

## О компьютерной проверке синтаксических конструкций в текстах на русском языке

Э.К.Лавошникова

*Научно-исследовательский вычислительный центр МГУ им. М.В.Ломоносова,  
Москва, Россия*

Поступила в редколлегию 25.07.2005

**Аннотация**—Рассматриваются некоторые проблемы автоматической генерации рекомендаций по стилистической и грамматической правке текстов на русском языке. Работа “русскоязычных” автокорректоров анализируется на примере самого распространенного из них — ОРФО, встроенного в текстовый редактор MICROSOFT WORD, версии 1997, 2000 и 2002 гг. (MS Office XP). Путем тестирования на специально подобранных примерах фраз воссоздается алгоритм одного из заявленных синтаксических правил “Неверное употребление деепричастного оборота” (расхожий пример: “Подъезжая к станции, с меня слетела шляпа”). Показана неполная адекватность алгоритма поставленной задаче и сообщениям программы-подсказки. Предложены рекомендации для разработки новых версий автокорректоров.

### 1. WORD'ОВСКИЕ ПРАВИЛА ПРОВЕРКИ ПРАВОПИСАНИЯ

Проблематику компьютерной проверки правописания логично обсуждать на примере наиболее распространенного **автокорректора, спеллера, спелл-чекера** (как справедливо отмечает В. А. Успенский [Успенский 2002: 915], общепринятого русского термина для таких программ нет). В текстовый редактор MICROSOFT WORD для русскоязычных пользователей встроен такой “монополист” — **автокорректор ОРФО**.

Мы попытались проанализировать работу автокорректора ОРФО с различными синтаксическими конструкциями в текстах на русском языке. Заметим при этом, что версии ОРФО 1997 и 2002 гг. почти идентичны версии 2000 г.

Не всегда погрешности в построении фраз происходят от слабого владения литературным языком. Синтаксис, т. е. сочетаемость и порядок следования слов внутри предложения, может нарушаться, в частности, и при невнимательной правке текста.

А. Майорова в статье “Доверяй, но проверяй. Системы контроля орфографии консервируют язык” (газета “Известия.Ру” от 18.03.03, сайт [www.izvestia.ru](http://www.izvestia.ru)) пишет: “Программу проверки правописания для MS Word разработала компания “Информатик” в 1987–1994 годах. Отцом проекта был аспирант ВЦ АН СССР Олег Григорьев. Игорь Ашманов и Нина Руссова, также сотрудники ВЦ АН, разработали словарь и машинную морфологическую модель, а также модель грамматического разбора предложения. Первая версия работала под DOS, ее продажи начались в 1991 году. Тогда же большая часть команды вместе с Олегом Григорьевым уехала в США. Игорь Ашманов не эмигрировал. Он набрал и возглавил новый коллектив, который в 1992–1993 годах выпустил систему для Windows. [...] Нынешняя версия программы в составе продуктов Микрософт — почти такая же, как и в 1994 году”.

К этому следует добавить, что автокорректор ОРФО, как и некоторые другие “русскоязычные” спеллеры, создан на основе первого издания (1977 года) “Грамматического словаря русского языка” А. А. Зализняка. Словарь был переведен на машинные носители в начале 80-х

годов под руководством В. М. Андриященко в Лаборатории автоматизированных лексикографических систем Научно-исследовательского вычислительного центра МГУ им. М.В. Ломоносова.

Заметим, что слова с орфографическими ошибками, а точнее, **словоформы** — грамматические формы словарного слова, — отсутствующие (не порождаемые) в системных словарях автокорректора ОРФО, в процессе проверки правописания в текстовом редакторе MS WORD подчеркиваются на экране красной волнистой линией, а “подозрительные” в отношении пунктуации, синтаксиса или стиля слова и конструкции — зеленой.

Предусматриваемые классы синтаксических ошибок в текстах на русском языке спеллером ОРФО отображены в виде правил как в списке опций “Грамматика”, так и в списке “Стиль”. Пользователю предоставляется возможность отключать любые проверки. Мы проводили тестирование как на правильных, так и на неправильно построенных предложениях, при полностью задействованном наборе правил, т. е. при заданном в “Параметрах” на кнопке “Правописание” режиме “Строго (все правила)”.

Очевидно, что спелл-чекер ОРФО проверяет корректность синтаксических конструкций только в границах одного предложения, не учитывая анафорические связи даже в пределах абзаца.

Мы можем изучать доступные нам не очень детализированные описания, но все же представляется нелишним протестировать работу спеллера с синтаксическими конструкциями на конкретных примерах.

Будем составлять фразы из известных автокорректору ОРФО слов.

**Примечание.** Фразы-примеры и конструкции, подчеркиваемые на экране в ходе заданной проверки правописания, мы будем также выделять подчеркиванием. Далее в тексте статьи будут также выделяться курсивом с подчеркиванием отдельные слова и словосочетания, взятые из таких фраз.

Посмотрим, например, как автокорректор ОРФО отреагирует на сложный предлог *начиная с*.

*Начиная с января у нас повысились тарифы на электроэнергию.* (1)

В этом предложении запятая не нужна, “слово *начиная* можно опустить без ущерба для смысла и структуры предложения” [Розенталь 2003: 65]. Однако спеллер подчеркивает первую часть фразы и выдает сообщение: Возможно, в конце деепричастного оборота пропущена запятая на отрезке предложения между словами: словом “Начиная” и словом “повысились”. Это сообщение ссылается на правило “**Запятые в деепричастных оборотах**” (из списка “Грамматика”).

Попробуем последовать рекомендации системы ОРФО. Вопреки правилам русской грамматики поставим запятую после конструкции *начиная с января*. В этом случае спеллер подчеркивает уже все предложение целиком и выдает сообщение о вероятном “противоречии сказуемого и деепричастия”, ссылающееся на правило “**Неверное употребление деепричастного оборота**”.

Работу этого, пожалуй, наиболее любопытного и нетривиального правила мы и попробуем проанализировать, попытаемся понять логику разработчиков алгоритма проверки на “противоречие сказуемого и деепричастия”.

Интересующее нас Правило в “Грамматических пояснениях” сформулировано следующим образом:

“В правильно построенном русском предложении сказуемое и деепричастие должны относиться к одному и тому же действующему субъекту. Например, в ошибочной фразе

“Подъезжая к станции, с меня слетела шляпа” подразумевается, что на самом деле подъезжает не шляпа, а ее владелец. При подобном несовпадении субъекта сказуемого и деепричастия предложение становится неуклюжим и сложным для понимания. Для исправления ошибки” . . . и т. д.

Как работает система ОРФО по данному Правилу, почему какие-то предложения пропускает, а какие-то подчеркивает, понять довольно трудно. Нам показалось интересным в этом разобраться. Мы решили провести некоторое тестирование, чтобы высветить логику алгоритма, связанного с этим Правилу, уяснить себе причины не всегда адекватных решений разработчиков этого алгоритма.

## 2. НЕОПОЗНАННЫЕ ПРОТИВОРЕЧИЯ СКАЗУЕМОГО И ДЕЕПРИЧАСТИЯ

Мы задали проверку (кнопка “Правописание”) некоторых предложений, которые по своей структуре сходны с примером, приведенным в тексте правила “Неверное употребление деепричастного оборота” (см. выше).

Заметим, что приводимый в экспликации (в “Справочнике по русскому языку” – ОРФО2002) пример Выражая в песнях, сказках, былинах мечту о счастливой доле, народом были созданы произведения большой глубины и силы пропускается автокорректором без каких-либо подчеркиваний зеленым и замечаний.

Следующая фраза спеллером ОРФО не подчеркивается и поэтому не вызывает никаких сообщений программы-подсказки.

*Подъезжая к станции и глядя на природу в окно, у меня слетела шляпа.* (2)

В этой (не вполне точной) чеховской цитате до запятой имеют место сразу два деепричастия. В таких предложениях распознавать “противоречия” система ОРФО пока не научилась. К следующему примеру у автокорректора также нет никаких замечаний.

*Накурившись, между солдатами завязался разговор.* (3)

Если в этой цитате из повести “Хаджи-Мурат” Л.Н. Толстого конструкцию *между солдатами* заменить сочетанием *у них* (но не *у солдат*), то предложение будет подчеркнута зеленым с выдачей сообщения о “противоречии сказуемого и деепричастия”.

Приведем неоднозначный пример.

*Заглядевшись на новые ворота, меня испугало стадо коров.* (4)

Если рассуждать формально, стадо тоже может “заглядеться” на что-либо. Однако не это соображение стало причиной отсутствия реакции спеллера — если изъять прилагательное *новые*, то предложение будет подчеркнута с выдачей сообщения, ссылающегося на Правило. Однако в таком сокращенном варианте фразы словоформу *меня* нельзя с тем же результатом заменить, например, словоформами *прохожего, его, ее, их* или *Ивана*. Вместо сочетания *на ворота* нельзя подставить *на них* — подчеркивание фразы снимается.

Рассмотрим вариант без прилагательного и слова *стадо*.

*Заглядевшись на ворота, меня испугала корова.* (5)

Подчеркивания не произойдет. Заменяем слово *корова* именем *Зорька* или однозначным (не имеющим в системном словаре ОРФО нарицательных омонимов) именем *Клеопатра*. Фраза в обоих случаях подчеркивается. Имена собственные, как и неодушевленное существительное *стадо* и ему подобные, спеллер почему-то не рассматривает как кандидатов на роль одушевленного подлежащего (см. ниже Условие Е).

## 3. ОПОЗНАВАЕМЫЕ СИСТЕМОЙ ОРФО “ПРОТИВОРЕЧИЯ”

Теперь рассмотрим такие некорректные предложения, которые автокорректор ОРФО вполне обоснованно подчеркивает — с выдачей сообщения, ссылающегося на тестируемое нами Правило.

“Пленив Москву”, на самозванца была возложена царская корона. (6)

Если вместо неодушевленного корона набрать корова или ворона, то “возражения” спеллера снимаются. Если вставить слово голову (на голову самозванца), то подчеркивания также не произойдет, так как словоформа голову — в отличие от взятой вне контекста словоформы самозванца — не может относиться к родительному падежу (об этом требовании алгоритма Правила см. ниже Условие D).

Построим более сложное предложение по схеме приведенного системой ОРФО примера. Мы уже знаем, что присутствие до первой запятой прилагательных препятствует дальнейшей проверке на рассматриваемое Правило. С удовлетворением отметим, что автокорректор ОРФО даже в таком нагромождении конструкций распознаёт противоречивость сказуемого и деепричастия, подчеркивая всю фразу целиком и отсылая пользователя к Правилу.

Оставшись вследствие разгула стихии без присмотра (без родителей), мальчика в густом и дремучем лесу, где было много диких зверей, выручали из беды только находчивость и смекалка. (7)

Но если в подчеркнутом предложении перед деепричастием (первая словоформа) вставить какое-либо слово, даже отрицание *не*, подчеркивание снимается.

Попробуем произвести некоторые замены, которые, казалось бы, не должны снимать противоречия между сказуемым и деепричастием.

Оказывается, конструкцию без родителей с тем же результатом (сообщением программы-подсказки) можно заменить сочетанием с родителями, но не вместе с родителями. Выяснилось также, что деепричастный оборот (а вернее — первая часть предложения) может содержать сочетание предлога с некоторыми именами собственными. До запятой можно вставить без Сашу, но не без Машу, с Марией, но не с Алей. Причина в том, что алгоритм проверки определяет словоформы Машу и Алей не только как имена собственные, но и как императивы от глаголов *махать* и *алеть*, а это противоречит выбранной схеме (см. ниже Условие С). При этом спеллер не замечает того, что эти слова набраны с прописной буквы, а главное, что непосредственно перед ними имеется предлог. По непонятным причинам нельзя подставить без Любви — подчеркивание снимается.

В примере (7) определение к слову мальчика можно поставить только в постпозиции, иначе предложение не будет подчеркиваться. Словоформу мальчика можно с тем же результатом исправить на тебя, подростка, девочек, но не на девочку, юношу или Виктора.

Придаточное предложение где было много диких зверей нельзя, как мы убедились, заменить придаточным где были дикие звери, иначе подчеркивание снимается. Существительное *звери* в именительном падеже алгоритм необоснованно считает возможным подлежащим всего сложноподчиненного предложения, а одушевленное подлежащее не вписывается в схему Правила (подробнее об этом см. ниже в Условии E).

## 4. ГИПЕРДИАГНОСТИКА ОРФОВСКОГО СПЕЛЛЕРА

Попробуем построить по образцу, заданному в тексте интересующего нас Правила (см. раздел 1), фразу, в которой сказуемое и деепричастие не вступают в противоречие, а относятся к одному и тому же слову — к неодушевленному подлежащему.

Зацепившись за ветку дерева, с меня свалилась шляпа. (8)

Предложение синтаксически правильное, пусть даже немного “неуклюжее” (по терминологии ОРФО). Однако выдается все то же сообщение, ссылающееся на Правило: Вероятно, в данном предложении имеется противоречие сказуемого и деепричастия — сказуемое подразумевает одного действующего субъекта, а деепричастие — другого, как в классическом примере “Подъезжая к станции, с меня слетела шляпа”...

В примере (8) такого противоречия нет.

Если с меня заменить сочетанием с головы, что, казалось бы, принципиально ничего не меняет, то полученная фраза пропускается спеллером без замечаний. Как показывают результаты тестирования, алгоритм Правила, несмотря на предшествующий словоформе головы предлог, необоснованно считает возможным отнесение ее к именительному падежу (мн. числа). Отсюда делается вывод, что эта словоформа в получившейся фразе может оказаться одушевленным подлежащим (в значении ‘должностное лицо’ слово голова склоняется по “одушевленному” типу склонения), что не подходит под заданную схему.

Теперь попробуем заменить слово шляпа несклоняемым существительным сомбреро. Подчеркивание фразы снимается. Спеллер не замечает даже расогласования с глагольной формой (свалилась). Все несклоняемые существительные алгоритм Правила необоснованно считает кандидатами на роль “одушевленного” подлежащего (см. ниже Условие Е).

Далее рассмотрим примеры без формального подлежащего, но достаточно корректные. Следующие три фразы подчеркиваются с сообщением, ссылающимся на Правило.

Учитывая успехи коллектива предприятия, работников цехов в связи с реализацией всех 100% продукции можно было бы поощрить 50-процентной надбавкой к зарплате. (9)

В этом предложении словоформу работников нельзя с тем же результатом (с сообщением о “противоречии”) заменить словоформой рабочих, так как алгоритм Правила воспримет ее как возможное прилагательное (см. ниже Условие D).

Будучи под воздействием наркотиков, никого не боятся. (10)

Если в примере (10) во вторую часть фразы вставить местоимение они, т. е. подразумеваемое одушевленное подлежащее выразить в эксплицитной форме, то “возражения” спеллера снимаются.

Говоря словами поэта, нас было много на челне. (11)

Вместо словоформы поэта нельзя с тем же результатом подставить форму имени собственного Пушкина (см. ниже Условие С).

## 5. ЗАГАДОЧНЫЕ ПРИМЕРЫ “ЛОЖНЫХ ТРЕВОГ”

Что касается следующих анекдотичных примеров, то их можно использовать как тест на сообразительность пользователей. Все эти фразы (хотя первые три — вообще без деепричастий) подчеркиваются с выдачей сообщения, ссылающегося на рассматриваемое нами правило “Неверное употребление деепричастного оборота”.

Слесаря, слесаря скорей! (12)

Здесь дело в том, что словоформа слесаря трактуется алгоритмом Правила неоднозначно: не только как форма существительного (после запятой), но и как ее омограф (до запятой), а

именно — как деепричастие от глагола *слесарить*. Однако если бы в системном словаре ОРФО имелась просторечная форма именительного падежа мн. числа *слесаря*, то не было бы подчеркивания и ссылки на Правило, так как его алгоритм прекращает проверку предложений, в которых после первой запятой имеется хотя бы одно одушевленное существительное в форме именительного падежа.

Далее еще подобные примеры.

Боря, из-за Льва Ивановича одни неприятности, правда? (13)

Здесь разгадка в том, что в словаре системы ОРФО представлен малоупотребительный (без постфикса *-ся*) глагол *бороть* с деепричастием *боря*. Кроме того, спеллер “думает”, что в этом примере речь идет о царе зверей, так как при подстановке вместо *Льва* любого другого однозначного имени (*Петра, Николая*) подчеркивание снимается.

Ежа, ужа и чижса пора бы уж выпустить на свободу. (14)

Догадаться, что слово *ежа* в системном словаре ОРФО определяется не только как форма существительного, но и как деепричастие настоящего времени *ёжа* от глагола *ёжить*, практически невозможно. О засоренности словаря автокорректора малоупотребительными словами мы писали в [Лавошникова 2002; 2003].

Оказывается, кроме того, что слово *уж* (как и словоформа *уже*) ни в каких контекстах не воспринимается системой ОРФО в качестве существительного (хотя *ужа, ужи* — воспринимаются), иначе подчеркивания бы не было (см. ниже Условие Е).

То же сообщение, ссылающееся на Правило, вызывает и фраза с деепричастным оборотом не в начале предложения.

Царя Петра, “мореплавателя и плотника”, отдавая должное заслугам этого самодержца, многие называют самым прогрессивным из русских царей. (15)

Пользователь, разумеется, воспримет сообщение спеллера о “противоречии сказуемого и деепричастия” как относящееся к деепричастию *отдавая*. Однако “второй” (а на самом деле — первый и единственный) деепричастный оборот алгоритм Правила просто не замечает. Деепричастным оборотом программа проверки считает сочетание *царя Петра*. Первая словоформа в качестве деепричастия входит в парадигму глагола *царить*. Непереходность этого глагола (помета **нп** в словаре Зализняка), как и в других случаях, системой ОРФО не учитывается. В первую часть предложения (до первой запятой) нельзя подставить цифры или латинские буквы (*Петра I*), иначе подчеркивание пропадает.

Чтобы уберечься от таких казусов, разработчикам спеллера следовало бы пропускать без проверки на Правило предложения с деепричастиями, совпадающими по написанию с другими словоформами, или проводить более детальный синтаксический анализ.

## 6. ВОССОЗДАНЫЙ АЛГОРИТМ ПРОВЕРКИ НА ПРОТИВОРЕЧИЕ СКАЗУЕМОГО И ДЕЕПРИЧАСТИЯ

В результате тестирования на большом количестве фраз и с использованием всевозможных подстановок мы воссоздали в общих чертах алгоритм проверки на “несовпадение субъекта сказуемого и деепричастия”, задействованный в автокорректоре текстового редактора MS WORD.

Каждое предложение из проверяемого текста, вызывающее сообщение о “противоречии сказуемого и деепричастия”, можно разделить на следующие **сегменты**:

1) “деепричастный оборот” (может быть, мнимый);

- 2) “действующий субъект” в косвенном падеже — с предлогом или без оного;
- 3) остаток (“хвост”), которого может и не быть, т. е. фраза может состоять из первых двух частей и не иметь даже сказуемого, противоречие с которым декларируется.

Как нам представляется, условия, проверяемые алгоритмом правила “Неверное употребление деепричастного оборота”, выглядят следующим образом.

**Условие А. Запятая обязательна.** В очередном проверяемом предложении (похоже, что практически любой разумной длины) между первой и последней словоформой должна быть хотя бы одна запятая.

**Условие В. Первое слово — деепричастие.** Первым словом предложения должно быть такое, которое (вне контекста) допускает идентификацию его как деепричастия.

**Исключения:** словоформы *для* и *благодаря* воспринимаются системой ОРФО только как предлоги, а не как деепричастия от *длить* и *благодарить*; словоформа *моя* не считается деепричастием от глагола *мыть*; словоформу *какая* алгоритм ни в каком контексте не воспринимает как деепричастие (хотя соответствующий глагол в системном словаре ОРФО имеется и никак в проверяемых текстах спеллером не подчеркивается — даже при настройке на “деловую переписку”).

Таким образом, перед деепричастием запрещается даже отрицание *не* (как и во всей первой части предложения — до запятой, см. ниже Условие С).

Кстати, можно было бы пропускать без проверки на Правило предложения с первым словом *зря* или *хотя* (но алгоритм считает их возможными деепричастиями от *зреть* в значении ‘видеть’ и *хотеть*).

**Примеры “неоднозначных” деепричастий, удовлетворяющих условию В:** омографы *буря* (от *бурить*), *Коля* (воспринимается спеллером не только как имя собственное, но и как деепричастие от *колоть*), *пища* (от *пицать*), а также *душа*, *чая* (от *чаять*), *горя*, *улав*, *сев*, *слив*, *обрыв* (от *обрывать*) и др.

**Условие С. Отсутствие “неподходящих” слов и символов до первой запятой.** То есть относится к первому сегменту фразы.

Между первым словом (деепричастием) и первой запятой не должно быть ни цифр, ни латинских букв (и некоторых других символов) и ни одной словоформы, которая (контекст алгоритмом не учитывается) могла бы квалифицироваться не как существительное, не как имя собственное, склоняющееся по образцу существительного, и не как предлог (пример: *Подвезжая к Мытищам...*). Таким образом, предлоги, существительные или имена собственные, если они входят в “деепричастный оборот”, не должны совпадать по написанию со словоформами, относящимися к остальным частям речи.

Если до первой запятой имеются “лишние” слова, то проверка предложения прерывается, т. е. сообщения о “противоречии” не будет.

**Исключения.** Предлог *при* не воспринимается алгоритмом как императив от глагола *переть*; допускаются также предлоги *на* (хотя есть частица *на* в значении ‘возьми’), *для*, *благодаря*.

Заметим, что предлоги *под* и *перед* в деепричастном обороте алгоритм допускает, в словаре ОРФО они совпадают только с существительными.

**Примеры словоформ и конструкций, противоречащих условию С:** *с Пушкиным* (эта фамилия склоняется не как нарицательное существительное), *рядом с* (*рядом* не только форма

существительного, но и наречие), *полтора*, *пол* (очевидно, в системном словаре — не только существительное, ср. *пол Московской области*), *посреди* (предлог и наречие), *вроде* (предлог и частица), а также сочетания *без устали*, *на правило*, *кроме Жени* — даже присутствие предлогов не мешает алгоритму считать вторые члены этих пар возможными глагольными формами от *устать*, *править* и *женить*. В конструкции *расплачиваясь за объём* спеллер не принимает во внимание даже букву *ё* в слове *объём* и считает его возможной личной формой глагола *обвести* (*Я вас не обведу?*).

**Условие D. Одушевленный “действующий субъект” в косвенном падеже после “деепричастного оборота”.** Относится ко второму сегменту фразы. Непосредственно после первой запятой и других разделителей (скобка, кавычка, тире и т. п.) может стоять любой несоставной предлог, даже предлог “о”, сочетающийся только с предложным падежом (*о Вас*, *о нас*, например).

**Сразу после предлога или вместо него** должно находиться слово, которое может быть квалифицировано как **нарицательное существительное** (написание с заглавной буквы игнорируется — см. пример (13) со словоформой *Льва*) или однозначно определяемое местоимение.

Существительное или местоимение не должно склоняться по образцу прилагательного, иначе проверка на Правило прерывается.

Существительное или местоимение должно иметь форму **одновременно родительного и винительного падежа**, что как бы обеспечивает одушевленность — по схеме *с меня (с путешественника) слетела*. При этом словоформа не должна совпадать с формой именительного падежа того же (например, как в несклоняемых существительных) или другого слова (как, например, словоформа *математика*).

Таким образом, сразу после деепричастного оборота допускается отсутствие предлога. Предположительно, схема такова: *Подвезжая к станции, меня (пассажира) удивило скопление народа*. Но не *его*, *ее*, *их удивило*, иначе подчеркивание снимается, так как эти местоимения одновременно являются притяжательными.

**Примеры словоформ, удовлетворяющих условию D:** *царя* (совпадение с деепричастием во втором сегменте не возбраняется), *жокея*, *кукол* (*кукла* — одушевленное существительное), *медсестер*, *старост*, *олений* и т. п. Местоимения: *кого*, *никого*, *тебя*. Сразу после деепричастного оборота в подчеркнутую фразу можно вставить *него*, *нее*, *них* даже без предлога и получить все тоже единственное сообщение о “противоречии сказуемого и деепричастия”, хотя эти формы употребляются только с предлогом. Такую простую проверку, очевидно, забыли предусмотреть.

**Примеры словоформ, одновременно относящихся и к родительному, и к винительному падежу, но не удовлетворяющих условию D:** *прохожего*, *заведующего*, *дежурного*, *учителя*, *критика*, *белок*, *чаек* (последние четыре совпадают по написанию с существительными в им. падеже), *голубей* (из-за совпадения не с императивом глагола *голубеть*, а со сравнительной степенью прилагательного *голубой*), *гуру*, *кенгуру*, *крокодилов*, *попов*, *дедов* (последние три словоформы могут быть притяжательными прилагательными — *крокодиловы слезы*, *попова дочка*, *дедово ружье*), а также *его*, *ее*, *их*, так как эти местоимения могут выполнять функцию определения, что не подходит под выбранную схему.

Сведения об одушевленности, имеющиеся в словаре А.А. Зализняка, в алгоритме проверки на рассматриваемое Правило не используются. Это приводит к тому, что, например, словоформы в винительном падеже *работницу*, *мужчину* под Условие D не подпадают (нет совпадения с род. падежом), а *работниц*, *мужчин* — подпадают. Это не выглядит логичным.



Таким образом, алгоритм прекращает проверку фраз, в которых второй сегмент начинается не с существительного или местоимения (с предлогом или без) в “родительно-винительном”, если так можно выразиться, падеже. Однако можно придумать много некорректных фраз с противоречием сказуемого и деепричастия, где “действующий субъект” стоит в другом косвенном падеже: творительном (*между солдатами*), как в примере (3), дательном (*Прибыв в Москву, туристам было предложено разместиться в общежитии*) или предложном (*... о туристах некому было позаботиться*).

**Условие Е. Отсутствие “претендента” на роль одушевленного подлежащего после первой запятой и до конца фразы.**

Предложение, в котором в правой части (второй и третий сегменты) можно по формальным признакам найти кандидата на роль одушевленного подлежащего, не подходит под схему выбранного разработчиками образца *... с меня слетела шляпа* и далее алгоритмом не рассматривается. Это условие не противоречит предыдущему Условию D.

Проверку на условие Е можно разбить на две составляющие.

1). В оставшейся части предложения — после первой запятой (завершающей “деепричастный оборот”) — не должно быть ни одной словоформы, которая совпадала бы по написанию с одушевленным нарицательным существительным в именительном падеже, не склоняющимся по образцу прилагательного.

Исключение. Слово *уж* в третьей части предложения (после “действующего субъекта в косвенном падеже”) допускается, так как алгоритм не считает его возможным существительным (см. выше пример (14)).

Одушевленным алгоритм считает любое нарицательное существительное, у которого хотя бы в одном из его значений (в одной из парадигм склонения) форма винительного падежа множественного (или единственного) числа совпадает с формой родительного падежа того же числа. Однако под это определение формально подпадают и все несклоняемые существительные, в том числе неодушевленные. Присутствие предлога перед существительным не принимается во внимание, поэтому любое несклоняемое существительное даже с предлогом (*у леди, вместо кимоно*) определяется как возможное подлежащее (что является явной недоработкой).

2). После первой запятой и до конца фразы (а вернее, во всей фразе — см. Условие С) не допускаются также и “одушевленные” местоимения в именительном падеже *кто, кое-кто, некто, никто, я, оно, мы* и др. (проверка прерывается). Алгоритм накладывает запрет на такие местоимения, которые в принципе могут выполнять функцию одушевленного подлежащего и не могут интерпретироваться как определения, эквивалентные прилагательным.

Очевидно, что “одушевленные” местоимения заданы списком, так как в издаваемых словарях они обычно не имеют помет об одушевленности.

**Примеры словоформ, формально не противоречащих условию Е:** *правофланговый, заведующая, многие, некоторые, все, сами* (адъективное склонение), *Анна, Иван, Смирновы*. Эти слова (падеж — именительный) могут находиться в третьем сегменте предложения, после “действующего субъекта”, не препятствуя дальнейшей проверке. Однако они вполне корректно могут выполнять функцию одушевленного подлежащего, чего алгоритм рассматриваемого нами Правила не учитывает.

Очевидно, что имена собственные в системных словарях текстового редактора MS WORD не снабжены пометами об их грамматической одушевленности или неодушевленности (эта информация имеется в [Зализняк 2003]).

**Примеры словоформ, не удовлетворяющих условию E:** *табу, кафе, виски* (есть несклоняемое, но есть и форма от *висок*), *пари* (даже если во фразе это императив от глагола *парить*), *проводник, стрелок, головы* (есть одушевленные значения), *лис, змей, доктора*, а также *матери, цари* (даже если во фразе это императивы от глаголов *материть* и *царить*). Алгоритм прекращает анализ предложения с такими словоформами, находящимися во втором и третьем сегментах (даже с предлогом и сразу после “деепричастного оборота” — см. Условие D).

**Условие F (немного туманное). Отсутствие во фразе некоторых конкретных цепочек символов.** Если в одном из трех сегментов предложения, т. е. до первой запятой, сразу после первой запятой (одна или две словоформы) или в оставшейся части встретились некоторые заданные тремя списками слова или конструкции, не запрещенные в предыдущих условиях, то сообщение о “противоречии сказуемого и деепричастия” не выдается.

**Примеры слов, не удовлетворяющих условию F:** после “действующего субъекта в косвенном падеже” (т. е. в третьем сегменте) — словоформы *его, ее, их, крокодилов, попов, дедов*. Этот запрет не выглядит обоснованным. Очевидно, разработчики алгоритма допустили небрежность и запретили такие словоформы во всем предложении, а не только во втором сегменте (см. выше Условие D).

Возможно, что запрещаются (препятствуют выдаче сообщения о противоречии) еще какие-либо словоформы, словосочетания или цепочки символов, которые нам пока не удалось обнаружить.

Итак, мы выявили шесть не противоречащих друг другу условий, из которых можно составить конъюнкцию. Таким образом, нам представляется истинным следующее утверждение.

**Если A & B & C & D & E & F, то выдается сообщение программы-подсказки, ссылающееся на правило “Неверное употребление деепричастного оборота”.**

Очень вероятно, что верна и обратная импликация, т. е. полученная конъюнкция из шести составляющих есть необходимое и достаточное условие для подчеркивания фразы и выдачи сообщения о “противоречии сказуемого и деепричастия”.

Приведем последний пример.

Царя, царя? (16)

Это образец самой короткой фразы — из тех, которые подчеркиваются со ссылкой на протестированное нами Правило. Здесь первая словоформа *царя* определяется алгоритмом Правила как деепричастие от глагола *царить*, а вторая — как форма винительного и одновременно родительного падежа одушевленного существительного *царь*. “Лишних” слов в примере нет, поэтому все шесть условий выполнены.

## 7. НЕКОТОРЫЕ ВЫВОДЫ

Итак, мы высветили содержимое “черного ящика”, т. е. в значительной степени уяснили себе алгоритм работы программы по выявлению предложений русского языка, на которые WORD’овский спелл-чекер считает нужным обратить внимание пользователя, так как предполагает в них “несовпадение субъекта сказуемого и деепричастия”.

Восстановленный алгоритм оказался не вполне адекватным задаче, поставленной в тексте заявленного системой ОРФО правила “Неверное употребление деепричастного оборота”, однако некоторые из недостатков этого алгоритма, на которые было указано выше, легко могут быть исправлены разработчиками в следующих версиях автокорректора.

Приведем еще одну цитату из вышеупомянутой статьи А. Майоровой: “Грамматические программы фиксируют норму, которая может и нормой-то не быть. Во-первых, эти программы пишутся вовсе не в Институте русского языка лингвистами, а в частных компаниях программистами. Во-вторых, нормы зачастую вообще не существует. Например, написание слов с дефисами в русском языке до сих пор является туманной областью, в том числе и для лингвистов. “Ваш покорный слуга, — говорит Игорь [Ашманов], — составил 150-тысячный словарь программы ОРФО, которая входит в русскую версию MS Office. Миллионы пользователей по сей день каждодневно правят орфографию так, как предписал им не только использованный мною словарь А.А. Зализняка, но и лично я в 1989–1994 годах. А ведь при разработке ОРФО было довольно много спорных случаев”. Разработчика можно понять. Не нанимать же лингвистов, привыкших к тонким исследованиям по одному слову в месяц и потому непригодных к сумасшедшим темпам ведения проекта”.

Если чей-то дядя “самых честных правил”, то спеллер текстового редактора MS WORD правит даже самых грамотных. Однако не всегда это у него получается без накладок.

Из всего вышеизложенного можно вывести следующие рекомендации для разработчиков новых версий автокорректоров:

- более подробно формулировать грамматические правила и исключения из них, обращая внимание пользователей на случаи, в которых программа-подсказка может выдавать не совсем правильные рекомендации;
- более детально анализировать синтаксические конструкции, отлаживая алгоритмы на примерах из лингвистической литературы;
- внести пометы об одушевленности в системный словарь имен собственных или использовать новый словарь имен собственных, которым дополнено последнее издание “Грамматического словаря” [Зализняк 2003], не забывая при этом об авторском праве;
- в алгоритмах проверки правильности синтаксических конструкций максимально использовать информацию об одушевленности/неодушевленности существительных и о переходности/непереходности глаголов, имеющуюся в словаре А.А. Зализняка.

Остается только выразить надежду, что следующие версии текстового редактора MS WORD будут больше опираться на достижения лингвистической науки.

#### СПИСОК ЛИТЕРАТУРЫ

1. Успенский В. А. *Труды по нематематике. С приложением семиотических посланий А. Н. Колмогорова к автору и его друзьям.* — М.: ОГИ, 2002. Т. 2.
2. Розенталь Д. Э. *Справочник по русскому языку. Пунктуация.* — М.: “ОНИКС 21 век”, 2003.
3. Лавошникова Э. К. О “подводных камнях” в компьютерных системах проверки правописания // *Вестник Московского университета. Серия 9. Филология.* 2002, № 6, стр. 151–162.
4. Лавошникова Э. К. О компьютерной коррекции “популярных” ошибок в текстах на русском языке // *Научно-техническая информация. Серия 2. Информационные процессы и системы.* 2003, № 9, стр. 28–34.
5. Зализняк А. А. *Грамматический словарь русского языка: Словоизменение. Ок. 110 000 слов. — 4-е изд., испр. и доп.* — М.: “Русские словари”, 2003.

*Статью представил к публикации член редколлегии В.И.Венец*