



Нестандартные количественные конструкции в русской речи носителей нанайского и ульчского языков

Non-standard syntax of numerals in the Russian of Nanai and Ulcha speakers

Natalia Stoynova¹ (Наталья Марковна Стойнова)

Accepted: 17 July 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Аннотация

В статье рассматривается оформление количественной конструкции в разновидностях русской речи, относящихся к пост-пиджинному континууму, а именно в речи старшего поколения двуязычных носителей южно-тунгусских языков (нанайского и ульчского). Наряду со стандартными употреблениями с генитивом типа *два дома, пять братьев* в их речи встречаются отсутствующие в литературном русском номинативные конструкции типа *два дом, пять братья*. Структурно они похожи на тунгусские количественные конструкции, однако можно предполагать и другие источники: наследование синтаксиса количественной конструкции из пиджина, из диалектной речи монолингвов или результат недоусвоения русской системы. Свидетельства в пользу того, что мы имеем дело именно с копированием тунгусской синтаксической модели, дает количественный анализ вариативности на корпусном материале. Частотное распределение между генитивной и номинативной моделями регулируется семантикой существительного и числительного: сочетания с существительными времени и меры, а также конструкции с большими числами склонны к стандартному генитивному оформлению. Все это контексты, более характерные для общения на русском языке, чем на южно-тунгусских, т.е. такие, в которых естественно ожидать меньшего контактного влияния. За распределение между двумя конкурирующими моделями с номинативом (*два брат vs. два братья*) отвечают те же факторы, что и в южно-тунгусских языках. Рассмотренные данные интересны в контексте более общего вопроса о контактных явлениях в грамматике. Полученные данные подтверждают предположение о том, что заимствованию легче подвергаются те грамматические черты языка-донора, которые соответствуют общим типологическими тенденциями и функционально мотивированы.

Abstract

The paper deals with non-standard encoding of numeral phrases in a post-pidgin variety of Russian, namely in the speech of older speakers of Southern Tungusic languages (Nanai

✉ N. Stoynova
stoynova@yandex.ru

¹ Institute of Linguistics, Russian Academy of Sciences & Russian Language Institute, Russian Academy of Sciences, Moscow, Russia

and Ulcha). Along with standard uses with the genitive (*dva doma* [two house.GEN.SG] ‘two houses’, *pjat’ brat’ev* [five brother.GEN.PL] ‘five brothers’), nonstandard nominative uses (*dva dom* [two house.NOM.SG] ‘two houses’, *pjat’ brat’ja* [five brother.NOM.PL] ‘five brothers’) are attested in their speech. These nominative constructions with numerals resemble Tungusic ones. However, some alternative hypotheses on their origin can be proposed: they might be inherited from the pidgin, from monolinguals’ dialectal speech, or be a result of incomplete acquisition of Standard Russian. Arguments in favor of structural borrowing from Tungusic come from a corpus-based quantitative study on variation. The variation between the genitive vs. nominative patterns is regulated by the semantics of the noun and the numeral: time and measure expressions, as well as expressions with large numbers tend to be encoded in a standard way. These are exactly those contexts, for which communication in Russian is more typical than communication in Tungusic; so they appear to be less affected by contact. The variation between two different nominative patterns, i.e. those with NOM.SG vs. NOM.PL (*dva brat* [two brother.NOM.SG] vs. *dva brat’ja* [two brother.NOM.PL] ‘two brothers’), is motivated by exactly the same factors as in Southern Tungusic. The data observed contribute to a more general question on contact-induced changes in grammar. They favor the idea that grammatical features of the donor language are more prone to undergo structural borrowing if they are in line with crosslinguistic tendencies and supported by general functional motivations.

1 Введение

В статье обсуждается нестандартное оформление количественной конструкции в русской речи носителей нанайского и ульчского языков. Близкородственные нанайский и ульчский языки относятся к южной подгруппе тунгусо-маньчжурских языков (Хабаровский край). Их носители в основном старшего поколения, все или почти все свободно владеют русским. О современных образцах русской речи нанайцев и ульчей можно говорить как о постпиджинном континууме (см. о посткреольском и постпиджинном континууме, возникающем, когда креольский язык или пиджин продолжает находиться в контакте с языком-лексификатором, De Camp, 1971; Bickerton, 1973; 1980; Rickford, 1987; Sebba, 1997, p. 210–224; Mühlhäusler, 1997, p. 211–221 и др.; дискуссию о случаях, когда его не наблюдается, в Schwegler, 2000; Deuber, 2006). Русский пиджин, ранее распространенный на этой территории, на данный момент уже не используется, однако русская речь самых старших носителей (1930-е гг. и старше) существенно отличается от речи монолингвов. Из ее грамматических особенностей можно назвать кодирование прямого объекта номинативом и некоторые другие нестандартные модели управления, опущение предлогов, рассогласование по роду, факультативное маркирование числа, особые глагольные конструкции (например, плюсквамперфектная с *быть*), см. обзор в Stoynova (2019). Русская речь более молодых носителей и полуносителей практически не отличается от русской речи монолингвов.

В литературном русском языке количественная конструкция имеет достаточно сложный и типологически необычный синтаксис (см. подробнее п. 2.1). Интересно проверить, какие изменения этот фрагмент грамматики будет претерпевать в речи билингвов. В речи старшего поколения нанайцев и ульчей, наряду со стандартными, фиксируются такие употребления, как (1)–(3):

- (1) *У нас тоже два собака*\ (gld, vsg)¹
- (2) *У мене уже четверо\ дети были* (gld, fna)
- (3) *Вот там пять дом!* (gld, vsg)

В работе будет рассмотрен вопрос о том, можно ли считать употребления типа (1)–(3) копированием модели (структурным заимствованием, синтаксической калькой) из южно-тунгусских языков, ср. пример (4) из нанайского языка:

- (4) *i ela enda bi-či-ni=go bun-du*
и три собака быть-PT-3SG=PTCL мы-DAT
'И три собаки у нас было'. (gld, lkb)

Как видно из (4), в южно-тунгусских языках количественные конструкции действительно оформляются похожим образом. С другой стороны, возможны и альтернативные объяснения. Можно видеть в (1)–(3) перестройку стандартной русской грамматической системы без копирования южно-тунгусского синтаксиса, свидетельствующую о недоусвоенности или нестандартном усвоении системы. Можно объяснять отступления от стандартной системы в терминах более общих прагматических или универсально-языковых механизмов: известно, что в языковых разновидностях, свободных от давления нормы, к которым можно отнести и рассматриваемую контактную разновидность русского, действие таких механизмов может быть более отчетливым, чем в литературном языке (см., например, Chambers, 2004). Употребления типа (1)–(3) могут также просто наследоваться из пиджина. Наконец, можно допустить, что они вообще не связаны с ситуацией языкового контакта, а проникают в речь двуязычных носителей из диалектной речи монолингвов.

Общая проблема, которую, таким образом, иллюстрирует это частное исследование, известна в разных областях контактной лингвистики (в т.ч. в исследованиях по креолам и пиджинам, в ареальной лингвистике): явление, которое на первый взгляд кажется копированием морфосинтаксической модели, часто оказывается устроено сложнее, имеет другую мотивацию или несколько разных, не всегда различных (см. о проблеме и возможных диагностиках грамматического заимствования, напр., Aikhenvald, 2007; Thomason, 2009; Poplack & Levey, 2010; Heine & Nomachi, 2013).

Русская речь двуязычных носителей малых языков России – удобное поле для ее обсуждения. В данном случае можно наблюдать ситуацию языкового контакта «в развитии». Это означает, что не нуждается в доказательстве сам факт языкового контакта и (хотя бы в общих чертах) известно, какие именно языковые разновидности контактировали друг с другом. С другой стороны, при такой ситуации мы имеем дело не с устойчивыми языковыми явлениями предположительно контактной природы, а с вариативностью – конкуренцией нестандартных грамматических явлений со стандартными, свойственными речи монолингвов.

В такой ситуации именно подробное количественное исследование вариативности – т.е. факторов, регулирующих частотное распределение конкурирующих стратегий морфосинтаксического оформления, – может дать дополнительные аргументы в дискуссии о копировании морфосинтаксической модели (см. об анализе вариативности и методах вариационной социолингвистики при исследовании языковых контактов, напр., Patrick, 2008; Meyerhoff, 2009, 2021; Légise & Chamoreau, 2013; Poplack,

¹Здесь и далее в скобках дается указание на источник примера в формате «ISO-код языка, код информанта» (gld – нанайский, ulc – ульчский). Все русские примеры взяты из Корпуса русской речи носителей языков Севера Сибири и Дальнего Востока.

2021). Если это те же факторы, которые (на уровне правил или тенденций) регулируют оформление количественной конструкции в нанайском и ульчском языках, то это будет аргументом за копирование тунгусской модели. Если это факторы, релевантные для русской грамматической системы, – это будет аргументом против. Важно также, какие более общие мотивации будут стоять за этими факторами – связанные с ситуацией языкового контакта, с упрощением системы, с типологическими тенденциями, прагматикой или иные.

Количественный анализ вариативности в оформлении конструкций с числительными был проведен на материале коллекции устной речи из Корпуса русской речи носителей языков Севера Сибири и Дальнего Востока (см. п. 3.1).

Анализ этих данных дает аргументы в пользу того, что наблюдаемые в речи носителей южно-тунгусских языков нестандартные количественные конструкции действительно копируют тунгусскую модель. При этом копирование модели поддерживается более общими универсально-языковыми принципами: двуязычные носители делают выбор в пользу более регулярного, более соответствующего общим типологическими тенденциями, функционально мотивированного синтаксического образца.

Статья организована следующим образом. В разделе 2 характеризуется и сопоставляется структура количественной группы в стандартном русском языке и в южно-тунгусских языках, а также приводится краткая информация о количественной группе в дальневосточном пиджине и в диалектной речи монолингвов. В разделе 3 рассматривается вариативность в оформлении конструкций с числительными, наблюдаемая в русской речи нанайцев и ульчей, в т.ч. приводится ее статистический анализ. В разделе 4 суммируются и обсуждаются результаты исследования, раздел 5 содержит краткое заключение.

2 Количественная группа в русском языке монолингвов и в южно-тунгусских языках

В этом разделе дано описание количественной конструкции в русском языке монолингвов (п. 2.1), с одной стороны, и в южно-тунгусских языках (п. 2.2), с другой стороны. В п. 2.3 русская и тунгусская стратегии оформления сопоставляются. В п. 2.4 кратко охарактеризована количественная конструкция в дальневосточном русском пиджине и в русских диалектах.

2.1 Количественная группа в русском языке монолингвов

Ниже кратко описаны синтаксические свойства русской количественной конструкции, релевантные для обсуждения. Более подробное описание синтаксиса русских числительных и его теоретическую трактовку в разных парадигмах см., например, в Мельчук (1985), Corbett (1993), Rappaport (2002), Franks (2002), Ionin и Matushansky (2018, р. 161–221), Циммерлинг (2019); см. также обзор подходов к описанию русских числительных в Сичинава (2018).

Во-первых, в номинативе и аккумулятиве (для тех существительных, для которых аккумулятив совпадает с номинативом²) количественная группа устроена не так, как в косвенных падежах. В косвенных падежах количественное числительное ведет себя как

²Далее везде речь идет о формах аккумулятива, тождественных номинативу. В соответствии с более общим правилом, это в полной мере касается неодушевленных существительных (*купил два кресла, пять кресел*), одушевленные в количественных конструкциях выступают в форме, тождественной генитиву

модификатор, согласуемый по падежу с вершиной-существительным (*двум домам*, ср. *большим домам*). В номинативе и аккузативе падеж принимает количественное числительное, а существительное оформляется родительным падежом, как зависимое в именной группе (*два дома*, ср. *друг дома*).

Во-вторых, разные классы числительных требуют в номинативе/аккузативе разных числовых форм существительного. Числительные малого количества (далее паукальные): *два*, *три*, *четыре*, производные от них составные числительные (*тридцать два*), а также *оба* и *полтора* – требуют генитива единственного числа (*три мешка*), числительные большого количества (*пять* и далее), собирательные и местоименные слова типа *сколько*, *много* – генитива множественного числа (*пять мешков*, *двое итанов*, *много детей*).³

Сочетания с паукальными числительными имеют несколько дополнительных особенностей. Во-первых, некоторые существительные могут выступать не в обычной форме генитива, а в форме с особым ударением: *до часа* vs. *четыре часа*. Во-вторых, числительные *два*, *оба*, *полтора* согласуются с существительным по роду (*два яблока* vs. *две яблони*). В-третьих, если в таком сочетании присутствует прилагательное, его форма отличается от формы существительного: это либо форма генитива мн.ч., либо форма номинатива мн.ч. (*два каменных дома*, *две мраморных ~ мраморные статуи*). В некоторых трактовках предлагается считать, что существительное при паукальном числительном стоит не в форме генитива, а в омонимичной ему (или, для некоторых существительных особой) падежной форме. См. дискуссию о конструкциях с паукальными числительными в Pereltsvaig (2011), Bailyn and Nevins (2008), Madariaga and Igartua (2017), Nessel and Nordrum (2019) и мн.др.

Отдельная сложность, важная в контексте статьи, состоит в том, что для части существительных форма, которая ожидается при числительном малого количества, совпадает с формой множественного числа номинатива: ср. *около башни* и *стоят башни*. Употребления типа *обе стороны* (ударение, характерное для NOM.PL) вместо ожидаемого *обе стороны* (GEN.SG) и форма номинатива мн.ч. прилагательного, выступающая в той же конструкции (*две большие башни*), позволяют предположить, что в части случаев форма номинатива мн.ч. может как минимум контаминироваться с формой, выступающей при паукальных числительных. См. подробное обсуждение в статье Nessel and Nordrum (2019).

Учитывая сказанное, далее мы, тем не менее, условно будем считать, что в общем случае в современном литературном русском языке мы имеем дело с формой генитива единственного числа.

Описанная система достаточно сложна и типологически нетривиальна, хотя и не уникальна. Так, например, в Kortjevskaja-Tamm и Wälchli (2001, p. 698–704) обсуждается сходство в оформлении количественных конструкций в русском (и, шире, славянских языках), балтийских, прибалтийско-финских и саамских языках, которое, по мнению авторов, не случайно, а является результатом языковых контактов в циркумбалтийском ареале. За пределами этого ареала похожая система (с числительным, управляющим падежом существительного) зафиксирована в шотландском (гэльском) языке, ср. обсуждение этой параллели в Stolz (2002, p. 383).

(*встретил двух девочек*) или также номинативу (*встретил пять девочек*), см. описание и трактовку этого распределения в (Mikaelian, 2013), об отдельных нестандартных употреблениях типа *на две девочки больше* также в (Yanko, 2004).

³Числительное *один*, которое во всех падежных формах синтаксически ведет себя как прилагательное, в рассмотрении не включалось. Также не рассматривались *тысяча*, *миллион*, *миллиард* и под., ведущие себя как существительные. Полностью исключены из рассмотрения порядковые числительные.

Ни выбор родовой формы, ни другие формальные характеристики самого числительного, ни форма прилагательного в количественной конструкции нас при обсуждении данных от носителей южно-тунгусских языков интересовать не будут. Речь пойдет только о форме существительного.

Ниже русская речь носителей южно-тунгусских языков будет условно сравниваться со стандартной устной речью монолингвов («литературным разговорным языком») в классификации Е. А. Земской (Земская, 1988), см. также обзор разновидностей русской речи, напр., в Mustajoki, 2010). Более корректным было бы сравнение с речью русских монолингвов из тех мест, где сделаны наши записи (подробными данными о которой мы, к сожалению, не располагаем). Однако для описываемой социалингвистической ситуации такое огрубление кажется допустимым, см. пп. 2.4 и 3.1.

2.2 Количественная группа в южно-тунгусских языках

В нанайском и ульчском языках количественная группа устроена следующим образом. Поведение числительного ничем не отличается от поведения прилагательного в атрибутивной позиции. Точно так же, как прилагательное, оно, как правило, выступает в препозиции к существительному и не принимает никаких словоизменительных показателей.⁴ Существительное в количественной группе оформляется показателями соответствующего падежа. Оно может принимать аффикс множественного числа (5) или выступать в немаркированной форме единственного числа (6). Подробнее о числительных в нанайском языке см. Аврорин (1959, с. 217–218; 232–238); в ульчском – Суник (1985, с. 38–39).

(5) *nadan žakpon žoan piktə-səl bi-či-či=goa*
семь восемь десять ребёнок-PL быть-PST-3PL=PTCL
'Семь, восемь, десять детей бывало'. (gld, lkb)

(6) *mī ənim-bi=təni nadan piktə bi-či*
1SG мать-1SG=a семь ребёнок быть-PST
'У моей мамы было семеро детей'. (gld, lkb)

Формы ед.ч. и мн.ч. существительного в количественной группе взаимозаменяемы (ср. примеры выше из одного и того же текста, которые различаются только числом).

В Таблице 1 представлены результаты анализа небольшой выборки нанайских текстов (амурские говоры).⁵

Анализ показывает, что наличие показателя множественного числа на существительном коррелирует с семантикой существительного и типом количественного слова. Одушевленные существительные значимо чаще оформляются показателем множественного числа, чем неодушевленные. При местоимениях-числительных (*əgži* 'много', *oi* 'мало', *xado* 'сколько') существительное значимо чаще принимает показатель множественного числа, чем при числительном. Для других нанайских диалектов и для ульчского языка точных подсчетов не проводилось, но, как кажется, там действуют те же механизмы.

⁴В более редких случаях (прежде всего связанных с особой коммуникативной структурой) числительное (как и прилагательное) может перемещаться в постпозицию и/или принимать падежные маркеры.

⁵Тексты собраны в 2011–2016 гг. в селах Хабаровского края, расшифрованы и отгlossированы С. А. Осольской и автором.

Таблица 1 Единственное vs. множественное число существительного в количественной группе: на-найский

параметр	значения	SG	PL	% PL	точный двусторонний критерий Фишера
одушевленность существительного	одушевленные	19	21	52,50%	p<0,0001
	неодушевленные	42	3	6,67%	
тип числительного	местоимения-числительные ^a	20	17	45,95%	p=0,003
	прочие числительные	41	7	14,58%	
всего		61	24	28,24%	

^aУпотребления с неисчисляемыми существительными (типа *много мяса*) не учитывались

Таблица 2 Количественная группа в русском и в южно-тунгусских языках

параметры	русский монолингвов	южно-тунгусские
порядок компонентов	NUM N	
синтаксическая структура	НОМ/АСС: числительное управляет падежом существительного (генитив) остальные падежи: числительное согласуется по падежу с существительным	числительное – неизменяемый модификатор, существительное изменяется по падежам
число существительного	зависит от падежа и класса числительного: в номинативе/аккузативе SG при паукальных числительных, PL при остальных	зависит от семантики существительного и класса числительного: более вероятно PL при одушевленных существительных и при местоимениях-числительных
правила выбора альтернативных стратегий оформления	выбор между GEN.SG и GEN.PL: жесткие правила	выбор между SG и PL: вариативность, количественные тенденции

2.3 Количественная группа в русском языке монолингвов vs. в южно-тунгусских языках: сопоставление

В Таблице 2 резюмируются сопоставительные сведения о структуре количественной группы в русском языке монолингвов и в южно-тунгусских языках. Как видно из таблицы, русская система устроена сложнее: она предполагает разное синтаксическое оформление для разных классов числительных (“privileging” в типологии Stolz, 2002), тогда как в тунгусской системе конструкции со всеми числительными оформляются единообразно (“egalitarian” по Stolz, 2002).

Двуязычный носитель имеет в качестве потенциальных ориентиров обе системы оформления количественной группы – русскую и тунгусскую. Как видно из Таблицы 2, системы обнаруживают достаточно сходств, чтобы обеспечить теоретическую возможность калькирования тунгусского образца в русской речи, но и достаточно различий, чтобы у двуязычного носителя возникала сама потребность копировать тунгусский образец.

Для анализа ниже были выбраны только количественные конструкции в номинативе/аккузативе. Во-первых, это одна из точек максимального расхождения между тунгусской и русской системами: именно в этой точке тунгусская и русская системы предсказывают разные синтаксические стратегии, и интересно проследить, какую из них выбирает двуязычный носитель. Во-вторых, такие употребления в наших данных гораздо частотней прочих и позволяют давать не только качественные, но и количественные оценки.⁶

2.4 Количественная конструкция в дальневосточном пиджине и в русских диалектах

Выше в качестве систем, на которые может ориентироваться носитель рассматриваемой разновидности русского языка, упомянута стандартная система русского языка монолингвов и система, представленная в южно-тунгусских языках. На самом деле есть еще два потенциальных ориентира – системы, представленные в пиджине и в диалекте.

Можно предположить, что оформление количественной конструкции, представленное в наших данных, наследуется из ныне не существующего базилектного варианта пиджина, который использовался более старшими поколениями носителей южно-тунгусских языков. Последовательных записей вариантов пиджина, распространенных во всех тех районах, откуда родом наши информанты, не сохранилось (предположительно, они были очень близки друг к другу). Достаточно детально описана только бикинская разновидность, использовавшаяся в Пожарском районе Приморского края носителями удэгейского языка и бикинского нанайского (данные от носителей которого в небольшом количестве вошли в нашу выборку), см. Перехвальская (2006, 2008).

По данным Е. В. Перехвальской, в базилектной речи присутствовали отдельные числительные, а в тех вариантах, которые автор относит к «расширенному» пиджину, также квантификаторы (*много*, *(u)ce/(u)cë* (рус. *все/всё*). Существительное, в свою очередь, не изменялось ни по числам, ни по падежам. Количественная конструкция, таким образом, образовывалась простым соположением числительного и существительного:

- (7) *Два мамка* была папа
‘У [моего] отца было две жены’ (пример из Перехвальская, 2006)

Также можно предположить, что нестандартная количественная конструкция возникает не или не только под влиянием тунгусо-маньчжурских языков, а наследуется из диалектной речи монолингвов. В некоторых русских говорах действительно встречаются нестандартные количественные конструкции с паукальными и собирательными числительными – с номинативом мн.ч. вместо генитива (типа *два мальчика*, *двое мальчики*, в разных говорах для этой модели наблюдаются разные ограничения на тип склонения существительного), см., напр., Аванесов и Орлова (1964, с. 174–175); ДАРЯ III(2), карта 7. Более младшее поколение наших двуязычных носителей с очевидностью ориентируется прежде всего на литературный русский, для более старших дей-

⁶Количественные конструкции в косвенных падежах встретились всего в 16 примерах из выборки, три из них обнаруживают нестандартный синтаксис (они обсуждаются в п. 3.2 в качестве дополнительного материала). Носители, речь которых вошла в рассмотрение, видимо, в целом избегают употребления числительных за пределами позиций субъекта и прямого объекта (отчасти это может быть связано с их двуязычием, отчасти с жанром и типом текстов).

ствительно можно допустить влияние диалектной речи. Однако рассмотренные территории заселялись русскими – переселенцами из самых разных мест – довольно поздно, наиболее интенсивный контакт приходится уже на XX век, так что указать какой-то определенный диалект, который мог бы оказать заметное влияние на все наши данные, скорее нельзя.

3 Количественная группа в русской речи носителей южно-тунгусских языков

В этом, основном, разделе обсуждаются корпусные данные русской речи носителей нанайского и ульчского языков. В п. 3.1 описан принцип отбора примеров и охарактеризована исследуемая выборка. В п. 3.2 перечислены все синтаксические модели оформления количественной группы, зафиксированные в данных. В п. 3.3 приводится и комментируется частотное распределение употреблений по этим типам. В пп. 3.4–3.5 обсуждаются характеристики числительного и существительного, с которыми коррелирует выбор между генитивным и номинативным кодированием существительного (*пять мальчиков* vs. *пять мальчики*). В п. 3.6 дан статистический анализ этих корреляций. В п. 3.7 показано, от чего зависит выбор между оформлением ед.ч. vs. мн.ч. номинатива (*пять мальчик* vs. *пять мальчики*). В п. 3.8 кратко обсуждаются возможные корреляции с социолингвистическими характеристиками носителей.

3.1 Данные

Исследование проведено на материале тунгусской коллекции Корпуса русской речи носителей языков Севера Сибири и Дальнего Востока (<http://web-corpora.net/ruscontact/>). Вошедшие в нее тексты записаны нами в качестве «побочного продукта» экспедиций по документации южно-тунгусских языков: нанайского (преимущественно амурские говоры, в небольшом количестве бикинский и кур-урмийский нанайский) и ульчского.⁷ Коллекция состоит из устных спонтанных текстов: это в основном короткие нарративы (сказки, случаи из жизни и под.) и этнографические описания. Коллекция расшифрована, часть коллекции снабжена автоматической морфологической разметкой и специальной ручной разметкой отклонений от русской речи монолингвов (см. о ней Khomchenkova et al., 2019). На момент исследования коллекция составляла 143 387 словоупотреблений.

Носители, речь которых представлена в коллекции, – с 1917 по 1961 г. р. Все они свободно владеют (владели) русским языком наряду с нанайским/ульчским. Русский язык был усвоен частью из них в школьном возрасте, частью – до школы, параллельно с нанайским/ульчским. В настоящее время русский язык используется всеми носителями активнее нанайского/ульчского. Более младшее поколение по-нанайски/по-ульчски, за редким исключением, свободно не говорит. См. подробнее о социолингвистической ситуации Герасимова (2002) (о нанайском и ульчском); Калинина и Оскольская (2016) (о нанайском); Сумбатова и Гусев (2016) (об ульчском). Русская речь, вошедшая в коллекцию, неоднородна с точки зрения близости к речи монолингвов: у части (самых

⁷Большинство текстов записаны автором совместно с С. А. Оскольской в 2011–2018 гг. в селах Хабаровского и Приморского края. Небольшая часть взята из архива полевых аудиозаписей Л. И. Сем. Тексты расшифрованы и размечены автором.

пожилых) носителей она близка к (ныне утраченному) пиджину, у части (как правило, самых молодых) носителей почти не отличается от литературного стандарта. См. более подробную характеристику носителей в п. 3.8.

Для исследования из корпуса была автоматически извлечена сплошная выборка примеров с числительными, всего 1 722 примера (включая примеры со словами, ошибочно автоматически распознанными как числительные). Выборка была отфильтрована вручную. В окончательную выборку попали примеры со следующими ограничениями.

- Включены только конструкции: с количественными, собирательными числительными и местоимениями-числительными (типа *много, сколько*), в т.ч. *оба, полтора*; исключены сочетания с порядковыми числительными и числительным *один*, а также с составными числительными, включающими *один* (типа *двадцать один*); исключены сочетания со словами *миллион, миллиард, тысяча*.
- Рассмотрены только количественные группы в номинативе/аккузативе (см. п. 2.3).
- Включены только сочетания с выраженным существительным или субстантивированным прилагательным: исключены сочетания с местоимениями (*много их, много чего*); исключены одиночные числительные (*двое в черном*).
- Не включены частотные сочетания с существительным *время* (*сколько время* – либо номинатив, либо, что вероятнее, нестандартная разговорная форма генитива, встречающаяся и у монолингвов).
- Остальные сочетания с морфологически неоднозначными формами существительного включены (*много масл[а], две тетради, двести грамм*) и условно трактуются как стандартные.
- Не включены примеры со смешением кодов (*пятнадцать сэ* ‘пятнадцать лет’).

Всего в окончательной выборке 758 вхождений количественной конструкции от 57 носителей.

При этом обнаруживается, что нестандартные употребления встретились только у 13 носителей из 57. Ниже подробно анализироваться будет подвыборка употреблений количественной группы только от этих носителей – всего 431 пример.⁸

3.2 Типы оформления количественной группы

Среди рассмотренных примеров зафиксированы все логически возможные типы оформления количественной конструкции – с генитивом и с номинативом, с единственным и множественным числом.

Два типа используют русскую модель с существительным в генитиве (в соответствии с правилами русского языка или нет):

- 1) GEN.PL (*двое детей, два коров*);
- 2) GEN.SG (*два пацана, двое пацанчика*);

Два типа с существительным в номинативе:

- 3) NOM.PL (*двое сыновья*);
- 4) NOM.SG (*два ведра*).

Ср. примеры на все типы:

- (8) *По два коров\ мы держали раньше* (gld, vsg)

⁸Все примеры с разметкой по исследуемым параметрам доступны в репозитории https://github.com/stoynova/tung_rus_numerals.

- (9) *Семья\ там уже. Двое пацанчика* (gld, vsg)
 (10) *А ведь двое\ сыновья – надо же и рыбачить* (ulc, oab)
 (11) *Он/ два ведро, я\ два ведро* (gld, vsg)

Далее мы будем говорить о «стандартных» употреблении количественной конструкции, совпадающих с ожидаемыми в русской речи монолингвов (*двое детей, два пацана*), и «нестандартных» – не характерных для русской речи монолингвов. Нестандартные употребления могут быть двух видов. Первый – использование не соответствующей правилам русской модели: генитив единственного числа вместо множественного и наоборот (*два коров, двое пацанчика*). Второй – использование номинатива вместо генитива (*двое сыновья, два ведро*).⁹

В употреблении с номинативом, как было сказано выше, можно подозревать структурную кальку (“pattern borrowing” в терминологии Sakel, 2007) с нанайского или ульчского, в которых числительное выступает в качестве неизменяемого приименного модификатора. Можно считать, что форма русского номинатива числительного в таких употреблениях выбирается в качестве дефолтной формы, эквивалентной неоформленному числительному в тунгусской модели.

Строго говоря, для большинства употреблений с номинативом/аккузативом нельзя по поверхностной форме определить, у какого из компонентов конструкции (существительного или числительного) номинатив/аккузатив приписан внешним синтаксическим контекстом, а у какого согласовательный или дефолтный. Это возможно только в тех немногочисленных случаях, когда аккузатив существительного формально не совпадает с номинативом, а именно для существительных *a*-склонения, если существительное употреблено в единственном числе.¹⁰ Если, в соответствии с тунгусской моделью, падеж приписан существительному, мы ожидаем примеров типа *две комнату*. Если, в соответствии с русской, числительному, то примеров типа *две комната*. В выборке употреблении с этим классом существительных единичны, но нашлись примеры на оба типа:

- (12) *Два титьку ки́сеха/ < ‘посол’ > – и наелась\ и всё* (gld, vsg)
 (13) *Три до́ска делают так на длину* (gld, vsg)

При этом употребления типа *три доска* в (13) естественнее трактовать все же не как случаи управления числительного номинативом существительного, а как случаи употребления номинатива существительного в объектной позиции (употребления типа *доска делают* в речи тех же носителей достаточно частотны, см. Stoynova, 2018).

Трактовку употреблений типа NUM.ном + N.ном как калькирующих тунгусскую модель с неизменяемым числительным в позиции приименного модификатора убедительно подтверждают нестандартные употребления в неноминативных контекстах типа (14). Они, однако, (как и в целом неноминативные употребления количественной группы) в нашей коллекции единичны (и в работе последовательно не рассматривались, см. п. 3.1).

- (14) *На четыре досках* (gld, vsg)

⁹Отклонения от речи монолингвов, не касающиеся выбора формы существительного, не учитывались. Так, например, сочетание *два ночи* далее рассматривается как стандартное, несмотря на нестандартный выбор формы рода числительного.

¹⁰Случай, когда аккузатив совпадает с генитивом, мы исключили из рассмотрения, наряду с собственно генитивом.

Таблица 3 Распределение нестандартных количественных конструкций по типам

тип	N	% от общего количества
GEN.SG (вместо GEN.PL)	5	4,59%
GEN.PL (вместо GEN.SG)	7	6,42%
NOM.SG (вместо GEN)	50	45,87%
NOM.PL (вместо GEN)	32	29,36%
NOM_uncount (вместо GEN) ^a	15	13,76%
всего GEN	12	11,01%
всего NOM	97	88,99%
всего	109	100,00%

^aПримеры с неисчисляемыми существительными (*много денег, столько жир*), для которых выбор единственного или множественного числа связан с индивидуальными особенностями существительного, были размечены отдельно

3.3 Распределение употреблений по типам

В общей выборке количественных конструкций нестандартных 14,38% (109/758). В подвыборке текстов от носителей, в речи которых засвидетельствована вариативность, нестандартных 25,29% (109/431). Ниже более подробные данные даются по этой подвыборке, Таблица 3.

Как видно из Таблицы 3, большинство нестандартных употреблений следуют номинативной модели (NUM + N.NOM):

- (15) *Ну пожилые\ были у нас, много\ пожилые* (gld, tmb)
 (16) *А сколько/ комната?* (gld, fna)
 (17) *Вот сколько добро\-то!* (gld, vsq)

Нестандартное использование генитивной модели представлено всего 12 примерами, при этом засвидетельствована мена модели в обе стороны – и использование ед.ч. генитива вместо мн.ч., и использование мн.ч. вместо ед.ч. Всем немногочисленным примерам можно дать правдоподобное индивидуальное объяснение – в терминах парадигматического или синтагматического влияния. Так, употребление ед.ч. *центнера* при числительном *двести* в (18) можно объяснить аналогией с более частотным числительным *два*, требующим единственного числа.

- (18) *Двести центнера* (gld, fna)

Мн.ч. *дней* при числительном *четыре* в (19) объясняется наличием сочиненного с ним числительного *пять*, требующего множественного числа.¹¹

- (19) *Дней четыре\ ли пять нул/* (gld, vsq)

Эти примеры ниже отдельно не рассматриваются: речь идет только о противопоставлении генитивной vs. номинативной стратегии оформления.

¹¹В литературном русском ожидалось бы *дня четыре или пять*. В этом примере можно усматривать случай нарушения не собственно синтаксиса количественной конструкции, а правил эллипсиса. Возможно, это вообще не контактно-обусловленное нестандартное употребление, а речевой сбой в ситуации сложного синтаксического контекста.

Таблица 4 (а) Выбор номинатива vs. генитива в зависимости от синтаксического типа числительного в русском языке монолингвов. **(б)** Выбор номинативной vs. генитивной модели в зависимости от морфосинтаксического класса числительного. **(в)** Выбор номинативной vs. генитивной модели в зависимости от семантического класса числительного

а				
синтаксический тип числительного в монолингвальном русском	NOM	GEN	% NOM	точный двусторонний критерий Фишера
тип GEN.SG	49	120	28,99%	значимо, p=0,0129
тип GEN.PL	48	214	18,32%	
всего	97	334	22,51%	
б				
класс числительного	NOM	GEN	% NOM	
большого количества (GEN.PL)		11	138	7,38%
собираательные (GEN.PL)		4	17	19,05%
местоимения (GEN.PL)		33	59	35,87%
паукальные (GEN.SG)		49	120	28,99%
всего		97	334	22,51%
в				
	NOM	GEN	% NOM	
местоимения	33	59	35,87%	
до четырех	53	130	28,96%	
от двадцати: круглые	4	36	10,00%	
пять – десять	7	76	8,43%	
одиннадцать – девятнадцать	0	24	0,00%	
от двадцати: некруглые	0	9	0,00%	
всего	97	334	22,51%	

3.4 Корреляции: формальные и семантические характеристики числительного

В Таблице 4а представлено распределение употреблений, следующих генитивной vs. номинативной стратегии, по синтаксическим типам числительных: требующие в русском языке монолингвов модели GEN.SG (паукальные: *два мальчика*) vs. требующие модели GEN.PL (все остальные: *пять / двое / несколько мальчиков*).

Числительные с ожидаемой моделью GEN.SG значительно чаще принимают вместо нее модель с номинативом, чем числительные с ожидаемой моделью GEN.PL.

Таблица 4б показывает распределение по более узким морфосинтаксическим классам числительных: паукальные (*оба, полтора, два, три, четыре, двадцать два* и под.) ~ собираательные (*двое, трое* и под.) ~ большого количества (*пять, шесть* и далее) ~ местоимения (*много, мало, сколько, столько* и их производные).

Данные Таблицы 4б свидетельствуют о том, что на самом деле не все числительные синтаксического типа GEN.PL ведут себя единообразно: наиболее заметно отличаются от типа GEN.SG (паукальных) числительные большого количества, в меньшей степени собираательные, а местоимения-числительные, наоборот, превосходят их по количеству употреблений, следующих номинативной модели. Это заставляет предположить, что наблюдаемые корреляции связаны с семантикой числительного, скорее чем с его формальным типом.

Таблица 5 Выбор номинативной vs. генитивной модели в зависимости от класса существительного

	NOM	GEN	% NOM
лица	31	55	36,05%
животные	7	13	35,00%
неодуш.: прочее	48	87	35,55%
мера	6	29	17,14%
время	5	150	3,23%
всего	97	334	22,51%

В Таблице 4в показано распределение примеров по семантическим классам числительного: до четырех (включая собирательные) ~ от пяти до десяти ~ от одиннадцати до девятнадцати ~ начиная с двадцати: круглые числа (*двадцать, двести* и др.) ~ прочие числительные начиная с двадцати (включая относящиеся к морфосинтаксическому классу паукальных: *двадцать два* и под.) ~ местоимения.

Данные Таблицы 4в подтверждают предположение о корреляции выбора номинативной модели с семантикой числительного. Наиболее склонны к нему местоимения-числительные и числительные с семантикой малого количества, в меньшей степени – другие числительные первого десятка и обозначающие «круглые числа», для остальных предпочитается генитив.

Вопрос о том, какой из рассмотренных взаимосвязанных параметров (синтаксический тип или семантика числительного) оказывается решающим при выборе номинативной модели, подробнее рассмотрен в п. 3.6.

3.5 Корреляции: семантика существительного

Таблица 5 показывает распределение количественных конструкций, оформленных генитивом vs. номинативом в зависимости от лексического класса существительного. Можно предположить, что, во-первых, на выбор модели будет влиять положение на шкале одушевленности, во-вторых, то, насколько данное существительное склонно к употреблению в количественной конструкции. В отдельные классы выделены: время (*год, час* и др., в этот же класс попало слово *раз*) ~ мера (прочие названия мер: *килограмм, рубль* и др., а также слова *пара* и *штука*) ~ прочие неодушевленные¹² ~ животные (*собака*) ~ лица (*жена*).

Из Таблицы 5 видно, что в количественных конструкциях с семантикой времени существительные оформляются преимущественно генитивом. Для существительных с семантикой меры процент номинативных употреблений также ниже, чем для прочих существительных, но неожиданно заметно выше, чем для обозначений времени. Это может быть связано с тем, что употребления типа *двести грамм* были размечены как генитивные (с особой счетной формой): возможно, часть из них на самом деле представляет собой употребления номинатива ед.ч. Если так, то количество генитивных употреблений для этого класса преувеличено. Заметных различий между названиями лиц, животных и неодушевленных существительных (кроме указанных групп) не наблюдается.

¹²Существительные, используемые в счетных конструкциях, но не являющиеся специфическими названиями мер (*ведро, бутылка, пачка, кусок* и под.), попали в этот класс.

3.6 Многофакторный анализ

Выявленные выше факторы, влияющие на выбор генитива vs. номинатива в конструкции с числительным: синтаксический тип числительного, семантика числительного и семантика существительного – логически не независимы. Синтаксический тип числительного определяется его семантикой: числительные, обозначающие малое количество, за исключением собирательных, требуют генитива ед.ч. Большинство числительных, обозначающих большое количество, за исключением некоторых составных (*сорок три*), — генитива мн.ч. Существительным с семантикой времени и меры свойственно встречаться в контексте числительного с семантикой большого количества, а, например, с собирательными числительными они не сочетаются вовсе. И наоборот, основной контекст употребления числительных с семантикой большого количества (особенно если речь не о числах первого десятка) – это как раз существительные со значением времени и меры: *сто литров, двадцать лет*. Чтобы проверить, какие из факторов в действительности определяют выбор между номинативом и генитивом, а какие показывают случайную корреляцию из-за связи с другими факторами, был проведен многофакторный статистический анализ данных.

Для анализа использовался метод биномиальной логистической регрессии (функция `glm()` в R), см. (Ваауен, 2008, р. 214–226; Gries, 2013, р. 293–310; Levshina, 2015, р. 253–276).¹³ В исходную модель были включены в качестве независимых переменных все три описанных выше потенциально релевантных фактора: синтаксический тип числительного (`num_type`), семантика числительного (`num_semantics`) и семантика существительного (`noun_semantics`). Зависимой переменной, для которой модель логистической регрессии оценивает шансы получить один из двух исходов, был, соответственно, тип синтаксического оформления: номинативный (NOM) или генитивный (GEN). Количество значений каждой независимой переменной было редуцировано до двух: переменные, их значения, распределение употреблений номинативного vs. генитивного синтаксического оформления для каждого значения независимой переменной показано в Таблице 6а.

В Таблице 6б приведены коэффициенты логистической регрессии для исходной модели. Строки соответствуют интересующим нас значениям каждой независимой переменной – тем, при которых ожидается склонность к оформлению номинативом. Строка “Intercept” соответствует значениям всех независимых переменных, выбранных в качестве т.н. “референциальных значений”: при них не ожидается склонность к оформлению номинативом (`num_type: gen_pl, num_semantics: large, noun_semantics: time&measure`). В столбце “Estimate” указан логарифм отношения шансов каждого из исходов (выбор тунгусского или русского оформления) для интересующих нас значений каждой независимой переменной. Отрицательная величина в этом столбце означает, что данное значение переменной располагает к выбору генитивного оформления, положительная величина – что оно располагает к выбору номинативного оформления, нулевая – что не располагает ни к тому, ни к другому выбору. В столбце “Std. Error” дано значение стандартной ошибки (высокие значения в этом столбце могут свидетельствовать о малом количестве данных или мультиколлинеарности, т.е. ситуации, когда несколько параметров разбивают данные на похожие группы); в столбце “z-value” – результат теста Вальда (высчитывается на основе значений из Estimate и Std. Error). Наконец, в последних двух столбцах указаны основные коэффициенты, показывающие значимость каждого из факторов, которые могут влиять на выбор номинатива vs.

¹³ Подробное описание кода доступно в репозитории https://github.com/stoynova/tung_rus_numerals.

Таблица 6 (а) Параметры, вошедшие в многофакторный анализ. (б) Логистическая регрессия: исходная модель. (в) Сравнение AIC (Akaike information criterion) для моделей с разным набором параметров. (г) Логистическая регрессия: оптимальная модель

а

переменная	значения	комментарий	NOM	GEN	всего	% NOM
num_type	gen_sg	требующие GEN.SG	49	120	169	28,99%
	gen_pl	требующие GEN.PL	48	214	262	18,32%
num_semantics	basic	до четырех и местоимения (сколько и др.)	86	189	275	31,27%
	large	остальные	11	145	156	7,05%
noun_semantics	non_time&measure	не время и мера	86	155	241	35,68%
	time&measure	время и мера	11	179	190	5,79%
всего			97	334	431	22,51%

б

	Estimate	Std. Error	z-value	Pr(> z)	Significance level
(Intercept)	-3,4598	0,3987	-8,677	<2e-16	*** (p<0,001)
num_type: gen_sg	0,1189	0,2793	0,426	0,6704	non-significant
num_semantics: basic	1,1229	0,3972	2,827	0,0047	** (p<0,01)
noun_semantics: non_time&measure	1,8683	0,3557	5,252	1,5e-07	*** (p<0,001)

в

модель	описание модели	AIC
numerals	исходная: с параметрами num_type, num_semantics, noun_semantics	392,9408
numerals1	оптимальная: с параметрами num_semantics, noun_semantics	391,1224
numerals2	с параметром noun_semantics	402,0939

г

	Estimate	Std. Error	z-value	Pr(> z)	Significance level
(Intercept)	-3,4419	0,3953	-8,706	<2e-16	*** (p<0,001)
num_semantics: basic	1,1956	0,3583	3,337	0,000847	*** (p<0,001)
noun_semantics: non_time&measure	1,8417	0,3503	5,257	1,46e-07	*** (p<0,001)

генитива: двустороннее p -значение, соответствующее данному z -значению при нормальном распределении, и его уровень значимости (большее количество звездочек соответствует большей уверенности в значимости фактора).

Так, в Таблице 6б отрицательное значение в столбце “Estimate” для строки (Intercept) означает, что когда числительное синтаксического типа GEN.PL и обозначает большое число, а существительное обозначает время или меру, то имеется тенденция к выбору генитивного оформления. Значения в последних двух столбцах показывают, что эта тенденция статистически значима. Положительные значения “Estimate” в остальных строках показывают, что при числительном синтаксического типа GEN.SG (паукальном), при числительном, обозначающем малое или неопределенное количество, и при существительном, не обозначающем меру или время, наблюдается тенденция к выбору номинативного оформления. Однако статистически значима эта тенденция

только для факторов “семантика числительного” и “семантика существительного”, но не “синтаксический тип числительного”.

Чтобы дополнительно проверить, какие из независимых переменных необходимы в модели логистической регрессии (т.е. какие из факторов вносят свой независимый вклад в выбор синтаксического оформления), был использован информационный критерий Akaike (AIC), см. см. (Gries, 2013, p. 266–267; Levshina, 2015, p. 266–267). Из нескольких моделей с разным набором независимых переменных оптимальной считается та, для которой на одной и той же выборке данных AIC принимает наименьшее значение. С помощью функции `drop1()` в R из исходной модели с тремя независимыми переменными переменные удалялись по одной: каждый раз удалялась та переменная, отсутствие которой давало наименьшее значение AIC (значения для трех сравниваемых моделей см. в Таблице 6в). Процедура проводится до тех пор, пока следующая модель дает меньшее значение AIC.

Оптимальной оказывается модель, показанная в Таблице 6г. Она содержит переменные `num_semantics` и `noun_semantics` (эффект обоих статистически значим), но не `num_type`. Ни одна из этих переменных не может быть удалена из модели, не ухудшая ее предсказательной силы.

Таким образом, многофакторный анализ показывает, что независимый значимый вклад при выборе номинативного vs. генитивного оформления существительного в количественной группе вносят лексический класс существительного (существительные с семантикой времени и меры предпочитают генитив) и лексический класс числительного (в группах с числительными с семантикой малого количества и местоимениями-числительными значимо чаще встречается существительное в номинативе). Синтаксический тип числительного отдельного эффекта не дает: отмеченная в п. 3.4 корреляция с ним, объясняется, видимо, связью этого параметра с семантикой числительного и отчасти с семантикой существительного.

3.7 Распределение между NOM.SG и NOM.PL

Оформление существительного в количественной конструкции номинативом встречается в двух разновидностях – с единственным и с множественным числом существительного (см. п. 3.2): ср. (20) и (21).

(20) *У нас тоже два собака* \ (gld, vsɡ)

(21) *Вот это дети ... трое* \ сыновья (ulc, oab)

Можно сформулировать два альтернативных предположения о том, что влияет на выбор между единственным и множественным числом. Первое – что на выбор влияет грамматика монолингвального русского, второе – что на него влияет грамматика южно-тунгусских языков. В соответствии с первым предположением, мы ожидаем единственного vs. множественного числа номинатива там же, где по правилам русского языка монолингвов было бы единственное vs. множественное число генитива. Если это так, то к единственному числу номинатива будут тяготеть конструкции с паукальным числительным. В соответствии со вторым предположением, мы ожидаем единственного vs. множественного числа там, где оно было бы на этом месте в южно-тунгусских языках. Если верно второе предположение, то к оформлению множественным числом будут тяготеть конструкции с одушевленным существительным и конструкции с местоимением-числительным (см. п. 2.2). Гипотезы и ожидания кратко сформулированы в Таблице 7а.

Таблица 7 (а) Распределение между NOM.SG и NOM.PL: русская vs. тунгусская гипотезы. (б) Выбор между NOM.SG и NOM.PL в зависимости от синтаксического типа числительного. (в) Выбор между NOM.SG и NOM.PL: одушевленность. (г) Выбор между NOM.SG и NOM.PL: местоимения-числительные vs. прочие числительные

а

	гипотеза	ожидание
«русская гипотеза»	распределение между ед.ч. и мн.ч. как в русском монолингвов	паукальное числительное => SG остальные классы => PL
«тунгусская гипотеза»	распределение между ед.ч. и мн.ч. как в южно-тунгусских	одушевленные существительные => PL местоимения-числительные => PL

б

тип числительного в русском монолингвов	NOM.SG	NOM.PL	% NOM.PL	точный двусторонний критерий Фишера
тип GEN.SG	36	13	26,53%	значимо, $p=0,006$
тип GEN.PL	14	19	57,58%	
всего	50	32	39,02%	

в

	NOM.SG	NOM.PL	% NOM.PL	точный двусторонний критерий Фишера
одуш.	10	28	73,68%	значимо, $p<0,0001$
неодуш.	40	4	9,09%	
всего	50	32	39,02%	

г

	NOM.SG	NOM.PL	% NOM.PL	точный двусторонний критерий Фишера
местоимения	5	13	72,22%	значимо, $p=0,002$
прочие	45	19	29,69%	
всего	50	32	39,02%	

В Таблицы 7б, 7в, 7г не включены примеры с неисчисляемыми существительными

В Таблице 7б дано распределение между множественным и единственным числом для разных синтаксических типов числительного.

Эти данные позволяют проверить русскую гипотезу. Если сравнивать распределение для паукальных числительных и для всех остальных, то действительно оказывается, что паукальные числительные, которые требуют единственного числа (генитива) в соответствии с русскими правилами, в речи билингвов больше других тяготеют к единственному числу номинатива существительного. При этом, как видно из Таблицы 7г ниже, числительные, требующие в литературном русском множественного числа (генитива) ведут себя неединообразно, что ожидалось бы, если бы распределение между SG и PL просто следовало русской модели. Возможно, выбор единственного числа связан не собственно с синтаксическим классом русского числительного, а с его се-

Таблица 8 (а) Выбор номинатива vs. генитива: уровень образования и год рождения. (б) Выбор номинатива vs. генитива: профиль носителя

а

параметр	значение	NOM	GEN	всего	% NOM	точный двусторонний критерий Фишера
образование	начальное	92	169	261	35,25%	значимо, $p < 0,0001$
	среднее	3	274	277	1,08%	
	высшее	2	218	220	0,91%	
год рождения	1917-1940	94	389	483	19,46%	значимо, $p < 0,0001$
	1941-1961	3	272	275	1,09%	
всего		97	661	758	12,80%	

б

носитель	% NOM (NOM GEN)	SyntIndex (ContSynt/tokens)	г.р.	образование	язык
vsg (жен.)	41,14% (72 103)	70506 (1567/22225)	1932	начальное	нан.
fna (жен.)	29,27% (12 29)	71549 (552/7715)	1930	начальное	нан.
oab (жен.)	19,05% (4 17)	40675 (217/5333)	1935	начальное	ульч.
aid (жен.)	0,00% (0 5)	11413 (42/3680)	1931	среднее специальное	ульч.
gip (жен.)	0,00% (0 13)	10395 (15/1443)	1934	среднее специальное	ульч.

мантикой (т.е. для числительных со значением малого количества более характерен выбор единственного числа), ср. п. 3.4.

В Таблицах 7в и 7г показано распределение по параметрам, отвечающим за выбор единственного vs. множественного числа в южно-тунгусских языках.

Таблицы 7в и 7г показывают, что, как и в южно-тунгусских языках, выбор формы множественного числа, значимо коррелирует и с одушевленностью существительного и с местоименным типом числительного.

В целом, таким образом, картина не вполне однозначная и требует более тщательной статистической проверки на большем объеме данных. На первый взгляд, имеются свидетельства в пользу обеих гипотез. Важно при этом, что вторая гипотеза (тунгусская) подтверждается совершенно отчетливо уже и на этой небольшой выборке: релевантными для выбора числа оказываются оба фактора, значимые для контактирующих с русским южно-тунгусских языков и нерелевантные для системы русского литературного языка.

3.8 Характеристики носителя

Как упоминалось выше (п. 3.1), немногие из носителей, речь которых представлена в корпусе, используют номинативное оформление количественной конструкции. Таблица 8а показывает, что использование номинатива в количественной конструкции коррелирует с годом рождения и уровнем образования: его используют преимущественно самые пожилые носители с начальным образованием.

По нашим наблюдениям, номинативное оформление характерно для тех носителей, речь которых в целом наиболее подвержена контактному влиянию тунгусских языков и наиболее отличается от речи монолингвов. Это требует дополнительной проверки. В Таблице 8б приведена информация о пяти носителях одного поколения (1930-е), которая отчасти подтверждает это наблюдение. Помимо общей социолингвистической

информации, в Таблице 8б дана количественная оценка того, насколько грамматика (синтаксис) данного носителя отличается от монолингвальной (столбец “SyntIndex”). Она подсчитана следующим образом: количество нестандартных синтаксических особенностей потенциально контактной природы, которые встретились в речи носителя (опущение предлогов, кодирование прямого объекта номинативом, нестандартная модель управления, нестандартный синтаксис отрицательных конструкций и проч., см. Khomchenkova et al., 2019) на миллион словоупотреблений. Подсчет синтаксических особенностей основан на ручной разметке, реализованной в корпусе для части коллекции; нестандартные количественные конструкции в подсчет не вошли.

Из Таблицы 8б видно, что (насколько можно судить по такому количеству данных) номинативная модель оформления количественной конструкции действительно характерна для речи носителей с наиболее нестандартным синтаксисом.

Все названные факторы: уровень образования, год рождения, общая нестандартность грамматики носителя – взаимосвязаны. Корреляции с характеристиками носителя требуют дальнейшего исследования на большем объеме данных, чем тот, что доступен на данный момент.

4 Результаты, обсуждение результатов

В этом разделе суммируются результаты анализа нестандартной количественной конструкции в русской речи нанайцев и ульчей. Они еще раз рассмотрены в контексте сформулированных выше гипотез о природе этих конструкций. Кратко повторим эти гипотезы еще раз:

- нестандартные количественные конструкции возникают в результате недоусвоенности / нестандартного усвоения системы литературного русского языка;
- копируют систему, представленную в южно-тунгусских языках;
- являются результатом действия более общих типологических тенденций и стоящих за ними универсальных когнитивных механизмов – можно ожидать, что они будут более отчетливо проявляться в устной речи двуязычных носителей, относительно свободной от давления литературной нормы;
- в неизменном виде наследуются из местного пиджина;
- не связаны непосредственно с контактным влиянием, а приходят из диалектной речи монолингвов.

4.1 Нестандартная количественная конструкция встречается у немногих носителей, но в их речи частотна

Рассмотренная в статье коллекция содержит образцы русской речи, в разной степени отклоняющиеся от речи монолингвов: можно сказать, что она отражает постпиджинный континуум. На одном его полюсе нестандартная речь с заметным контактным влиянием, близкая к мезолекту одной из разновидностей дальневосточного (сибирского) пиджина, который в базилектном варианте, видимо, можно считать вымершим. На другом полюсе – речь, мало отличающаяся от стандартной русской речи монолингвов.

Если располагать в этом континууме контактно-обусловленные грамматические явления, то нестандартные употребления количественной группы окажутся ближе к левому полюсу. Нестандартные употребления засвидетельствованы лишь у небольшой части носителей. Это довольно неожиданно, учитывая, насколько сложно и нетривиально устроена русская количественная конструкция.

Можно сравнить это с данными от носителей эритажного русского. Так, в (Polinsky, 2018, р. 90–95) показано, что в естественных текстах носители эритажного русского с доминирующим английским практически всегда используют стандартную количественную конструкцию, даже если в целом им свойственно нестандартное употребление падежных форм, в эксперименте на порождение также редко отступают от стандартного синтаксического образца (в работе также приводятся и обсуждаются ссылки на аналогичные результаты для носителей эритажного русского с доминирующим немецким).

Носители, использующие нестандартную количественную конструкцию, – это преимущественно самые пожилые носители с начальным образованием. По нашим наблюдениям, это как раз те носители, речь которых в целом наиболее отличается от речи монолингвов. При этом в речи таких носителей нестандартная количественная конструкция может быть довольно частотна (для некоторых ок. 30–40% всех употреблений). Тип носителей, для которых характерны нестандартные количественные конструкции, вписывается во все сформулированные выше предположения о природе этих конструкций. Это носители, в наименьшей степени и не с раннего детства имевшие доступ к литературному русскому языку (предположение о недоусвоенности русской системы). Они с большей вероятностью, чем более молодые поколения носителей, ориентирующиеся на диалектную речь, с одной стороны, и на пиджин, с другой. Наконец, это те носители, на речь которых в наименьшей степени влияет прескриптивная норма и, соответственно, в речи которых можно ожидать изменений, вызванных непосредственным действием универсально-языковых тенденций.

4.2 Использование номинативной модели, а не нестандартное использование генитивной

Основную массу примеров нестандартного употребления количественной конструкции составляют употребления, в которых и числительное, и существительное стоит в форме номинатива единственного или множественного числа (*три мальчик, три мальчишки*). Такие употребления можно интерпретировать как структурную кальку с нанайского/ульчского, в которых числительное выступает в качестве неизменяемого приименного модификатора. Можно считать, что форма русского номинатива числительного в таких употреблениях выбирается в качестве дефолтной формы, эквивалентной неоформленному числительному в тунгусской модели.

В свете предположения о том, что синтаксис количественной конструкции в речи билингвов будет стремиться к когнитивно мотивированному и/или типологически более распространенному образцу, можно отметить, что система тунгусского типа устроена проще и последовательнее русской.

Если бы синтаксис группы числительного в неизменном виде наследовался в рассматриваемую разновидность из пиджина или из диалекта монолингвов, мы бы ожидали последовательного употребления номинатива ед.ч. (в случае пиджина) или мн.ч. (в случае диалекта монолингвов).

В соответствии с предположением о недоусвоенности системы литературного русского, можно было бы ожидать нестандартных конструкций другого типа – несоответствующее правилам русского языка использование генитивной модели: генитив множественного числа вместо единственного или единственного вместо множественного (*три мальчиков*). Калькированием тунгусской модели такие употребления объяснить нельзя. Они в наших данных действительно есть, но очень редки, и всем таким примерам можно найти какое-либо частное объяснение. Наши данные заметно отличаются,

например, от данных детской речи, применительно к которой можно говорить именно о недоусвоенности. В речи детей довольно долго, примерно до 5 лет по данным (Gagarina & Voeikova, 2009, p. 199), фиксируется выбор генитива ед.ч. вместо генитива мн.ч. (*пять мешка*) и наоборот (*два мешков*).

4.3 Факторы, влияющие на выбор между номинативной и генитивной стратегиями: семантические, а не морфосинтаксические

Как показывает анализ корпусных данных, для выбора между номинативной и генитивной (характерной для литературного русского) стратегиями оформления количественной группы релеванты семантика существительного и семантика числительного. Стандартные употребления по русской генитивной модели значимо чаще встречаются в конструкциях с существительными, обозначающими время и меру, номинативная модель предпочтительна для прочих классов существительных. Номинативная модель значимо частотнее в конструкциях с числительными, обозначающими числа до пяти и местоимениями-числительными *сколько, столько, много, мало*. Синтаксический тип числительного (требующие в русском языке монолингвов GEN.SG vs. GEN.PL) отдельного влияния на выбор билингвами тунгусской vs. русской модели не оказывает.

Эти корреляции можно объяснить в рамках предположения о копировании тунгусской модели. Что касается семантики существительного, во-первых, существительные времени и меры – это те существительные, для которых контекст количественной конструкции наиболее характерен. Количественные конструкции с ними наиболее частотны в инпуте; некоторые из них, возможно, запоминаются как застывшие фразы. Отсюда меньшее отклонение от стандартного монолингвального образца для количественных конструкций с такими существительными. Во-вторых, время и мера относятся к кругу «русских реалий» и кругу тем, скорее обсуждаемых по-русски, чем по-нанайски/по-ульски. Примечательно, что значительная часть названий мер и отрезков времени в нанайском и ульском языках заимствована из русского (*час, минута, килограмм, метр* и под., традиционные нанайские/ульские меры длины, веса и под. в быту не используются, и их названия полностью или полузабыты). Схожим образом можно объяснить тенденцию к использованию тунгусской модели в конструкциях с числительными, обозначающими малое количество, и местоимениями-числительными: это те числительные, которые наиболее частотны, более широко используются в повседневном быту и раньше усваиваются. Числительные, обозначающие большое количество, начинают активно использоваться прежде всего в русскоязычной школе и используются в тех контекстах, которые более характерны для русской, а не нанайской/ульской речи.

Похожие факторы отмечались в литературе как релевантные при заимствовании числительных: ср. типологические шкалы заимствования числительных Я. Матраса (Matras, 2007, p. 51), (22).

- (22) легче заимствуются ↔ устойчивы к заимствованию
 а. numerals in more formal contexts > **in less formal contexts**
 б. higher numerals 1000, 100 > above 20 > above 10 > above 5 > **below 5**

Слева на шкале (22а) числительные, которые легче заимствуются, справа – более устойчивые к заимствованию. Слева на шкале (22б) контексты, в которых числительные легче заимствуются, справа – контексты, в которых они устойчивы к заимствованию. В нашем случае можно сказать, что левый конец этих же шкал легче подстраивается под монолингвальный русский синтаксический образец, а правый более устойчив

(или, скорее, инертен) в том смысле, что сохраняет влияние синтаксического образца родного языка.

В случае простого наследования синтаксической модели из пиджина специальных оснований ожидать именно такого распределения скорее нет: в базилекте пиджина последовательно используется номинативная модель. Впрочем, тот факт, что русская модель предпочитается в тех контекстах, для которых больше вероятность, что носители слышат их от монолингвов, важен и в этом случае.

В случае заимствования модели из диалекта монолингвов можно было бы ожидать разницы в поведении морфосинтаксических, а не семантических типов числительных: в русских говорах номинативную модель могут принимать те числительные, которые в литературном русском требуют генитива ед.ч., и собирательные. Апелляции к морфосинтаксическим типам, имеющимся в самой русской системе, мы бы скорее ожидали и при недоусвоенности/нестандартном усвоении системы литературного русского. Синтаксис русской конструкции с паукальными числительными (NUM + N.GEN.SG) кажется более сложным и непрозрачным, чем синтаксис конструкции с другими типами числительных. Эксперимент (Xiang et al., 2011) со взрослыми монолингвами показал, что для всех конструкций с числительными, кроме конструкции с паукальными числительными, время на обработку грамматичной конструкции меньше, чем время на обработку неграмматичной, а для конструкции с паукальными числительными время на обработку грамматичной конструкции выше и сопоставимо со временем на обработку неграмматичной. По мнению авторов, этот результат свидетельствует о том, что конструкция с паукальным числительным сложнее для восприятия. Исходя из этого, можно было бы предположить, что она будет вызывать большие трудности у билингвов и более активно заменяться на конструкцию тунгусского типа. Однако подобного эффекта в наших данных не наблюдается. Факторы, располагающие к частотности стандартных / нестандартных употреблений, универсально-семантические и не связаны непосредственно с морфосинтаксисом русской количественной группы.

Для предположения о формировании у билингвов системы, следующей универсальным тенденциям, релевантно типологическое обобщение о том, что в грамматических системах, где часть числительных проявляет свойства прилагательных, а часть – свойства существительных, на прилагательные по свойствам будут больше похожи числительные малого количества, а на существительные – числительные большого количества, но не наоборот (см. Corbett, 1978, где это обобщение впервые предложено как типологическая универсалия, и Hurford, 1987, p. 187–213, где оно уточняется и ему дается ряд прагматических объяснений). Например, в системе русского литературного языка (Corbett, 1978, p. 356–359) *один* синтаксически полностью ведет себя как прилагательное, *два* управляет числительным, как существительное, но согласуется с ним по роду, как прилагательное, для *два–четыре* можно говорить о согласовании по одушевленности, как у прилагательного, в остальном они ведут себя как существительные, числительные больше четырех проявляют все свойства существительных, кроме изменения по числу, а *тысяча*, *миллион* изменяются и по числам. Тот же самый универсальный принцип, но иначе проявляется на уровне частотной тенденции в речи тунгусских билингвов: числительные, обозначающие малые числа, склонны принимать номинативную модель с числительным-атрибутом (поведение прилагательного), числительные, обозначающие большие числа, – склонны управлять генитивом (как существительные).

4.4 Факторы, влияющие на выбор между единственным и множественным числом номинатива: те же, что в южно-тунгусских языках

Отдельно было проанализировано распределение между NOM.SG и NOM.PL в номинативных употреблениях количественной конструкции.

Как было сказано выше, вариативность в числе свидетельствует против предположений о наследовании модели из диалекта монолингвов (где ожидалось бы только мн.ч.) или из пиджина (тогда ожидалось бы только ед.ч.).

Частично подтверждается предположение о том, что форма числа существительного в нестандартной номинативной конструкции следует числовому распределению, которое задается стандартной русской генитивной конструкцией (NOM.PL в тех случаях, когда в русском монолингвов GEN.PL, NOM.SG в тех случаях, когда в русском монолингвов GEN.SG). Это как будто бы свидетельство в пользу гипотезы о нестандартном усвоении системы литературного русского языка.

Убедительно подтверждается предположение о том, что за распределение между формами единственного и множественного числа отвечают те же факторы, что и в южно-тунгусских языках. И в нанайском/ульчском, и в нестандартных количественных конструкциях в русской речи нанайцев/ульчей релевантным оказывается положение существительного на шкале одушевленности (больше мн.ч. для одушевленных) и семантика числительного (больше мн.ч. для местоимений-числительных). Повторение не просто синтаксической модели языка-источника, а наблюдаемой в нем вариативности – самый весомый аргумент в пользу гипотезы о копировании тунгусской модели. Особенно показательно, что в русской речи нанайцев и ульчей работают оба независимых фактора, релевантных для нанайского и ульчского.

При этом различие в маркировании числа в зависимости от позиции на шкале одушевленности – не уникальная черта южно-тунгусских языков, а типологическая устойчивый образец, ср. описание таких случаев в языках мира в (Smith-Stark, 1974; Corbett, 2000, p. 54–132). В (23) представлена типологическая шкала из (Corbett, 2000, p. 90): при дифференцированном маркировании существительного по числу левый край шкалы тяготеет к выражению множественным числом, а правый – к выражению единственным.

$$(23) \quad \text{PL} \leftrightarrow \text{SG} \\ \text{speaker} > \text{addressee} > \text{3rd person pronoun} > \text{kin} > \text{human} > \text{animate} > \text{inanimate}$$

В этом можно усматривать действие более общих когнитивных предпосылок. К оформлению множественным числом более склонны существительные с большей степенью индивидуализации ~ активности ~ топикальности ~ дискурсивной выделенности (см. обсуждение подобных шкал и их действия в разных фрагментах грамматики в Silverstein, 1976; Moravcsik, 1978; Hopper & Thompson, 1980; Kuno & Kaburaki, 1977; DeLancey, 1981; Malchukov, 2008; Croft, 2003, 130 и мн. др., а также обзоры в Filimonova, 2005; Haude & Witzlack-Makarevich, 2016).

Что касается оформления множественным числом существительного в конструкциях с местоимениями-числительными (*много дети*), эта тенденция, в отличие от первой, как кажется, не претендует на статус устойчивого типологического образца, но для нее тоже можно предложить, помимо объяснения через контактное влияние, правдоподобную прагматическую мотивацию. Важный контекст для местоимений типа *много* (в отличие от обычных числительных) – контекст неисчисляемых существительных. Более последовательное использование мн.ч. за его пределами позволяет более отчетливо развести исчисляемый vs. неисчисляемый контексты (т.е. противопоставить, например, значения типа ‘много рыб’ и ‘много рыбы’).

Таким образом, можно считать, что контактная мотивация здесь дополнительно поддерживается универсально-языковыми тенденциями.

4.5 Обобщение: копирование тунгусской модели и следование более общим типологическим тенденциям

В Таблице 9 суммируются изложенные выше аргументы в пользу разных предположений о природе нестандартных количественных конструкций в русской речи двуязычных нанайцев и ульчей. Более темной заливкой выделены более сильные аргументы в пользу каждой из гипотез.

Из Таблицы 9 видно, что больше всего подтверждений находит предположение о том, что в русской речи двуязычных носителей происходит копирование синтаксической модели южно-тунгусских языков. При этом копирование модели оказывается дополнительно подкреплено универсально-языковыми тенденциями.

5 Заключение

В работе были проанализированы случаи нестандартного синтаксического оформления количественной конструкции в русской речи двуязычных носителей южно-тунгусских языков.

В речи носителей южно-тунгусских языков в этом фрагменте грамматики наблюдается вариативность между стандартным генитивным (*два кувшина*) и номинативным оформлением (*два кувшин, два кувшины*), но при этом он организован достаточно системно, и можно выявить морфосинтаксические и семантические факторы, предсказывающие выбор между вариантами на уровне частотности.

Основной целью работы было проверить, можно ли интерпретировать номинативные количественные конструкции (*два кувшин, два кувшины*) как прямое копирование южно-тунгусской модели. В качестве альтернативного объяснения можно было бы предположить, что они наследуются в неизменном виде из базилекта пиджина, из диалектной речи монолингвов, являются результатом недоусвоения системы русского литературного языка или возникают под действием универсально-языковых механизмов, не связанных непосредственно с языковым контактом.

Корпусный количественный анализ вариативности отчасти дает возможность разграничить эти случаи. Наиболее убедительные аргументы обнаруживаются в пользу гипотезы о копировании тунгусской синтаксической модели. Особенно показателен тот факт, что вариативность между единственным и множественным числом номинатива (*два кувшин vs. два кувшины*) регулируется теми же факторами, что и аналогичная вариативность в южно-тунгусских языках. В Meyerhoff (2009) обсуждается возможность копирования вариативности, а не только категориальных грамматических черт из языка-источника применительно к креольским языкам (и их субстратным языкам, соответственно). Контактно-обусловленные варианты языка, близкие к речи монолингвов, такие как описываемый в работе тунгусский русский, дают не меньше материала для исследования копирования вариативности. По предположению М. Мейерхофф, при копировании в язык-реципиент вариативность претерпевает ту или иную трансформацию (меняется сам набор регулирующих ее факторов, их сила или направление действия). Интересно было бы проверить на большем объеме данных, так ли это в нашем случае. Предварительный анализ, приведенный в статье, показывает, что

Таблица 9 Гипотезы об источниках нестандартных количественных конструкций в русской речи носителей южно-тунгусских языков: обобщение

	недоусвоенность/нестандартное усвоение системы литературного русского	копирование тунгусской системы	система, следующая типологическим тенденциям/мотивированная общими когнитивными механизмами	наследование системы из пиджина	наследование системы из диалекта монолингвов
социолингвистические факторы	да: нестандартная количественная конструкция более всего характерна для двуязычных носителей самого старшего поколения, с начальным образованием, с самым нестандартным русским				нет: нет свидетельств о влиянии конкретного диалекта с нестандартной количественной конструкции на всей описанной территории
	носители, имеющие минимальный доступ к русскому языку	носители, для которых нанайский/ульчский первый и основной язык	носители, речь которых минимально подвергается давлению нормы	ближайшее поколение к носителям пиджина	носители с доступом скорее к диалектному, чем к литературному русскому
тип нестандартной конструкции	нет/да: использование GEN.SG вместо GEN.PL и наоборот есть, но очень редко	да: использование NOM.SG и NOM.PL, как в южно-тунгусских	да/нет: номинативная модель => единообразное поведение числительных в прямых и косвенных падежах	да/нет: использование не только NOM.SG, как в пиджине, но и NOM.PL	да/нет: использование не только NOM.PL, как в русских говорах, но и NOM.SG
факторы, располагающие к выбору номинативной/генитивной модели	нет: морфосинтаксический тип числительного не влияет	да: генитивная модель в конструкции с большими числами и названиями мер и отрезков времени (т.е. в контекстах, характерных для речи по-русски), номинативная в остальных (т.е. в контекстах, характерных для речи по-нанайски/по-ульчски)	да/нет: малые числа => числительные ведут себя как прилагательные, большие числа => числительные ведут себя как существительные (типологическая тенденция)	да/нет: генитивная модель в тех контекстах, которые чаще можно услышать от монолингвов; при этом в базилекте пиджина последовательно номинативная модель	нет: морфосинтаксический тип числительного не влияет (в диалектах для паукальных и собирательных)

Таблица 9 (Continued)

факторы, отвечающие за распределение NOM.SG и NOM.PL	да/нет: корреляция с синтаксическим типом числительного (требуется проверка)	да: местоимения-числительные и одушевленность	нет	нет
		как в южно-тунгусских		

для русской речи билингов факторы, релевантные для южно-тунгусских языков, оказываются, видимо, не единственными.

Описанные данные демонстрируют в разных аспектах устройства количественной группы не просто следование структурному образцу южно-тунгусских языков, но в то же время и следование типологически более распространенному и функционально мотивированному образцу (и можно, видимо, сказать, что более простому в восприятии и усвоении). Рассмотренного материала недостаточно для того, чтобы говорить, что в нем прослеживается жесткая закономерность типа «структурный образец копируется в том случае, если он оказывается типологически более распространенным / когнитивно более мотивированным / более простым в усвоении / требующим меньших усилий на порождение и/или восприятие». Однако это наблюдение кажется неслучайным и релевантным в контексте более общей проблематики структурных заимствований. Оно согласуется, в частности, с идеей (в общем виде высказанной, например, в Thomason, 2009) о том, что контактные изменения, как правило, имеют то же направление, что и естественноразовые, оба типа изменений регулируются одними и теми же функциональными мотивациями, не всегда различимы и, более того, не всегда оправданы попытки их различить.

Сокращения

- 1, 3 – 1, 3 лицо
- ACC – аккузатив
- DAT – датив
- GEN – генитив
- N – существительное
- NOM – номинатив
- NUM – числительное
- PL – множественное число
- PST – прошедшее время
- PTCL – частица
- SG – единственное число

Благодарности Автор выражает признательность анонимным рецензентам журнала *Russian Linguistics*, С. А. Оскольской, И. А. Хомченковой, П. С. Плешак, участникам Циркумпольного проекта в ИЯЗ РАН и участникам конференции «Русский язык в многоязычном мире – 2019», а главное, всем своим информантам – носителям нанайского и ульчского языков.

Вклад авторов Неприменимо.

Финансовая поддержка Работа подготовлена при поддержке гранта РФФИ No 17-18-01649.

Доступность данных Корпус, использованный в исследовании, доступен по адресу <http://web-corpora.net/ruscontact/>.

Доступность материалов Таблицы с выборкой размеченных примеров доступны в репозитории https://github.com/stoynova/tung_rus_numerals (tung_rus_numerals.csv, tung_rus_numerals_all_data.csv).

Доступность кода Описание кода в R, использованного для статистического анализа данных, Таблицы с выборкой размеченных примеров доступны в репозитории https://github.com/stoynova/tung_rus_numerals (tung_rus_numerals_R_code.rmd, tung_rus_numerals_R_code.pdf).

Этические декларации/Дополнительные декларации для статей, в которых сообщаются результаты исследований с участием людей и/или животных

Конфликт интересов Исследование является частью работы в Институте языкознания РАН по гранту РНФ No 17-18-01649.

Этическое одобрение Неприменимо.

Согласие на участие Рассказчики, речь которых анализировалась в исследовании, дали согласие на аудиозапись и ее использование в научных целях, в т.ч. в научных публикациях (в анонимизированном виде).

Согласие на публикацию Рассказчики, речь которых анализировалась в исследовании, дали согласие на аудиозапись и ее использование в научных целях, в т.ч. в научных публикациях (в анонимизированном виде).

Литература

- Аванесов, Р. И., & Орлова, В. Г. (1964). *Русская диалектология*. Москва: Наука.
- Аврорин, В. А. (1959). *Грамматика нанайского языка*. Т. I. Москва/Ленинград: Изд-во Академии наук СССР.
- Герасимова, А. Н. (2002). Нанайский и ульчский языки в России: сравнительная характеристика социолингвистической ситуации. *Языки коренных народов Сибири*, 12, 246–257.
- ДАРЯ – Мораховская, О. Н. (Отв. ред.) (1996). *Диалектологический атлас русского языка. Центр Европейской части СССР. Выпуск III: Синтаксис. Лексика*. Москва: Наука.
- Земская, Е. А. (1988). Городская устная речь и задачи ее изучения. В Д. Н. Шмелев & Е. А. Земская (Отв. ред.), *Разновидности городской устной речи* (с. 5–43). Москва: Наука.
- Калинина, Е. Ю., & Оскольская, С. А. (2016). Нанайский язык. В В. Ю. Михальченко (Гл. ред.), *Язык и общество. Энциклопедия* (с. 293–296). Москва: Азбуковник.
- Мельчук, И. А. (1985). *Поверхностный синтаксис русских числовых выражений* (Wiener Slawistischer Almanach. Sonderband, 16). Wien.
- Перехвальская, Е. В. (2006). *Сибирский пиджин: дальневосточный вариант: формирование. История. Структура*. Автореферат диссертация д.ф.н., Санкт-Петербург: СПбГУ.
- Перехвальская, Е. В. (2008). *Русские пиджины*. Санкт-Петербург: Алетейя.
- Сичинава, Д. В. (2018). Числительные. В *Материалы к корпусной грамматике русского языка. Выпуск III. Части речи и лексико-грамматические классы* (с. 193–297). Санкт-Петербург: Нестор-История.
- Сумбатова, Н. Р., & Гусев, В. Ю. (2016). Ульчский язык. В В. Ю. Михальченко (Гл. ред.), *Язык и общество. Энциклопедия* (с. 513–515). Москва: Азбуковник.
- Суник, О. П. (1985). *Материалы по ульчскому языку и фольклору*. Ленинград: Наука.
- Циммерлинг, А. В. (2019). Корпусная грамматика количественных групп в русском языке. In *Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2019»* (с. 742–760). Москва: РГГУ.
- Aikhenvald, A. Y. (2007). Grammars in contact: a cross-linguistic perspective. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *Grammars in contact. A cross-linguistic typology* (pp. 1–66). Oxford: Oxford University Press.
- Vaayen, H. R. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bailyn, J. F., & Nevins, A. (2008). Russian genitive plurals are impostors. In A. Bachrach & A. Nevins (Eds.), *Inflectional identity* (pp. 237–270). Oxford: Oxford University Press.

- Bickerton, D. (1973). On the nature of the creole continuum. *Language*, 49(3), 640–669. <https://doi.org/10.2307/412355>.
- Bickerton, D. (1980). Decreolisation and the creole continuum. In A. Valdman & A. R. Highfield (Eds.), *Theoretical orientations in creole studies* (pp. 109–127). New York: Academic Press.
- Chambers, J. K. (2004). Dynamic typology and vernacular universals. In B. Kortmann (Ed.), *Dialectology meets typology: dialect grammar from a cross-linguistic perspective* (pp. 127–145). Berlin/New York: Mouton de Gruyter.
- Corbett, G. G. (1978). Universals in the syntax of cardinal numerals. *Lingua*, 46, 355–368.
- Corbett, G. G. (1993). The head of Russian numeral expressions. In G. G. Corbett, N. M. Fraser, & S. McGlashan (Eds.), *Heads in grammatical theory* (pp. 11–35). Cambridge: Cambridge University Press.
- Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.
- Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge: Cambridge University Press.
- De Camp, D. (1971). Towards a generative analysis of a post-creole speech continuum. In D. Hymes (Ed.), *Pidginization and creolization of languages, proceedings of a conference held at the University of the West Indies, Mona, Jamaica, April 1968* (pp. 349–370). Cambridge: Cambridge University Press.
- DeLancey, S. (1981). An interpretation of split ergativity and related patterns. *Language*, 57, 626–657. <https://doi.org/10.2307/414343>.
- Deuber, D. (2006). Aspects of variation in educated Nigerian Pidgin: verbal structures. In A. Deumert & St. Durrleman-Tame (Eds.), *Structure and variation in language contact* (pp. 243–326). Amsterdam/Philadelphia: John Benjamins.
- Filimonova, E. (2005). The noun phrase hierarchy and relational marking: problems and counterevidence. *Linguistic Typology*, 9, 77–113. <https://doi.org/10.1515/lity.2005.9.1.77>.
- Franks, S. (2002). A Jakobsonian feature-based analysis of the Slavic numeric quantifier genitive. *Journal of Slavic Linguistics*, 10, 141–181.
- Gagarina, N., & Voekova, M. D. (2009). Acquisition of case and number in Russian. In U. Stephany & M. D. Voekova (Eds.), *Development of nominal inflection in first language acquisition: a cross-linguistic perspective* (pp. 179–216). Berlin/New York: Mouton de Gruyter.
- Gries, St. Th. (2013). *Statistics for linguistics with R. A practical introduction* (2nd revised ed.). Berlin/New York: Mouton de Gruyter.
- Haude, K., & Witzlack-Makarevich, A. (2016). Referential hierarchies and alignment: an overview. *Linguistics*, 54(4), 433–441. <https://doi.org/10.1515/ling-2016-0008>.
- Heine, B., & Nomachi, M. (2013). Contact-induced replication. Some diagnostics. In M. Robbeets & H. Cuyckens (Eds.), *Shared grammaticalization: with special focus on the Transeurasian languages* (pp. 67–100). Amsterdam/Philadelphia: John Benjamins.
- Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 56, 251–299. <https://doi.org/10.2307/413757>.
- Hurford, J. (1987). *Language and number: the emergence of a cognitive system*. Oxford: Blackwell.
- Ionin, T., & Matushansky, O. (2018). *Cardinals: the syntax and semantics of cardinal-containing expressions*. Cambridge/London: MIT Press.
- Khomchenkova, I., Pleshak, P., & Stoyanova, N. (2019). The corpus of contact-influenced Russian of Northern Siberia and the Russian Far East. In *Computational linguistics and intellectual technologies: proceedings of the international conference "Dialogue 2019"* (pp. 276–287). Moscow: RSUH.
- Koptjevskaja-Tamm, M., & Wälchli, B. (2001). The circum-Baltic languages: an areal-typological approach. In Ö. Dahl & M. Koptjevskaja-Tamm (Eds.), *Circum-Baltic languages. Vol. II. Grammar and typology* (pp. 615–750). Amsterdam/Philadelphia: John Benjamins.
- Kuno, S., & Kaburaki, E. (1977). Empathy and syntax. *Linguistic Inquiry*, 8(4), 627–672.
- Léglise, I. & Chamoreau, C. (Eds.). (2013). *The interplay of variation and change in contact settings*. Amsterdam/Philadelphia: John Benjamins.
- Levshina, N. (2015). *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins.
- Madariaga, N., & Igartua, I. (2017). Idiosyncratic (dis)agreement patterns: the structure and diachrony of Russian paucal subjects. *Scando-Slavica*, 63(2), 99–132. <https://doi.org/10.1080/00806765.2017.1390922>.
- Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking. *Lingua*, 118, 203–221. <https://doi.org/10.1016/j.lingua.2007.02.005>.
- Matras, Y. (2007). The borrowability of structural categories. In Y. Matras & J. Sakel (Eds.), *Grammatical borrowing in cross-linguistic perspective* (pp. 31–73). Berlin: Mouton De Gruyter.
- Meyerhoff, M. (2009). Replication, transfer and calquing: using variation as a tool in the study of language contact. *Language Variation and Change*, 21, 1–21. <https://doi.org/10.1017/S0954394509990196>.
- Meyerhoff, M. (2021). Variation in pidgin and creole languages. In U. Ansaldi & M. Meyerhoff (Eds.), *The Routledge handbook of Pidgin and Creole languages* (pp. 348–362). London/New York: Routledge.

- Mikaelian, I. (2013). Cardinal numeral constructions and the category of animacy in Russian. *Russian Linguistics*, 37(1), 71–90.
- Moravcsik, E. (1978). On the distribution of ergative and accusative patterns. *Lingua*, 45, 233–279. [https://doi.org/10.1016/0024-3841\(78\)90026-8](https://doi.org/10.1016/0024-3841(78)90026-8).
- Mühlhäusler, P. (1997). *Pidgin and creole linguistics* (2nd ed.). London: University of Westminster Press.
- Mustajoki, A. (2010). Types of non-standard communication encounters with special reference to Russian. In M. Lähteenmäki & M. Vanhala-Aniszewski (Eds.), *Language ideologies in transition: multilingualism in Russia and Finland* (pp. 35–56). Frankfurt am Main etc.: Peter Lang.
- Neset, T., & Nordrum, M. (2019). Do Russian paucal numerals govern the genitive? Evidence from stress placement. *Russian Linguistics*, 43, 87–105. <https://doi.org/10.1007/s11185-019-09209-7>.
- Patrick, P. (2008). Pidgins, creoles, and variation. In S. Kouwenberg & J. V. Singler (Eds.), *The handbook of pidgin and creole studies* (pp. 461–488). Hoboken (New Jersey): Wiley-Blackwell.
- Pereltsvaig, A. (2011). As easy as two, three, four? In W. Browne et al. (Eds.), *Formal approaches to slavic linguistics (FASL-18). The second Cornell meeting 2009* (Michigan slavic materials, 56, pp. 417–434). Ann Arbor: University of Michigan.
- Polinsky, M. (2018). *Heritage languages and their speakers*. Cambridge: Cambridge University Press.
- Poplack, Sh. (2021). A variationist perspective on language contact. In E. Adamou & Y. Matras (Eds.), *The routledge handbook of language contact* (pp. 46–62). London/New York: Routledge.
- Poplack, Sh., & Levey, St. (2010). Contact-induced grammatical change: a cautionary tale. In P. Auer & J. E. Schmidt (Eds.), *Language and space: an international handbook of linguistic variation. Volume I: theories and methods* (pp. 391–419). Berlin/New York: Mouton De Gruyter.
- Rappaport, G. C. (2002). Numeral phrases in Russian: a minimalist approach. *Journal of Slavic Linguistics*, 10(1–2), 327–340.
- Rickford, J. (1987). *Dimensions of a creole continuum*. Stanford: Stanford University Press.
- Sakel, J. (2007). Types of loan: matter and pattern. In Y. Matras & J. Sakel (Eds.), *Grammatical borrowing in cross-linguistic perspective* (pp. 15–29). Berlin/New York: Mouton De Gruyter.
- Schwegler, A. (2000). The myth of decreolization. The anomalous case of Palenquero. In I. Neumann-Holzschuh & E. W. Schneider (Eds.), *Degrees of restructuring in creole languages* (pp. 409–436). Amsterdam/Philadelphia: John Benjamins.
- Sebba, M. (1997). *Contact languages. Pidgins and creoles*. London: Macmillan.
- Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). New Jersey: Humanities Press.
- Smith-Stark, T. C. (1974). The plurality split. In M. W. La Galy, R. A. Fox, & A. Bruck (Eds.), *Papers from the tenth regional meeting, Chicago Linguistic Society, April 19–21* (pp. 657–671). Chicago: Chicago Linguistic Society.
- Stolz, Th. (2002). Is 'one' still 'one' in 'twenty-one'? On agreement and government properties of cardinal numerals in the languages of Europe. *STUF*, 55(4), 354–402.
- Stoynova, N. (2018). Differential object marking in contact-influenced Russian speech: evidence from the corpus of contact-influenced Russian speech of Russian Far East and Northern Siberia. In *Computational linguistics and intellectual technologies, 17. Papers from the annual international conference "Dialogue"* (pp. 721–734). Moscow: RSUH.
- Stoynova, N. (2019). Russian in contact with southern Tungusic languages: evidence from the contact Russian corpus of Northern Siberia and the Russian Far East. *Slavica Helsingiensia*, 52, 9–36.
- Thomason, S. G. (2009). Why universals versus contact-induced change. In M. Filppula, J. Klemola, & H. Paulasto (Eds.), *Vernacular universals and language contacts: evidence from varieties of English and beyond* (pp. 349–364). London: Routledge.
- Xiang, M., Harizanov, B., Polinsky, M., & Kravtchenko, E. (2011). Processing morphological ambiguity: an experimental investigation of Russian numerical phrases. *Lingua*, 121(3), 548–560. <https://doi.org/10.1016/j.lingua.2010.10.016>.
- Yanko, T. (2004). Russian numerals with nouns denoting human beings. *General Linguistics*, 43(1–4), 61–84.