

УДК 81'276

РУССКИЙ ЯЗЫК ПОВСЕДНЕВНОГО ОБЩЕНИЯ: ГРАММАТИЧЕСКИЙ АСПЕКТ¹

Наталья Викторовна Богданова-Бегларян

д. филол. н., профессор кафедры русского языка

Санкт-Петербургский государственный университет

199034, г. Санкт-Петербург, Университетская наб., 7/9. n.bogdanova@spbu.ru

Ольга Владимировна Блинова

к. филол. н., доцент кафедры общего языкознания

Санкт-Петербургский государственный университет

199034, г. Санкт-Петербург, Университетская наб., 7/9. o.blinova@spbu.ru

Григорий Яковлевич Мартыненко

д. филол. н., профессор кафедры математической лингвистики

Санкт-Петербургский государственный университет

199034, г. Санкт-Петербург, Университетская наб., 7/9. g.martynenko@spbu.ru

Татьяна Юрьевна Шерстинова

к. филол. н., доцент кафедры математической лингвистики

Санкт-Петербургский государственный университет

199034, г. Санкт-Петербург, Университетская наб., 7/9. t.sherstinova@spbu.ru

Статья посвящена описанию отдельных особенностей устной речи у представителей различных социальных групп носителей современного русского языка. Исследование выполнено на материале аннотированной части корпуса «Один речевой день», который содержит звукозаписи повседневной коммуникации информантов, принадлежащих к разным гендерным, возрастным, профессиональным и статусным группам. Обсуждаются признаки, которые характеризуют речь представителей разных социальных групп. Согласно полученным данным, существуют инвариантные свойства устной русской речи, в равной мере присущие представителям этих групп. Среди них распределение частей речи, а также наиболее частотных синтаксических структур и синтаксических нерегулярностей. Кроме того, существует значительный пласт «потенциально инвариантных» характеристик устной русской речи: эти характеристики на проанализированном речевом материале показали некоторые количественные различия у разных социальных групп, однако степень этих различий не позволяет на настоящий момент считать их статистически достоверными. Наконец, можно выделить ряд характеристик речи, которые присущи отдельным социальным слоям, т. е. являются диагностическими с точки зрения социолингвистики.

Ключевые слова: русский язык; устная речь; повседневная речь; речевой корпус; грамматика; морфология; синтаксис.

Введение

Поиск специфических черт в речи представителей различных групп носителей современного русского языка и описание этих различий продолжают оставаться актуальными задачами лингвистики. В настоящей статье эта задача решается на материале большого корпуса «Один речевой день» (ОРД), создаваемого в Санкт-Петербургском государственном университете (подробнее о нем см. [Богданова-Бегларян и др.

2015, 2017а, 2017б; Русский язык повседневного общения... 2016]). Исследование выполняется на материале морфологически и синтаксически аннотированной части корпуса ОРД.

Сегодня ОРД содержит звукозаписи повседневной коммуникации 130 информантов, принадлежащих к разным гендерным, возрастным, профессиональным и статусным группам [Bogdanova-Beglarian et al., 2017]. Корпус достиг размера в 1 250 часов звучания, насчитывает

более 1 млн словоупотреблений в текстовых расшифровках.

Для индикации принадлежности информантов к различным социальным группам в корпусе используется многоуровневая классификация, включающая обозначения для двух гендерных, десяти профессиональных, пяти статусных и трех возрастных групп².

1. Характеристика аннотированного подкорпуса

Аннотированный подкорпус ОРД объемом в 125 тыс. словоупотреблений содержит речь 100 основных информантов и 154 их собеседников (коммуникантов). Размер общей выборки составляет 125 437 слов, из них 47 135 – из речи мужчин, 78 302 – из речи женщин. Возрастные группы представлены подкорпусами соответственно в 46,3 тыс. слов (младшая), 51,4 тыс. слов (средняя) и 23,5 тыс. слов (старшая). Кроме того, в подкорпусе представлена речь коммуникантов-детей, которая не была предметом специального анализа.

Морфологическая разметка проводилась автоматически с помощью морфоанализатора TreeTagger с последующей ручной коррекцией.

Кроме частеречных помет³ и помет, отражающих информацию о словоизменительных характеристиках словоформ, учитывалось наличие в речи информантов аграмматичных и ненормативных форм. Такие формы размечались в транскрипте вручную и маркировались расшифровщиками знаком (*). Синтаксическая разметка также выполнена вручную, ее параметры подробно описаны в [Богданова-Бегларян и др., 2017a].

2. Морфологический анализ

2.1. Встречаемость частеречных классов

Анализ частотных списков, фиксирующих абсолютное и относительное количество употреблений (в % относительно общего количества употреблений в аннотированном подкорпусе), показал, что самые частотные части речи в составе подкорпуса – это глагол (17,43%), существительное (15,29%), местоимение-существительное (14,13%), частица (13,35%) и союз (9,47%). При этом различия в употреблении слов различных классов в речи представителей разных социальных групп в целом незначительны (см. табл. 1).

Таблица 1

Встречаемость слов разных частеречных классов в речи представителей различных социальных групп⁴, %

Часть речи	Гендер				Возраст					
	Муж.		Жен.		age1		age1		age1	
	%	Ранг	%	Ранг	%	Ранг	%	Ранг	%	Ранг
V	17,44	1	17,42	1	17,21	1	17,65	1	17,33	1
S	15,68	2	15,06	2	14,86	2	15,56	2	15,30	2
S-PRO	13,31	4	14,63	3	14,35	3	13,75	3	14,21	3
PART	13,40	3	13,32	4	13,80	4	12,79	4	13,83	4
CONJ	9,19	5	9,64	5	9,31	5	9,31	5	10,24	5
PR	7,25	6	7,00	6	6,99	6	7,19	6	7,17	6
ADV-PRO	5,68	7	5,08	7	5,32	7	5,52	7	4,88	7
ADV	4,41	8	4,57	8	4,64	8	4,53	8	4,42	9
A-PRO	4,29	9	4,31	9	4,14	9	4,34	9	4,52	8
A	3,92	10	3,61	10	3,72	10	3,93	10	3,41	10

Мужчины несколько чаще употребляют частицы (в их речи частица имеет 4 ранг, местоимение-существительное – 5), **женщины** несколько чаще употребляют местоимения-прилагательные (в их речи порядок рангов обратный). Числительные более характерны для мужской речи, междометия – для женской. Видно, таким образом, что верхние зоны частотных списков, учитывающих употребительность частей речи,

у мужчин и женщин различаются весьма незначительно.

Ранговое распределение частей речи для верхней зоны частотных списков в речи представителей трех рассматриваемых **возрастных** групп также практически идентично. Различия наблюдаются в рангах наречия и местоимения-прилагательного: у младших и средних по возрасту говорящих наречие и местоимение-

прилагательное имеют 8 и 9 ранг соответственно. У старших картина обратная. При этом относительные количества употреблений упомянутых частей речи различаются незначительно – на десятки доли процента.

Интересно, что материал речи попавших в выборку детей не показывает никаких существенных отличий от речи взрослых говорящих – ни в отношении относительных цифр, отражающих употребительность слов различных частеречных классов, ни в отношении их рангового распределения. Различия в основном таковы: в речи детей не встретилось ни одного местоимения-предикатива; дети в три раза реже, чем взрослые, употребляют вводные слова, несколько чаще употребляют числительные-прилагательные (0,43%). Впрочем, в рабочем подкорпусе на детскую речь приходится лишь 4200 слов. Для обоснованных выводов относительно детской речи требуется расширение размеров подкорпуса.

В речи всех выделенных **профессиональных групп** глагол имеет первый ранг. Исключение составляет группа «естественников», у которых наиболее употребительной частью речи оказалось существительное. Верхние зоны частотных списков, составленных для различных профессиональных групп, демонстрируют в том числе такие различия: у офисных работников местоимение-существительное имеет более высокий ранг (2), чем существительное (3). При этом можно заметить, что относительные величины, отражающие употребительность, различаются лишь на десятые доли процента: S-PRO (14,97%), S (14,63%). В речи инженеров и рабочих частица имеет более высокий ранг, чем в речи других профессиональных групп (3, а не 4, как у остальных).

Сравнительно многочисленные различия относительно общей картины по подкорпусу наблюдаются в распределении рангов частей речи только у представителей творческих специальностей. Так, нижняя часть списка «топ-10 частей речи» для этой группы выглядит следующим образом: наречие (7), местоимение-наречие (8), прилагательное (9), местоимение-прилагательное (10).

Сколько-нибудь заметные различия в употребительности слов разных частеречных классов в речи представителей разных социально-профессиональных групп также наблюдаются лишь при анализе рангового распределения частеречных тегов в частотном списке. Например, обращает на себя внимание более высокий (3) ранг частицы у неработающих и пенсионеров, относительно более высокий (9) ранг прилагательного

в речи бизнесменов, относительно более высокий ранг местоимения-прилагательного в речи руководителей.

2.2. Встречаемость аграмматичных, ненормативных, «редких», «сложных» и других форм

Как было указано выше, в транскриптах ОРД вручную размечены ненормативные и аграмматичные формы. Таких форм в составе аннотированного подкорпуса немного. При этом все употребления можно разделить на две категории: во-первых, это «настоящие» аграмматичные формы, которые говорящие употребляют в силу речевых привычек, неумения образовать корректные формы склонения или спряжения; во-вторых, это «игровые» формы, которые говорящие употребляют, намеренно нарушая известные им языковые правила.

Языковая игра с намеренно неправильными формами склонения и спряжения свойственна прежде всего речи младших говорящих, например:

- *с(:) этим () с пальтом** (Ж, 25)⁵.

См. также ироничное высказывание говорящего младшего возраста, противопоставляющее методы лечения молодого и старшего поколений:

- *даже даже мелкого там (м-м) бабушки / кричали (...) слезьми* обливались / типо чего ж вы ребёночку дайте парацетамольчик* (S109, М, 28, инженер, специалист).

«Игровые» употребления грамматических форм часто сопровождаются смехом, при этом собеседники продолжают начатую языковую игру, например:

- *я боюся* // да / боюся* я // *С* (S34, М, 77, творческий работник, специалист);
- (S67, Ж, 67) *без житонов* (sic! ударение на первом слоге) *док... док... доклады* мне //* (общий смех), (F2, коммуникант, смеясь) *всё! Элеонора% сбила с толку* (F1, коммуникант) *пост номер три / постовой Букварёва% пост сдала.*

В то же время ясно, что если в речи говорящего встречаются элементы просторечия и неединичные показатели низкого уровня речевой культуры, то повторяющиеся и ненормативные формы, скорее всего, не следует относить к «игровым»:

- *у ней* будет паспорт в субботу, <...> тоже у ней* иногда что-то мы не понимаем да?* (S67, Ж, 67, работник сферы обслуживания, специалист);
- *<...> да на девятьсот с чем-то у ней* Виктор Петрович%* (S67, Ж, 67, работник сферы обслуживания, специалист).

Различить употребления первой («ошибки») и второй («игра») категорий не всегда просто. Показателем ошибки может служить самокоррекция или коррекция, выполненная слушающим. Но говорящий может употребить неверную форму, осуществить самокоррекцию, а затем намеренно повторять неверную форму уже как проявление языковой игры, например:

- (э-э) () *вино с лёдом* со льдом придётся / <...> холодное с... сейчас лёду* () набросаю // <...> вот с лёдом* // вообще с(:)казка* (S85, М, 46, работник сферы обслуживания, специалист).

По результатам анализа контекстов, к ошибкам, кроме уже упомянутых, можно отнести следующие употребления:

- *ну вот все жаждят** (F3-S74, Ж, 30);
- *подожди / когти вытеру** (F1-S28, Ж, 50);
- *а ты куда ложись* ?* (M1-S20, М 30);
- *развешать* сушить ?* (M1-S91, М, 47, инженер, специалист)⁶.

Интересно, что среди «игровых» форм преобладают формы склонения (*пальтом, слезьми, лёдом*), а среди «настоящих» ошибок – формы спряжения (*вытеру, жаждят, развешать*).

Ошибки допускают говорящие из всех возрастных и различных профессиональных и статусных групп, т. е. выделить группу, в речи которой ошибок значимо больше, не представляется возможным. При этом ясно, что использовать ненормативные формы в ходе языковой игры склонны, скорее, младшие говорящие.

Что касается относительно «редких» грамматических форм, в подкорпусе для форм склонения наблюдается следующая картина.

Формы **второго родительного** (автоматически размечаемого с применением тега «gen2») встретились 41 раз. Это прежде всего формы партитива *чаю* (#10), *чайку* (#2), *супчику* (#1), *перцу* (#1), *кипяточку* (#1), а также компоненты выражений *ни разу, по второму разу* (#7), *сбить с толку, что толку, никакого толку* (#7), *много народу* (#2), *с левого боку* (#1). Формы gen2 – в том числе все формы партитива от диминутивов – употребляют преимущественно женщины (32 употребления из 41).

Формы **второго предложного** (или локатива, автоматически размечаемого с применением тега «loc2», 68 употреблений) встречаются прежде всего в составе выражений *иметь в виду* (#25) и *в ... году* (#21), а также *на полу* (#4), *в углу* (#3) и др. Распределение форм локатива в речи говорящих из разных социальных групп статистически не показательно.

Формы **второго винительного** (автоматически размечаемого с применением тега «acc2») встретились в подкорпусе 16 раз. Это употребления *в гости* (#14), *в начальники* (#1) и *в друзья* (#1). Распределение этих форм в речи говорящих из разных социальных групп также непоказательно.

Распределение в речи говорящих – представителей разных социальных групп – «сложных» форм (причастий и деепричастий) оказалось показательным и заслуживает отдельного обсуждения. Так, в составе аннотированного подкорпуса указанного объема формы причастий употреблены 311 раз (из них полные формы – всего 75 раз), формы деепричастий – 78 раз.

Наблюдение над употреблением форм полных причастий в речи различных **возрастных групп** дало следующие результаты: дети употребили форму полного причастия дважды, говорящие младшей возрастной группы – 15 раз, средней возрастной группы – 40 раз, старшей возрастной группы – 20 раз. Если учесть разницу размеров выборки в словах, то становится понятно, что говорящие младшей возрастной группы употребляют полные причастия существенно реже, чем говорящие старшего и среднего возраста.

Что касается форм деепричастий, то частота их употребления в речи младшей и средней возрастных групп примерно одинакова, при этом говорящие старшего возраста употребляют деепричастие примерно в три раза реже, чем остальные: age1 – #35, age2 – #30, age3 – #13. Кроме того, деепричастия употребляют прежде всего говорящие, имеющие статус «специалист» (#52 из общего #78).

Выявить иную «привязку» «сложных» форм к гендерным или социально-профессиональным группам не удалось.

«Прагматически маркированная» звательная форма представлена в составе подкорпуса 239 употреблениями. Прежде всего, это формы *мам* (#63), *пап* (#13), формы личных имен типа *Алён, Вить* (#149), а также восклицание *боже* (#13) и форма *отче* (#1) (употребленные не в качестве обращения и относимые к звательным по формальным основаниям). Звательную форму используют преимущественно женщины (#15 – 65,7 % всех употреблений) и дети (#40). На долю мужчин в подкорпусе приходится лишь 42 подобных употребления. Эмоциональное восклицание *боже* встречается почти исключительно в женской речи (#12 92 % употреблений).

Таким образом, проведенное исследование подтвердило, что распределение «сложных» форм (причастий и деепричастий) коррелирует с

возрастом и статусом говорящих, а распределение «прагматически маркированных» форм, в частности звательной формы, коррелирует с полом и возрастом говорящих.

3. Синтаксический анализ

На синтаксическом уровне анализировались следующие особенности:

- 1) линейные структуры глагольных кустов;
- 2) количество левых и правых членов глагольных кустов;
- 3) непроективные синтаксические конструкции (явления инверсии);
- 4) линейные структуры именных групп;
- 5) случаи парцелляции, эллипсиса, обрывов и самокоррекции.

Синтаксическая разметка выполнялась вручную, размечено 13 200 синтаксических структур.

Самыми частотными в речи представителей всех социальных групп являются одноэлементные синтаксические структуры:

D – частица / дискурсивное слово (3,7%) – *вот //, так //, да //, ладно //* и т. п.;

V – глагол-сказуемое (1,9%) – *звони ! //, поехали //;*

S – существительное-подлежащее (2,3 %) – *понедельник //, супчик ? //;*

Q? – вопросительное слово (0,8 %) – *зачем ? //, куда ? //.*

Кроме того, относительно частотны оказались простые предложения структуры SV (существительное-подлежащее + глагол-сказуемое – *я перезвоню //; ты посмотри ! //* (1,1 %) – или группы частиц {D} (1,95 %) – *вот так //; ну вот //; да да да //.*

Непроективные (инверсированные) структуры (*разные есть краткие версии; твой у меня номер*) чаще встречаются в речи мужчин, говорящих старшего возраста, руководителей и представителей естественных областей знания. Последние два вывода, впрочем, следует рассматривать как наблюдаемую тенденцию, которая требует статистического подтверждения на большем объеме эмпирического материала.

Согласно полученным данным, ранговые порядки отдельных синтаксических нерегулярностей инвариантны для всех рассмотренных социолектов (см. табл. 2). Обрывы фраз, эллипсис, явления парцелляции выявлены прежде всего в речи говорящих старшей возрастной группы (в том числе неработающих пенсионеров), а также военнослужащих. Таким образом, просматривается общая тенденция к увеличению доли синтаксических нерегулярностей с возрастом говорящих, однако ее статистическая состоятельность требует специальной проверки.

Таблица 2

Встречаемость синтаксических нерегулярностей в речи представителей различных социальных групп⁷, %

Синтаксические явления	Всего	Ранг	Гендер		Возраст		
			Муж.	Жен.	age1	age2	age3
CUT	3,66	1	3,38	4,06	2,98	3,72	4,45
EL	1,80	2	1,82	1,79	1,84	1,58	2,13
COR	1,34	3	1,36	1,32	1,08	1,41	1,59
PARC	0,48	4	0,48	0,48	0,39	0,50	0,60

Заключение

Полученные данные позволяют сделать вывод, что существуют инвариантные свойства устной русской речи, в равной мере присущие представителям всех исследованных социальных групп. Среди них, в частности, распределение частей речи, дистрибуция наиболее частотных синтаксических структур и синтаксических нерегулярностей.

Кроме того, существует значительный пласт «потенциально инвариантных» характеристик устной русской речи. Эти характеристики показали на проанализированном речевом материале

некоторые количественные различия в употребительности разными социальными группами, однако степень этого различия не позволяет на настоящий момент считать их статистически достоверными.

Наконец, были выявлены характеристики, по которым речь представителей разных социальных групп действительно различается, например, «прагматически маркированные» формы коррелируют с полом и возрастом говорящих (например, звательную форму типа *Надь* употребляют в основном женщины и дети).

Примечания

¹ Работа выполнена при поддержке гранта РНФ, проект № 14-18-02070 «Русский язык повседневного общения: особенности функционирования в разных социальных группах».

² Профессиональные группы: РАБ – рабочие; ИНЖ – инженеры; СИЛ – военнослужащие; ЕСТ – представители естественных наук; ГУМ – представители гуманитарных наук; ОБР – работники образования; СО – представители сферы обслуживания; ИТ – IT-специалисты; ОФ – офисные служащие; ТВОР – творческая интеллигенция. Статусные группы: УЧ – студенты и учащиеся; СП – наемные работники и специалисты; РУК – руководящие работники; БИЗ – бизнесмены и частные предприниматели; НР – неработающие и пенсионеры. Возрастные группы: age1 – младшая (18–30 лет), age2 – средняя (31–54 года), age3 – старшая (более 55 лет).

³ Использовались следующие частеречные пометы: V – глагол; S – существительное; S-PRO – местоимение-существительное; PART – частица; CONJ – союз; PR – предлог; ADV-PRO – местоимение-наречие; ADV – наречие; A-PRO – местоимение-прилагательное; A – прилагательное; PRAEDIC – предикатив; INTJ – междометие; NUM – числительное; PARENTH – вводное слово; ANUM – числительное-прилагательное; PRAEDIC-PRO – местоимение-предикатив.

⁴ В таблице используются обозначения, оговоренные в примечаниях 2–3.

⁵ О специальных обозначениях в расшифровках материала ОРД см. подробнее: [Шерстинова, Рыко, Степанова 2009; Русский язык повседневного общения... 2016: 242–243]. В скобках после примера дается атрибуция информанта: его номер (S1, S2...), пол, возраст, а также профессиональная и статусная принадлежность.

⁶ Данные реплики – не из речи информантов, а из речи их коммуникантов, поэтому в атрибуции к ним указан номер самого коммуниканта с фиксацией его пола (F1, M1 и под.), номер информанта, с которым происходит разговор, а также возраст коммуниканта.

⁷ В таблице используются следующие сокращения: CUT – обрыв фразы; EL – эллипсис; COR – самокоррекция; PARC – парцелляция, а также сокращения, оговоренные в примечании 2.

Список литературы

Богданова-Бегларян Н.В. и др. Звуковой корпус русского языка: новая методология анализа устной речи / Богданова-Бегларян Н.В., Асиновский А.С., Блинова О.В., Маркасова Е.В., Рыко А.И., Шерстинова Т.Ю. // Язык и метод: Русский язык в лингвистических исследованиях XXI века / ред. Д. Шумска, К. Озга. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2015. Вып. 2. С. 357–372.

Богданова-Бегларян Н.В. и др. Некоторые инвариантные характеристики русской разговорной речи: фонетика, морфология, синтаксис Богданова-Бегларян Н.В., Блинова О.В., Мартыненко Г.Я., Шерстинова Т.Ю. // Компьютерная лингвистика и интеллектуальные технологии: по матер. ежегодной междунар. конф. «Диалог» / ред. В.П. Селегей. М.: РГГУ, 2017а. Вып. 16(23). Т. 2. С. 75–86.

Богданова-Бегларян Н.В. и др. Корпус «Один речевой день» в исследованиях социолингвистической вариативности русской разговорной речи Богданова-Бегларян Н.В., Шерстинова Т.Ю., Блинова О.В., Мартыненко Г.Я. // Анализ разговорной русской речи (АР³ – 2017): тр. седьмого междисциплинарного семинара / науч. ред. Д.А. Кочаров, П.А. Скредин. СПб.: Политехника-принт, 2017б. С. 14–20.

Русский язык повседневного общения: особенности функционирования в разных социальных группах / отв. ред. Н.В. Богданова-Бегларян. СПб.: ЛАЙКА, 2016. 244 с.

Шерстинова Т.Ю., Рыко А.И., Степанова С.Б. Система аннотирования в звуковом корпусе русского языка «Один речевой день» // Формальные методы анализа речи: материалы XXXVIII Междунар. филол. конф. СПб.: Ф-т филологии и искусств СПбГУ, 2009. С. 66–75.

Bogdanova-Beglarian N. et al. Linguistic Features and Sociolinguistic Variability in Everyday Spoken Russian / Bogdanova-Beglarian N., Sherstina T., Blinova O., Martynenko G. // Speech and Computer: 19th International Conference, SPECOM 2017 / A. Karpov, R. Potapova, I. Mporas, Iosif (Eds.). Hatfield, UK, 2017P. 503–511.

RUSSIAN EVERYDAY SPEECH FROM GRAMMAR PERSPECTIVE

Natalia V. Bogdanova-Beglarian
Professor, Russian Language Department
Saint Petersburg State University

Olga V. Blinova
Assistant Professor, General Linguistics Department
Saint Petersburg State University

Grygory Ya. Martynenko
Professor, Mathematical Linguistics Department
Saint Petersburg State University

Tatiana Yu. Sherstinova
Assistant Professor, Mathematical Linguistics Department
Saint Petersburg State University

The paper discusses some peculiar features which characterize oral Russian speech of various social groups representatives. The study was carried out on the basis of the morphologically and syntactically annotated part of the «One Day of Speech» Corpus which contains sound recordings of everyday communication of 130 informants belonging to different gender, age, professional and status groups. The obtained results showed the existence of invariant properties of oral Russian speech equally similar for representatives of all the groups considered. Among them there are the distribution of parts of speech, as well as of most frequent syntactic structures and syntactic irregularities. Furthermore, there is a significant set of “potentially invariant” features of oral Russian speech; the analyzed speech material proved the existence of some qualitative differences for various social groups, though the degree of these differences does not allow at present to regard them as statistically reliable. Finally, a set of speech features characteristic for particular social layers has been revealed.

Keywords: spoken Russian language; everyday speech; speech corpus; grammar; morphology; syntax.