

UDC 81'33

DOI: 10.18413/2313-8912-2023-9-1-1-2

Kseniia A. Studenikina 

**Parametrizing number variation in Russian noun phrases  
with experimental studies and language modeling**

Lomonosov Moscow State University  
GSP-1, Leninskie Gory, Moscow, 119991, Russia  
*E-mail:* [xeanst@gmail.com](mailto:xeanst@gmail.com)

*Received 24 January 2023; accepted 16 March 2023; published 30 March 2023*

**Acknowledgments.** This work was supported by Non-commercial Foundation for Support of Science and Education «INTELLECT».

**Abstract.** The study deals with the problem of agreement with coordinated constructions and number form alternation. The target, agreeing with two conjoined singular nouns, copies either plural or singular number feature. This paper focuses on the syntax of Russian noun phrases with coordinated modifiers which demonstrate the number variation of the agreement controller. If two conjoined singular adjectives have split interpretations, both singular and plural nouns are acceptable. In contrast to previous studies relying on introspection and corpus, we parametrize the number variation based on the results of the self-paced acceptability experiments (Likert scale 1-7). We compare the data about human perception with the language probabilities predicted by a neural model for text generation ruGPT-3. Two case studies were conducted to analyze morphological and syntactic factors parametrizing variation. The first study considers the impact of noun morphology on number alternation. The second study examines the effect of the premodifier attributive agreement. Both offline acceptability scores and online reading time demonstrate that the observed morphological and syntactic factors should be considered while parametrizing number variation in Russian noun phrases with coordinated modifiers. The ruGPT-3 language model, trained on a vast collection of Russian texts, manages to predict the correct probability for highly acceptable and highly unacceptable sentences, but it fails to assign accurate probability values to the cases of variation.

**Keywords:** Coordination; Agreement variation; Russian; Experimental syntax; Language modeling; Acceptability; Perplexity

**How to cite:** Studenikina, K. A. (2023). Parametrizing number variation in Russian noun phrases with experimental studies and language modeling, *Research Result. Theoretical and Applied Linguistics*, 9 (1), 192-205. DOI: 10.18413/2313-8912-2023-9-1-1-2

УДК 81`33

DOI: 10.18413/2313-8912-2023-9-1-1-2

Студеникина К. А. 

**Параметризация числовой вариативности для русских  
именных групп с помощью экспериментальных  
исследований и языкового моделирования**

Московский государственный университет имени М. В. Ломоносова  
Ленинские горы, 1, Москва, 119991, Россия  
E-mail: [xeanst@gmail.com](mailto:xeanst@gmail.com)

*Статья поступила 24 января 2023 г.; принята 16 марта 2023 г.;  
опубликована 30 марта 2023 г.*

**Информация об источниках финансирования или грантах, благодарности:**  
Исследование выполнено при финансовой поддержке Некоммерческого Фонда развития науки и образования «Интеллект».

**Аннотация.** В данном исследовании рассматривается проблема вариативности числа при согласовании с сочиненными конструкциями. Мишень, согласующаяся с двумя сочиненными существительными единственного числа, может копировать признак единственного или множественного числа. Статья посвящена изучению синтаксиса русских именных групп с сочиненными модификаторами, которые демонстрируют вариативность числа контроллера согласования. Если два сочиненных прилагательных единственного числа имеют расщепленную интерпретацию, для вершины-существительного допустимо и единственное, и множественное число. В отличие от предшествующих работ, которые использовали интроспекцию и корпусный метод, мы осуществляем параметризацию вариативности числа с опорой на результаты экспериментальных исследований. В качестве методики для экспериментов были выбраны оценка приемлемости по шкале Ликерта от 1 до 7 и чтение с саморегуляцией скорости. Также мы сравниваем данные человеческого восприятия с вероятностными метриками, предсказанными нейронной языковой моделью для генерации текста ruGPT-3. Для изучения морфологических и синтаксических факторов, которые влияют на выбор стратегии согласования, было проведено два исследования. Первое исследование рассматривает роль числовой морфологии существительного, второе исследование анализирует влияние атрибутивного согласования премодификатора. Посредованные данные о приемлемости и непосредственные данные чтения показывают, что изучаемые морфологические и синтаксические факторы оказывают значимое влияние на выбор числа существительного в русских именных группах с сочиненными модификаторами. Рассматриваемая языковая модель ruGPT-3, обученная на большом количестве неразмеченных русских текстов, предсказывает корректные вероятностные метрики для строго приемлемых и строго неприемлемых предложений, однако не способна определить верную стратегию согласования для контекстов вариативности.

**Ключевые слова:** Сочинение; Вариативность согласования; Русский язык; Экспериментальный синтаксис; Языковое моделирование; Приемлемость; Перплексия

**Информация для цитирования:** Студеникина К. А. Параметризация числовой вариативности для русских именных групп с помощью экспериментальных исследований и языкового моделирования // Научный результат. Вопросы теоретической и прикладной лингвистики. 2023. Т. 9. № 1. С. 192-205. DOI: 10.18413/2313-8912-2023-9-1-1-2

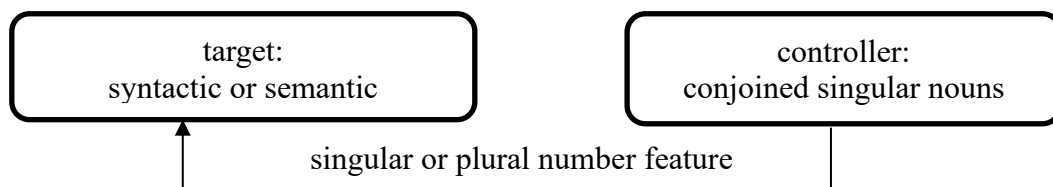
### Introduction

Agreement with coordinated structures frequently permits number form alternation. The syntax of coordinated structures has been widely examined in different languages (Wilder, 1997; Hartmann, 2000; Barros and Vicente, 2011; Grosz, 2015; Shen, 2018,

among others). The target (e.g., verb or adjective), agreeing with two conjoined singular nouns, copies either singular or plural number feature corresponding to a syntactic and semantic agreement, respectively (Figure 1).

**Figure 1.** Number agreement variation

**Рисунок 1.** Вариативность числового согласования



The number choice in coordinated constructions is parametrized by morphological, syntactic, and semantic factors (Corbett, 1979). English conjoined expressions provide confirmation for this thesis. With conjoined inanimate nouns, the

attributive must be singular (1) while the predicate can be of either number (2). Agreement with conjoined animate nouns shows a shift toward syntactic agreement. The singular is still required in the attributive position, while the plural is required in the predicate (3).

- (1) *this/ \*these frost and freezing fog*
- (2) *Frost and freezing fog has/ have affected most of the country today.*
- (3) *This/ \*these man and woman were/ \*was living in the moat.* (Corbett, 1979: 6)

The agreement controller may also demonstrate number variation which is observed in Russian noun phrases with

coordinated modifiers. If two conjoined singular adjectives have split interpretations, both singular and plural nouns are acceptable (4).

- (4) *v imenitel'n-om i vinitel'n-om padezh-e/ padezh-akh*  
in nominative-LOC.SG and accusative-LOC.SG case-LOC.SG case-LOC.PL  
'in nominative and accusative cases' (Kodzasov, 1987: 210)

Previous studies revealed factors parametrizing variation relying on introspection (Kodzasov, 1987) and corpus data (Pekelis, 2013)<sup>1</sup>. However, these

methods have several drawbacks (Talmy, 2018). Introspection enables receiving data only about the author's idiolect, making this method subjective and inaccurate. Corpus studies provide only data on the frequency of

<sup>1</sup> Pekelis, O. E. (2013). *Sochinenie* [Coordination], [Materials for the project of Russian corpus grammar](#),

Manuscript copyright.

acceptable sentences but do not present information about the perception of unacceptable examples.

This paper uses an experimental method and language modeling to parametrize the number variation. There are two goals of the study. Firstly, we aim to parametrize number variation in Russian noun phrases with coordinated modifiers based on the results of syntactic experiments. Experimental method allows us to reveal particular factors which influence the number choice in coordinated constructions. Secondly, we aim to compare the human perception of alternative number forms with the neural models' probabilities. The modern neural models extract language knowledge from large amounts of unlabeled data and can be evaluated by comparison with human acceptability judgments. The methods and materials of the study will be considered in the next section.

### Materials and methods

In order to parametrize the number variation, we conducted syntactic experiments which combined two tasks: the acceptability judgments on a Likert scale from 1 to 7 (Likert, 1932) and self-paced reading (Aaronson and Scarborough, 1976). In contrast to introspection and corpus studies, experimental methods enable us to gather controlled data from many speakers and identify subtle differences in acceptability. The usage of these two methods enables the retrieval of not only offline data on the acceptability of a particular structure but also online data on possible delays and difficulties in the perception. We have chosen the 7-point Likert scale for formal judgment collection since this method demonstrates the greatest statistical power compared to magnitude estimation and two alternative forced choices based on a random sample (Sprouse et al., 2013).

The acceptability scores collected from the experiment must be normalized in order to solve the problem of scale bias. The respondents might use the scale in different ways: avoid the highest (7) and the lowest (1)

scores, and use only the top or bottom of the scale. The calculation of normalized scores or z-scores is represented in (5), where  $Z_{ij}$  is the normalized score  $j$  of respondent  $i$ ,  $X_{ij}$  is the score  $j$  of respondent  $i$  on the 1-7 scale,  $\bar{X}_i$  is the mean score of the respondent  $i$  and  $\sigma_i$  is the standard deviation of the respondent  $i$  (Schütze and Sprouse, 2014). Normalization was not applied to reading time, but values less than 100 ms and more than 10000 ms were excluded (Gold, 2021).

$$(5) Z_{ij} = \frac{(X_{ij} - \bar{X}_i)}{\sigma_i}$$

Besides experimental syntax, acceptability judgments have been broadly applied in natural language processing (Sprouse et al., 2018; Warstadt et al., 2019; Brunato et al., 2020 among others). They are used to test language models' robustness and probe their acquisition of grammatical phenomena. Modern neural language models can generate text almost indistinguishable from human texts (Dou et al., 2022). They are trained on large language datasets without explicit grammar information but somehow generate grammatical texts. In this study, we calculated sentence probabilities by ruGPT-3<sup>2</sup>, a neural model for text generation in Russian.

We used perplexity to measure how well a model predicts a sentence (Lau et al., 2020). Perplexity is an evaluation metric for language models and correlates well with human judgments (Lau et al., 2017). Perplexity can be interpreted as the inverse probability of a sentence, normalized by the number of words (6), where  $PP$  is perplexity measure,  $W$  is a sentence,  $w_1, w_2, \dots, w_n$  are the words of a sentence and  $n$  is the number of words in that sentence. Since we are taking the inverse probability, a lower perplexity indicates a higher acceptability. The sentences may have different lengths, so the metric must be independent of its size. It could be

<sup>2</sup> [https://huggingface.co/sberbank-ai/rugpt3large\\_based\\_on\\_gpt2](https://huggingface.co/sberbank-ai/rugpt3large_based_on_gpt2)

obtained by normalizing the probability of the sentence by the total number of words.

$$(6) PP(W) = \frac{1}{\sqrt{P(w_1, w_2, \dots, w_n)}}$$

Two platforms were used to conduct the experiments: IbexFarm (Drummond, 2013)<sup>3</sup> and Pen Controller for IBEX (Zehr and Schwarz, 2018)<sup>4</sup>. The library Transformers (Wolf et al., 2020) on Python programming language was used to launch ruGPT-3 model. The R programming language was used for the statistical analysis of the results (R Development Core Team, 2009)<sup>5</sup>. Linear mixed effects models were implemented to

(7) *slab-yj*      *vostochn-yj*      *veter*  
weak-M.SG      east-M.SG      wind.M.SG  
'A weak east wind brought a thick fog.'

(8) *parizhsk-aja*      *nacional'n-yj*  
Paris.ADJ-F.SG      national-M.SG  
*sezon.*  
season.SG.ACC

'The Paris National Opera has opened a new season.'

Using grammatical and ungrammatical fillers has a certain goal (Belova et al., 2021). The fillers are used to mark the boundaries of the scale and verify the scores of the stimuli since the fillers do not include context variations. We expect low perplexity values and high acceptability scores for grammatical fillers and high perplexity values and low acceptability scores for ungrammatical fillers. If our expectation is met, the perplexity values and acceptability scores for stimuli will be correctly localized on the scale.

The next two sections describe experimental studies and language model evaluation in detail.

identify subtle differences in acceptability (Gries, 2021). To determine the significance level of fixed factors, the *lmerTest* library was used (Kuznetsova et al., 2017)<sup>6</sup>. After selecting a suitable model, a pairwise comparison of conditions was carried out with the *emmeans* package (Lenth et al., 2019)<sup>7</sup>.

Two case studies are described in this paper. The first study reveals the role of morphological factors while the second study deals with the effect of syntactic context. For each study, 64 sentences were created: 32 stimuli and 32 fillers. One half of the fillers were grammatical (7), another half were ungrammatical and included agreement mistakes (8).

*nagnal*      *gust-oj*      *tuman.*  
bring.PST.M      thick-SG.ACC      fog.SG.ACC

*oper-a*      *otkryl-a*      *nov-yj*  
opera-F.SG      open.PST-F      new-SG.ACC

### The effect of noun morphology on the number variation

The first study examines the effect of noun morphology on the number variation. The role of number morphology in Russian noun phrases with coordinated modifiers has not been studied before. The correlation between the morphology and the agreement was described for Bulgarian (Harizanov and Gribanova, 2015). The regular number morphology of the noun enables only semantic strategy (9) while the irregular number morphology enables only syntactic agreement (10).

<sup>3</sup> Drummond, A. (2013). *Ibex Farm*. <http://spellout.net/ibexfarm/>

<sup>4</sup> Zehr, J. and Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. <https://farm.pcibex.net/>

<sup>5</sup> R Development Core Team. (2009). *A language and environment for statistical computing*, R Foundation for Statistical Computing.

<sup>6</sup> Kuznetsova, A., Brockhoff, P. B. and Christensen, R. H. (2017). *lmerTest* package: tests in linear mixed effects models, *Journal of statistical software*, 82, 1–26.

<sup>7</sup> Lenth, R., Singmann, H., Love, J., Buerkner, P. and Herve, M. (2019). *Emmeans: Estimated marginal means, aka least-squares means*, R package version 1.4.5. <https://cran.r-project.org/web/packages/emmeans/index.html>



- (9) *bălgarsk-i-ja* *i* *rusk-i* ***narod-i/ \* narod***  
 bulgarian-SG.M-DEF and russian-SG.M nation-PL/ nation.SG  
 ‘the Bulgarian and Russian nations’  
 (10) *naj-nisk-o-to* *i* *naj-visok-o* ***dete/ \* deca***  
 most-short-SG.N-DEF and most-tall-SG.N child/children  
 ‘the shortest and tallest child’

The current study tested this assumption in Russian and included two factors: the noun number (singular / plural) and the morphological noun type. The nouns of four morphological types were examined:

- (11) *vysok-ij* *i* *nizk-ij* ***chelovek/ ljudi*** *perebezh-al-i*  
 tall-SG and short-SG man people run.PST-PL  
*cherez dorog-u.*  
 across road-SG.ACC  
 ‘A tall and a short man ran across the road.’

- (12) *bur-yj* *i* *bel-yj* ***medvezhonok/ medvezhata adaptirovalis***  
 brown-SG and white-SG bear.SG bear.PL adapt.PST.PL  
*v zoopark-e.*  
 in zoo-SG.LOC  
 ‘The brown and the white bears adapted in the zoo.’

- (13) a. *tvorcheska-ja* *i* *nauchna-ja* ***molodjoz*** *spel-i*  
 creative-SG and scientific-SG youth sing.PST-PL  
*na prazdnik-e.*  
 at festival-SG.LOC  
 ‘Creative and scientific youth sang at the festival.’

- b. *bolsh-ie* *i* *malen’k-ije* ***ochki*** *zavalilis* *pod krovat*.  
 big-PL and small-PL glasses fall.PST.PL under bed.SG.ACC  
 ‘Big and small glasses fell under the bed.’

- (14) *pozhil-oj* *i* *molod-oj* ***prepodavatel’/ prepodavatel-i*** *voshl-i*  
 elderly-SG and young-SG teacher.SG teacher-PL go.PST-PL  
*v auditori-ju.*  
 in classroom-SG.ACC  
 ‘An elderly and a young teachers entered the classroom.’

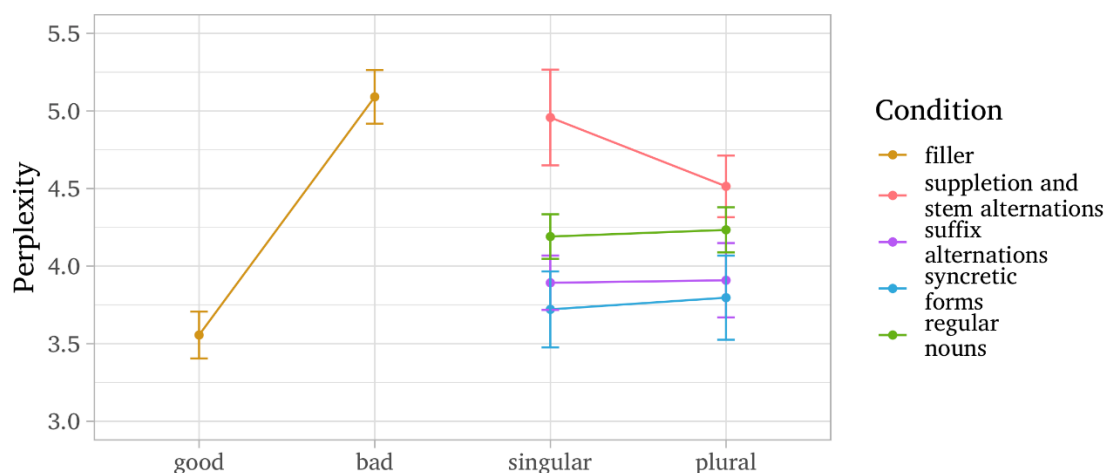
Figure 2 demonstrates the perplexity measures of the ruGPT-3 language model. As mentioned before, the lower the perplexity (y-axis), the more probable or acceptable the sentence is. The grammatical fillers receive low perplexity values, while the perplexity measures for the ungrammatical fillers are high. It corresponds to our expectations and allows us to compare stimuli perplexity value with some benchmark. Thus, the stimuli with the singular suppletive nouns receive almost

suppletion and stem alternations (11), suffix alternations (12), syncretic forms, namely singularia tantum (13a) and pluralia tantum (13b), regular nouns (14).

as high a perplexity value as ungrammatical fillers. The singular number is significantly less probable than the plural number for the sentences with suppletion and stem alternations ( $p < .05$ ). The other noun types do not demonstrate a significant difference in number probability ( $p > .05$ ). According to ruGPT-3, the suppletive nouns are the least probable, the regular nouns are more probable than the nouns with suffix alternations, and the syncretic nouns are the most probable.

**Figure 2.** Mean perplexity values for fillers and stimuli

**Рисунок 2.** Средние значения перплексии для филлеров и стимульных предложений



The experimental study processes the results of 68 respondents. The mean age of the participants was 37 years ( $sd=10$ ). There were 32 women and 36 men among the respondents, including 5 people whose profession is related to language (linguist, philologist), and 63 people who are not related to language by profession. A resource Yandex.Toloka<sup>1</sup> was used to involve the respondents. Initially, 71 people participated in the experiments, but the responses of 3 participants were excluded. One respondent used only extreme scores: 1 and 7. Since such responses cannot sufficiently reflect possible contrasts in acceptability, this data has been deleted. In addition, the winsorization by the sum of square deviations was applied to search for outliers. This method was described by J. Sprouse<sup>2</sup>. If the scores for fillers differed from the expected ones (2 for ungrammatical and 6 for grammatical) by more than 2 standard deviations, then the data of such respondents were deleted. After applying this procedure, the results of 2 participants were excluded.

Figure 3 shows the acceptability measures of human judgments. In contrast to

perplexity, the higher the normalized score (y-axis), the most acceptable the sentence is. Consequently, the grammatical fillers receive high scores, and ungrammatical fillers are scored low. The acceptability judgments also reveal a significant difference between the singular and plural number of the suppletive nouns ( $p<.0001$ ). However, the correlation is different: the singular nouns turn out to be more acceptable, while the plural nouns are almost as unacceptable as ungrammatical fillers. There is also a significant difference in the number of regular nouns ( $p<.01$ ). The plural forms are more acceptable than the singular forms, but the singular forms are still more acceptable than ungrammatical fillers. There is no statistically significant difference in number for the nouns with suffix alternations ( $p>.05$ ) and for the syncretic forms ( $p>.05$ ). The syncretic forms, both in singular and plural, receive higher scores than other noun types. The other noun types are equally acceptable in singular number but differ in plural: the most acceptable are regular nouns, followed by forms with suffix alternations, and then by nouns with stem alternations.

<sup>8</sup> <https://toloka.yandex.ru/en/>

<sup>9</sup> <https://www.jonsprouse.com/courses/experimental-syntax/scripts/7.4.identify.and.remove.outliers.R>

**Figure 3.** Mean normalized acceptability scores for fillers and stimuli

**Рисунок 3.** Средние нормализованные оценки для филлеров и стимульных предложений

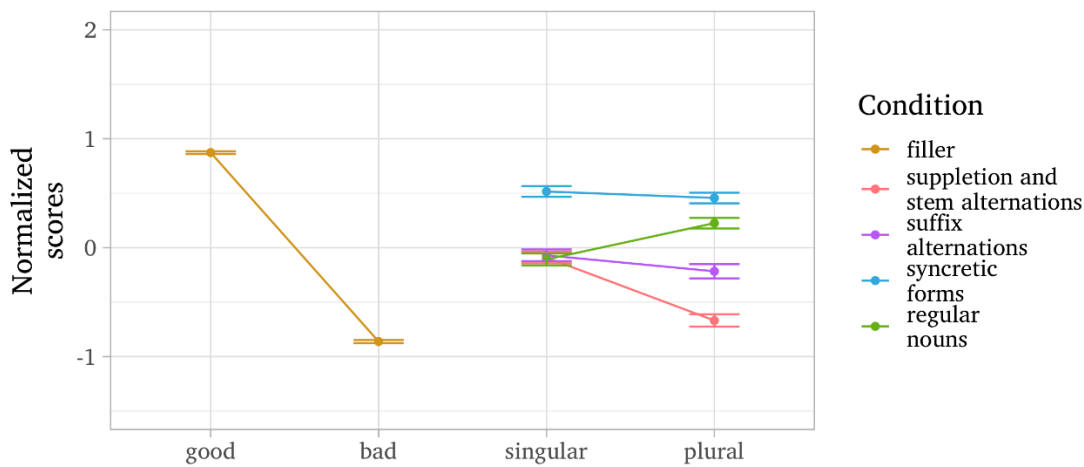
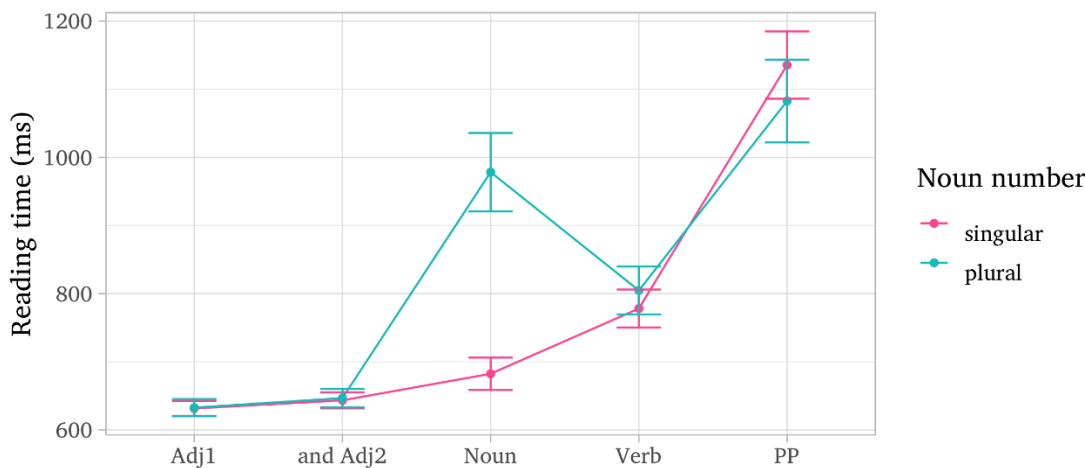


Figure 4 presents the results of the self-paced reading task for sentences with suppletion and stem alternations. The statistical analysis reveals a significant difference in reading time on the noun only

for this morphological type ( $p < .01$ ). Thus, the perception difficulties of plural suppletive forms are confirmed not only by low acceptability scores but also by reading delays.

**Figure 4.** Mean reading time for the suppletive nouns and the forms with stem alternations

**Рисунок 4.** Средние значения времени чтения для супплетивных существительных и форм с чередованием в корне



To summarize the results, we need to compare the predictions of the language model and the human judgments. Both perplexity values and normalized scores assign high acceptability to grammatical fillers and low acceptability to ungrammatical fillers. According to human judgments, nouns with stem alternations prefer singular number, resulting in higher acceptability scores and

less reading time. The nouns with suffix alternations demonstrate free variation. So, stem rather than suffix alternation influences the number choice. The sentences with singularia tantum and pluralia tantum demonstrate the syncretism effect: they receive the highest scores and show no reading delay on the noun since there is no number choice needed. The ruGPT-3



language model also predicts the highest probability for syncretic nouns. Moreover, it assigns an equal probability for singular and plural forms of the nouns with suffix alternations and syncretic nouns. These results correspond with human acceptability judgments. However, ruGPT-3 fails to predict the preference for singular forms in sentences with suppletion and stem alternations.

In this section, we discussed the effect of noun morphology on the number variation in Russian noun phrases with coordinated

(15) *atlantichesk-ij i tikh-ij okean-y slavjats'a*  
Atlantic-SG and Pacific-SG ocean-PL are\_famous.PL  
*svo-imi uragan-ami*  
their-PL.INSTR hurricane-PL.INSTR  
'The Atlantic and Pacific oceans are famous for their hurricanes.'

The second study considered in this section deals with the effect of the syntactic context on the number variation. Specifically, we examine the attributive agreement of the premodifier. The study tested three factors: the premodifier number (singular/plural), the

modifiers. The impact of the syntactic context will be analyzed in the next section.

### The impact of syntactic context on the number variation

According to S. Kodzasov (1987), the subject position and the plural predicate agreement cause the preference for the plural forms (15). This effect of syntactic context was also observed in the first experimental study described in the previous section. The regular plural forms were significantly more acceptable than singular nouns.

noun number (singular/plural), and the case of the noun phrase (direct/oblique). The stimuli were preceded by a context (16) to provide a split interpretation when each adjective characterizes its object. An example of a stimulus is presented in (17-18).

(16) *prodavec predlozhit dva kostjum-a na vybor.*  
assistant.SG offer.PST two suit-PL for choice  
'The assistant offered two suits for choice.'

(17) *Vanja primeril et-ot/ et-i sin-ij i*  
Vanja try\_on.PST this-ACC.SG/ this-ACC.PL blue-ACC.SG and  
*korichnev-yj kostjum/ kostjum-y v magazine-e odezhd-y.*  
brawn-ACC.SG suit-ACC.SG/ suit-ACC.PL in store-LOC.SG clothes-GEN.SG  
'Vanya tried on these blue and brown suits in a store.'

(18) *Vanja zaljubovalsja et-im/ et-imi sin-im i*  
Vanja admire.PST this-INSTR.SG/ this-INSTR.PL blue-INSTR.SG and  
*korichnev-ym kostjum-om/ kostjum-ami v magazine-e odezhd-y.*  
brawn-INSTR.SG suit-INSTR.SG/ suit-INSTR.PL in store-LOC.SG clothes-GEN.SG  
'Vanya admired these blue and brown suits in a store.'

It was noted in Russian that many structures are permissible only in the context of the direct case (Gerasimova, 2018). For instance, masculine nouns denoting female referents allow either masculine or feminine attributive agreement in the direct case (19a).

(19) a. *nov-yj/ nov-aja vrach ann-a petrovn-a*  
new-SG.NOM.M new-SG.NOM.F doctor.NOM.M Anna-NOM.F Petrovna-NOM.F

In the oblique case, only masculine attributive agreement is possible (19b). Therefore, the direct case was expected to allow a number mismatch of the premodifier and the noun, while the oblique case was expected to disallow a number mismatch.

‘the new doctor Anna Petrovna’ (direct case)

b. *nov-omu/ \*nov-oj vrach-u ann-e petrovn-e*  
 new-SG.DAT.M new-SG.DAT.F doctor.DAT.M Anna-DAT.F Petrovna-DAT.F

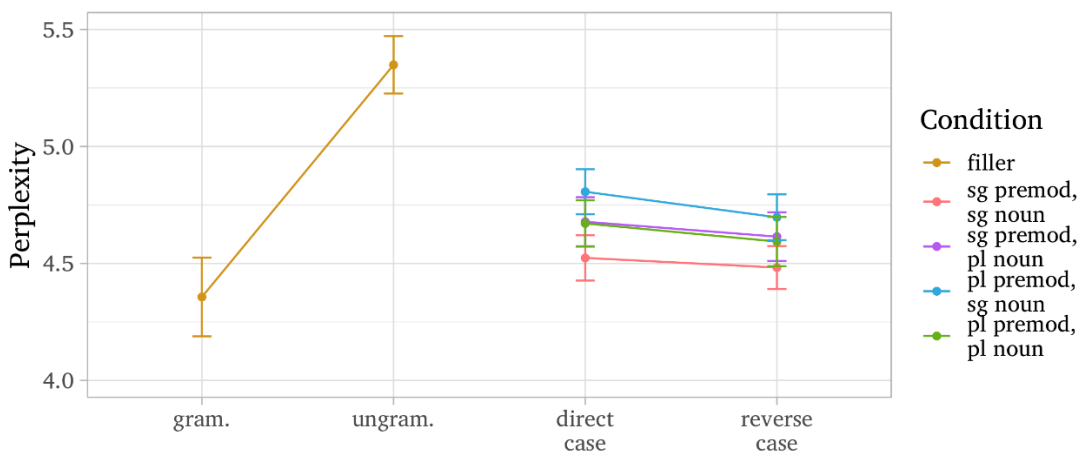
‘to the new doctor Anna Petrovna’ (oblique case)

Figure 5 shows the mean perplexity values generated by the ruGPT-3 language model. As in the first study, the perplexity measures for the grammatical fillers are low, and the ungrammatical fillers receive high perplexity values. According to the language model, there is no significant difference in probability for the direct and the oblique case ( $p > .05$ ). A statistically significant difference in probability occurs between singular and

plural premodifiers for the singular noun ( $p < .01$ ) but not between other conditions ( $p > .05$ ). Thus, the sentences with plural premodifier and singular noun are the least probable, the sentences with singular or plural premodifier and a plural noun are more probable, and the sentences with singular premodifier and singular number are the most probable.

**Figure 5.** Mean perplexity values for fillers and stimuli

**Рисунок 5.** Средние значения перплексии для филлеров и стимульных предложений



In the experimental study, the results of 73 people were analyzed. The mean age of the respondents was 36 years ( $sd=13$ ). There were 25 women and 48 men among the participants; 20 people whose profession is related to language (linguist, philologist), and 53 people who are not related to language by profession. Social networks VKontakte<sup>1</sup> and d3<sup>2</sup> were used to invite the respondents. Initially, 74 people took part in the experiment, but the data of one respondent was excluded. We used the winsorization method to search for outliers described in the previous section.

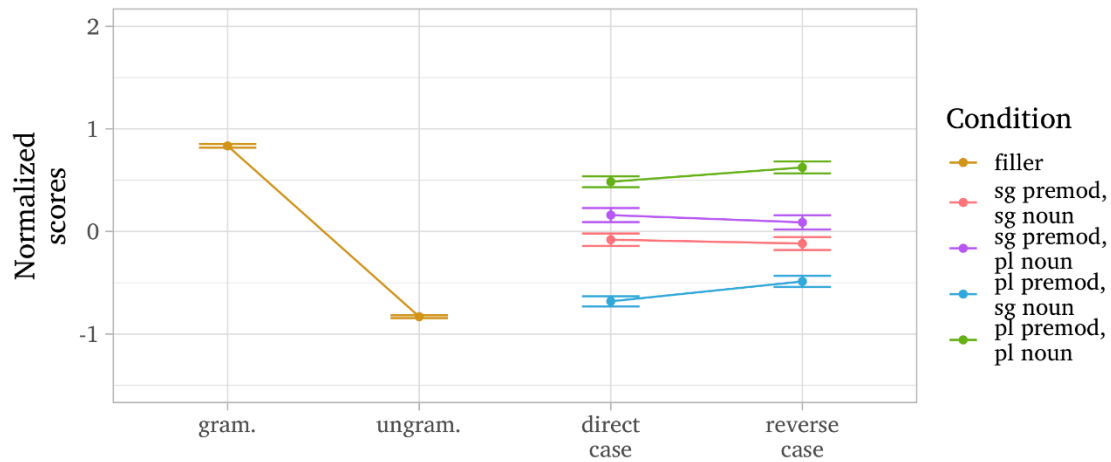
Figure 6 demonstrates the results of human acceptability judgments. As for the perplexity measures, the case does not influence the acceptability significantly ( $p > .05$ ). Beyond that, the statistical analysis reveals a significant difference between all four agreement strategies ( $p < .01$ ). The construction with the plural premodifier and the plural noun receives the highest scores and turns out to be the most acceptable. The combination of the singular premodifier and the plural noun is less acceptable. The acceptability of the singular premodifier and the singular noun is lower. The sentences with the plural premodifier and the singular noun are the least acceptable, which was also predicted by perplexity values.

<sup>10</sup> <https://vk.com/>

<sup>11</sup> <https://d3.ru/>

**Figure 6.** Mean normalized acceptability scores for fillers and stimuli

**Рисунок 6.** Средние нормализованные оценки для филлеров и стимульных предложений

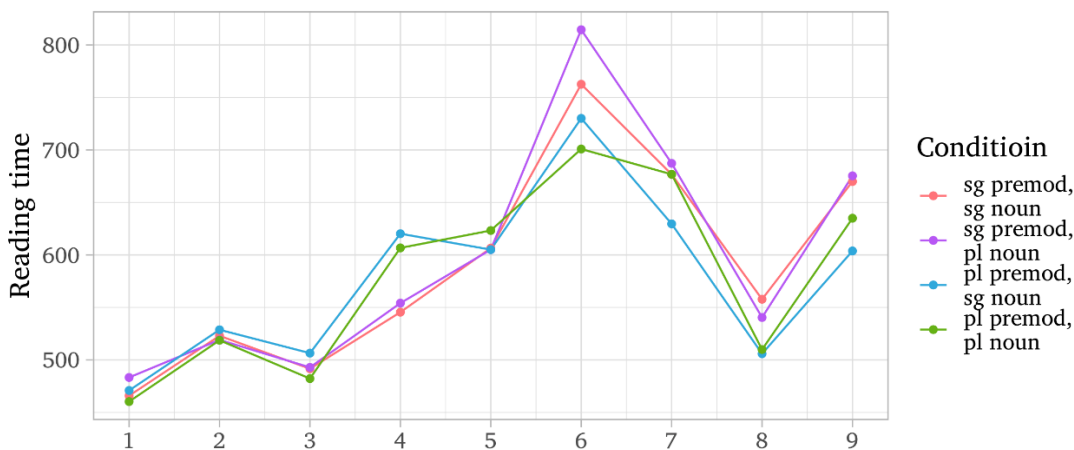


The mean reading time is illustrated in Figure 7. The results for self-paced reading reveal a significant reading delay ( $p < .05$ ) on the plural noun (6<sup>th</sup> word) after the singular premodifier (3<sup>rd</sup> word). Although this

experimental condition receives high acceptability scores, the reading delay can be attributed to the perception difficulties of the sentences with the singular premodifier and the plural noun.

**Figure 7.** Mean reading time for each word

**Рисунок 7.** Средние значения времени чтения по словам



We can now compare the results for the ruGPT-3 language model and the human judgments. As in the first study, the probability of the fillers meets our expectations: the grammatical fillers turn out to be highly acceptable, while the acceptability of the ungrammatical fillers is very low. According to the human acceptability scores, the premodifier number feature implies the same noun number feature.

The highest scores were given to the sentences with the plural premodifier and the plural noun, as well as to the constructions with the singular premodifier and the singular noun. The number mismatch turns out to be acceptable for the plural noun. However, relatively high acceptability scores are accompanied by significant reading delays. The number mismatch is unacceptable for the singular noun, which causes low acceptability

scores. Similarly, the language model predicts the lowest probability for the constructions with the plural premodifier and the singular noun. Besides that, ruGPT-3 assigns high probability to constructions with the same number of premodifiers and nouns. These results correspond with human acceptability judgments. However, the language model fails to predict the preference of the plural premodifier and the plural noun in contrast to the singular premodifier and the plural noun.

### Conclusions

This paper considered the number variation in Russian noun phrases with coordinated modifiers. The coordinated singular adjectives with the split interpretation allow either singular or plural number feature for the head noun. The current study examined morphological and syntactic factors influencing the number choice in this construction. We have conducted two self-paced acceptability experiments to parametrize number variation and identified subtle differences in the acceptability of possible agreement strategies. Beyond that, we compared the results for human judgments with the perplexity values predicted by the neural ruGPT-3 language model. Since the language model was trained on a large amount of text, we expected that its measures will correspond with human judgments and it would predict the correct choice of agreement strategy.

The results of the study are as follows. According to human judgments, stem alternation significantly affects the choice of the number form and causes the preference for the singular number form. The attributive agreement also turns out to be significant since the number of the premodifier implies the same number feature of the noun. Even though the language model did not reveal these contrasts, the perplexity values predicted by ruGPT-3 partially correspond with human judgments. The syncretic forms singularia tantum and pluralia tantum receive the highest normalized scores and the lowest perplexity for their high acceptability and probability. The constructions with the plural

premodifier and the singular noun receive the highest perplexity values and the lowest normalized scores, arguing their low acceptability and probability.

To summarize, the offline acceptability scores and the online reading time demonstrate that the observed morphological and syntactic factors are significant and should be considered while parametrizing number variation in Russian noun phrases with coordinated modifiers. The ruGPT-3 language model, trained on a vast collection of Russian texts, predicts the correct probabilities for highly acceptable and unacceptable sentences (both fillers and stimuli); however, it fails to assign correct probability values to the cases of variation.

### Abbreviations

ACC – accusative, ADJ – adjective, DAT – dative, DEF – definite, INSTR – instrumental, F – feminine, GEN – genitive, LOC – locative, M – masculine, N – neutral, PST – past tense, PL – plural, SG – singular.

### References

- Aaronson, D. and Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence, *Journal of Experimental Psychology: Human perception and performance*, 2 (1), 56–70. <https://doi.org/10.1037/0096-1523.2.1.56> (In English)
- Barros, M. and Vicente, L. (2011). Right node raising requires both ellipsis and multidomination, *University of Pennsylvania working papers in linguistics*, 17 (1), 1–9. <https://doi.org/10.22099/jill.2022.41795.1275> (In English)
- Belova, D., Demina, J., Gerasimova, A., Lyutikova, E., Morgunova, E., Petelin, D., Studenikina, K. and Voznesenskaya, A. (2021). *Russkije ostrova v svete eksperimental'nyh dannyh* [Russian islands in the light of experimental data], E. Lyutikova, A. Gerasimova (eds.), Buki Vedi, Moscow, Russia. (In Russian)
- Brunato, D., Chesi, C., Dell'Orletta, F., Montemagni, S., Venturi, G. and Zamparelli, R. (2020). AcCompl-it @ EVALITA2020: Overview of the Acceptability & Complexity Evaluation Task for Italian, *International Workshop on*

*Evaluation of Natural Language and Speech Tools for Italian*, 1–8. (In English)

Corbett, G. G. (1979). The agreement hierarchy, *Journal of linguistics*, 15 (2), 203–224. <https://doi.org/10.1017/S0022226700016352> (In English)

Dou, Y., Forbes, M., Koncel-Kedziorski, R., Smith, N. A. and Choi, Y. (2022). Is GPT-3 Text Indistinguishable from Human Text? Scarecrow: A Framework for Scrutinizing Machine Text, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Dublin, Ireland, 7250–7274. <http://dx.doi.org/10.18653/v1/2022.acl-long.501> (In English)

Gerasimova, A. (2018). Rassoglasovanie po rodu v ruskom yazyke (eksperimental'noe issledovanie) [Mixed agreement patterns in Russian (experimental study)], *Izvestiia Rossiiskoi akademii nauk. Seriya literatury i iazyka*, 77 (1), 65–71. (In Russian)

Gold, J. W. (2021). Locus of Gender Resolution: on Goal or on Probe? *Workshop on Agreement in Multivaluation Construction (AMC 2021)*. (In English)

Gries, S. T. (2021). *Statistics for Linguistics with R: A Practical Introduction*, De Gruyter Mouton, Berlin/Boston, Germany/USA. <https://doi.org/10.1515/9783110718256> (In English)

Grosz, P. G. (2015). Movement and agreement in right-node-raising constructions, *Syntax*, 18 (1), 1–38. <https://doi.org/10.1111/synt.12024> (In English)

Harizanov, B. and Gribanova, V. (2015). How across-the-board movement interacts with nominal concord in Bulgarian, *Proceedings from the Annual Meeting of the Chicago Linguistics Society 49*, University of Chicago, IL, Chicago Linguistics Society, 115–129. (In English)

Hartmann, K. (2000). *Right node raising and gapping: Interface conditions on prosodic deletion*, John Benjamins, Amsterdam, Netherlands. <https://doi.org/10.1075/z.106> (In English)

Kodzasov, S. (1987). Chislo v sochinitel'nykh konstrukciyax [Number in coordinated structures], in Kodzasov, S., Laufer, N. and Savina, E. (eds.), *Modelirovanije jazykovoj dejatel'nosti v intellektual'nykh sistemax*, Moscow, Russia, 204–219. (In Russian)

Lau, J. H., Armendariz, C., Lappin, S., Purver, M. and Shu, C. (2020). How furiously can

colorless green ideas sleep? Sentence acceptability in context, *Transactions of the Association for Computational Linguistics*, 8, 296–310. [https://doi.org/10.1162/tacl\\_a\\_00315](https://doi.org/10.1162/tacl_a_00315) (In English)

Lau, J. H., Clark, A. and Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge, *Cognitive science*, 41 (5), 1202–1241. <https://doi.org/10.1111/cogs.12414> (In English)

Likert, R. (1932). A technique for the measurement of attitudes, *Archives of Psychology*, 140, 1–55. (In English)

Schütze, C. and Sprouse, J. (2014). Judgment data, in Sharma, D. and Podesva, R. (eds.), *Research methods in linguistics*, Cambridge: Cambridge University Press, 27–50. (In English)

Shen, Z. (2018). *Feature arithmetic in the nominal domain*, Ph.D. Thesis, University of Connecticut, Storrs, CT. (In English)

Sprouse, J., Schütze, C. T. and Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001–2010, *Lingua*, 134, 219–248. <https://doi.org/10.1016/j.lingua.2013.07.002> (In English)

Sprouse, J., Yankama, B., Indurkha S., Fong, S. and Berwick, R. C. (2018). Colorless green ideas do sleep furiously: gradient acceptability and the nature of the grammar, *The Linguistic Review*, 35 (3), 575–599. <https://doi.org/10.1515/tlr-2018-0005> (In English)

Talmy, L. (2018). Introspection as a Methodology in Linguistics, *Ten lectures on cognitive semantics*, Brill, Leiden, 218–262. [https://doi.org/10.1163/9789004349575\\_007](https://doi.org/10.1163/9789004349575_007) (In English)

Warstadt, A., Singh, A. and Bowman, S. R. (2019). Neural network acceptability judgments, *Transactions of the Association for Computational Linguistics*, 7, 625–641. [https://doi.org/10.1162/tacl\\_a\\_00290](https://doi.org/10.1162/tacl_a_00290) (In English)

Wilder, C. (1997). Some properties of ellipsis in coordination, in Alexiadou, A. and Hall, T. A. (eds.), *Studies in universal grammar and typological variation*, Amsterdam, 59–108. <https://doi.org/10.1075/la.13.04wil> (In English)

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A. and Rush, A. M. (2020). Transformers: State-of-the-art natural language processing, *Proceedings of the 2020 conference*



*on empirical methods in natural language processing: system demonstrations*, Association for Computational Linguistics, 38–45. <http://dx.doi.org/10.18653/v1/2020.emnlp-demos.6> (In English)

*Все авторы прочитали и одобрили окончательный вариант рукописи.*

*All authors have read and approved the final manuscript.*

*Конфликты интересов: у автора нет конфликтов интересов для декларации.*

*Conflicts of interests: the author has no conflicts of interest to declare.*

**Kseniia A. Studenikina**, Programmer at the Laboratory for Computational Lexicography, Research Computing Center, Lomonosov Moscow State University, Ph.D. Student of the Department of Theoretical and Applied Linguistics, Philological Faculty, Lomonosov Moscow State University, Moscow, Russia.

**Ксения Андреевна Студеникина**, программист Лаборатории автоматизированных лексикографических систем Научно-исследовательского вычислительного центра Московского государственного университета имени М. В. Ломоносова, аспирант кафедры теоретической и прикладной лингвистики филологического факультета Московского государственного университета имени М. В. Ломоносова, Москва, Россия.