possible. However, these analyses are not of direct relevance for my research and will therefore not be explored in the following.

In order to investigate the distribution of adjective forms empirically, I created a database with examples from the Russian National Corpus (main corpus) from 1825-2012.³ All examples involving the paucal numerals *dva/dve* ‘two’, *tri* ‘three’ and *četyre* ‘four’ followed by an adjective and a noun were extracted from the corpus and exported to a spreadsheet, where the examples were annotated manually for a number of factors that have been suggested to be relevant in the scholarly literature on Russian numeral constructions.⁴ These factors will be presented in detail in section 3 below. In order to have a database with independent observations, one example per text was singled out for further analysis. The database, which consists of more than 6,000 examples, was analyzed statistically by means of CART (“Classification and Regression Tree”) and Random Forest (Strobl et al. 2009), as well as logistic regression. The database and code for statistical analysis are available at the Tromsø Repository of Language and Linguistics (TROLLing).⁵

The contribution of the study can be summarized as follows. First, the analysis shows that the distribution of the adjective forms has changed radically over the last two centuries. Second, it is shown that a split took place in the second half of the twentieth century; in combination with masculine and neuter nouns, adjectives in the genitive have ousted their nominative/accusative rivals, while the opposite development has taken place in constructions with feminine nouns, where examples like (2b) are in the process of being marginalized. Third, it is argued that this development is the last step in a long “birth process” of gender-specific paucal constructions, which started with the loss of the dual in the Middle Ages. Fourth, it is suggested that the changes under scrutiny display a cascading effect, whereby a new change starts when a previous change is approaching culmination. Finally, the study addresses the hypothesis that S-curves represent a template for language change. Although the changes under scrutiny show some resemblance to S-curves, the analysis also evokes general problems with testing the S-curve hypothesis empirically.

We will consider the statistical analysis of the corpus data in sections 3-4, and S-curves in section 5. However, before we can analyze the corpus data, a brief exposition of the early development of paucal constructions is in order, since this sets the stage for the changes under scrutiny in the present study.

² I use the masculine and neuter form *dva* as the citation form to represent the whole lexeme ‘two’, which also has the feminine form *dve* attested in (2). *Tri* and *četyre* have the same forms for all genders. Paucal constructions with *poltora* ‘one and a half’ and *oba* ‘both’ are not scrutinized in the present study.

³ The corpus goes further back in time, but older examples are too scarce for statistical analysis.

⁴ As pointed out by an anonymous reviewer, the infinite *rešat’* ‘decide’ in (2a) is imperfective, while (2b) contains a verb of perfective aspect (the infinitive *rešit’* ‘decide’). We cannot *a priori* exclude that verbal aspect might be of importance for the choice between adjectives in *-ye* or *-yx*, but since it was not possible to include verbal aspect as a variable in my statistical investigation, this question will have to be left for future research. To the best of my knowledge, the relationship between verbal aspect and adjective endings in paucal constructions has not been investigated earlier, but Nessel and Makarova (2018) concluded that aspect was *not* relevant for Russian constructions denoting decades, which involve ordinal numerals. A complicating factor is the fact that in both (2a) and (2b), the inflected verbs are imperfective, viz. *pozvoljat’* ‘allow’ in (2a) and *moč’* ‘be able to’ in (2b).

⁵ The TROLLing post is available here: <https://doi.org/10.18710/54ZJGQ>.

2. Historical background

In order to appreciate the changes in the adjectives in Russian paucal constructions, we first need to consider earlier changes in the nouns, which are interesting for two reasons. First, they document the emergence of a uniform paucal construction, and second they demonstrate the increasing relevance of grammatical gender, for which adjectives also provide evidence.

Old Russian nouns had a tripartite morphological number opposition between singular, dual and plural.⁶ The numeral *d''va* ‘two’ combined with nouns in the dual, while *tri* ‘three’ and *četyre* ‘four’ combined with nouns in the plural. Syntactically, all the paucal numerals displayed adjective-like behavior since they agreed with the head noun in gender and case. Higher numerals like *pjat'* ‘five’, on the other hand, were historically nouns and governed the genitive case in an accompanying noun.

During the Old Russian period, the morphological dual was lost, and this launched a development towards a unified paucal construction encompassing the numerals 2-4 plus accompanying nouns. The development involves extensive variation, but for present purposes it is sufficiently precise to consider three stages that are visualized in Figure 1. For each stage, the figure gives the noun endings for the most relevant declensions when combined with the numerals ‘two’ (top row) and ‘three’ and ‘four’ (bottom row).⁷

	Stage 1 (dual intact)				Stage 2 (dual lost)				Stage 3 (unified paucal constr.)		
	ǫ _M :	ǫ _N :	ā:	→	ǫ _M :	ǫ _N :	ā:	→	ǫ _M :	ǫ _N :	ā:
2	-a	-ě	-ě		-a	-a	-y		-a	-a	-y
3 and 4	-i	-a	-y		-y	-a	-y		-a	-a	-y

Figure 1: The development of a uniform paucal construction in three stages; ǫ_M = ǫ-declension (masculine nouns), ǫ_N = ǫ-declension for neuter nouns, ā = ā-declension

Stage 1 in Figure 1 represents the situation before the loss of the dual. At this stage, all genders took different endings after ‘two’ on the one hand vs. ‘three’ and ‘four’ on the other, since ‘two’ combined with nouns in the dual, whereas ‘three’ and ‘four’ combined with nouns in the plural. The shaded cells indicate endings that were syncretic with the genitive singular, a point we will return to shortly.

When the morphological dual was still intact, masculine ǫ-declension noun forms like *roga* with the dual ending *-a* would be interpreted as ‘two horns’. When the morphological dual was lost, the form in isolation would no longer receive the interpretation ‘two horns’, but as shown by Žolobov (2002 and 2003), the forms in *-a* were still widely used with the numeral *d''va* ‘two’. It is an open question how the speakers analyzed this ending, but it is a fact that *-a* was also the ending of the genitive singular, and it is likely that *d''va + ...a* was reanalyzed as the genitive singular, as shown by the successive development of the nouns in the neuter ǫ-declension as well as in the ā-declension.

⁶ I use the traditional term “Old Russian”, although it is a misnomer to the extent that the relevant language is the ancestor of all the East Slavic languages, not just Russian. For discussion of alternative terms, see Nessel (2015: 10).

⁷ Old Russian inherited six declension classes for nouns from Common Slavic, but here it is sufficient to consider the ǫ- and ā-declensions. For the ǫ-declension it is customary to distinguish between two subclasses, one for masculine and one for neuter nouns. The endings in the table represent stems with non-palatalized final consonants.

While masculine *ŏ*-declension nouns had the ending *-a* in the nominative dual, the ending for neuter and feminine nouns was *-ě*.⁸ This ending was not syncretic with the genitive singular, but it was gradually replaced by the ending *-a* in the neuter *ŏ*-declension and *-y* in *ā*-declension, as shown in stage 2 in Figure 1. The new endings were the same as the genitive singular, so it is possible that speakers had made the generalization that masculine nouns occur in the genitive singular after the numeral ‘two’ and that they extended this to the neuter *ŏ*-declension and the *ā*-declension. However, another motivation is also possible. The *-a* ending for the neuter *ŏ*-declension and the *-y* ending for the *ā*-declension are also identical to the endings of the nominative plural, which was the case/number form used in combination with the numerals ‘three’ and ‘four’ in Old Russian, as shown in the figure. It remains unclear whether it was the analogy to the masculine nouns in the *ŏ*-declension or the analogy to ‘three’ and ‘four’ that motivated this change, but a likely guess is that both motivations worked in tandem. Either way, the result is a system where neuter and feminine nouns had the same endings for all paucal numerals. Masculine nouns in the *ŏ*-declension, on the other hand, still had different endings for ‘two’ on the one hand and ‘three’ and ‘four’ on the other.

A final change completes the picture and brings us to stage 3, which represents the situation that emerges in the 1600s (Žolobov 2002). Now the opposition between ‘two’ and ‘three’/‘four’ became neutralized for masculine *ŏ*-declension nouns too, since the *-a* ending replaced the original nominative/accusative plural ending *-y* after ‘three’ and ‘four’. This produced a system with a unified paucal construction where the numeral was always followed by a noun with the ending of the genitive singular: *-a* in the *ŏ*-declension, and *-y* in the *ā*-declension. More than one interpretation is conceivable. It is possible that language users analyzed the relevant endings as the genitive singular, i.e. that paucal numerals governed nouns in the genitive singular. However, it is also possible that they analyzed the endings as a separate paucal number or numerative case that happened to be syncretic with the genitive singular. Either way, stage 3 in the figure testifies to the existence of a “uniform” paucal construction, i.e. a construction where all paucal numerals combine with nouns with the same set of endings.

Žolobov (2002: 4) maintains that the development described above implies a strengthening of the opposition between the grammatical genders. This is true in the sense that the *ŏ*-declension encompasses masculine and neuter nouns, whereas the *ā*-declension is dominated by feminine nouns. Thus, in paucal constructions at stage 3 masculine and neuter nouns generally took the ending *-a*, while feminine nouns combined with *-y*.

What was the situation for adjectives? Žolobov’s (2002, 2003) thorough studies lend some support to the idea that the gender opposition was as important for the development of adjectives as it was for nouns in paucal constructions. The predecessors of both the modern adjectival patterns in (2) are well attested in the Middle Russian period after the loss of the dual. In (3a), we have an adjective in the nominative plural, while the adjectives in (3b) are in the genitive plural:

- (3) a. **Dvě rybki zoločony**_{NOM/ACC}
 ‘two golden fishes.’ (*Spiritual and contractual documents of grand princes and appanage princes, 14th-16th centuries*, Žolobov 2003: 179)

⁸ Old Russian had a phonemic contrast between /e/ and /ě/, the latter representing a somewhat higher vowel than /e/ (Nesset 2015: 245).

b. Slity byša **tri kolokoly bolšix_{GEN} da dva malyx''_{GEN}**.

'Three big and two small bells were cast.' (*Moscow chronicle collection, end of 15th century*, Žolobov 2003: 179)

The historical origin of the nominative pattern is most likely constructions with *tri* 'three' and *četyre* 'four', which combined with nominative adjectives in Old Russian, while genitive adjectives were probably taken over from constructions with numerals higher than *četyre*, which in Old Russian had adjectives in the genitive plural (Žolobov 2003: 179-181). Table 1, which is adapted from Žolobov (2003: 182), shows the distribution of the two adjectival patterns in combination with masculine/neuter nouns (to the left) and feminine nouns (to the right). Two observations can be made on the basis of these (admittedly not very large) numbers. First, masculine and neuter nouns show a certain affinity to the genitive pattern, while the nominative pattern is more characteristic of feminine nouns (Žolobov 2003: 182). Second, the difference between masculine and neuter genders on the one hand and the feminine gender on the other is more distinct in the 1600s than in the 1500s. Notice that (3a) with the nominative pattern has a feminine noun (*rybka* 'fish'), while the masculine noun *kolokol''* 'bell' is used in (3b) as illustration of the genitive pattern.

	Masculine/neuter nouns			Feminine nouns		
	#Nom	#Gen	%Gen	#Nom	#Gen	%Gen
1500s	27	20	43%	26	8	24%
1600s	12	37	76%	28	10	26%

Table 1: Development in the 1500s and 1600s for nouns of different genders (adapted from Žolobov 2003: 182)

To summarize, the available evidence from Middle Russian testifies to the relevance of the gender opposition for the rivalry between nominative/accusative and genitive adjectives in paucal constructions. At the same time, it seems that neither rival managed to oust its competitor completely (Iordanskij 1958: 62), and the rivalry continued for all genders into Modern Russian. In the two next sections, we will explore the development in the nineteenth, twentieth and twenty-first centuries, which are the main focus of the present study.

3. Corpus investigation 1: Random Forest and logistic regression

In the scholarly literature on Russian paucal constructions, a number of factors have been suggested to affect the choice of adjective form. Based on a Random Forest analysis of these factors, it will be shown in the following that three factors are responsible for the distribution of adjective forms in the nominative/accusative vs. genitive in the dataset under analysis: gender, period, and (to a lesser extent) numeral.

The factors that were examined are listed in Table 2, which also provides the values of the factors included in the database. I will comment on each factor before I present the statistical analysis of their interaction. As mentioned in the previous sections, the grammatical gender of the noun is often mentioned as a key factor, and the nouns in each example were therefore tagged as masculine, feminine or neuter. Notice that gender is understood as "agreement class" (Corbett 1991). Accordingly, nouns in *-a* such as *papa* 'dad' are classified as masculine since they show masculine syntactic agreement (e.g. *xorošij_{MASC} papa* 'good dad'), although they belong to a declension class that is dominated by feminine nouns.

Variables	Values
Gender (of quantified noun)	Masculine Neuter Feminine
Period (when example was produced)	25-year periods from 1825
Numeral	<i>Dva/dve</i> ‘two’ <i>Tri</i> ‘three’ <i>Četyre</i> ‘four’
Modifier (pronoun, adjective, participle)	Preposed nominative Preposed genitive Postposed nominative Postposed genitive
Predicate (of numeral phrase as subject)	Preposed plural Preposed singular Postposed plural Postposed singular
Noun stress (of quantified noun)	Immobile (gen sg = nom pl) Mobile (gen sg ≠ nom pl)
Preposition (governing numeral phrase)	No preposition Preposition
Numeral case	Nominative Accusative
Conjoined subject	Conjoined Not conjoined
Complex numeral	Complex Not complex

Table 2: Overview of variables and values included in the Random Forest analysis

It has been proposed in the scholarly literature that the Russian paucal constructions have changed over time (see e.g. Pereltsvaig 2010). In order to investigate this factor, the examples were classified according to when they were produced. Based on the production years provided by the Russian National Corpus, the examples were grouped in twenty-five year periods, and the factor is therefore referred to as “period” in Table 2. As mentioned in section 1, the corpus contains examples from the 1700s until 2012, but the searches returned too few hits to facilitate reliable analysis of examples from the periods before 1825.

A third factor that has been considered relevant in the literature is the numeral itself. For instance, Gallis (1947: 70) found that the relative frequency of adjectives in the genitive is higher for ‘three’ and ‘four’ than for ‘two’ (see also Suprun 1957: 77, Mel’čuk 1985: 127, and Corbett 1993: 24–25). Each example in the database was accordingly tagged for the relevant numeral, so that this factor could be included in the statistical analysis.

In addition to the adjective following directly after the numeral, Russian numeral phrases can combine with preposed pronominal determiners or adjectival modifiers, i.e. elements that precede the numerals. These elements may occur in the nominative plural as in (4a) or in the genitive plural as in (4b):

- (4) a. Sam Putin vydělil **sledujuščie**_{NOM} **tri glavnyx**_{GEN} **problemy** [...].
‘Putin himself singled out the following **three main problems** [...].’ (*Lebed’*, 2003)

- b. Teper' my imeem **celyx**_{GEN} **tri ežednevnyx**_{GEN} **rejsa** v Švejcariju.
 'Now we have as much as **three daily flights** to Switzerland.' (*Domovoj*, 2002)

Postposed participial or adjectival modifiers, i.e. elements following the quantified noun, are also attested in the nominative plural or the genitive plural, as shown in (5a-b):

- (5) a. V Samaru prišli **četyre novye**_{NOM} **aviakompanii**, prežde zdes' ne **rabotavšie**_{NOM}.
 'Four new aviation companies have appeared in Samara that were not active here previously.' (*Delo*, 2002)
- b. **Dva sočnyx**_{GEN} **jabloka**, s' **edennyx**_{GEN} ežednevno, pomogut predupredit' razvitie ateroskleroza.
 'Two juicy apples, eaten daily, will help preventing the development of atherosclerosis.' (*Istorii iz žizni*, 2004)

In order to find out whether preposed or postposed elements have an impact on the nominative/accusative vs. genitive rivalry under scrutiny in the present study, all examples in the database were tagged manually for the four options in (4) and (5): preposed and postposed modifiers in the nominative or genitive plural. Notice that "modifier" is used in a broad sense here as a cover term for pronominal determiners, adjectives and participles.

If the numeral phrase is the subject of the sentence, the predicate can occur in the plural or the singular.⁹ Examples (6a-b) illustrate this:

- (6) a. Byli_{PL} obrazovany_{PL} **dve sudebnye**_{NOM/ACC} **sistemy**: obščaja i mestnaja.
 'Two legal systems were established: a general one and a local one.' (*Otečestvennye zapiski*, 2003)
- b. V programme večera bylo_{SG} **dva krupnyx**_{GEN} **sočinenija** Šnittke.
 'On the program this evening were two large works by Schnittke.' (*Rossijskaja muzykal'naja gazeta*, 2003)

Since predicates can precede or follow the numeral phrase, the database was tagged for four options: preposed and postposed predicates in the plural or singular. This made it possible to investigate the potential impact of predicates.

Russian nouns have complex stress patterns where some nouns have mobile stress with different stress placement in the singular and plural (Zaliznjak 1967 and 1985). Thus, the noun *nogá* 'leg' has the genitive singular *nogí* with stress on the ending, but the nominative plural form *nógi* with stress on the stem. Nouns with immobile stress, on the other hand, keep the stress on the same syllable throughout the inflectional paradigm, as in *kvartíra* 'apartment' (genitive singular: *kvartíry*, nominative plural: *kvartíry*). Whether stress is relevant for the nominative/accusative vs. genitive rivalry is a controversial question (see e.g. Pereltsvaig 2010: 426 and Šaronov 2014 for discussion), but some scholars have proposed that stress is relevant. For instance, Wade (2011: 215) states that a "genitive plural adjective is preferred with a feminine noun after 2–4 when there is a stress difference between the genitive singular and nominative plural of the noun" (see also Rozental' 1987: 276 and Graudina et al. 2001: 41). In order to test this hypothesis, all the quantified nouns in the database were manually tagged for stress pattern.

Numeral phrases can be governed by prepositions, e.g. *na* '(here) into':

⁹ Non-past tense forms of verbs are in the 3rd person plural or the 3rd person singular, while past tense forms of verbs occur in the plural or the neuter singular. Predicative adjectives are in the nominative plural or the nominative neuter singular.

(7) Skauty deljatsja **na tri vozrastnye_{NOM} gruppy**.

‘Scouts are divided into **three age groups**.’ (*Tramvaj*, 1990)

The following prepositions were attested in the database: *čerez* ‘through’, *na* ‘on(to), o ‘against’, *po* ‘distributed’, *pod* ‘under’, *pro* ‘about’, *s* ‘approximately the size of’, *skvoz’* ‘through’, *v* ‘in(to)’, and *za* ‘behind, in (about time)’. The possible impact of prepositions, especially the distributive *po*, on the rivalry between adjectives in the nominative/accusative or genitive has been discussed in the literature (see e.g. Suprun 1957: 79, Gorbačevič 1971: 261, Rozental’ 1987: 277). For this reason, all numeral phrases in the database were tagged for the presence or absence of a governing preposition.

Three final variables were included in the analysis, although they seem to have received little attention in the literature. First, the numeral in the phrases under scrutiny in the present study occur in the nominative or the accusative. Thus, all numerals were tagged for nominative or accusative case in order to clarify whether this variable has any impact on the choice between nominative/accusative or genitive form of the adjective. Second, numeral phrases are sometimes conjoined (“five small dogs and two big cats”), and in order to find out if this phenomenon would be relevant for the choice of adjective form, it was included in the analysis. Last but not least, Table 2 mentions “complex numerals”, i.e. numerals like *sorok dva* ‘forty-two’ that consist of an additional numeral (here: *sorok* ‘forty’) preceding the paucal numeral (here: *dva* ‘two’, see Graudina et al. 2001: 41 for a short comment, and Žolobov 2003: 180 on the possible relevance of this factor in Old and Middle Russian).

In order to investigate the interaction of all these variables and find out which of them have an impact on the distribution of the adjective forms in the paucal constructions, I carried out a Random Forest analysis with all the variables in Table 2 as independent variables and the case of the adjective as the dependent variable. Random Forest (Strobl et al. 2009) is a technique that constructs a large number of bootstrap samples and then for each of them creates a decision tree. Based on this forest of trees, votes are collected from each individual tree, which gives a reliable picture of the relative importance of the relevant variables (Baayen et al. 2013: 265).

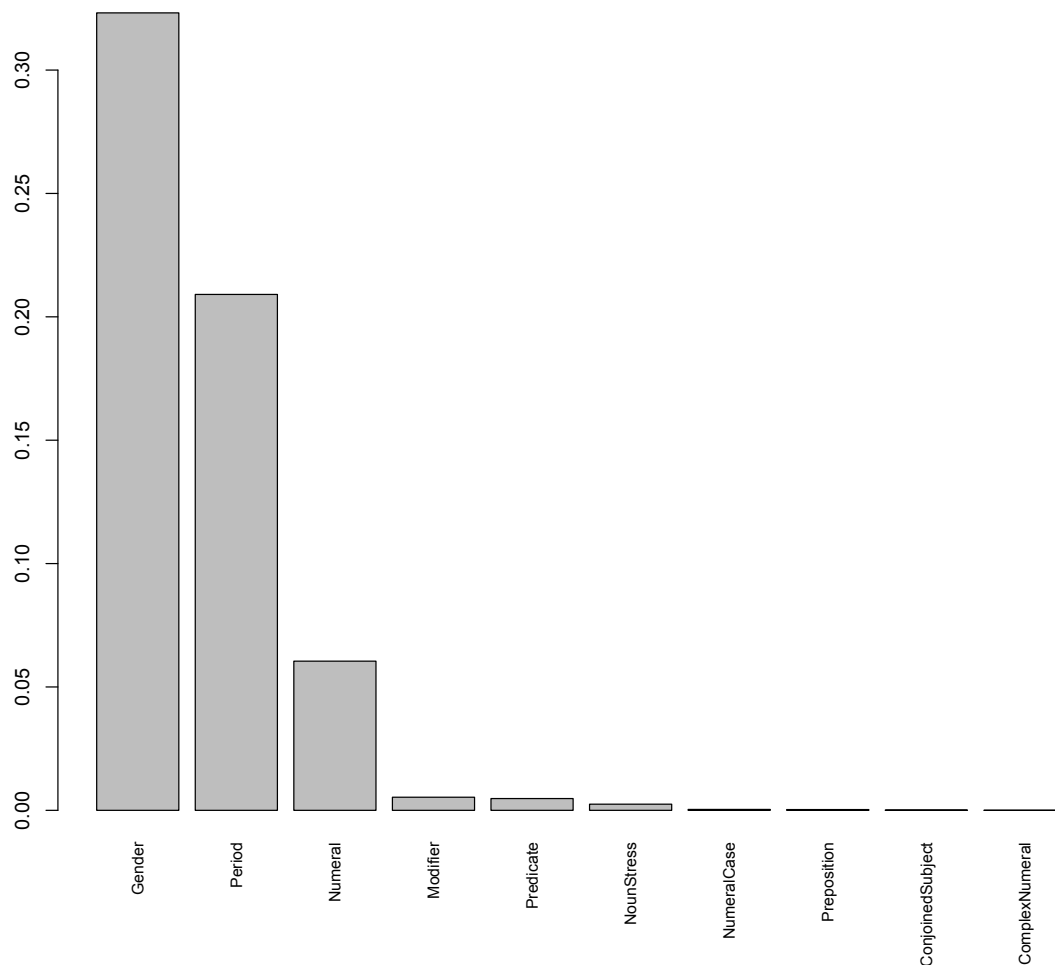


Figure 2: Variable importance plot from Random Forest

The Random Forest model returned the variable importance plot in Figure 2, which shows the relative importance of the various independent variables in Table 2. As shown in the figure, gender of the nouns turned out to be the most important variable, followed by period (i.e. the twenty-five year period the example was produced). The numeral ('two', 'three' or 'four') is also important, but much less so than period and especially gender. All the remaining independent variables turned out to have negligible impact on the choice between nominative/accusative and genitive in the adjective. It is worth pointing out that this does not mean that the relevance of these factors can be ruled out completely. However, in order to investigate factors beyond gender, period and numerals, one would need a much larger database (cf. Nessel 2019 for discussion). Unfortunately, no such database can be constructed at present, since we do not have larger diachronic corpora of Russian.

A logistic regression analysis corroborates the result of the Random Forest analysis reported above. I constructed a model with the case of the adjective as the dependent variable and gender, period, and numeral as independent variables. The optimal model, which includes both main effects and interactions between all three independent variables, correctly predicts the case of the adjective for 5,923 (91.5%) of the total 6,475 observations.¹⁰ Importantly, removal of any factors and/or interactions weakens the performance of the

¹⁰ The optimal regression analysis returned the following values: $R^2=0.779$, $C=0.955$, and $D_{xy}=0.911$.

regression model, reducing the values of R^2 , C, and Dxy, and increasing the Akaike Information Criterion, which is lowest for the model that includes interactions between all the three independent variables under scrutiny.

Before we leave the Random Forest and logistic regression analyses, it is worth pointing out that all variables mentioned above concern language internal factors. To what extent are external factors such as the impact of the literary norm on language usage relevant for the problem under scrutiny in the present study? Characteristic of the Soviet period were a strong central power and a strong literary norm, which may have an inhibitory effect on language change. Subsequent to a complex transitional period after the fall of the Soviet Union, the awareness of the strong literary norm has been restored. Recent years have seen vivid meta-linguistic discussions of the relationship between language use and literary norm, but these discussions tend to address issues such as borrowing of foreign words and the use of substandard language in public discourse (Ryazanova-Clarke 2006). Purely grammatical changes like the one under scrutiny in the present study, seem less controversial, although it is worth noting that websites that offer advice on normative (“correct”) language use occasionally receive questions about the nominative/accusative vs. genitive rivalry in paucal constructions. Interestingly, both the authoritative state service *gramota.ru* and the regional service “Yekaterinburg speaks correctly” maintain that both nominative/accusative and genitive adjectives are part of the literary norm.¹¹ Although a systematic investigation of the impact of the literary norm and other language external factors on paucal constructions is an interesting endeavor, this question must be left for future research, since the Russian National Corpus, on which the present study is based, is best suited for the investigation of language internal factors (Nesset and Makarova 2018).

4. Corpus investigation 2: CART and logistic regression

Now that we have seen that three variables have an impact on the nominative/accusative vs. genitive rivalry in the dataset under scrutiny, the next question is how these three variables interact in predicting the choice of adjective form. In order to find out, I carried out a CART analysis. As we will see in the following, the CART analysis corroborates the conclusion from Random Forest with regard to the relative importance of variables and furthermore shows that a split took place in the second half of the twentieth century, which led to the emergence of gender-specific paucal constructions. The results from the CART analysis are supported by the logistic regression analysis.

CART, which is short for “Classification and Regression Tree” (Strobl et al. 2009), is a statistical model that facilitates analysis of interacting predictor variables, i.e. situations where a number of independent variables (in our case gender, time period and numeral) exercise influence on the choice between two or more outcomes (in our case nominative/accusative vs. genitive on the adjective). The methods were introduced to linguistics by Tagliamonte and Baayen (2012; see also Levshina 2015: 291). Baayen et al. (2013) have demonstrated that by most measures CART performs almost identically compared to traditional regression models. However, compared to traditional regression models, CART has the advantage of providing tree diagrams that make it easy to interpret the results.

¹¹ See <http://new.gramota.ru/spravka/buro/search-answer?s=%D0%9F%D1%80%D0%B8+%D1%81%D1%83%D1%89%D0%B5%D1%81%D1%82%D0%B2%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D1%8B%D1%85+%D0%B6%D0%B5%D0%BD%D1%81%D0%BA%D0%BE%D0%B3%D0%BE> and <http://www.ekburg.ru/news/18/64014-kak-pravilno-dve-novykh-ili-dve-novye-elki/>. Both sites accessed on May 28, 2019.

The CART analysis of my database returned the decision tree in Figure 3, which contains 15 numbered nodes. The model tries to predict the choice of case for the adjective by performing binary splits based on the available information about gender, time period and numeral, as well as the other predictors summarized in Table 2 in section 3. Each split yields the cleanest possible separation of the adjective case based on the available information. As shown in node 1 at the top of the tree diagram, the model first takes gender and sorts masculine and neuter nouns in one group (the left branch from node 1) and feminines in another group (the right branch from node 1). The regression analysis discussed in the previous section supports this, insofar as masculine and neuter genders are shown to prefer adjectives in the genitive, whereas feminine gender favors adjectives in the nominative.

Let us first consider the masculine and neuter nouns. As shown by the branches under node 2, the CART model does not distinguish between masculine and neuter nouns, which indicates that these two genders develop similarly. Under node 2 we see a split that groups examples from before and after 1925 in different groups. The situation before 1925 involves further splits for the three numerals *dva* 'two', *tri* 'three' and *četyre* 'four' (spelled "chetyre" in the figure in order to avoid diacritics in the statistical software).

The terminal nodes are bar diagrams which indicate the relative distribution of adjective forms in the nominative/accusative (represented as "Nom" in the figure) and genitive ("Gen"). As can be seen from nodes 7-8, the likelihood of the genitive is higher for 'four' and 'three' than for 'two'. This is as predicted by Gallis (1947), as mentioned in the previous section, and the result is furthermore supported by the regression analysis mentioned above, according to which 'two' favors the nominative, while 'three' and 'four' tend to favor the genitive in the period before 1925.

After 1925 the situation changes considerably, as shown in nodes 4-5. In the 1925-49 period, the genitive is used in about 75% of the examples (node 5). After 1950, the genitive is used almost exclusively (node 4). Once again, the interactions from the regression analysis show the same trend; after 1925 the genitive is favored by all numerals combining with masculine or neuter nouns.

We now turn to the right portion of the tree diagram (nodes 9-15), which represent the distribution for feminine nouns. While the masculine and neuter nouns showed a gradual increase of genitive adjectives, the figure shows that for feminine nouns it is the nominative/accusative that takes over. The nodes under node 13 represent the situation between 1850 and 1950. During this period, 'four' had nominative adjectives in about 60% of the examples (node 15), whereas 'two' and 'three' strongly preferred the genitive (node 14).

Nodes 11-12 represent the situation before 1850 and after 1950. In node 11, we see that nominative adjectives were strongly preferred. Node 11 also covers the period 1950-1999, thus suggesting a bell-shaped development, whereby the post-war period goes back to the distribution from before 1850. The increasing use of the nominative continues; in the twenty-first century the nominative is used in more than 80% of the examples, as shown in node 12. This result is quite close to the findings of Pereltsvaig (2010: 425), who reports 70% nominative, and Madariaga and Igartua (2017: 109), who found nominative adjectives in 60%-65% of the examples. Both Pereltsvaig's and Madariaga and Igartua's findings are based on surveys, not corpus data. The regression analysis provides a similar picture; until 1950 the genitive is still favored, especially in combination with 'two', but then the nominative gradually takes over.

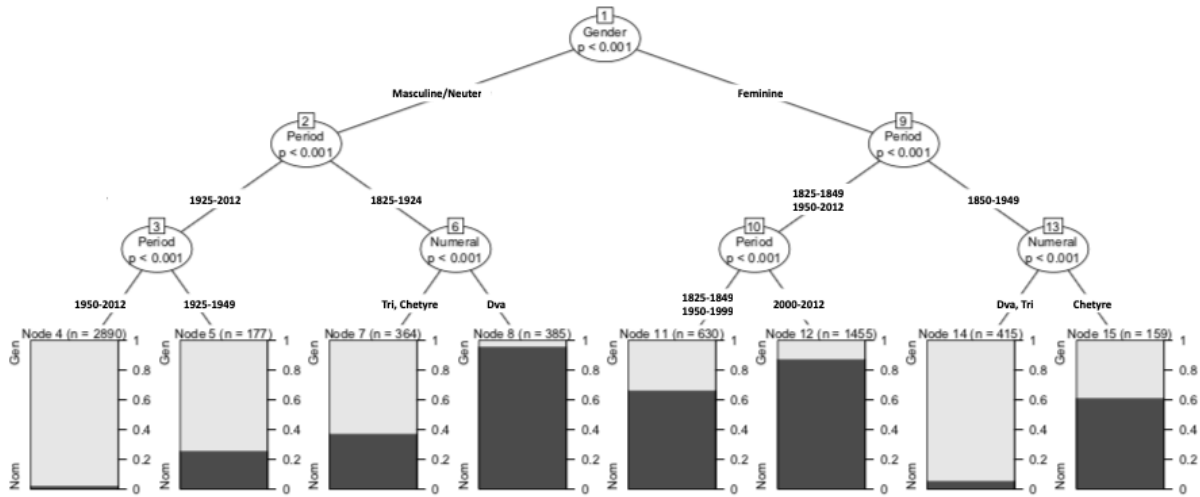


Figure 3: CART with three independent variables (gender, period, and numeral)

The following conclusions can be drawn from the CART analysis. First, the fact that gender and period populate the nodes high up in the tree diagram in Figure 3, while the numerals are found further down, reinforces the result of the Random Forest analysis (see Figure 2 in section 3), which showed that gender and period are the best predictors. Second, the fact that other predictors do not appear in the decision tree at all supports the conclusion from section 3 that these factors cannot predict the distribution of the adjective forms in the dataset under scrutiny in the present study. Third, the tree diagram clearly shows that the adjective rivalry has undergone radical change. In particular, we see that constructions with masculine and neuter nouns have nearly completely ousted adjectives in the nominative/accusative (see node 4), while the opposite development has taken place for feminine nouns, where nominative/accusative adjectives now dominate, as shown in node 12. Clearly, we are dealing with a split, as a result of which gender-specific paucal constructions emerge. In section 2, we saw that the opposition between the feminine and masculine/neuter genders has been relevant for the development of paucal constructions ever since the loss of the dual in medieval times. In this sense, the gender-specific paucal constructions have gone through a truly long birth process. I hasten to add that the statement about the long birth process is a purely descriptive statement capturing the relevance of gender over a long period of time. The statement is not intended to suggest causal links between changes distant in time, nor does the statement entail a teleological view of language change, whereby diachronic change takes place in order to fulfill a future goal.

The emergence of gender-specific paucal constructions raises an interesting question: “how can there be a gender distinction in a plural context?” (Bailyn and Nevins 2008: 264). In Russian, the opposition between the masculine, neuter and feminine genders is neutralized in the plural, where nouns of all three genders display identical agreement patterns. It is not possible to do justice to the relationship between the categories of gender and number, since this would require detailed discussions of theoretical positions that are beyond the scope of the present study. I will therefore limit myself to the following brief remarks. If one is not willing to accept the paucal constructions as exceptions to the generalization that the gender opposition is neutralized in the plural, there are two approaches one may pursue.

First, one may dispute the correctness of the statement that Russian has neutralized the gender opposition in the plural. While there is no distinction between masculine, neuter and feminine agreement in the plural, Russian has different agreement patterns in plural contexts

for animate and inanimate nouns. Corbett (1991: 161-168) has shown that animacy is an important part of the Russian gender system, and gender oppositions are therefore still found in plural contexts.

A second possible approach is to dispute the correctness of the claim that paucal constructions involve plural contexts. As mentioned in section 1, the traditional analysis of paucal constructions is that the noun is in the genitive *singular*. In recent years, this analysis has been defended from very different theoretical points of view by Andersen (2006) and Pesetsky (2013). If one accepts that the paucal constructions involve singular nouns, the fact that masculine/neuter and feminine nouns tend to combine with adjectives with different agreement endings (–ye vs. –yx) appears less surprising.

5. S-curves in language change – a cascading effect?

The idea that language change can be modeled as S-curves has figured prominently in diachronic linguistics in recent years, and the question therefore arises as to whether the changes explored in the previous sections follow S-curves. In the following, we will see that there is some support for S-curves, but that the analysis also enables us to address general problems concerning scarcity and interpretation of data. Furthermore, the available data suggest a cascading effect, whereby a new change starts when the previous change approaches culmination.

S-curves are curves that have a relatively flat shape in the beginning, then rise steeply, before they flatten out towards the end: “[a] given change begins quite gradually; after reaching a certain point [...], it picks up momentum and proceeds at a much faster rate; and finally tails off slowly before reaching completion” (Bailey 1973: 77). Mathematically, it has been suggested that it is the logistic function that underlies S-curves (Kroch 1989a-b). According to “Piotrowski’s law”, S-curves are crucial in language change (Leopold 2005), and it has been argued that S-curves represent a “kind of template for [language] change” (Chambers 2002: 361, see also Nessel and Makarova 2018 for discussion). Blythe and Croft (2012: 280), who have investigated a considerable number of changes of the S-shaped type, go as far as to claim that “there are no clearly documented cases of change going toward completion that follows either a simple linear trajectory or an exponential curve”. Importantly, S-shaped curves have proved relevant for a wide variety of language changes, ranging from sound change to syntactic change. For example, Labov (1994:65-75) relates S-curves to vowel change in Philadelphia, while Kroch (1989a-b, 2003) discusses the S-shaped nature of syntactic change regarding *do*-insertion in Old English.

Against this background, we would expect the changes in the adjective rivalry under scrutiny to follow S-curves. In order to test this hypothesis, we will take a closer look at the interaction of gender and time period, which are the most important predictors, as shown in the previous sections. Table 3 summarizes the development for masculine, neuter and feminine nouns. For each period and gender, the table gives the number of examples with adjectives in the nominative/accusative (#Nom), the number of examples with the genitive (#Gen), as well as the percentage of genitive examples (%Gen).

	Masculine nouns			Neuter nouns			Feminine nouns		
	#Nom	#Gen	%Gen	#Nom	#Gen	%Gen	#Nom	#Gen	%Gen
1825–49	82	12	13	38	5	12	10	5	33
1850–74	113	18	14	37	16	30	17	20	54
1875–99	87	18	17	36	30	45	15	65	81
1900–24	88	85	49	27	62	70	44	193	81
1925–49	34	68	67	15	64	81	34	177	84
1950–74	8	99	93	5	122	96	152	90	37
1975–99	13	183	93	1	277	100	251	119	32
2000–12	21	1545	99	11	611	98	1268	185	13

Table 3: Development over time for nouns of different genders

The data from Table 3 are visualized in Figure 4, which provides one curve for each gender. Do these curves have the expected S-shape? Constructions with masculine nouns (represented as a dashed line in the figure) display a flat contour in the beginning and then a steep rise at the end of the nineteenth and in the first half of the twentieth century, before the curve flattens out towards the end of the twentieth century. This appears to be a good example of an S-curve.

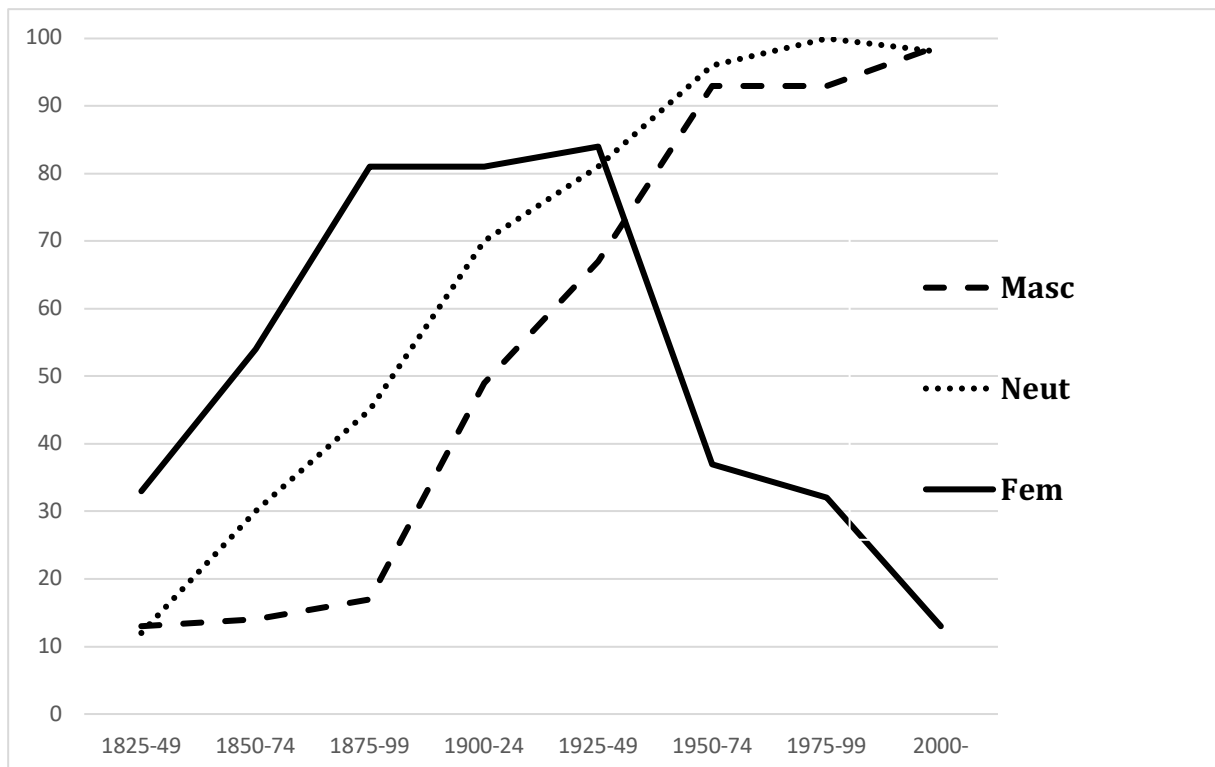


Figure 4: Development over time for nouns of different genders (measured in percentage of genitive)

Neuters follow a similar path. We observe a steep rise in the nineteenth and twentieth centuries followed by a flatter shape at the end of the twentieth and the beginning of the twenty-first century. However, the neuters lack the flat part at the beginning of the curve, so we are not dealing with a perfect example of an S-curve, if this curve can be considered S-shaped at all. At least three possible reasons for the deviations from the S-shape come to mind. First, it is possible that the dataset under scrutiny is too small to show the precise shape

of the curve. If we had more data, one might think, the development would come out as a perfectly S-shaped curve. A second possibility is that we do not have data far enough back in time; if we had reliable data from before 1825, we might be able to see the expected flat contour in the beginning. A third possible reason for the observed deviations concerns the interaction between gender and numeral. Table 3 and Figure 4 conflate the data for all three numerals (*dva* ‘two’, *tri* ‘three’ and *četyre* ‘four’), although the CART analysis explored in section 4 suggests differences between constructions with different numerals. Ideally, we should have been able to draw nine curves in Figure 4, three for each numeral in the three genders. Then we would be able to evaluate each curve to determine to what extent it conforms to the S-shape. Unfortunately, this is not possible because we do not have enough data for the earlier periods. While we can say with a high degree of confidence that all numerals in all genders have ended up as indicated in Figure 4 (cf. nodes 4 and 12 in Figure 3), we have less data for the earlier periods.

All three potential reasons boil down to one fundamental problem that goes beyond Russian numerals and is familiar to most historical linguists, viz. scarcity of data (Berdičevskis and Eckhoff 2014), which makes a reliable test of the S-curve hypothesis difficult. The development of the feminines raises additional problems, since in Figure 4 the feminines follow a bell-shaped rather than an S-shaped curve, a difference that cannot be due to scarcity of data. The curve clearly shows that the feminines diverge from the masculines and neuters, but is it possible to reconcile the development of the feminines with the S-curve hypothesis? Two things would need to be done – one trivial and one controversial. Consider Figure 5, where both manipulations have been carried out.

The trivial manipulation is that in Figure 5 the development of the feminines is measured as the percentage of nominative/accusative adjectives (which increases), instead of the percentage of genitive adjectives (which decreases). Thus, in Figure 5 the gray dashed and dotted lines indicate the rise of genitive adjectives in combination with masculine and neuter nouns, while the black solid line represents the increase of nominative/accusative adjectives for the feminine nouns. This makes it easier to relate their development of all genders to the S-curve hypothesis, but does not involve any substantial changes.

The more controversial manipulation is that the oldest part of the development for the feminines has been omitted. The rationale is as follows. In the beginning, the feminines followed a development that is roughly parallel with the S-curves of the other genders, and the development of the feminines was therefore part of a general change that affected all genders. Arguably, therefore, the (part of the) curve that is relevant for the development of the feminines starts when the feminines begin to deviate from the general trend. Therefore, the black solid curve in Figure 5 shows the development from the time when the feminines start to diverge from the masculines and neuters and begin to live a life of their own, i.e. the development that created gender-specific paucal constructions in Russian.

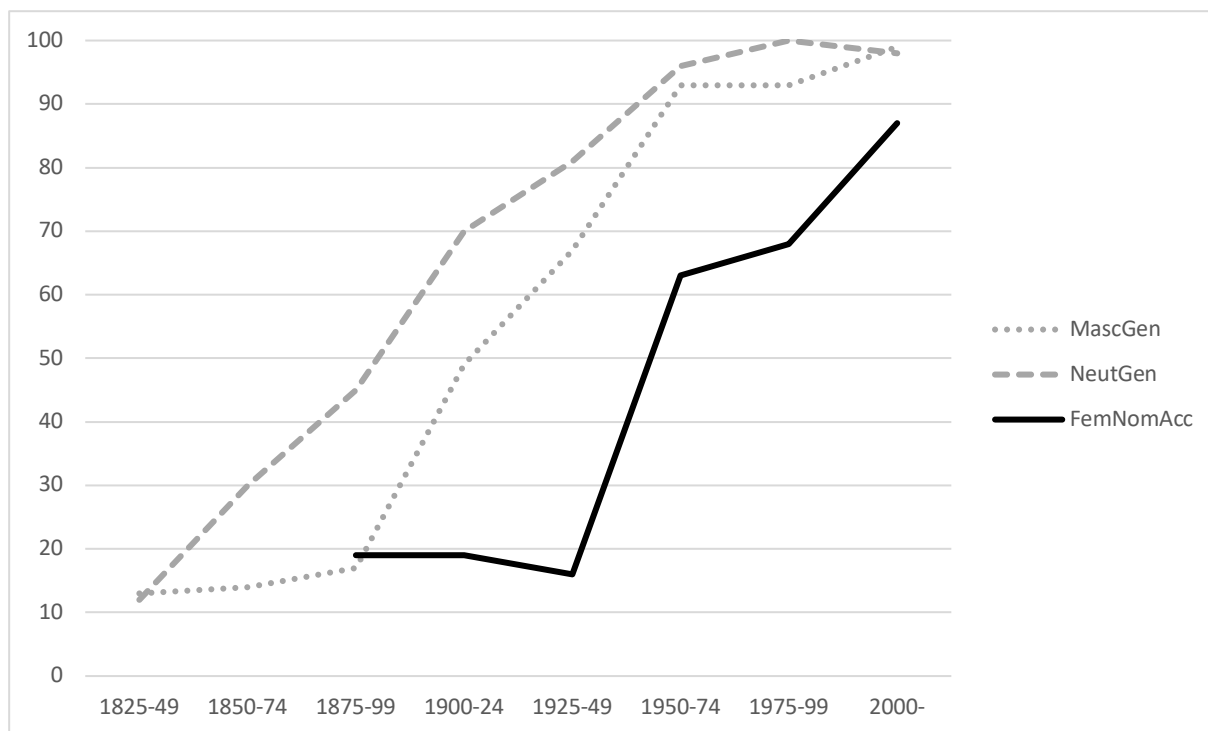


Figure 5: Development over time for nouns of different genders. The curves in gray indicate the growing use of the genitive, while the black curve measures the increasing use of the nominative/accusative.

The black curve resulting from the two manipulations resembles an S-curve, since it is flat in the beginning and then shows a steep rise. Whether it will flatten out towards the end remains to be seen – in a few decades we will know. As presented in Figure 5, the curve for the feminines does not contradict the S-curve hypothesis, but are the manipulations in Figure 5 legitimate? It is not a daring guess that many historical linguists would not approve. However, we are dealing with a matter of interpretation, for which it is hard to establish objective and unambiguous criteria. Thus, the manipulated curve in Figure 5 illustrates that the problems of testing the S-curve hypothesis go beyond scarcity of data, since in addition the hypothesis raises complex issues of interpretation of the data. While the problems explored above pertain to numerals in Russian, similar issues of data interpretation are likely to be relevant for the analysis of other phenomena in other languages as well.

The juxtaposition of the black solid curve and the gray dotted and dashed curves in Figure 5 brings out one aspect of the change in the paucal constructions that cannot be seen so clearly in the other visualizations in Figures 3 and 4. What we see in Figure 5 is that the feminine curve starts rising when the masculine and neuter curves are approaching their peaks and before they start to flatten out. I speculate that this is not a coincidence. It stands to reason that the use of genitive adjectives in the paucal construction must have been established as the general rule before the nominative/accusative could be monopolized by the feminines. Only when nominative/accusative adjectives were no longer widely used with masculine and neuter nouns, were the adjectives in the nominative/accusative “freed up” for use in the feminine gender, and could gradually become the rule for this gender. Stated in more general terms, only when a general rule is in the process of being established, I suggest, is it possible to deviate from it and create a new pattern. This may give rise to a cascading effect, whereby a new S-curve starts when the previous curve is about to flatten out at its peak. Before we leave the cascading effect, two remarks are in order. First, I hasten to add that that my account of the cascading effect is speculative; more evidence from other

examples of language change are needed before firm conclusions can be drawn.¹² Second, although the S-curves in Figure 5 provide a straightforward visualization of the cascading effect, the effect does not depend on S-curves. It is possible to describe the effect without reference to S-curves: the situation first stabilizes for masculine and neuter nouns and then for feminines.

6. Concluding remarks

On the basis of the present study of more than 6,000 corpus examples the following conclusions may be drawn. First, detailed statistical analysis has documented that the distribution of adjective forms in Russian paucal constructions has undergone radical change over the last two hundred years. Second, we have seen that a split took place in the twentieth century; in combination with masculine and neuter nouns, adjectives in the genitive have ousted their nominative/accusative rivals, while the opposite development has taken place in constructions with feminine nouns. Third, it has been argued that this development is the last step in a long “birth process” of gender-specific paucal constructions, which started with the loss of the dual in the Middle Ages. Fourth, it has been proposed that we may be witnessing a cascading effect of language change, whereby a new change accelerates when the previous change is approaching culmination. Fifth, although the changes under scrutiny have been shown to resemble S-curves, we have seen that empirical test of the S-curve hypothesis faces some challenges due to scarcity of data and issues of interpretation – challenges that are likely to generalize beyond the study of Russian paucal constructions.

The proposed analysis opens up a number of alleys for future research. While I have shown that three factors (gender of quantified noun, time period and the numeral itself) are sufficient to account for the nominative/accusative vs. genitive rivalry, we do not have enough data to analyze the impact of the numeral in detail. It is furthermore possible that a larger database might testify to the relevance of further variables in different periods of time. In particular, it would be interesting to further investigate the feminine nouns in present-day Russian. As shown, the genitive is still used in 13% of the examples after the year 2000, and it is conceivable that the genitive will carve out a niche for itself instead of dying out completely. Finally, my proposal about a cascading effect needs further testing on other examples of language change, in particular from languages other than Russian. However, these issues are beyond the scope of the present study and must be left for future research.

References

- Andersen, Henning (2006): Some Thoughts on the History of Russian Numeral Syntax. *Harvard Ukrainian Studies* 28.1–4: 57–67.
- Baayen, R. Harald, Anna Endresen, Laura A. Janda, Anastasia Makarova and Tore Nessel (2013): Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37.3: 253–291.
- Bailey, Charles-James (1973): *Variation and linguistic theory*. Washington: Center for Applied Linguistics.
- Bailyn, John F. and Andrew Nevins (2008): Russian genitive plurals are impostors. In: Asaf Bachrach and Andrew Nevins (eds.): *Inflectional identity*. Oxford: Oxford University Press, 237–270.

¹² It is worth mentioning that the cascade effect in Russian numerals resembles the development of vowels in Philadelphia, as described in Labov (1994: 67).

- Berdičevskis, Aleksandrs and Hanne Eckhoff (2014): Verbal constructional profiles: possibilities and limitations. In: Verena Henrich, Erhard Hinrichs, Daniël de Kok, Petya Osenova and Adam Przepiórkowski (eds.): *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT 13)*. Tübingen: University of Tübingen.
- Blythe, Richard A. and William Croft (2012) S-curves and the mechanisms of propagation in language change. *Language* 88.2: 269–304.
- Chambers, J.K (2002): Patterns of variation including change. In J.K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.): *Handbook of language variation and change*. Oxford: Blackwell, 349–372.
- Corbett, Greville G. (1991): *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. (1993): The head of Russian numeral expressions. In Greville G. Corbett, Norman M. Fraser and Scott McGlashan: *Heads in Grammatical Theory*. Cambridge: Cambridge University Press, 11–35.
- Gallis, Arne (1947): Tallordenes syntaks i russisk. In Erik Krag and Arne Gallis (eds.): *Festskrift til professor Olaf Broch på hans 80-årsdag: fra venner og elever*. Oslo: Det norske videnskapsakademi, 63–75.
- Gorbačevič, Kirill S. (1971): *Izmenenie norm russkogo literaturnogo jazyka*. Leningrad: Prosveščenie.
- Graudina, Ljudmila K., Viktor A. Ickovič and Lija P. Katlinskaja (2001): *Grammatičeskaja pravil'nost' russkoj reči*. Moscow: Astrel'.
- Lordanskij, Anatolij M. (1958): Istorija upotreblenija soglasovannyx opredelenij pri sočetačnijax čislitel'nyx dva, tri, četyre s imenami suščestvitel'nymi v russkom jazyke. *Učenyje zapiski Vladimirskogo gosudarstvennogo pedagogičeskogo instituta* 4: 54–82.
- Išačenko, Alexander V. (1962): *Die russische Sprache der Gegenwart*. Halle: Niemeyer.
- Kroch, Anthony (1989a): Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change* 1: 199-244.
- Kroch, Anthony (1989b): Function and grammar in the history of English: periphrastic *do*. In Ralph W. Fasold and Deborah Schiffrin (eds.), *Language change and variation*. Amsterdam and Philadelphia: John Benjamins, 133-172.
- Kroch, Anthony (2003): Syntactic change. In Mark Baltin and Chris Collins (eds.), *The handbook of contemporary syntactic theory*. Oxford: Blackwell, 699-729.
- Labov, William (1994): Principles of language change. Volume 1: Internal factors. Oxford: Blackwell.
- Leopold, E. (2005). Diachronie: Grammatik. In R. Köhler, G. Altmann & R. Piotrowski (eds.), *Quantitative Linguistics: An International Handbook*. Berlin and New York: De Gruyter, 607-633.
- Levshina, Natalia (2015): *How to do linguistics with R*. Amsterdam: John Benjamins.
- Madariaga, Nerea and Iván Igartua (2017): Idiosyncratic (Dis)agreement Patterns: The Structure and Diachrony of Russian Paucal Subjects. *Scando-Slavica* 63.2: 99–132. <https://doi.org/10.1080/00806765.2017.1390922>
- Mel'čuk, Igor A. (1985): *Poverxnostnyj sintaksis russkix čislovyx vyraženij*. Wiener *Slawistischer Almanach Sonderband* 16, Vienna.
- Neset, Tore (2015): *How Russian Came to Be the Way It Is*. Bloomington: Slavica Publishers.
- Neset, Tore (2019): Big data in Russian linguistics? *Zeitschrift für Slawistik* 64.2: 157–174. <https://doi.org/10.1515/slwa-2019-0012>

- Nesset, Tore and Anastasia Makarova (2018): The decade construction rivalry in Russian: Using a corpus to study historical linguistics. *Diachronica* 35.1: 71–106. doi 10.1075/dia.16043.nes
- Pereltsvaig, Asya (2010): As Easy as Two, Three, Four. *Formal Approaches to Slavic Linguistics* 18: 418–435.
- Rappaport, Gilbert C. (2002): Numeral phrases in Russian: A minimalist approach. *Journal of Slavic Linguistics* 10.1-2: 327–340.
- Rozental', Ditmar Ė. (1987): *Praktičeskaja stilistika ruskogo jazyka* (5th edition). Moscow: Vysšaja škola.
- Ryazanova-Clarke, Lara (2006): "The Crystallization of Structures": Linguistic Culture in Putin's Russia. In: Ingunn Lunde & Tine Roesen (eds.): *Landslide of the Norm: Language Culture in Post-Soviet Russia*. = *Slavica Bergensia* 6. Bergen: University of Bergen. Available at <http://boap.uib.no/books/sb/catalog/book/3>.
- Šaronov, Igor' (2014): Narodnaja ètimologija i količestvennye sočetanija v rusckom jazyke. *Antropologičeskij forum* 21: 137–144.
- Strobl, Carolin, James Malley and Gerhard Tutz (2009): An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14.4: 323–348.
- Suprun, A. (1957): K upotrebleniju roditel'nogo i imenitel'nogo padežej množestvennogo čisla prilagatel'nyx v sočetanijax s čislitel'nymi dva, tri, četyre v sovremennom rusckom jazyke. *Kirgizskij gosudarstvennyj zaočnyj pedagogičeskij institut, učenye zapiski* 3:72–84.
- Tagliamonte, Sali A. and R. Harald Baayen (2012): Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24.2: 135–178.
- Wade, Terence (2011): *A comprehensive Russian grammar* (Third edition). Oxford: Blackwell.
- Zaliznjak, Andrej A. (1967): *Russkoe imennoe slovoizmenenie*. Moscow: Nauka.
- Zaliznjak, Andrej A. (1985): *Ot praslavjanskoj akcentuacii k rusckoj*. Moscow: Nauka.
- Žolobov, Oleg F. (2002): Morfosintaksis čislitel'nyx dva, tri, četyre: k istorii malogo kvantitativa. *Russian Linguistics* 26.1: 1–27.
- Žolobov, Oleg F. (2003): K istorii malogo kvantitativa: adnumerativnye formy prilagatel'nyx i suščestvitel'nyx. *Russian Linguistics* 27.2: 177–197.

Author address:

Department of Language and Culture, UiT The Arctic University of Norway, NO-9037 Tromsø, Norway.

E-mail: tore.nesset@uit.no