Tore Nesset* and Laura A. Janda

A network of allostructions: quantified subject constructions in Russian

Abstract: This article contributes to Construction Grammar, historical linguistics, and Russian linguistics through an in-depth corpus study of predicate agreement in constructions with quantified subjects. Statistical analysis of approximately 39,000 corpus examples indicates that these constructions constitute a network of constructions ("allostructions") with various preferences for singular or plural agreement. Factors pull in different directions, and we observe a relatively stable situation in the face of variation. We present an analysis of a multidimensional network of allostructions in Russian, thus contributing to our understanding of allostructional relationships in Construction Grammar. With regard to historical linguistics, language stability is an understudied field. We illustrate an interplay of divergent factors that apparently resists language change. The syntax of numerals and other quantifiers represents a notoriously complex phenomenon of the Russian language. Our study sheds new light on the contributions of factors that favor singular or plural agreement in sentences with quantified subjects.

Keywords: agreement; allostruction; Construction Grammar; corpus; language change; quantifiers; Russian

1 Introduction

Modern Russian allows two rival agreement patterns for sentences with quantified subjects. In both examples below, the subject is *šest' čelovek* 'six persons', but in (1a) the verb is in the singular, while (1b) has the same verb in the plural:[1]

---

1 Examples with singular agreement are in the neuter gender in the past tense, while singular verbs in the non-past tense are in third person. We do not refer to gender or person, since number (singular vs. plural) is sufficient to distinguish between the rival constructions under scrutiny. All numbered examples are from the Russian National Corpus (main corpus) available at www.ruscorpora.ru and

*Corresponding author: Tore Nesset, UiT The Arctic University of Norway, Tromsø, Norway,
E-mail: tore.nesset@uit.no. https://orcid.org/0000-0003-1308-3506
Laura A. Janda, UiT The Arctic University of Norway, Tromsø, Norway, E-mail: laura.janda@uit.no.
https://orcid.org/0000-0001-5047-1909

(1) a. V   okončatel'n-om   sostav-e   bratstv-a   **osta-l-o-s'**
      [in   final-M.LOC.SG   set-LOC.SG   brotherhood-GEN.SG   remain-PST-N-REFL
      šest'   čelovek.   (Zaxarov 1988–2000)
      six.NOM   person.GEN.PL]
      'In the end six persons remained in the brotherhood.'

   b. Posle   ix   uxod-a   na   l'din-e   **osta-l-i-s'**
      [after their   departure-GEN.SG on   ice.block-LOC.SG remain-PST-PL-REFL
      šest'   čelovek […].   (*Znanie-Sila* 1997)
      six.NOM   person.GEN.PL]
      'After their departure six persons remained on the block of ice.'

We analyze the singular versus plural alternation of verbs in the Russian quantified subject construction on the basis of nearly 39,000 corpus examples. We find that this alternation has remained largely stable at the rate of about 44% singular versus 56% plural for the past two centuries. The distribution of singular versus plural verb forms is associated with several other factors: the presence of premodifiers (modifiers immediately preceding the quantifier), Quantifier Type and Frequency, Word Order, and Animacy of the subject, in addition to the individual preferences of specific quantifiers and verbs.

This study contributes to the theoretical framework of Construction Grammar by presenting an in-depth empirical study of the role of allostructions in a grammatical alternation. The focus on Russian furthermore expands the horizon of allostructional analysis, which has mainly been concerned with data from English and, to some extent, other Germanic languages (Cappelle 2006; Diessel 2019: 199–200; Grafmiller et al. 2018; Ungerer 2021; Van de Velde 2014).[2] Since the allostructional network we explore includes both vertical and horizontal relations among allostructions, we offer empirical evidence for the importance of such relations in Construction Grammar.

The present study also sheds new light on the role of stasis (as opposed to S-curve progression, Blythe and Croft 2012) in diachrony. In addition to a theory of language change, we need a theory of language stability, i.e., situations where language does not change. Our analysis of allostructional relations contributes to this field since we

---

are supplied with the year the example was created and the name of the author (of works of fiction) or the periodical where it was published (for non-fiction). For the convenience of the reader, the quantified subject and the predicate are boldfaced in each example.

2 The strong focus on English is also characteristic of syntactic alternations in general, where there is a long tradition for analyzing the so-called dative, genitive, and locative alternations. An extensive discussion and bibliography are beyond the scope of the present article (but see, e.g., Bresnan 2007; Gries and Stefanowitsch 2004; Iwata 2005; Pinker 1989; Rappaport-Hovav and Levin 2008 and references therein).

show that a multidimensional network of constructions need not serve as a context for language change (cf. Traugott 2018).

From a descriptive perspective, this study gives better focus to our understanding of the number alternation with quantified subjects in Russian. This topic has attracted considerable scholarly attention, but no previous analysis has taken into account the combination of factors and their interactions using confirmatory statistics. We demonstrate that premodifiers enable us to state a nearly categorical rule, whereby premodifiers in the nominative case yield plural agreement. However, this factor is shown to have low cue availability, and is therefore of limited importance for the system as a whole. For the remaining data we show that Quantifier Type and Frequency, Word Order, and Animacy form the basis for statistical tendencies, rather than categorical rules. Word Order, Animacy, and Quantifier Type vary in a series of closely-related quantified subject constructions, or "allostructions" (see definition below), that admit the singular versus plural alternation.

Construction Grammar is the focus of Section 2. Section 3 presents a description of the quantified subject construction in Russian, the various factors that previous scholars have suggested might influence the choice of number in this construction, and our hypothesis based on this background. Our database, its source and structure are described in Section 4. A mixed-effects logistic regression model tests our hypothesis in Section 5. Discussion of the results of the statistical model are presented in Section 6. Conclusions are offered in Section 7.

2 Allostructions in Construction Grammar

Construction Grammar (Croft 2001; Fillmore 1988; Fried and Östman 2004; Goldberg 1995, 2006), which has co-evolved with Cognitive Linguistics (Croft and Cruse 2004; Dąbrowska and Divjak 2015; Geeraerts and Cuyckens 2007; Lakoff 1987; Langacker 1987, 1991a, 2008), takes the construction as the essential unit of language and explains linguistic behavior in terms of general cognitive strategies.

Construction Grammar is a usage-based approach that acknowledges that linguistic structure is shaped by language use (Barlow and Kemmer 2000; Bybee 2001; Diessel 2019; Divjak 2019; Langacker 1991b, 1999; Schmid 2020). Therefore, linguistic productions such as utterances and corpus examples are data that can serve as the basis for analysis. While it is impossible to state with certainty that corpus data reflect the cognitive reality of language, there is considerable evidence that native speakers are sensitive to the statistical distributions of phenomena in their languages, and that the results of corpus analysis and parallel experimental results tend to converge (Bresnan 2007; Divjak et al. 2016; Klavan and Divjak 2016; Taylor 2012). In other words, statistical models of corpus data are repeatedly found to be consistent

with experimental findings, although of course we must acknowledge some degree of interspeaker variation (Dąbrowska 2015).

A "construction" can be defined as an entrenched language-specific form-meaning pairing at any level of linguistic complexity. "Entrenched" means that the form-meaning pairing is repeated and learned (Langacker 1987, 2008; Schmid 2020). A construction can be entrenched regardless of where it falls along a scale ranging from relatively compositional (as in the Russian quantified subject construction, which transparently consists of a quantifier, an NP, and a verb) to non-compositional (as in idiomatic constructions like English *by and large*). The presence of the singular versus plural alternation is both repeated and learned as part of the Russian quantified subject construction. While the form of a construction is always language-specific, at a schematic level many constructions may be universal. It is reasonable to expect that most (maybe all) languages will have some means to express quantification of the subject of a verb. "Meaning" is broadly understood as a communicative function, thus including both lexical and grammatical meaning (Langacker 2008). "Complexity" spans the range of linguistic units, including morphemes, words, multi-word phrases, and discourse structures. Construction grammar endeavors to describe the entirety of a language in terms of constructions. From this perspective a language is a "constructicon" (see, e.g., Lyngfeldt et al. 2018).

In keeping with the framework of Cognitive Linguistics, the constructicon of a language is organized in accordance with what is known about general cognitive mechanisms, namely that human categorization is structured in networks of associations. This means that we expect to find complex relationships among the members of the constructicon. Diessel (2019) details Construction Grammar in terms of "constructional relations" that yield networks of constructions. Two types of relations are particularly relevant for our analysis of Russian quantified subjects: 1) hierarchical relations ranging from schematic "macro-constructions", through mid-level "meso-constructions", to individual "micro-constructions" (terms introduced by Traugott 2008a, 2008b, see also Pijpops et al. 2021 for discussion), and 2) lateral relations between "allostructions" (cf. Cappelle 2006) that perform nearly synonymous functions. Traugott (2008b: 238, 242) characterizes macro-constructions as "highly abstract schemas" that are at once both "primitive" and "presumably universal", as opposed to the lower-level meso- and micro-constructions which are language-specific.

We define "allostruction" as a member of a network of two or more grammatical constructions with (nearly) synonymous function and very similar form. A set of allostructions exists in a multidimensional network, with vertical relations to a more abstract macro-construction (schema) and horizontal relations among allostructions. Our data investigate allostructions that share the same function, namely that of signaling a quantified subject, and neighboring allostructions differ minimally in form in various ways, such as singular versus plural verb agreement, Word Order, Animacy of the subject, and Quantifier Type. Other studies of allostructional variation have focused primarily on English syntax, e.g., the ditransitive alternation (Bresnan 2007;

Gries 2003; Stefanowitsch 2006), the 's-genitive/of-genitive (Gries 2002; Rosenbach 2003), and the locative alternation (Boas 2006; Goldberg 1995, 2006).

While the status of hierarchical relations has always been uncontroversial in Construction Grammar, lateral (horizontal) relations among constructions have been subjected to considerable discussion. We may define "lateral relation" as a connection between constructions of the same level of specificity in a construction network. In Langacker's Cognitive Grammar such relationships are accounted for in terms of "extension relations" (Langacker 1987: 70). However, as pointed out by Diessel (2019: 199), the importance of lateral relations was downplayed in the early years of Construction Grammar. In opposition to generative linguistics, which assumed transformational relationships between constructions (e.g., active and passive, Radford 1988: 420–435), early proponents of Construction Grammar tended to analyze each construction in its own right (Goldberg 2002; Langacker 1991a: 464–471, see also Gries and Stefanowitsch 2004 for discussion). Cappelle (2006) argued that lateral relations are important; for him, allostructions are a means to represent alternations within the framework of Cognitive Linguistics, obviating any need to posit one variant as more "basic" than another (cf. also Goldberg 2002). Traugott (2018) models the vertical and horizontal relationships among related constructions as a multidimensional network.

Allostructional alternations have received considerable attention in recent years (Grafmiller et al. 2018), but, as mentioned in Section 1, most work focuses on English and relatively simple situations with only two allostructions. The strong English bias is unfortunate, since, as pointed out by Schmid (2020: 11), English is "quite exceptional, e.g., regarding its rudimentary system of grammatical categories marked by inflectional morphology." We explore a complex multidimensional network of allostructions. Our study lends empirical support to the importance of such relations in Construction Grammar.

3 Numeral syntax in Russian: previous scholarship and hypotheses

The syntax of numerals above '1' in Russian is notoriously complex (Babby 1987), owing in part to the fact that numerals did not constitute a separate part of speech in the antecedent language termed Late Common Slavic (Townsend and Janda 1996: 190–197). This complexity is partly due to the historical loss of the Dual number, and partly due to other diachronic changes (see Janda 1996: Ch. 4). Numerals for '2', '3', and '4' stem etymologically from adjectives and the numeral for '2' (along with 'both') continues to agree with nouns in gender: *dva/oba* '2/both (masculine or neuter)' versus *dve/obe* '2/both (feminine)'. Dual morphology for nouns spread also to *tri* '3', *četyre* '4', and *poltora/poltory* 'one and a half', expanding the syntactic pattern for what is now known as the "paucal" class of numerals. Dual endings for nouns

underwent reinterpretation, although the result is still a matter of debate: many formerly Dual endings can be interpreted in Modern Russian as either Genitive Singular or as Nominative Plural, and some scholars suggest that they should be given their own "paucal" designation (see, e.g., Corbett 1993: 16; Igartua and Madariaga 2018; Pereltsvaig 2010: 427). The inflection of adjectives that modify nouns quantified by paucal numerals has undergone historical change in the past two centuries, and at present the trend is to use Genitive Plural adjective forms with masculine and neuter nouns, but Nominative Plural adjective forms with feminine nouns (Nesset 2020).

Numerals for '5' and above stem historically from nouns, and when quantified by such numerals, both nouns and the adjectives that modify them appear in their Genitive Plural forms (with the exception of Nominative Plural premodifiers, see Section 3.1.1). In effect, Russian does not have a direct equivalent for a quantified phrase like English *five things*, but instead expresses *pjat' vešč-ej* [five.NOM thing-GEN.PL], literally '(a) five of things'. In addition to the paucal numerals and numerals for '5' and above, Russian has so-called collective numerals for groups from two to ten, like *troe* 'threesome' used primarily for counting pluralia tantum objects and groups of human beings (Isačenko 1982: 540; Timberlake 2004: 196), as well as indefinite quantifiers like *mnogo* 'many, a lot'. Both collective numerals and indefinite quantifiers follow the syntactic pattern of numerals for '5' and above, with the entire noun phrase appearing in the Genitive Plural. Thus *troe det-ej* [three.NOM child-GEN.PL], the normal way to say 'three children' is literally '(a) threesome of children' and likewise *mnogo det-ej* [many.NOM child-GEN.PL] 'many children' is literally '(a) lot of children'.

This situation becomes more complex when the quantified noun phrase is the grammatical subject and one has to choose a verb form. All four types of numerals – paucals, numerals '5' and above (henceforth merely "numerals"), collectives, and indefinites – admit both singular and plural verb forms of the types illustrated in examples (1a–b).

3.1 Previous scholarship and observations

In the copious scholarly literature on the syntax of Russian quantifiers, the singular-plural agreement rivalry in sentences with quantified subjects has received considerable attention. In addition to relatively brief descriptions in grammars (e.g., Sičinava 2012; Švedova 1980: 242–243; Timberlake 2004: 357–361; Wade 2011: 226–228,) and prescriptive works (e.g., Gorbačevič 1989; Graudina et al. 2001; Rozental' and Telenkova 1976), numerous studies explore the theoretical implications of the singular-plural agreement rivalry, mostly for various versions of generative grammar (e.g., Franks 1995; Glushan 2013; Madariaga and Igartua 2017; Pereltsvaig 2006; Pesetsky 2013). Among the more empirically oriented studies of the factors that influence the choice between singular or plural agreement, Corbett's (1983, see also

2000, 2006) pioneering study identified a number of factors relating to the grammatical subject of the sentence (the agreement "controller" in Corbett's terminology), while Robblee (1993) studied factors pertaining to the predicate of the sentence (the agreement "target"). The following subsections address the factors that might contribute to the singular versus plural alternation with Russian quantified subjects based on previous scholarship and observations.

3.1.1 Premodifiers

We use the term "premodifier" to refer to an adjectival or pronominal element preceding the quantifier, as illustrated by *èti* 'these' and *celye* 'whole', both in Nominative Plural in (2a–b):[3]

(2) a. **Èt-i**   **pjat'**   **predloženij**   **da-l-i-s'**   mne
         [this-NOM.PL   five.NOM   sentence.GEN.PL   give-PST-PL-REFL   I.DAT
         tjažel-ee,   čem   ves'   ostal'n-oj   tekst.
         difficult-CMP   than   whole.M.NOM.SG   remaining-M.NOM.SG   text.NOM.SG]
         (*Russkij reporter* 2013)
         'These five sentences were harder to process than all the rest of the text.'

    b. […] **cel-ye**   **tri**   **knig-i**   **okazyva-jut-sja**
         [whole-NOM.PL   three.NOM   book-GEN.SG   turn.out.to.be-PRS.3PL-REFL
         pod   zapret-om.   (Rassadin 2004–2008)
         under   prohibition-INS.SG]
         'As many as three books turn out to be forbidden.'

Many researchers have argued that nominative premodifiers favor plural agreement and claimed that this is a categorical rule. For instance, Crockett (1976: 335) writes that "when a phrase which contains a numeral in the nominative case also contains a pluralized attributive in the nominative case, then a verb associated with it must be pluralized" (see also *inter alia* Kuz'minova 2004: 41; Pereltsvaig 2006: 441; Rozental' and Telenkova 1976: 246; Švedova 1980: 243; Timberlake 2004: 359). If plural agreement with nominative premodifiers is indeed a categorical rule devoid of variation, it will make sense to focus the analysis on examples without nominative premodifiers since those are the examples with the relevant singular versus plural variation.

    We assess the presence of a nominative plural premodifier in terms of cue validity. Cue validity is the product of cue reliability – the percentage of cases in which a cue gives the expected result, multiplied by cue availability – the percentage of cases where the cue is present (cf. Dittmar et al. 2008; Goldberg 2006: 105–126;

---

3 For discussion of the syntactic constituency of premodifiers, see Babby (1987: 117–136) and Pereltsvaig (2006: 430–432).

MacWhinney et al. 1984; Perek and Goldberg 2017). In our dataset of 38,988 examples (described in detail in Section 4), there are 2,806 examples that contain premodifiers in the nominative. Of these, 2,801 examples occur with plural verb forms, while only five examples[4] occur with singular verb forms. It is noteworthy that all five exceptional examples instantiate the construction with a singular verb form before an

---

4 The five examples are:

(i)　　Konečno,　prošč-e　　vs-ego　　　　napisa-t':　v　berezov-yx　les-ax
　　　　[of.course　easy-CMP　all-N.GEN.SG　write-INF　in　birch-LOC.PL　forest-LOC.PL
　　　　obita-et　　　　**vs-e**　　　　**tri**　　　**vid-a**　　　　berezov-yx
　　　　dwell-PRS.3SG　all-NOM.PL　three.NOM　type-GEN.SG　birch-GEN.PL
　　　　grib-ov […].　　　　　　　　　　　　　　　　　　　　　(Solouxin 1967)
　　　　mushroom-GEN.PL]
　　　　'Of course, the easiest thing is to write: in the birch forests dwell all three kinds of shelf fungi […].'

(ii)　　Zanovo　**by-l-o**　　**vozved-en-o**　　　**vs-e**　　　**devjat'**　　**vokzal-ov**,
　　　　[again　be-PST-N　reconstruct-PST.PTCP-N　all-NOM.PL　nine.NOM　station-GEN.PL
　　　　gde　　ostanavliva-jut-sja　　poezd-a.　　　　　　　　　(*Znanie-sila* 2009)
　　　　where　stop-PRS.3PL-REFL　　train-NOM.PL]
　　　　'Once again, all nine stations where trains stop were reconstructed.'

(iii)　　Proved-en-o　　　zasedani-e　　　konsul'tativn-ogo　　sovet-a
　　　　[carry.out-PST.PTCP-N　meeting-NOM.SG　consulting-M.GEN.SG　board-GEN.SG
　　　　aviakompanij,　　　　na　kotor-om　　　v　okončatel'n-om
　　　　airline.company.GEN.PL　on　which-N.LOC.SG　in　final-M.LOC.SG
　　　　variant-e　　　**predlož-en-o**　**vs-e**　　**šest'**
　　　　variant-LOC.SG　offer-PST.PTCP-N　all-NOM.PL　six.NOM
　　　　versij　　　　samolet-ov　　　semejstv-a […].
　　　　version.GEN.PL　airplane-GEN.PL　family-GEN.SG]
　　　　　　　　　　　　　　　　　　　　　(*Vestnik aviacii i kosmonavtiki* 2004)
　　　　'A meeting of the consulting board of the airline companies was held, at which all six versions of the family of airplanes were offered in their final variant.'

(iv)　　Čt-o　　**znač-it**　　**kak-ie-to**　　　**poltor-y**　　　**butylk-i**?
　　　　[what-ACC　mean-PRS.3SG　some-NOM.PL-that　one.and.a.half-F.NOM　bottle-GEN.SG]
　　　　　　　　　　　　　　　　　　　　　　(Averčenko 1910–1911)
　　　　'What does as little as one and a half bottles mean?'

(v)　　V　ee　al'bom-e　　emu　**by-l-o**　　**posvjašč-en-o**　　**cel-ye**
　　　　[in　her　album-LOC.SG　he.DAT　be-PST-N　dedicate-PST.PTCP-N　whole-NOM.PL
　　　　dv-e　　**stranič-k-i.**　　　　　　　　　　　　　(Simonova 2002)
　　　　two-F.NOM　page-DIM-GEN.SG]
　　　　'In her album all of two little pages were dedicated to him.'

inanimate subject. However, there are also 592 examples of premodifiers with the same combination of verb-before-subject word order and an inanimate subject where the verb form is plural. Overall, the cue reliability of the presence of a nominative plural premodifier is excellent, since in over 99.8% of such cases the verb appears in a plural form. We can indeed consider the prediction of a plural verb form in the presence of a nominative plural premodifier to be for all practical purposes a reliable rule. However, the cue availability is poor, since less than 7.2% of examples contain a Nominative Plural premodifier. Multiplying 99.8% cue reliability by 7.2% cue availability yields less than 7.2% cue validity. Since this is a low value because premodifiers are rare in our data, we leave aside the examples with nominative premodifiers in the remainder of our analysis and focus instead on the remaining 36,182 examples that represent nearly 93% of the total dataset where there are no such straightforward predictors. We need to look elsewhere to account for the observed alternation.

3.1.2 Diachrony

In view of the fact that the presence of competing variants is often indicative of language change (Chambers 2002), and the specific fact that other variations in Russian numeral constructions are known to be undergoing language change (for an example see Nesset 2020), it is surprising that the diachrony of the singular versus plural alternation has received little attention in the extensive literature on Russian quantifiers. Some prescriptivist sources claim that plural agreement has increased (Gorbačevič 1971: 256; Rozental' 1974: 218–228). Corbett (1981), who surveyed the – admittedly limited – empirical data available at the time, found no evidence for an increasing use of plural agreement. Against this background, it is reasonable to examine our data from a longitudinal perspective, and we therefore include the year each example was created in our investigation.

3.1.3 Quantifiers

The corpus frequencies of numerals vary, with a tendency for lower numerals to have higher frequencies than higher numerals (for example *dva* 'two' is more frequent than *dvadcat'* 'twenty', which is more frequent than *dvesti* 'two hundred'). An often-made observation is that subjects with lower numerals (paucals and collectives) show a tendency to combine with verbs in the plural (e.g., Corbett 1983: 230, 2000: 214; Madariaga and Igartua 2017: 102; Rozental' and Telenkova 1976: 246; Timberlake 2004: 359). Thus, the numeral *dva* 'two' is expected to prefer a verb in the plural, while higher numerals more frequently show singular agreement. Most indefinite quantifiers are said to prefer singular agreement (Kuz'minova 2004: 40; Švedova 1980: 243), although some scholars have pointed out that the quantifier

neskol'ko 'some' may occur with plural agreement. According to Gorbačevič (1989: 192), in the 1930s L. Ščerba considered such plural agreement incorrect, but more recent prescriptivist sources accept both singular and plural agreement (e.g., Graudina et al. 2001: 36; Rozental' and Telenkova 1976: 248). In order to sort out whether the distribution can be attributed to frequency alone or is also affected by the type of quantifier or any preferences specific to a given quantifier, all three kinds of information are included in our analysis.

3.1.4 Animacy

A relevant factor frequently discussed in the literature is the Animacy hierarchy (Silverstein 1976, also referred to as the "individuation hierarchy", see, e.g., Sasse 1993; Timberlake 1985). Animacy is grammatically relevant for a variety of phenomena in Russian, most particularly in the marking of accusative case, where both masculine animate nouns in the singular and all animate nouns in the plural have differential marking syncretic with the genitive case, meaning that the marking of animate nouns is distinct from the accusative case marking of inanimate nouns. While different researchers assume different versions of the hierarchy, the consensus is that – other things being equal – animate nouns are more likely to cooccur with plural agreement than are inanimate nouns (Andersen 2006: 65; Corbett 1983: 143, 2000: 214; Glushan 2013: 259; Gorbačevič 1989: 193; Madariaga and Igartua 2017: 115). Animacy of the quantified noun is therefore included in our analysis.

3.1.5 Word order as it relates to animacy

Previous researchers have pointed out that Word Order, specifically the location of the subject relative to the verb, is a relevant factor. Corbett (1983: 150) stated that "subject-predicate word-order makes the semantically justifiable agreement [i.e., plural] more likely than in predicate-subject order" (see also Corbett 2000: 214). Other scholars who have explored the impact of Word Order for some or all quantifiers include Gorbačevič (1989: 193), Graudina et al. (2001: 37), Kuz'minova (2004: 40), Madariaga and Igartua (2017: 102), and Švedova (1980: 242). They all agree that subject-verb (SV) Word Order makes plural agreement more likely, while verb-subject (VS) Word Order motivates singular agreement.[5]

It is well known that Russian Word Order interacts with the Animacy of the subject (see for example Lobanova 2011: 141), such that animate subjects show a

---

[5] In Russian, word order is related to information structure, but a comprehensive discussion of this complex relationship goes beyond the scope of this article. We refer the reader to Timberlake (2004: Chapter 7.3).
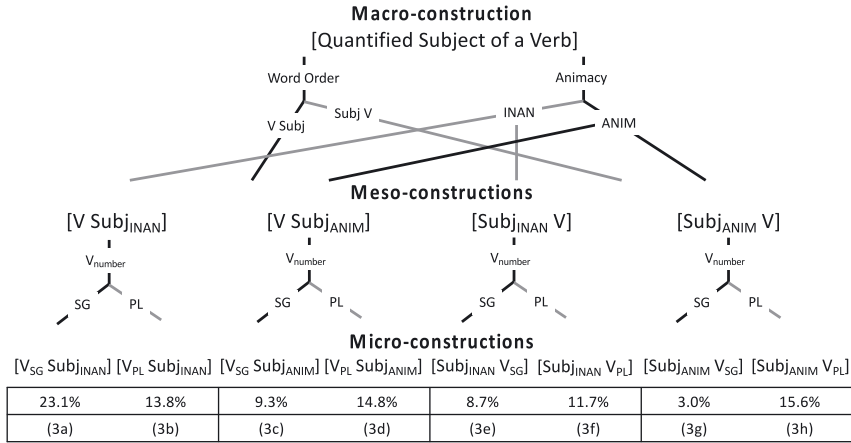
Macro-construction
[Quantified Subject of a Verb]

Word Order   Animacy

V Subj   Subj V   INAN   ANIM

Meso-constructions

[V Subj$_{INAN}$]   [V Subj$_{ANIM}$]   [Subj$_{INAN}$ V]   [Subj$_{ANIM}$ V]

V$_{number}$ V$_{number}$ V$_{number}$ V$_{number}$

SG   PL   SG   PL   SG   PL   SG   PL

Micro-constructions

[V$_{SG}$ Subj$_{INAN}$] [V$_{PL}$ Subj$_{INAN}$] [V$_{SG}$ Subj$_{ANIM}$] [V$_{PL}$ Subj$_{ANIM}$] [Subj$_{INAN}$ V$_{SG}$] [Subj$_{INAN}$ V$_{PL}$] [Subj$_{ANIM}$ V$_{SG}$] [Subj$_{ANIM}$ V$_{PL}$]

23.1%	13.8%	9.3%	14.8%	8.7%	11.7%	3.0%	15.6%
(3a)	(3b)	(3c)	(3d)	(3e)	(3f)	(3g)	(3h)

Figure 1: Variation of Word Order, Animacy, and verb number within the Russian quantified subject construction at the level of macro-, meso-, and micro-constructions. Numbers at the bottom of the figure indicate the relative distribution of micro-constructions and labels (3a)–(3h) indicate corresponding corpus examples.

preference for subject-verb order, while inanimate subjects show a preference for verb-subject word order. These preferences are tendencies, rather than rules, and all possible combinations of Animacy, Word Order, and Predicate Number are well attested in our data. This interaction can be illustrated and visualized as in Figure 1.

The percentages at the bottom of Figure 1 show the distribution of corpus examples for each of the eight micro-constructions in our database, while the labels (3a)–(3h) refer to the following examples:

(3)   a.   [V$_{SG}$ Subj$_{INAN}$]

Do   zvonk-a   **osta-l-o-s'**   **desjat' minut**.   (Gubarev 1951)
[until ring-GEN.SG remain-PST-N-REFL ten.NOM minute.GEN.PL]
'Ten minutes remained until the bell rang.'

b.   [V$_{PL}$ Subj$_{INAN}$]

Za   ee   predel-ami   **osta-l-i-s'**   **st-o**
[beyond her limit-INS.PL remain-PST-PL-REFL hundred-NOM.SG
million-ov   častn-yx   krest'jansk-ix xozjajstv.
million-GEN.PL private-GEN.PL peasant-GEN.PL farm.GEN.PL]

(*Nauka i žizn'* 2009)

'Outside its borders remained a hundred million private farms.'

c.   [V$_{SG}$ Subj$_{ANIM}$]

V   konečn-om   sčet-e   ot   vs-ego   polk-a
[in final-M.LOC.SG account-LOC.SG from all-M.GEN.SG regiment-GEN.SG

osta-l-o-s'   **sem'**   **čelovek.**   (Axmedov 2011)
remain-PST-N-REFL   seven.NOM   person.GEN.SG]
'At the end, from the whole regiment there remained seven persons.'

d.   [V_PL Subj_ANIM]
Tam   **osta-l-i-s'**   **tridcat'**   **babušek**[…].   (*Zavtra* 2003)
[there   remain-PST-PL-REFL   thirty.NOM   grandmother.GEN.PL]
'There remained thirty grandmothers.'

e.   [Subj_INAN V_SG]
U   nego   tol'ko   **šest'**   **patron-ov**   **osta-l-o-s'**!   (Šoloxov 1932)
[by   he.GEN   only   six.NOM   bullet-GEN.PL   remain-PST-N-REFL]
'He had only six bullets left!'

f.   [Subj_INAN V_PL]
Pjat'   **mašin**   **osta-l-i-s'**   na   perv-om
[five.NOM   car.GEN.PL   remain-PST-PL-REFL   on   first-M.LOC.SG
učastk-e.   (*Za rulem* 2004)
lot-LOC.SG]
'Five cars remained in the first lot.'

g.   [Subj_ANIM V_SG]
U   nas   v   rezerv-e   vs-ego   **pjatnadcat'**
[by   we.GEN   in   reserve-LOC.SG   all-N.GEN.SG   fifteen.NOM
čelovek   **osta-l-o-s'**!   (Krasnov 1922)
person.GEN.PL   remain-PST-N-REFL]
'We had only fifteen men in reserve!'

h.   [Subj_ANIM V_PL]
Semnadcat'   **čelovek**   **osta-l-i-s'**   leža-t'   v   pol-e
[seventeen.NOM   person.GEN.PL   remain-PST-PL-REFL   lie-INF   in   field-LOC.SG
do   več er-a,   do   temnot-y.   (Gel'fand 1941–1943)
until   evening-GEN.SG   until   darkness-GEN.SG]
'Seventeen men remained lying in the field until the evening, until dark.'

All of these examples contain past tense forms of the same finite verb, *ostat'sja* 'remain, be left' in either the singular (*ostalos'*) or plural (*ostalis'*). The subject is either animate (and in these examples always human) or inanimate and quantified by an ordinary numeral, although other types of quantifiers, namely paucal numerals, collective numerals, and indefinite quantifiers are also possible.

Note that the designations in Figure 1 of levels as "macro-", "meso-", and "micro-" are meant only to indicate vertical relationships in the network of allostructions. They are not meant to suggest any rigid classification. Further differentiation is of course possible, for example by taking into account the various types of quantifiers.

3.1.6 Verb

The role of the verb in Russian quantified subject constructions was studied in detail by Robblee (1993, for discussion see also Glushan 2013: 183; Timberlake 2004: 357–360). Robblee (1993: 425–426) considered three classes of predicates, which she referred to as "agentive predicates", "intransitive predicates", and "inversion predicates". Among the "agentive" verbs in Robblee's classification, we find what she refers to as "semi-transitive" verbs (e.g., *rabotat'* 'work') and "transitive" (e.g., *udarit'* 'hit'). Under the label "intransitive", Robblee considered stative verbs such as *nravit'sja* 'appeal to' and postural verbs (e.g., *stojat'* 'stand'). Inversion predicates (a term borrowed from Relational Grammar, Perlmutter 1984) include *byt'* 'be', as well as "existentials, locatives, modals, quantificational predicates and perceptual predicates". Examples of inversion predicates are *proizojti* 'occur' and *naxodit'sja* 'be located'.

According to Robblee, plural agreement is preferred by agentive predicates, while inversion predicates favor singular agreement. Intransitive predicates occupy an intermediate position. Simplifying somewhat, Robblee's findings can be summarized as follows, where the arrow means "tend to occur with" (Robblee 1993: 437):

Agentive predicates → plural agreement

Intransitive predicates → singular or plural agreement

Inversion predicates → singular agreement

Robblee's study was based on a relatively small dataset, and she did not provide a systematic analysis of the interaction of her three classes of predicates with other predictors, such as those discussed in Sections 3.1.1 through 3.1.5 of the present study. We are in a position to test Robblee's claims against a larger set of verbs and data, and they motivate the hypothesis that various verbs may have individual preferences.

3.2 Hypothesis

Based on previous research cited above we advance the following hypothesis concerning factors that may affect the choice of singular versus plural verb agreement (predicate number) in Russian quantified subject constructions in the absence of nominative premodifiers:

Relevant factors may include the year of origin; the frequency, type, and specific preferences of the quantifier; an interaction between animacy and word order; and specific preferences of the verb:

- For year of origin, we expect the proportion of plural agreement to increase.
- For frequency of quantifier, we expect a correlation between frequency and the use of plural, such that quantifiers of low frequency will have a higher preference for singular agreement, while quantifiers of high frequency will have a higher preference for plural agreement.
- For type, we expect paucal numerals to have a stronger tendency to use plural agreement than other quantifiers.
- For animacy, we expect animate subjects to prefer plural agreement.
- For word order, we expect VS word order to prefer singular agreement.
- We expect an interaction between animacy and word order, since animate subjects tend to appear in SV word order, while inanimate subjects tend to appear in VS word order.

This hypothesis is tested in Section 5 based on data presented in Section 4. Importantly, the statistical model we present in Section 5 allows us to test not only the relevance of the individual factors, but also their interplay and relative importance. These are facets of the singular versus plural alternation that are understudied, since no one has undertaken a large-scale corpus study of the alternation before.

4 Database and variables

The database is aggregated from a series of searches in the main portion of the Russian National Corpus, which is annotated for part of speech, inflectional morphology, and semantic features, but not syntactically parsed. Search queries targeted strings containing a nominative numeral or indefinite quantifier governing a noun in the genitive. All four types of numerals (paucal, collective, indefinite, and other) were included in the searches, which thus included all quantities between 'one and a half' and '900'. In order to keep the database to a manageable size, complex numerals such as *sorok dva* 'forty-two' that combine more than one numeral were not included. It was necessary to search both for sentences where the verb precedes and for sentences where the verb follows the quantified noun. Separate searches were carried out for animate and inanimate nouns. The results of all searches were exported to a spreadsheet and conflated into a single database, which was then cleaned manually. In particular, it was necessary to weed out numerous examples where the quantifier was actually in the accusative case, but misidentified by the corpus as nominative because nominative and accusative are often syncretic in Russian. In order to facilitate statistical analysis of independent observations, for each combination of a verb, a quantifier, and an animate or inanimate noun we included only one example per author.

After the exclusion of examples with nominative plural premodifiers (see Section 3.1.1), our database contains 36,182 example sentences. This database is annotated for the following variables and their values, arranged in Table 1 according to how they figure into the analysis in Section 5:

Table 1: Annotation of variables in database.

Type of variable	Name	Abbreviation	Description
Result variable	Predicate number	PredNumber	Singular (sg), plural (pl)
Fixed effects numerical variables	Year created	YearCreated.sc	Year of origin scaled (*z*-scored) such that 1800 = −2.97 (minimum), 1949 = 0 (mean), 2017 = 1.37 (maximum)
	Quantifier frequency	LogQuantFrequ.sc	The corpus frequency of each numeral, logarithmically transformed and scaled (*z*-scored), such that 97 = −5.65 (minimum), 159,318 = 0 (mean), 352,781 = 1.07 (maximum)
Fixed effects categorical variables	Quantifier type	QuanType	Levels: paucal, collective, numeral, indefinite
	Animacy	Animacy	Levels: inanimate (inan), animate (anim)
	Word order	WordOrder	Levels: Verb before Subject (VS), Subject before Verb (SV)
Random effects variables	Quantifier	Quantifier	Lemmas of numerals (e.g., *dva* 'two', *dvoe* 'two(some)', *sto* 'one hundred', *mnogo* 'many')
	Verb lemma	VerbLemma	Lemmas for verbs attested over 100 times (57% of total, remainder listed as NotLemmatized)

Predicate Number is called the result variable because it is the outcome that we are trying to predict (the dependent variable) based on the other (independent) variables. The remaining variables are of three types. Fixed effects numerical variables have values that are numbers, in this case the Year Created and Quantifier Frequency. Because corpus frequency follows a skewed (zipfian) distribution, it is logarithmically transformed. In addition, both numerical variables have been *z*-scored, meaning that they are scaled according to their standard deviations around a mean of zero. Fixed effects categorical variables have a limited set of categorical levels: Quantifier Type has four levels (paucal, collective, numeral, indefinite), Animacy has two levels (inan, anim), and Word Order has two levels (VS, SV). The purpose of random effects is to represent the contributions of specific preferences from individuals (here quantifiers and verbs) that belong to an open class.

5 Analysis

A mixed effects logistic regression model for predicting Predicate Number was fitted using the glmer() function in R (R Core Team 2022) with the following formula:[6]

> PredNumber ~ 1 + YearCreated.sc + LogQuantFreq.sc + QuanType + Animacy * WordOrder + (1| Quantifier) + (1|VerbLemma)

This formula can be interpreted as: "Predicate Number is predicted in relation to an overall intercept (1), with main effects of Year Created, Quantifier Frequency, and Quantifier Type, an interaction between Animacy and Word Order, and random effects of Quantifier and Verb Lemma".

A drop1() function carried out to test whether any of the predictor variables could be felicitously removed from the model showed that all variables should be retained. The R2 values for the model are good, indicating that the model accounts for 42–68% of the variance in the data.[7] The C score which evaluates model fit is 0.922 (a value above 0.8 is considered good or excellent, cf. Gries 2021: 335–336). A test for collinearity of variables confirms that this is an appropriate model.[8]

The baseline for evaluating the predictions of the model is the maximum proportion of the response variable Predicate Number. That value is 0.5596, or 55.96% plural. Table 2a compares the observed values for Predicate Number in the rows with the predicted values in the columns of a confusion matrix. For example, the model predicted 13,034 uses of singular where singular was indeed observed, but the model also predicted 2,899 uses of plural where singular was observed. The top left and bottom right cells in the confusion matrix show where the model made correct predictions, whereas the other cells show where it failed to make correct predictions. Table 2a additionally shows the overall classification accuracy of the model.

---

6 Calculations and figures were made in accordance with methods and code in Gries (2021).

7 The R2 values calculated using the r.squaredGLMM() function are:

	R2m	R2c
theoretical	0.4463087	0.6868255
delta	0.4174733	0.6424507

The rows show values for marginal and conditional R-squared GLMM calculated with different methods, such as "delta", "log-normal", "trigamma", or "theoretical" for models of binomial family.

8 Variance inflation factors (VIF) for all variables are under 2.0; cf. Gries (2021: 300–302) suggests that VIF is only a problem when it exceeds 10.

Table 2a: Confusion matrix and classification accuracy for the regression model.

	Confusion matrix	
	Singular (predicted)	**Plural (predicted)**
Singular (observed)	13,034	2,899
Plural (observed)	2,817	17,432
	Classification accuracy 84.2%	

Table 2b: Precision and recall for the regression model.

	Precision	**Recall**
Singular	82.23%	81.81%
Plural	85.74%	86.09%

With correct predictions for 84.2% of the data, the model is well above the baseline, which is correct 55.96% of the time if one always guesses plural. A comparison of the model's level of successful prediction with the baseline shows that the difference is highly significant and the probability that this difference could occur by chance is zero.[9]

Precision and recall for both singular and plural are displayed in Table 2b. Precision measures how often the model was correct for specific values. 82.23% of predictions of singular corresponded to singular observations, and 85.74% of predictions of plural were indeed plural. Recall looks at the total number of observations of a certain type and measures how well the model succeeded in finding them. The model correctly found 81.81% of observed singulars and 86.09% of observed plurals.

The results of the statistical analysis for each variable are presented according to the types of the variables, with fixed effects variables in Section 5.1 and random effects variables in Section 5.2.

5.1 Results for fixed effects variables

Table 3 summarizes the results of the logistic regression model for fixed effects variables.

---

[9] This is the formula to calculate the probability given the baseline of 0.5596429:

    sum(dbinom(x = 30,466:36,182, size = 36,182, prob = baseline))

From this formula, R calculates the probability to be zero:

    [1] 0

Table 3: Results of logistic regression model for fixed effects. Three stars indicated significance below 0.001. A period represents a value close to, but slightly above 0.05.

	Estimate	Std. error	z value	Pr(>|z|)	
(Intercept)	1.20724	0.30691	3.933	8.37e-05	***
YearCreated.sc	0.08970	0.01610	5.571	2.53e-08	***
LogQuantFreq.sc	0.39513	0.11103	3.559	0.000373	***
QuanTypecollective	−0.88051	0.46554	−1.891	0.058576	.
QuanTypenumeral	−2.03981	0.34382	−5.933	2.98e-09	***
QuanTypeindefinite	−4.60623	0.41864	−11.003	<2e-16	***
Animacyanim	0.98308	0.04317	22.771	<2e-16	***
WordOrderSV	1.28468	0.04896	26.239	<2e-16	***
Animacyanim:WordOrderSV	0.59289	0.07251	8.177	2.91e-16	***

Table 3 represents the prediction of plural as opposed to singular with reference to the Intercept. The Estimate is an effect size, and positive values of the Estimate indicate relative preference for plural, while negative values indicate relative preference for singular. The Intercept represents the situation that holds when the values of YearCreated.sc and LogQuantFreq.sc are at their mean (zero), the quantifier is a paucal numeral, the subject is inanimate, and the verb precedes the subject. All comparisons are with this Intercept, which is why "Wordorder [VS]", "Animacy [inan]" and "QuanType [paucal]" do not appear in Table 3. At the Intercept, the Estimate is positive, indicating some preference for plural. All of the variables are highly significant, as indicated by their *p*-values in the fifth column of Table 3. The model predictions for each variable are investigated in turn below.

5.1.1 Year created

Figure 2 is a visualization of the effect of Year Created on the prediction of plural for Predicate Number. Year Created is on the *x*-axis, and recall that this variable has been scaled so that $1800 = -2.97, 1949 = 0, 2017 = 1.37$. The "rug" at the bottom indicates the presence of observations. The *y*-axis is the prediction of the value for Predicate Number, which is singular for values below 0.5 and plural for higher values. The blue line shows the model predictions for Predicate Number, and the shaded region shows the 95% confidence interval for those predictions.

The predictions of YearCreated.sc yield a nearly flat line, and many entirely flat lines fit within the range of the confidence interval, almost including even the baseline of 0.5596, which barely falls below the confidence interval's lowest level (0.5682) at the extreme right. Note also that the Estimate of the effect of YearCreated.sc in our model is quite small: 0.08970 (see Table 3). It is remarkable how little change has taken place in a period of over two centuries.
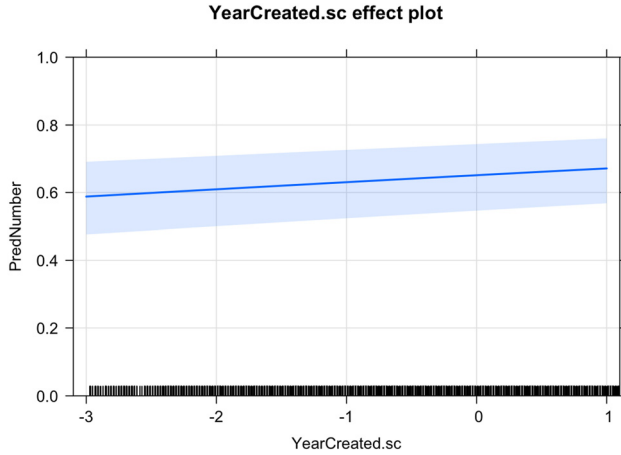
YearCreated.sc effect plot

Figure 2: Effect of Year Created on prediction of plural for Predicate Number.

5.1.2 Quantifier frequency

Figure 3 is similar to Figure 2, this time with Quantifier Frequency on the *x*-axis. Recall that Quantifier Frequency has been logarithmically transformed and scaled so that the minimum value of 97 = −5.65, the mean of 159,318 = 0, and the maximum value of 352,781 = 1.07. Here we see a clear trend: low frequency numerals are more likely to be associated with singular Predicate Number, while high frequency numerals are more likely to be associated with plural.

LogQuantFreq.sc effect plot
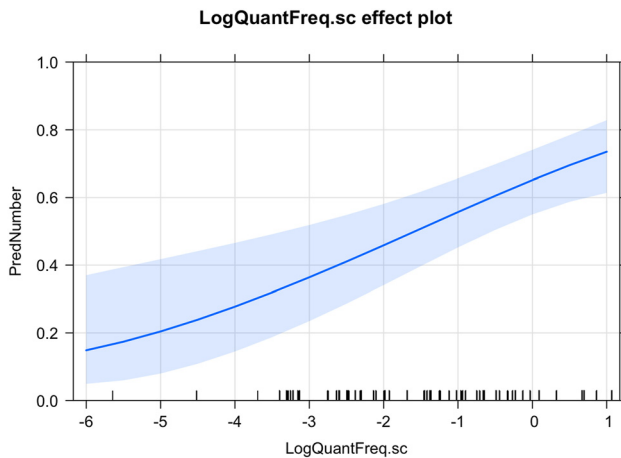
Figure 3: Effect of Quantifier Frequency on prediction of plural for Predicate Number.
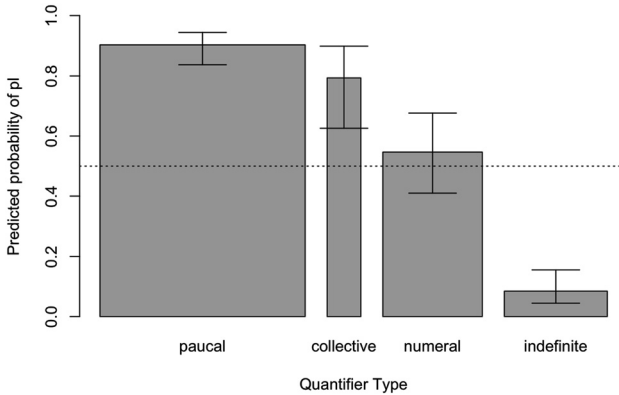
Figure 4: Effect of Quantifier Type on prediction of plural for Predicate Number.

5.1.3 Quantifier type

Figure 4 visualizes the model predictions for the various types of quantifiers. The *y*-axis shows the prediction of plural in the same increments as in Figures 2 and 3. A dotted line is added at 0.5 to show where the prediction goes from a preference for singular (under 0.5) to plural (over 0.5). Bar width indicates the relative amount of data for each type of quantifier and the brackets show the 95% confidence interval.

There are pronounced differences in the effect of Quantifier Type on the choice of Predicate Number. Paucal numerals are most frequent, comprising 46.5% of the data, and have the strongest preference for plural (0.90 prediction). Collective numerals are relatively infrequent (7.6% of data) but also prefer plural (0.79 prediction). Other numerals comprise 22.5% of the data and slightly prefer plural (0.54 prediction of plural), but prediction of singular is also within the confidence interval. Indefinite quantifiers are of similar frequency (23.3%) and have a stronger preference for singular (0.08 prediction of plural).

5.1.4 Animacy and word order

Figure 5 presents the model predictions for the interaction of Animacy and Word Order. The *y*-axis shows the prediction of plural similar to Figures 2–4. The plot has two panels, the left panel with predictions for verb-before-subject (VS) Word Order, and the right panel with predictions for subject before verb (SV). Each panel compares the prediction for inanimate subjects with the prediction for animate subjects. Points show predictions and brackets show the 95% confidence levels for those predictions.
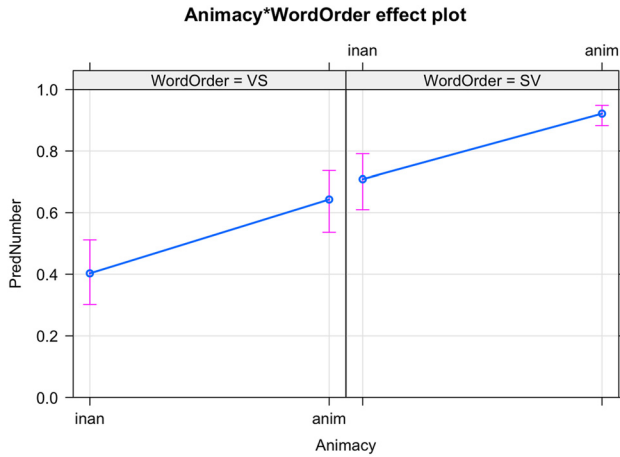
Figure 5: Effect of interaction between Animacy and Word Order on prediction of plural for Predicate Number.

Only in the case of VS Word Order and inanimate subjects is singular predicted (0.40 prediction of plural). For all other combinations, plural is preferred, ranging from 0.64 for SV order with inanimate subjects to 0.92 for SV order with animate subjects.

5.2 Results for random effects variables

Table 4 displays the variance and standard deviation for the random effects variables. As with the fixed effects, positive values indicate relative preference for plural, while negative values indicate relative preference for singular, with the difference that values are here associated with individual lemmas. The variance and standard deviation indicate how spread the distribution is around zero. The quantifiers show both a lower variance (0.5347) and a lower standard deviation (0.7313) than the verb lemmas, which are nearly twice as numerous.

Table 4: Variance and standard deviation for the random effects variables.[a]

Groups	Number of lemmas in each group	Variance	Std. dev.
Quantifier	57	0.5347	0.7313
VerbLemma	98	1.9918	1.4113

[a]Plots of the random effects are available at the following TROLLing post https://doi.org/10.18710/4D2QII (Janda and Nesset 2023).

5.2.1 Quantifier

The quantifiers are overall mostly quite similar to each other in their preferences, with values close to zero. The value closest to zero is for *malovato* 'rather few', with −0.0016, but since there is only one observation of this numeral, this value is not very certain, and the 95% confidence interval is very wide. The next closest value is for *pjat'* 'five', at −0.0079, a more certain value since it is based on 1,351 observations. The 95% confidence interval for the random effects for most of the quantifiers (43 of them) includes zero.

There are strong differences within the classes of paucal and indefinite numerals. Both types of quantifiers break into two groups, one that prefers plural and one that prefers singular (upper and lower parts of Table 5 respectively), with the bulk of the remaining collectives and other numerals distributed between these two groups. The two most extreme values at both ends of the distribution are all paucal numerals, and indefinites come in third place at both ends. Table 5 shows the random effects values for all paucal and indefinite quantifiers (except *malovato* 'rather few').

This distribution suggests that within the paucal and indefinite types, numerals referring to 'both', 'two [feminine]' or a small indefinite number prefer plural, whereas others prefer singular. However, random effects for numerals are interpreted with respect to Quantifier Type. The indefinite *neskol'ko* 'some' does not in itself

Table 5: Random effects values for paucal and indefinite numerals.

Quantifier type	Quantifier	Random effects value	Rank from extreme end
Paucal	*obe* 'both [feminine]'	2.4349	1st
Paucal	*oba* 'both [masculine or neuter]'	2.2346	2nd
Indefinite	*neskol'ko* 'some'	1.2978	3rd
Indefinite	*nemnogo* 'not many'	0.4129	5th
Paucal	*dve* 'two [feminine]'	0.3518	10th
Indefinite	*nemalo* 'not few'	0.3174	13th
Indefinite	*malo* 'few'	0.1404	22nd
– 22 intervening collectives and other numerals –			
Indefinite	*stol'ko* 'so many'	−0.4453	11th
Paucal	*četyre* 'four'	−0.7163	7th
Paucal	*dva* 'two [masculine or neuter]'	−0.7315	6th
Indefinite	*mnogo* 'many'	−0.7537	5th
Paucal	*tri* 'three'	−0.8353	4th
Indefinite	*skol'ko* 'how many'	−0.8590	3rd
Paucal	*poltory* 'one and a half [feminine]'	−0.9854	2nd
Paucal	*poltora* 'one and a half [masculine or neuter]'	−1.8791	1st

strongly prefer plural – it occurs with plural in only 33% of examples in our data. But *neskol'ko* 'some' does stand out as allowing much more plural than other indefinites, which range from *skol'ko* 'how many' with 12% of examples with plural to *mnogo* 'many' with only 5% of examples with plural. Similarly, *tri* 'three' is not especially drawn to singular – it occurs with plural in 70% of examples in our data – but this rate is lower than that of the paucals meaning 'both' and 'two', which range from 99 to 89% plural. By contrast, the paucal *poltora* 'one and a half [masculine or neuter]' really does appear toward the bottom of the range for proportion of plural, with only 15%.

5.2.2 Verb

The verbs that appear in the quantified subject construction have varying preferences for singular versus plural. The distribution of random effects values for individual verbs is a gradual cline from *nesti* 'bring' (2.4128) with strongest preference for plural, through *pribyt'* 'arrive, increase' (−0.0373) with near-zero preference, to *ispolnit'sja* 'complete' (−4.4962) with strongest preference for singular. The random effects values lend some support to Robblee's (1993) suggestion that more agentive verbs tend to prefer plural, while inversion verbs prefer singular. For example, the verbs with the highest positive values following *nesti* 'bring' are indeed agentive according to Robblee's definition: *razgovarivat'* 'converse with' (2.1958), *vesti* 'lead, conduct' (2.1659), and *smotret'* 'watch' (2.1632). And the verbs with the most extreme negative values after *ispolnit'sja* 'complete' are *minut'* 'pass by [said of time]' (−3.6278), *prixodit'sja* 'be necessary' (−3.1647), and *potrebovat'sja* 'be needed' (−2.8782). However, the verbs do not cluster in three groups as predicted by Robblee, and some of the verbs that Robblee would classify as inversion predicates have very high positive random effects values, in particular *sostavljat'* 'comprise, be' (ranks fifth of all verbs, 1.9845) and *javljat'sja* 'be' (ranks ninth of all verbs, 1.5902). The situation is clearly more complex than Robblee's analysis suggests.

Among the verbs with strong preference for singular verb forms, we find several that appear in what we may call a "measurement construction", illustrated in examples (4), (5), and (6):

(4) Kol-e   Kuznecov-u   **ispoln-it-sja**   **šest'**   **let**
     [Kolja-DAT.SG   Kuznecov-DAT.SG   fill-FUT.3SG-REFL   six.NOM   year.GEN.SG
     v   ma-e.   (*Domovoj* 2002)
     in   May-LOC.SG]
     'Kolja Kuznecov turns six years old in May.'

(5)   Kal'biev   ne   zna-l   sam-ogo   interesn-ogo –
        [Kal'biev.NOM.SG   not   know-PST.M   most-N.GEN.SG   interesting-N.GEN.SG
        projd-et   **dv-e**   **nedel-i**,   i   na osvobodi-vš-ee-sja
        pass-FUT.3SG two-F.NOM   week-GEN.SG and on free-PST.PTCP-N.ACC.SG-REFL
        mest-o   naznač-at   ego,   Kal'biev-a.   (Ibragimov 1969)
        position-ACC.SG appoint-FUT.3PL he.ACC Kal'biev-ACC.SG]
        'Kal'biev didn't know the most interesting part: two weeks would go by, and
        he, Kal'biev, would be appointed to the vacant position.'

(6)   Na   odin   tol'ko   remont […]   **uš-l-o**   **dv-a**
        [on   one.M.ACC.SG   only   renovations.ACC.SG   spend-PST-N   two-M/N.NOM
        million-a   dollar-ov.   (*Èkspert* 2015)
        million-GEN.SG   dollar-GEN.PL
        'Two million dollars were spent just on renovations.'

In these examples the quantified subject typically refers to duration (e.g., *nedelja* 'week' in (5)) or other resources (e.g., money in (6)). The fact that the "measurement construction" is attested in our database with singular agreement is not unexpected. The Russian Academy grammar (Švedova 1980: 243) states that singular agreement is used in examples of this type (see also Robblee 1993: 431–433; Rozental' and Telenkova 1976: 245–249).[10]

6 Summary of findings

The findings support the hypothesis presented in Section 3.2. While plural number is slightly preferred overall, singular number becomes more likely when a quantifier is of low frequency, the quantifier is indefinite, or the subject is inanimate with VS word order. The diachronic effect is very small, indicating relative stasis. Both quantifiers and verbs have individual preferences for singular versus plural agreement.

---

10 We analyze the quantified NP as the grammatical subject in the measurement construction. A conceivable alternative would be to analyze such NPs as measure adverbials in the accusative case. However, such an analysis would run into problems in sentences with the quantifier *odin* 'one':

(i)   Odnako,   **proš-l-a**   **odn-a**   **nedel-ja** …   (Kurganskij 1930)
      [however   pass-PST-F   one-F.NOM.SG   week-NOM.SG]
      'However, one week went by …'

Here, the quantified NP is clearly in the nominative case (the accusative would be *odnu nedelju*), and since the quantified NP shows gender agreement with the verb, it behaves like the grammatical subject of the sentence. It seems reasonable to adopt this analysis for sentences with other quantifiers as well.

In historical linguistics, questions about language change have been at center stage: Why and how do languages change? This focus on change makes sense. As Friðriksson (2008: 35) points out, "[c]hange is by nature more dynamic than stability and it is easy to see how observing features in motion is generally more interesting than observing static features." However, in addition to a theory of language change, we also need a theory of language stability. Our study of Russian numeral syntax is a good example of relative stasis. In addition to language external factors such as social networks, nationalism, purism, language attitudes and language planning that may contribute to language stability (Friðriksson 2008), we suggest that language-internal factors are also relevant. In particular, we propose that relations among constructions may create alternations that preserve variation at a nearly static level. While our data indicate very slight growth in the use of plural (see Figure 2 in Section 5.1.1), what is most remarkable is how little the balance of 44% singular versus 56% plural has changed across more than two centuries: an entirely flat line is within the 95% confidence interval.

Given that the morphosyntax of Russian quantifiers has undergone radical change since medieval times and is still in the process of changing (Nesset 2020; Pereltsvaig 2010), the stability of the singular versus plural alternation demands an explanation. Blythe and Croft (2012) hypothesize that the difference between a stable variation between linguistic alternants and a diachronic change is motivated by differential weighting of the alternants based on social factors (see also Blythe and Croft 2021; Labov 2001; Milroy and Milroy 1992; Nettle 1999). In other words, language change takes place when a linguistic unit becomes connected with a social value, such as prestige. While we do not dispute the relevance of social weighting for language change, we propose that linguistic factors may also be relevant. We suggest that the opposing preferences of allostructions such as [V_{SG} Subj$_{INAN}$] and [Subj$_{ANIM}$ V_{PL}] may counterbalance each other and thus help to maintain an alternation in a stable state over time, even in a linguistic system with variation that we would expect to be conducive to language change.

Our investigation of Quantifier Type and individual preferences for Quantifiers reveals additional complexity within types. The paucal numerals are a rather heterogeneous group: numerals associated with definiteness ('both') and feminine gender (*dve* 'two [feminine]') have a strong preference for plural; *dva* 'two [masculine or neuter]', *tri* 'three', *četyre* 'four' have a weaker preference for plural, and *poltora/poltory* 'one and a half' has a strong preference for singular. The gender effect comports well with previous findings by Nesset (2020), who argues that numerals that express feminine show special morphosyntactic behavior in related constructions. The quantifier *neskol'ko* 'some' deviates from the strong preference for singular observed for the indefinite type, and behaves more like a numeral (from 'five' and up).

7 Conclusion

Our corpus study of predicate agreement in Russian constructions with quantified subjects contributes to Construction Grammar, historical linguistics, and Russian linguistics. An important question in Construction Grammar in recent years concerns the role of relations among constructions in a multidimensional network. Our study lends empirical evidence to the importance of such relations. Much of the discussion of such relations has revolved around phenomena in English, so empirical evidence from a morphosyntactically more complex language represents an additional perspective.

We argue that a theory of language change should be supplemented with a theory of language stability. Our empirical study targets a stable phenomenon embedded in a part of Russian morphosyntax that has been undergoing language change since medieval times. Our findings suggest that not only extralinguistic factors such as social networks, nationalism, purism, language attitudes and language planning may contribute to language stability, but that under certain conditions language-internal factors are also relevant. Specifically, we propose that multidimensional relations among allostructions can resist language change.

The syntax of quantifiers is a complex phenomenon that has received considerable attention in Russian linguistics, but no large-scale empirical investigation of corpus data has been carried out before. The present study fills this gap by offering a statistical analysis of a database of nearly 39,000 examples. Our contribution to the study of Russian quantifiers can be summarized as follows. Nominative premodifiers are found to predict plural agreement as a (virtually) categorical rule. However nominative premodifiers are themselves infrequent, present for only about 7% of quantified subjects. For the remaining 93% of data, there is variation between plural (56%) and singular (44%) agreement, and this distribution has remained quite stable over the past two hundred years. Aside from individual preferences of specific quantifiers and verbs, the factors that contribute to the prediction of plural versus singular agreement are the frequency and type of the quantifier and an interaction between animacy and word order. While there is a negative relationship between the quantity of a numeral and its frequency (numerals with high values tend to have lower frequency) there is a positive correlation between the frequency of the numeral and the probability that plural will be chosen. This means that lower numerals, which tend to be more frequent, are associated with plural, while higher numerals which tend to be less frequent, are associated with singular. Paucal and collective numerals are clearly associated with plural number, indefinite numerals are clearly associated with singular number, and other numerals show a balance (54% preference for plural) close to the overall baseline (56% plural). There is an

interaction between animacy and word order, such that only inanimate subjects with VS word order prefer singular, whereas all other combinations prefer plural, with the strongest preference for animate subjects with SV word order. The traditional classification of some Russian quantifiers as paucal and indefinite might be in need of revision, since these groups are not coherent in their behavior.

Our mixed-effects logistic regression model provides a more nuanced portrayal of the interplay of numerous factors in a multidimensional network of allostructions, and a more comprehensive representation of the Russian quantified subject construction.

Data availability statement

All data (and related metadata) and code supporting the results presented in this article are publicly available in TROLLing at https://doi.org/10.18710/4D2QII (Janda and Nesset 2023).

References

Andersen, Henning. 2006. Some thoughts on the history of Russian numeral syntax. *Harvard Ukrainian Studies* 28(1–4). 57–67.

Babby, Leonard H. 1987. Case, pre-quantifiers, and discontinuous agreement in Russian. *Natural Language & Linguistic Theory* 5. 91–138.

Barlow, Michael & Suzanne Kemmer (eds.). 2000. *Usage-based models of language*. Stanford, CA: CSLI Publications.

Blythe, Richard A. & William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88(2). 269–304.

Blythe, Richard A. & William Croft. 2021. How individuals change language. *PLoS One* 16(6). e0252582.

Boas, Hans Christian. 2006. A frame-semantic approach to identifying syntactically relevant elements of meaning. In Petra C. Steiner, Hans Christian Boas & Stefan J. Schierholz (eds.), *Contrastive studies and valency. Studies in honor of Hans Ulrich Boas*, 119–149. Frankfurt/New York: Peter Lang.

Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 75–96. Berlin: Mouton de Gruyter.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Cappelle, Bert. 2006. Particle placement and the case for "allostructions". *Constructions* SV1-7/2006. Available at: www.constructions-online.de.

Chambers, Jack K. 2002. Patterns of variation including change. In Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *Handbook of language variation and change*, 349–372. Oxford: Blackwell.

Corbett, Greville G. 1981. Agreement with quantified subjects in Russian: A fictitious linguistic change? *Russian Linguistics* 5. 287–290.

Corbett, Greville G. 1983. *Hierarchies, targets and controllers: Agreement patterns in Slavic*, 11–35. London: Croom Helm.

Corbett, Greville G. 1993. The head of Russian numeral expressions. In Greville G. Corbett, Norman M. Fraser & Scott McGlashan (eds.), *Heads in grammatical theory*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Crockett, Dina B. 1976. *Agreement in contemporary standard Russian*. Cambridge, Mass.: Slavic Publishers.

Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.

Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.

Dąbrowska, Ewa. 2015. Individual differences in grammatical knowledge. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of cognitive linguistics*, 650–668. Berlin: De Gruyter Mouton.

Dąbrowska, Ewa & Dagmar Divjak (eds.). 2015. *Handbook of cognitive linguistics*. Berlin: De Gruyter Mouton.

Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.

Dittmar, Miriam, Kirsten Abbot-Smith, Elena Lieven & Michael Tomasello. 2008. German children's comprehension of word order and case marking in causative sentences. *Child Development* 79(4). 1152–1167.

Divjak, Dagmar. 2019. *Frequency in language: Memory, attention and learning*. Cambridge: Cambridge University Press.

Divjak, Dagmar, Eva Dąbrowska & Antti Arppe. 2016. Machine meets man: Evaluating the psychological reality of corpus-based probabilistic models. *Cognitive Linguistics* 27(1). 1–33.

Fillmore, Charles J. 1988. The mechanisms of "Construction Grammar". In *Proceedings of the fourteenth annual meeting of the Berkeley Linguistics Society*, 35–55. Berkeley: Berkeley Linguistics Society.

Franks, Steven. 1995. *Parameters of Slavic morphosyntax*. New York, NY: Oxford University Press.

Friðriksson, Finnur. 2008. *Language change versus stability in conservative language communities: A case study of Icelandic*. Gothenburg: Gothenburg University PhD dissertation.

Fried, Mirjam & Jan-Ola Östman. 2004. Construction Grammar: A thumbnail sketch. In Mirjam Fried & Jan-Ola Östman (eds.), *Construction Grammar in a cross-linguistic perspective*, 11–86. Amsterdam & Philadelphia: John Benjamins.

Geeraert, Dirk & Hubert Cuyckens (eds.). 2007. *The Oxford handbook of cognitive linguistics*. Oxford: Oxford University Press.

Glushan, Zhanna A. 2013. *The role of animacy in Russian morphosyntax*. Storrs: University of Connecticut PhD dissertation.

Goldberg, Adele E. 1995. *Constructions*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13(4). 327–356.

Goldberg, Adele E. 2006. *Constructions at work*. Oxford: Oxford University Press.

Gorbačevič, Kirill Sergeevič. 1971. *Izmenenie norm russkogo literaturnogo jazyka*. Leningrad: Prosveščenie.

Gorbačevič, Kirill Sergeevič. 1989. *Normy sovremennogo russkogo literaturnogo jazyka*, 3rd edn., corrected. Moscow: Prosveščenie.

Grafmiller, Jason, Benedikt Szmrecsanyi, Melanie Röthlisberger & Benedikt Heller. 2018. General introduction: A comparative perspective on probabilistic variation in grammar. *Glossa: A Journal of General Linguistics* 3(1). 1–10.

Graudina, Ljudmila K., Viktor A. Ickovič & Lija P. Katlinskaja. 2001. *Grammatičeskaja pravil'nost' russkoj reči: stilističeskij slovar' variantov*. Moscow: Nauka.

Gries, Stefan Th. 2002. Evidence in linguistics: Three approaches to genitives in English. *LACUS Forum* 28. 17–31.

Gries, Stefan Th. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1. 1–27.

Gries, Stefan Th. 2021. *Statistics for linguistics with R*, 3rd rev. & ext. edn. Berlin: De Gruyter.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.

Igartua, Ivan & Nerea Madariaga. 2018. The interplay of semantic and formal factors in Russian morphosyntax: Animate paucal constructions in direct object function. *Russian Linguistics* 42. 27–55.

Isačenko, Aleksandr V. 1982. *Die russische Sprache der Gegenwart: Formenlehre*. Munich: Max Hueber Verlag.

Iwata, Seizi. 2005. Locative alternation and two levels of verb meaning. *Cognitive Linguistics* 16(2). 355–407.

Janda, Laura A. 1996. *Back from the brink*. Munich and Newcastle: Lincom Europa.

Janda, Laura A. & Tore Nesset. 2023. Replication data for: A network of allostructions: Quantified subject constructions in Russian. Available at: https://doi.org/10.18710/4D2QII.

Klavan, Jane & Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50(2). 355–384.

Kuz'minova, Elena A. 2004. Soglasovanie podležaščego i skazuemogo pri vyraženii podležaščego čislitel'nymi i količestvennymi sočetanijami. In Alla Vasil'evna Veličko (ed.), *Kniga o grammatike I*, 28–34. Moscow: Moskovskij gosudarstvennyj universitet.

Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors*. Oxford: Blackwell.

Lakoff, George. 1987. *Women, fire, and dangerous things*. Chicago & London: The University of Chicago Press.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*, vol. 1. Stanford: Stanford University Press.

Langacker, Ronald W. 1991a. *Foundations of Cognitive Grammar*, vol. 2. Stanford: Stanford University Press.

Langacker, Ronald W. 1991b. *Concept, image, and symbol: The cognitive basis of grammar*. Berlin & New York: Mouton de Gruyter.

Langacker, Ronald W. 1999. *Grammar and conceptualization*. Berlin: Mouton de Gruyter.

Langacker, Ronald W. 2008. *Cognitive Grammar: A basic introduction*. Oxford: Oxford University Press.

Lobanova, Anna. 2011. The role of prominence scales for the disambiguation of grammatical functions in Russian. *Russian Linguistics* 35. 125–142.

Lyngfeldt, Benjamin, Lars Borin, Kyoko Ohara & Tiago Timponi Torrent (eds.). 2018. *Constructicography: Constructicon development across languages*. Amsterdam: John Benjamins.

MacWhinney, Brian, Elisabeth Bates & Reinhold Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23(2). 127–150.

Madariaga, Nerea & Iván Igartua. 2017. Idiosyncratic (dis)agreement patterns: The structure and diachrony of Russian paucal subjects. *Scando-Slavica* 63(2). 99–132.

Milroy, Lesley & James Milroy. 1992. Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21. 1–26.

Nesset, Tore. 2020. A long birth: The development of gender-specific paucal constructions in Russian. *Diachronica* 37(4). 514–539.

Nettle, Daniel. 1999. Using social impact theory to simulate language change. *Lingua* 108. 95–117.

Perek, Florent & Adele E. Goldberg. 2017. Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition* 168. 276–293.

Pereltsvaig, Asya. 2006. Small nominals. *Natural Language & Linguistic Theory* 24. 433–500.

Pereltsvaig, Asya. 2010. As easy as two, three, four? In Wayles Browne, Adam Cooper, Alison Fisher, Esra Kesici, Nikola Predolac & Draga Zec (eds.), *Annual workshop on formal approaches to Slavic linguistics: The second cornell meeting 2009*, 418–435. Ann Arbor: Michigan Slavic Publications.

Perlmutter, David M. 1984. Working 1s and inversion in Italian, Japanese, and Quechua. In David Perlmutter & Carol Rosen (eds.), *Studies in relational grammar*, vol. 2, 292–330. Chicago: The University of Chicago Press.

Pesetsky, David. 2013. *Russian case morphology and the syntactic categories* (Linguistic Inquiry Monographs 66). Cambridge: MIT Press.

Pijpops, Dirk, Dirk Speelman, Freek van de Velde & Stefan Grondelaers. 2021. Incorporating the multi-level nature of the constructicon into hypothesis testing. *Cognitive Linguistics* 32(3). 487–528.

Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, Mass.: MIT Press.

R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Radford, Andrew. 1988. *Transformational grammar*. Cambridge: Cambridge University Press.

Rappaport-Hovav, Malka & Beth Levin. 2008. The English dative alternation: The case for verb sensitivity. *Journal of Linguistics* 44. 129–167.

Robblee, Karen. 1993. Individuation and Russian agreement. *The Slavic and Eastern European Journal* 37(4). 423–441.

Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of grammatical variation in English*, 379–411. Berlin: Mouton de Gruyter.

Rozental', Ditmar È. 1974. *Praktičeskaja stilistika russkogo jazyka*. Moscow: Vysšaja Škola.

Rozental', Ditmar È. & Margarita A. Telenkova. 1976. *Praktičeskaja stilistika russkogo jazyka*. Moscow: Russkij jazyk.

Sasse, Hans-Jürgen. 1993. Syntactic phenomena in the world's languages I: Categories and relations. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax. Ein internationales Handbuch zeitgenössischer Forschung*, 646–686. Berlin: Mouton de Gruyter.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization and entrenchment*. Oxford: Oxford University Press.

Sičinava, Dmitrij V. 2012. Čislitel'noe. In *Russkaja korpusnaja grammatika*, 193–257. Available at: http://rusgram.ru/.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.

Stefanowitsch, Anatol. 2006. Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2(1). 61–77.

Švedova, Natalija Ju (ed.). 1980. *Russkaja grammatika*, vol. 2. Moscow: Nauka.

Taylor, John R. 2012. *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.

Timberlake, Alan. 1985. Hierarchies in the genitive of negation. In Richard D. Brecht & James S. Levine (eds.), *Case in Slavic*, 338–360. Columbus, Ohio: Slavica Publishers.

Timberlake, Alan. 2004. *A reference grammar of Russian*. Cambridge: Cambridge University Press.

Townsend, Charles E. & Laura A. Janda. 1996. *Common and comparative Slavic: Phonology and inflection with special emphasis on Russian, Polish, Czech, Serbo-Croatian, and Bulgarian*. Bloomington: Slavica Publishers.

Traugott, Elizabeth Closs. 2008a. The grammaticalization of NP of NP patterns. In Alexander Bergs & Gabriele Diewald (eds.), *Constructions and language change*, 23–45. Berlin & New York: Mouton de Gruyter.

Traugott, Elizabeth Closs. 2008b. Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English. In Regine Eckhardt, Gerhard Jager & Tonjes Veenstra (eds.), *Variation, selection, development: Probing the evolutionary model of language change*, 219–250. Berlin & New York: Mouton de Gruyter.

Traugott, Elizabeth Closs. 2018. Modeling language change with constructional networks. In Salvador Pons Bordería & Óscar Loureda Lamas (eds.), *Beyond grammaticalization and discourse markers: New issues in the study of language change*, 17–50. Boston: BRILL.

Ungerer, Tobias. 2021. Using structural priming to test links between constructions: English caused-motion and resultative sentences inhibit each other. *Cognitive Linguistics* 32(3). 389–420.

Van de Velde, Freek. 2014. Degeneracy: The maintenance of constructional networks. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the scope of Construction Grammar*, 141–179. Berlin: De Gruyter.

Wade, Terence. 2011. *A comprehensive Russian grammar*. Oxford: Wiley-Blackwell.