

Я. Э. Ахапкина

*Национальный исследовательский университет «Высшая школа экономики»
(Москва, Россия)
yakhapkina@hse.ru*

ЭРРАТОЛОГИЧЕСКАЯ РАЗМЕТКА КОРПУСА РУССКИХ УЧЕБНЫХ ТЕКСТОВ: ТАКТИЧЕСКИЕ РЕШЕНИЯ

Созданный школой лингвистики ФГН НИУ ВШЭ Корпус русских учебных текстов включает спонтанные (выполненные в аудитории) и подготовленные (домашние) работы студентов бакалавриата в жанрах ответа на вопрос, аргументированного суждения, эссе, курсовой работы. Студенты, осваивая академический регистр, проходят разные стадии формирования представлений об организации учебно-научного текста. На каждом этапе интерференция стилей и жанров, разнородный характер получаемых речевых образцов и слабая редакторская рефлексия приводят к возникновению неизбежных системных ошибок в сфере грамматики и грамматической стилистики, семантики и прагматики текста. Отклонения от речевого стандарта отражают как стадиальность формирования навыка исследовательского письма, так и процессы, характерные для динамики речевой системы в целом, становления новых узуальных норм на осколках норм устаревающих, демонстрируют границы вариативности слово- и формоупотребления. Эти отклонения маркируются системой тэгов, разработанной и оптимизированной командой Корпуса (Н. А. Зевахиной, С. С. Джакуповой, Ю. М. Кувшинской, С. Ю. Пужаевой при активном участии коллег и студентов). Эрратологическая разметка включает лексические, грамматические, дискурсивные пометы. Грамматический блок отражает частотные отклонения от морфологических и синтаксических канонов, связанные с расшатыванием ряда конструкций. Так, отдельную проблему представляет расширение круга легких глаголов (десемантизованных единиц, обслуживающих синтаксические потребности высказывания и делегирующих лексическое значение управляемому слову), выбор падежа управляемого имени, компаративные и интенсифицирующие конструкции, анафорические употребления. В статье рассмотрены конкретные примеры, снабженных тэгом «нарушение в согласовании»

* Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ в 2019 году.

(agr). Обсуждается мотивировка разметки при выборе маркера для фрагмента высказывания.

Ключевые слова: лингвистический корпус, корпус нестандартной речи, русские учебные тексты, согласование, вариативность, рекомендательная норма

Разметка коллекции Корпуса русских учебных текстов (http://web-corpora.net/learner_corpus/) включает маркирование речевых сбоев относительно рекомендательной нормы. Нарушением согласования считается не только отступление от норм соотношения по роду, числу и падежу зависимого слова с главным, но и отступление от правил координации, тяготения, аппликации [Валгина 2003: 55–58; Грамматика 1970: 548, 633; Булаховский 1952, 1: 287; Распопов 1970: 40]. В эту же рубрику попадает нарушение анафорической связи (семантическое согласование, согласование по смыслу) [Розенталь и др. 1999; Лаптева 2003; Голуб 2010]. Маркируется как грубое нарушение нормы, так и вариативное грамматическое оформление.

Анализ нестандартной речевой продукции, неизбежно содержащей как осознанные, так и не отрефлексированные пишущим отклонения от рекомендательной нормы, позволяет выявить тенденции развития современной речи [Пужаева и др. 2015; Dzhakupova et al. 2014; Zevakhina et al. 2015a, b]. Кроме того, анализ рефлексии разметчика позволяет увидеть, какие нормы в большей или меньшей степени осознаны современным носителем языка.

В Корпусе русских учебных текстов объемом 3115 188 слов нарушения в согласовании отмечаются в 587 документах и 971 контексте. В ряде случаев эрратологическое маркирование (выделение ошибок) опирается на многофакторный анализ высказывания.

- (1) *<В-пятых, с развитием новых технологий человеческие знания заменяются компьютерными, поэтому с каждым днем люди все меньше нуждаются в использовании своих навыков.> Использование известной информации становится недостаточным, поэтому их (ref, проп; нарушение кореференции) нужно подкреплять другими знаниями, которые обеспечивают (infl; именное и глагольное словоизменение) **технологий** (agr; согласование) каждый день. *<Например, с появлением новых цифровых фотоаппаратов людям приходится запоминать новые методы их использования, потому что знание об использовании фотоаппаратов прошлого поколения не позволяет пользоваться продуктом правильно.>**

В примере (1) из эссе социолога-первокурсника можно усмотреть нарушение согласования-координации при инвертированном прочтении субъектно-объектной пары (*технологии обеспечивают что? знания*), ошибку управления и падежную конкуренцию родительного / винительного / именительного при неинвертированном прочтении: *знания обеспечивают что? технологии* (в этом случае возможна описка, связанная с диакритикой: мена и/й), избыточный эллипсис в случае опущения управляющего слова: *<развитие> технологий*).

Дистантная координация при однородных зависимых в форме мн. ч. в разрывающем контексте приводит к тому, что прайминг плюральных форм заставляет разметчика игнорировать корректное согласования рамки, опирающееся на генерализованное прочтение формы ед. ч. подлежащего:

- (2) *Запоминание картинки, набора слов или звуков, напротив, **уязвимо** к любым изменениям.*

Множественность объектов в этом случае имплицитно для разметчика и множественность процессов, словно речь идет о разных видах запоминаний. В этой связи показательны рассуждения И. Н. Смирнова о влиянии множественного объекта на аспектуальную характеристику действия как повторяющегося и, соответственно, не локализованного во времени, в отличие от локализованного во времени единичного конкретного действия, совершаемого над единичным конкретным объектом [Смирнов 2008], ср. в примере (2а) аналогичную примеру (2) интерпретацию имени *вопрос* при конкретизирующих распространителях, включающую признание пишущим генерализованного употребления имени и отрицание такого прочтения разметчиком:

- (2а) *В данном тексте Ричард Фейнман ставит **вопрос** (agr) узкости мышления, доминирования зубрежки над пониманием, неспособности мозга самостоятельно работать, придумывать новое, а только лишь воспроизводить заученное.*

Избыточное согласование по числу в коннекторе сентенциального атрибута в примере (3) вызвано доминированием влияния левого контекста над правым, стремлением уподобить определение определяемому в ущерб структуре основы определительного придаточного. Дополнительным фактором, вызывающим этот сбой, становится разрыв основы придаточного вводной конструкцией.

- (3) *Здесь можно выделить несколько факторов, **которых** как нам кажется, влияют на наш выбор в какой-либо ситуации.*

В этом случае нарушения нормативной координации можно говорить о широко понимаемом прайминге [Русакова 2013: 371–410], то есть избыточном влиянии предшествующего образца (формы род. пад. мн. ч. *факторов*) на конструирование последующей формы *которых*, ср. пример (4) из аннотации первокурсника:

- (4) *Книга посвящена анализу событий, **которых** автор текста называет черными лебедями, это события, происходящие спонтанно и непредсказуемо.*

В этом случае коннектор выступает не подлежащим, а прямым дополнением, и грамматически актуализуется проблема выбора одушевленного или неодушевленного варианта маркирования винительного падежа. Корректная форма *которые*, омонимичная именительному, игнорируется в пользу формы, омонимичной родительному, поскольку левый контактный контекст задает генитивную форму определяемым словом *событий*.

Зоной особого риска становятся существительные со значением символической собирательности (*народ, аудитория, электорат, власть, профессура* и под.). При необходимости местоименной замены таких имен, как и в случае присоединения определительных оборотов, уточняющих конструкций или сентенциальных атрибутов, повышается вероятность смыслового согласования по числу в ущерб грамматическому. Множественная референциальная соотнесенность таких имен превалирует в сознании пишущего над сингулярностью грамматической формы, см. в примере (5) анафорическую замену *раса* — *их*:

- (5) *Так, например, основываясь на неприятном случае с афро-американцем (ограбление, избиение и т. д.), простой человек может стать расистом, обвиняя во всём чернокожую расу и их (agr ref pron) дикость.*

К широкой сфере семантического согласования можно отнести и выбор плюральной формы глагола при наличии зависимых у подлежащего, как в примере (6):

- (6) *Здесь ты найдешь ответы на эти и многие другие вопросы, узнаешь, какую роль играет экономическая свобода и каким образом в умах людей могут одновременно присутствовать вера в две совершенно разные и несочетаемые вещи.*

Невозможность формы мн. ч. у существительного *pluralia tantum* осознана пишущим, однако идея двойственности распространителя диктует форму мн. ч. глаголу как бы сквозь имя *вера*. В определенном смысле эта стратегия пишущего корреспондирует со стратегией разметчика в примере (2), стремящегося увидеть разные виды запоминаний в соответствии с указанием на объекты действия. Можно предположить, что генерализованное употребление имени функционирует по тем же законам, что и семантически (но не грамматически) собирательное имя, и формальная множественность предиката обусловлена референциальной множественностью существительного в таком употреблении.

При постпозиции однородных подлежащих допустимо как ед., так и мн. ч. сказуемого [Розенталь и др. 1999]. Однако однородность может ослабляться контекстом. Так, в примере (7) появление иерархического указателя *как следствие*, устанавливающего семантическую обусловленность второго подлежащего первым, заставляет разметчика склоняться к предпочтению сингулярной формы предиката, несмотря на кодифицированную вариативность:

- (7) *<Во-вторых, знания, полученные тщательным и фундаментальным разбором, дают толчок к появлению новых идей, к исследованию более глубоких и сложных проблем.> Сам Ричард Фейнман убеждает нас в том, что не только оробевшим новичкам но и известным ученым свойственны (agr) механическое запоминание информации и, как следствие, непрочные знания. <На примере того, как физик, ассистент Эйнштейна, не узнал в задаче элементарного физического закона из-за немного переделанных условий,*

мы лишний раз убеждаемся в том, что понимание темы позволяет применять знания в любых ситуациях.>

Однако противоположная стратегия пишущего, координирующего сказуемое с первым членом однородного ряда постпозитивных подлежащих, также не вызывает понимания у разметчика:

- (8) *Во-вторых, для того, чтобы дать ученикам импульс к стремлению понять материал, учителю **необходима** грамотно выстроенная методика преподавания и великолепное владение своим предметом, а так бывает далеко не всегда.*

Семантика предикативного комплекса *сложное пошаговое действие* в примере (9), включающая референциальную множественность, представляется разметчику недостаточным условием соотнесенности с однородными подлежащими при прямом порядке слов даже в случае контекстной иерархизации перечня, выраженной в предпочтении союза *а также* (маркирующего дополнительный статус второго подлежащего по отношению к первому):

- (9) *Я думаю, что так происходит, потому что осмысление полученных знаний, а также их практическое применение — это гораздо более **сложное многошаговое действие** чем заучивание наизусть.*

Неинвертированность высказывания и семантическая стадийность первичного *осмысления* и последующего *применения* побуждают разметчика ожидать форму мн. ч. у сказуемого, следующего за двумя подлежащими. Здесь можно усмотреть отказ интерпретировать имя *действие* в форме ед. ч. как генерализованное, с множественной референтной соотнесенностью.

В примере (10) разметчик предполагает взаимнооднозначное соотнесение *задачи* и ее *условия*, отрицая возможность выбора формы мн. ч. для терминологического имени в специализированном контексте при описании одной *задачи*:

- (10) *Во-вторых, если не понимать смысл выученного, то при изменении условий задачи, становится не ясно, как её решать.*

Между тем употребление имен в контексте примера (9) — анализа этюда Ричарда Фейнмана о запоминании и понимании — не строго терминологично, речь может идти о расширенном и в определенном смысле метафорическом понимании *задачи* как проблемы и *условий* как новых вопросов в проблемной области.

Квалификация ошибки как нарушения согласования может конкурировать с нарушением управления в тех случаях, когда гипотетическая форма мн. ч. омонимична родительному падежу ед. ч., как в случае конкуренции номинатива и генитива при отрицании (притом, что имя *суть* в норме не имеет формы мн. ч.):

- (11) *Как люди воспринимают науку, и способен ли человек принять знание и найти грамотный выход из ситуации, не осознавая суть (agr) явлений и законов природы?*

В примере (12), с одной стороны, можно усмотреть аналогию с примером (11), поскольку конструкция с негацией допускает возможность конкуренции падежей (гипотетический вариант правки *подключения* можно трактовать как форму им. мн. ч. и как род. ед. ч.), с другой — с примером (2) по признаку наличия серии зависимых (и тогда предполагается возможный вариант корректировки не *подключения*, а *подключений* — с вытекающей из такой правки проблемой отнесения имени к классу обладающих парадигмой мн. ч.):

- (12) *Механическое зазубривание не подразумевает подключение (agr) воображения, зрительной, слуховой памяти или любых других инструментов мышления.*

Согласование по числу может провоцировать сбой, связанный с полисемией определяемого, как в примере (14). Отвлеченное значение существительного *материал* ('учебные разработки, сведения') может восприниматься как принадлежащее имени *singularia tantum* и требующее согласования по ед.ч. Тем не менее, в этом значении форма мн. ч. имени употребительна, в НКРЯ находится 52 вхождения по запросу «учебный + материал (мн.ч.)», наиболее раннее датируется 1963 годом:

- (13) *Учебные материалы иного рода в книге не затрагивались, если не считать двух-трех беглых упоминаний.* [Григорий Коган. Работа пианиста (1963)]

При этом для ед. ч. этого сочетания обнаруживается 204 вхождения, начиная с 1862 года (первый пример принадлежит К. Д. Ушинскому).

- (14) *Также принудительное изучение материалов, который преподают в школах, не позволяет развиваться ученикам в отдельных областях, ведь учебная программа идет дальше темы сменяются и изученный материал остается позади.*

Возможно, форма *который* стала результатом своеобразной коррекции замысла пишущего, не удовлетворенного выражением *изучение материалов, которые преподают в школе* по причине механического двойного прочтения субъектно-объектной пары в этой конструкции (вариант исправления с пассивным оборотом *которые преподаются* оказался при этом не востребованным).

При дистантном расположении определяемого и определения разметчика смущает форма последнего имени, предшествующего определительному обороту. Строго говоря, контактность расположения определяемого и определения, действительно, предпочтительна, и обретает силу правила в отношении придаточных определительных. Однако для определений другой синтаксической природы аналогичной жесткости не предусмотрено, и в этом смысле пример (15) может рассматриваться как корректный:

- (15) *<В конце хотелось бы сказать, что лучший способ запоминания информации — понимание и осознание полученных знаний.> Такой подход имеет*

массу преимуществ перед зазубриванием, рассмотренных выше. <Поэтому я считаю, что проблема, которую поднял Ричард Фейнман в книге «Вы, конечно, шутите, мистер Фейнман» актуальная и должна быть освещена.>

Не предполагается, вопреки решению разметчика, фиксация сбоя и при контактном расположении придаточного определительного к номинативно-генитивному сочетанию:

- (16) *Каждый день человек, от самого маленького и до самого взрослого, получает огромный поток информации, **который** он должен запомнить.*

«В некоторых случаях между подлежащим и сказуемым усматривается еще одна разновидность синтаксической связи — тяготение: в предложениях типа *Сестра лежала больная; Отец вернулся веселый (веселым)*. Тяготением связь называется потому, что именная часть сказуемого соотносится с подлежащим через посредство третьего компонента. Такая связь обнаруживается при двойном отношении: прилагательное связано и с глаголом и с именем (или местоимением)» [Валгина 2003: 57]. В примере (17) игнорируется допустимость именительного и творительного падежей прилагательного при тяготении. Действительно, не любое лексическое наполнение современной конструкции допускает творительный, ср. архаичное **Сестра лежала больной*, но разметчик усомнился в правомерности творительного при глаголе *становиться*:

- (17) *Но пример с ассистентом Эйнштейна показывает, что помимо постоянного просто использования, знание необходимо подставлять и в нестандартные ситуации, чтобы оно становилось более **гибким** (агг) в плане приращения.*

При глаголе *становиться*, по данным НКРЯ, творительный падеж доминирует (14642 против 1212 вхождений с именительным). Именительный не редкость и сохраняется до современности (правда, в отсутствие сравнительной степени в контексте), ср. в НКРЯ:

- (17а) *Он выскочил из хаты. Дни **становились теплые**. Из-за Авдулиных бугров, из-за Пугачева горба заметно тянуло весной. [Г.П. Данилевский. Воля (1863)]*
- (17б) *Кременецкий тоже говорил, что возможны очень серьезные осложнения, и вид у него при этом **становился озабоченный**. [М.А. Алданов. Бегство (1930)]*
- (17в) *Раз с человека держава взять ничего не может, он **становится бесполезный**. [Василий Гроссман. Все течет (1955–1963) // «Октябрь», 1989]*
- (17г) *Матвей с Костином быстренько «взлетали» на шкаф. Обстановка **становилась напряженная**. Один лишь Сверчок радовался, затевал с Лари игру в догонялки. [Н.Ф. Королева. Другая собака // «Наука и жизнь», 2007]*
- (17д) *Когда актеры играли хоть чуть-чуть интересно, режиссер залиvisto хохотал. И лица у актеров **становились довольные**. На меня актеры уже*

не смотрели. [Саша Денисова. Театральный роман // «Русский репортер», № 15 (143), 22–29 апреля 2010, 2010]

- (17е) *Стекло холодное, и лоб от него тоже становится холодный.* [Дина Сабитова. Где нет зимы (2011)]
- (17ж) *Траектория развития такого брака... Его психика не выдерживает такого давления, и он регулярно уходит из дома «в отпуск» (живет у родителей). Жена после такого «отпуска» становится шелковая. На какое-то время.* [Алексей Фролов. Очередной случай из практики... (2014.10.20)]

Несколько архаичным воспринимается пример (17в), возможно из-за одушевленности субъекта состояния. Но в (17ж) при определяемом женского рода и разговорном регистре архаизации уже не усматривается, хотя субъект состояния одушевленный.

Несмотря на сохранность именительного предикативного в таких предложениях, примат творительного, вопреки мнению разметчика, в описываемой конструкции очевиден как по данным НКРЯ, так и по данным интернет-узуса, ср. запрос в Google *становится гибкий* — 126 результатов, *становится гибким* — 225 результатов.

Конкуренция родительного и творительного субъекта или дательного адресата (в примере (18) — *понимание кого и кем, понятен кому*) наблюдается в современном узусе, демонстрирующем постепенную экспансию генитива в сферу других косвенных падежей (у этой тенденции есть и встречная компенсирующая, в частности, связанная с захватом дательным падежом некоторых функций родительного, см. [Ахапкина 214]). В примере (18) разметчик счел выбор родительного падежа ошибочным (вероятно, тэг нарушения управления в этом случае в большей степени отвечал бы замыслу).

- (18) *Текст в большинстве своём чётко и понятен для понимания практически любого читателя.*

Падежная конкуренция может быть вызвана разными сопутствующими факторами, среди которых фиксируют порядок слов (ср. *понимание текста читателем / читателем текста; поддакивание секретаря начальнику / поддакивание начальнику секретарем*, о стратегии выбора падежа в последнем случае см. [Герасимова 2017]).

НКРЯ демонстрирует вариативность в этой зоне. Генитивные примеры встречаются в разные периоды и не зависят от падежа управляющего.

- (18а) *Но далее отрезки расходятся, и нос двоится не только в стиле автора и в глазах Ковалева, но и в понимании читателя.* [В. В. Виноградов. Эволюция русского натурализма. Гоголь и Достоевский (1920–1929)]
- (18б) *Они приспособлены ко вкусу и пониманию читателя ниже чем среднего.* [В. Ф. Ходасевич. О Чехове (1929)]
- (18в) *Почти все они доступны пониманию читателей, обладающих подготовкой в объеме неполной средней школы.* [С. А. Шорыгин. Что читать об Арктике // «Наука и жизнь», 1936]

- (18г) *Мы полностью полагались на эзопов язык и на понимание читателя, умевшего читать между строк.* [Валерий Аграновский. Вторая древнейшая. Беседы о журналистике (1976–1999)]
- (18д) *Наверно, надо начать с того, что каждый стихотворец ... имеет свое понимание читателя.* [Алексей Биргер. Страстный пилигрим (1998) // «Профессионал», 1998.07.01]

В случае реализации ролей субъекта и объекта в форме косвенных падежей (*понимание чего / кого кем // понимание кем чего / кого*) субъект закономерно маркирован творительным, объект — винительным.

- (18е) *И насчет невозможности понимания обыкновенными людьми великих, читателями — писателей, Марк загнул по обыкновению.* [Юлий Даниэль. Письма из заключения (1966–1970)]
- (18ж) *Привычное и внутренне присущее христианскому мировоззрению разделение мира на полюсы: небо и земля, рай и ад, бог и диавол, добро и зло, праведники и грешники, последовательно проводимое через весь «Элucidарий», несомненно, в высшей степени благоприятствовало пониманию и усвоению читателями и слушателями преподносимых им истин.* [А. Я. Гуревич. Популярное богословие и народная религиозность средних веков (1976)]
- (18з) *Вместе с тем воздействие литературы на жизнь общества, ее понимание и осмысление читателями (иначе говоря — литература в меняющихся социально-культурных контекстах ее восприятия) является предметом одной из литературоведческих дисциплин — историко-функционального изучения литературы (термин предложен М. Б. Храпченко в конце 1960-х годов).* [В. Е. Хализев. Теория литературы (1999)]

Собственно эффект прайминга (уподобления образцу) на формальном уровне наблюдается в примерах (19–20):

- (19) *Задумывались вы о причинах появления тоталитарных режимов.*
- (20) *У вас есть шанс раз и навсегда изменить свое отношение к миру и происходящих в нем событий и по-новому посмотреть на все глобальные процессы в нашей жизни.*

М. В. Русакова отмечает, что замена родительного мн.ч. предложным у существительных вызвана ориентацией на флексию предшествующего прилагательного [Русакова 2008].

В сфере согласования наиболее болезненной оказывается семантико-грамматическая зона числа, вызывающая сложность как у пишущих, так и у разметчиков. Это связано как с семантической и референциальной характеристикой определяемого (несовпадением плюрального содержания с сингулярной формой), так и синтаксическими особенностями фразы, включающей зависимые от существительного компоненты, часто в плюральных формах, а также разрывающие элементы согласованного сочетания вставные структуры. В сфере падежного маркирования размываются границы согласования и управления (в частности, в случае

предпочтения творительного именительному при так называемом тяготении). Одним из факторов, влияющих на сбой в согласовании, становится эффект копирования предшествующего образца.

Описанные наблюдения позволяют предложить три дополнительные процедуры для разметки коллекции корпуса. Во-первых, выстроить субкатегоризацию внутри сбоя «нарушение в согласовании» (в частности, выделив подтип нарушенного согласования — в роде, числе, падеже или нескольких категориях одновременно — и указав возможную причину рассогласования, такую как референциальная множественность определяемого или влияние разрывающего контекста, плюральная форма зависимых от определяемого имен). Во-вторых, снабдить маркированные фрагменты вариантом правки в окне комментария, чтобы сравнение корректного варианта с ошибочным позволяло анализировать ошибку. В-третьих, противопоставить собственно нарушение нормы (речевой сбой) вариативности и шкалировать вариативность от равноправных, предпочтительных и допустимых вариантов до маргинальных и тяготеющих к сбою.

Литература

Ахапкина Я. Э. Причина пожара: приименной дательный в позиции родительного // Слово. Словарь. Словесность: Русский язык в научном, культурном и образовательном пространстве (к 190-летию со дня рождения К. Д. Ушинского). СПб., 2014. С. 38–42.

Булаховский Л. А. Курс русского литературного языка. Т. 1. Киев, 1952.

Валгина Н. С. Современный русский язык: Синтаксис: учебник. 4-е изд., испр. М., 2003.

Герасимова А. А. Дифференциальное падежное маркирование и линейная позиция аргументов в русских событийных номинализациях // Acta Linguistica Petropolitana. Труды института лингвистических исследований. 2017. Т. XIII. Ч. 3. С. 265–281.

Голуб И. Б. Литературное редактирование: учебное пособие. М.: Логос, 2010.

Грамматика современного русского литературного языка. М., 1970.

Лаптева О. А. Живая русская речь с телеэкрана. Разговорный пласт телевизионной речи в нормативном аспекте. 5-е изд. М., 2003.

Пужаева С. Ю., Зевахина Н. А., Джакупова С. С. Контаминация конструкций в речи нестандартных русскоговорящих на материале корпуса русских учебных текстов // Труды Международной научной конференции «Корпусная лингвистика-2015». СПб., 2015. С. 390–397.

Распопов И. П. Структура простого предложения в современном русском языке. М., 1970.

Розенталь Д. Э., Джанджакова Е. В., Кабанова Н. П. Справочник по правописанию, произношению, литературному редактированию. М., 1999.

Русакова М. В. Элементы антропоцентрической грамматики русского языка. М., 2013.

Смирнов И. Н. Выражение повторяемости и обобщенности действия в современном русском языке. СПб., 2008.

Dzhakupova S., Zevakhina N. Case (non-)coincidence in elliptical coordinated constructions: Learner texts of Russian native speakers // *Slavica Helsingiensia 45 Instrumentarium of Linguistics: Language errors and multilingualism.* Helsinki, 2014. С. 35–49.

Zevakhina N., Dzhakupova S. Corpus of Russian student texts: design and prospects // Труды 21-й Международной конференции по компьютерной лингвистике «Диалог». М., 2015а.

Zevakhina N., Dzhakupova S. Russian metalinguistic comparatives: a functional perspective. Working papers by NRU HSE. Series WP BRP «Linguistics». 2015b. No. 39.

Yana E. Akhapkina

National Research University Higher School of Economics (HSE)

(Moscow, Russia)

yana.akhapkina@gmail.com

ERRATORICAL MARKING OF THE TEXTBOOK CORPORA: TACTICAL SOLUTIONS

Created by the School of Linguistics of the Faculty of Humanities at the National Research University Higher School of Economics, the Corpus of Russian Student Texts (CoRST) includes texts belonging to such genres as answers to various questions, argumentative statements, essays, course papers etc., which were written either spontaneously (in the classroom) or as prepared texts (at home) by students in Bachelor's degree programs.

In the process of studying academic writing, students pass through different stages of understanding how to structure academic texts.

At each stage, the interference of different styles and genres, the heterogeneous nature of received speech patterns as well as low levels of self-correction lead to inevitable systemic errors in grammar and grammatical stylistics, semantics and text pragmatics.

The deviations from standard speech reflect both the stadial nature of academic writing skills and the processes characteristic of speech system dynamics in general; the formation of new customary (usual) norms on the remains of obsolete (conservative) norms demonstrates the limits of variability in the usage of words and word forms.

These deviations are marked by a system of tags developed and optimized by the Corpus team (N. A. Zevakhina, S. S. Dzhakupova, Yu. M. Kuvshinskaya, S. Yu. Puzhaeva, with active assistance from colleagues and students).

The error markup contains lexical, morphological, and discursive information.

The grammatical section shows the frequency of deviations from morphological and syntactical patterns are connected with the slackening of a number of constructions.

For example, there are such challenges as the broadening of a number of 'light' verbs (units devoid of semantic value and satisfying the syntactic needs of a statement, whose lexical meaning is delegated to a governed word); the choice of case for governed nouns; comparative and intensifying constructions; and anaphoric usage.

The article considers specific examples marked with the tag “agreement error” (*agr*). The motivation for markup when choosing a marker for a speech fragment is discussed.

Keywords: linguistic corpus, corpus of non-standard speech, Russian learners' texts, agreement, variability, recommended norms

References

Ахапкина Я. Е. Причина пожара: приименной дател'nyj в позиции родител'nogo. *Slovo. Slovar`. Slovesnost`*: *Russkij yazyk v nauchnom, kul'turnom i obrazovatel'nom prostranstve (k 190-letiyu so dnya rozhdeniya K. D. Ushinskogo)*. St. Petersburg, 2014, pp. 38–42. (In Russ.)

Bulaxovskij L. A. *Kurs russkogo literaturnogo yazyka*. T. 1. Kiev, 1952. (In Russ.)

Dzhakupova S., Zevakhina N. Case (non-)coincidence in elliptical coordinated constructions: Learner texts of Russian native speakers. *Slavica Helsingiensia 45 Instrumentarium of Linguistics: Language errors and multilingualism*. Helsinki, 2014, pp. 35–49.

Gerasimova A. A. Differential'noe padezhnoe markirovanie i linejnaya poziciya argumentov v russkix sobytijnyx nominalizacijax. *Acta Linguistica Petropolitana. Trudy instituta lingvisticheskix issledovanij*. 2017. T. XIII. Ch. 3, pp. 265–281. (In Russ.)

Golub I. B. *Literaturnoe redaktirovanie: uchebnoe posobie*. Moscow: Logos, 2010. (In Russ.)

Grammatika sovremennogo russkogo literaturnogo yazyka. Moscow, 1970. (In Russ.)

Lapteva O. A. *Zhivaya russkaya rech s teleekrana. Razgovornyj plast televizionnoj rechi v normativnom aspekte*. 5-e izd. Moscow, 2003. (In Russ.)

Puzhaeva S. Yu., Zevaxina N. A., Dzhakupova S. S. Kontaminaciya konstrukcij v rechi nestandardny'x russkogovoryashhix na materiale korpusa russkix uchebny'x tekstov. *Trudy Mezhdunarodnoj nauchnoj konferencii «Korpusnaya lingvistika-2015»*. St. Petersburg, 2015, pp. 390–397. (In Russ.)

Raspopov I. P. *Stroenie prostogo predlozheniya v sovremennom russkom yazyke*. Moscow, 1970. (In Russ.)

Rozental' D. E., Dzhandzhakova E. V., Kabanova N. P. *Spravochnik po pravopisaniyu, proiznosheniyu, literaturnomu redaktirovaniyu*. Moscow, 1999. (In Russ.)

Rusakova M. V. *Elementy antropocentricheskoj grammatiki russkogo yazyka*. Moscow, 2013. (In Russ.)

Smirnov I. N. *Vyrazhenie povtoryaemosti i obobshhennosti dejstviya v sovremennom russkom yazyke*. St. Petersburg, 2008. (In Russ.)

Valgina N. S. *Sovremennyj russkij yazyk: Sintaksis: uchebnik*. 4-e izd., ispr. Moscow, 2003. (In Russ.)

Zevakhina N., Dzhakupova S. Corpus of Russian student texts: design and prospects. *Trudy 21-j Mezhdunarodnoj konferencii po komp'yuternoj lingvistike «Dialog»*. Moscow, 2015a.

Zevakhina N., Dzhakupova S. *Russian metalinguistic comparatives: a functional perspective*. Working papers by NRU HSE. Series WP BRP «Linguistics». 2015b. no. 39.