

Э.К. Лавошникова

## WORD И МАЛОУПОТРЕБИТЕЛЬНАЯ ЛЕКСИКА В СЛОВАРЯХ ТЕКСТОВОГО РЕДАКТОРА

Рассматривается работа текстового редактора MS Word 2016 и его программы-«подсказки» под названием «корректор» (модификация 2017–2018 гг.). На многочисленных примерах показано, что перегруженность внутренних системных словарей низкочастотной лексикой при не вполне совершенном синтаксическом контроле снижает эффективность работы spellера. Это утверждение иллюстрируется на специально (в целях тестирования системы) сконструированных фразах, изобилующих намеренными ошибками и опечатками. Приводятся также списки малоупотребительных слов и архаизмов, которые с большей вероятностью могут появиться в результате ошибок или невнимательности при наборе современных текстов в русском регистре, но пропускаются Word без замечаний.

**Ключевые слова:** Microsoft Word; WinWord; компьютерная проверка правописания; текстовый редактор; spellер; устаревшая лексика; орфографические ошибки; нормативное написание; архаизмы; русский язык.

1. Текстовый редактор Word в разных версиях широко распространен, эффективно применяется для компьютерной проверки правописания и заслуженно пользуется большим авторитетом (иногда даже непререкаемым, что, как мы увидим далее на конкретных примерах, не всегда оправдывается). Поставив перед собой задачу протестировать новую версию (модификацию 2017–2018 гг.) системы Microsoft Word 2016, будем сопоставлять результаты работы его spellера по выявлению ошибок и опечаток с рекомендациями наиболее авторитетных словарей русского языка, изданных в последнее время.

Написание приводимых примеров будем сверять с Русским орфографическим словарем [1] (далее РОС), Орфографическим словарем русского языка Б.З. Букчиной, И.К. Сазоновой и Л.К. Чельцовой [2] (далее БСЧ), а также (в некоторых случаях) с 6-м изданием Грамматического словаря русского языка А.А. Зализняка [3] (далее ГСРЯ) и толковым двухтомным Новым словарем русского языка Т.Ф. Ефремовой [4] (далее НСРЯ).

Заметим при этом, что словарь БСЧ и одно из предыдущих изданий словаря Зализняка входят в список четырех словарей, грамматик и справочников, рекомендованных в 2009 г. Межведомственной комиссией по русскому языку при Минобрнауки и содержащих нормы современного русского литературного языка.

Необходимо отметить, что один из самых первых spellеров, получивших широкое распространение под названием «Орфо», базировался на 1-м издании словаря А.А. Зализняка (1977 г.). Слова во всех изданиях этого словаря расположены в обратном (инверсионном) порядке, по алфавиту конечных букв<sup>1</sup>. Каждое слово снабжено формализованной пометой, отсылающей к грамматическим таблицам. В 1980-х гг. коллективом Лаборатории автоматизированных лексикографических систем НИВЦ МГУ им. М.В. Ломоносова словарь был размечен – при скудном, в отличие от типографского, наборе символов на тогдашних ЭВМ – и перенесен на машинные носители (бобины с магнитной пленкой). Электронная версия словаря Зализняка легла в основу большинства современных компьютерных программ, работающих с русской морфологией, в том числе и системы MS Word.

Справедливости ради следует отметить, что в текстовом редакторе Word 2016 устранены многие огрехи, имевшие место в предыдущих версиях системы. Однако к компьютерному анализу согласования слов (во фразе и в отдельных словосочетаниях) вопросов у пользователей систем Word 2013 и последующих стало только больше – по сравнению с предыдущими версиями, основанными на программе «Орфо» (см., например, [5. С. 59]<sup>2</sup>). И даже в выявлении орфографических ошибок (наиболее разработанное направление), а также в рекомендациях по их исправлению, выдаваемых программой-«подсказкой» «Корректор», остается немало разного рода недочетов, что будет далее показано на примерах.

2. При компьютерной проверке правописания в режиме реального времени spellер текстового редактора MS Word (WinWord) *красной волнистой линией* подчеркивает словоформы (или иногда даже части слова при написании через дефис), *отсутствующие* в его системных словарях. Тем самым пользователю предлагается обратить внимание на подчеркнутое красным слово – нет ли в нем ошибки или опечатки. Программа-«подсказка» при ее вызове выдает варианты исправления неопознанного слова, если может их найти в своих внутренних словарях. *Синие линии* в текстовом редакторе Word 2016, а также в Word 2013 (или *зеленые* в предыдущих версиях, разработанных на основе системы «Орфо») отмечают *предположительные* синтаксические огрехи – ошибки в пунктуации или согласованности словоформ в словосочетаниях.

Подчеркиваемые *красной волнистой линией* слова из наших примеров мы *тоже будем подчеркивать*. Если словоформа нами не подчеркнута, то это означает, что она опознается и пропускается spellером без замечаний, так как содержится во внутренних системных словарях версий 2016–2018 гг. текстового редактора MS Word.

Программа «Корректор» (т.е. «подсказка» в модифицированной версии 2017–2018 гг.) с помощью замен, вставок, удаления и перестановок букв в слове, *подчеркнутом красной волнистой линией*, ищет в своих словарях полученные такими способами цепочки символов. Если находит, то выдает первые

найденные словоформы как возможные варианты исправления этого не опознанного спеллером слова. Текстовый редактор MS Word 2016 после обновлений 2017 г. стал предлагать *не больше трех* вместо выдававшихся в предыдущих версиях пяти вариантов исправления, но при этом иногда добавляются пояснения – синонимы или слова соответствующей тематики. Нередко среди предлагаемого подсказкой вылезают «подводные камни» – ошибочно включенные или порожденные в системных словарях словоформы<sup>3</sup>.

Качество проверки синтаксиса и стилистики автокорректором Word 2016, к сожалению, оставляет желать лучшего, что на многочисленных примерах показано в статье [6]. Но даже в выявлении орфографических ошибок остается немало разного рода недочетов.

3. Как будет показано далее на конкретных примерах, «богатство» внутренних словарей текстового редактора, включение в него некоторых редко встречающихся в *современных* текстах слов и словоформ иногда не только не способствует, но *препятствует* выполнению прямой задачи спеллера – выявлению ошибок и опечаток в текстах пользователей.

Задача автоматизированной проверки орфографии распадается на две подзадачи: а) указывать на ошибки и опечатки; б) по возможности не «дискредитировать» словоформы, набранные правильно (исходя из современных норм правописания). Эти два требования зачастую входят в противоречие друг с другом. Какая подзадача важнее? Думается, что первая, хотя подчеркивание красным некоторых слов и словоформ с нормативным написанием может раздражать пользователя или заставлять его сомневаться в своей грамотности. При этом подчеркивание *коротких* слов «напрягает» меньше, в то время как в длинных словоформах (не только в сложносоставных, но и, например, в причастиях или прилагательных превосходной степени) труднее заметить ошибку. Отсюда вывод – желательно делать упор на пополнение системы сложно образованными словами и с осторожностью относиться к включению во внутренние словари коротких низкочастотных слов и словоформ (и не только архаизмов), что, как мы увидим далее, может при проверке правописания приводить к пропуску ошибок и опечаток. Особенно это относится к тем словам, которые графически совпадают с началами или концами более длинных словоформ (незапланированные пробелы или дефисы в текстах пользователей не так уж редко встречаются).

С удовлетворением можно отметить, что из системного словаря наконец убраны некоторые низкочастотные слова, о чем мы писали в своих работах [7 и др.]: *переплотить* (РОС: «к плот»); *поленица* (РОС: «богатырь», наряду со словом *поленица* от *полено*); *пощепать* (от *щепка*, РОС: «к щепать», но малоупотребительное *щепать*, которое может появиться также и в результате ошибки в глаголе *щипать*, пропускается спеллером без подчеркивания); *рожество* (в РОС этого написания нет); *холява* (РОС: «заготовки стекольного производства; голенище»)<sup>4</sup>; *эстокада* (РОС: «в фехтовании»). При этом во внутреннем словаре текстового редактора остается, например, существительное *мысле* (РОС: «название буквы»). По-

этому при не вполне совершенном синтаксическом контроле система не замечает ошибки в глагольной форме *мыслите* («Зачем вы так парадоксально мыслете?») – ожидаемого синего подчеркивания как указания Word'a на неправильный синтаксис этой фразы на настоящий момент нет).

4. К сожалению, не внесены в системные словари и подчеркиваются красной волнистой линией некоторые слова (представленные в РОС), которые могут быть набраны пользователями с ошибками или опечатками. Это включение в систему особенно желательно для сравнительно длинных или заимствованных слов, например: *бэбиситтер* (*бэбиситтер* тоже подчеркивается), *блицинттервью* (*блицинттервью*, как и многие дефисные образования, часто даже неправомерно порожденные самой системой, пропускается без замечаний, но «Корректор» предлагает разделение на два слова, а не дефис), *велфер*, *возжжённный* (и *возжжённный*), *дежавю*, *имейл* (дефисное *и-мейл* не подчеркивается), *кремлиннолог* (в БСЧ такое же написание), *лабораторский*, *микроЭВМ* (и *суперЭВМ*), *недобдеть* (*перебдеть* не подчеркивается), *нерукопожатный*, *прокремлёвский* (однако *прокремлевский* в системном словаре имеется), *стопятидесятилетие* (склоняется без буквы «а»: «к *стопятидесятилетию*», «о *стопятидесятилетии*», но «Корректор» выдает только один вариант «исправления» – пробел после «сто»), *телегеничность*, *харассмент*, *чрезвычайщина* и др.

Спеллер текстового редактора Word отмечает как ошибочное рекомендуемое современными словарями (в том числе ГСРЯ, БСЧ, РОС) написание *мелочовка* и *плащовка*. При вызове программы «Корректор» предлагается бывшее не так давно нормативным *мелочёвка* (и *мелочевка*). А для *плащовка* эта подсказка выдает максимально возможные три варианта «исправления»: *плащика*, *плутовка*, *плюсовка*.

Подчеркиваются красным и некоторые достаточно часто встречающиеся существительные с префиксоидом «не-», в которых программа «Корректор» 2017–2018 гг. (как и подсказка-2016) предлагает «не» отделять пробелом. Примеры: *неаппетитность*<sup>5</sup>, *невозрастание*, *негуманность* (РОС), *неединственность*, *неидеальность*, *несжимаемость* (соответствующие прилагательные с «не-» в системном словаре имеются). Для слова *невоспроизводимость* (РОС) у спеллера «нет ни одного предложения», хотя слово без «не» *воспроизводимость* в его системном словаре имеется. Не включены во внутренний системный словарь некоторые «женские» варианты существительных, в которых пользователями могли бы быть допущены ошибки, примеры: *риелторша* и *халявищица* (представлены в РОС).

В системном словаре отсутствуют (не порождаются) многие формы превосходной степени, хотя в таких длинных словах более вероятны опечатки. Примеры: *аппетитнейший*, *бездарнейший* («Корректор» предлагает «исправление» *бездомнейший*), *изысканнейший*, *наивыгоднейший* (РОС), *обаятельнейший*, *отменнейший*, а также *свежайший* (РОС) и *свежеший* (второй вариант в РОС), но только *наисвежайший* пропускается без подчеркивания.

Конечно, пользователь может пополнить словарь для себя лично (в Word предоставлена такая возможность) любыми словоформами – *терминами*, не включенными в системный словарь, и прочими используемыми им низкочастотными или особенными словами, если их подчеркивание при проверке текстов его не устраивает. Приведем пример: существует зоологический термин *бадяга* (РОС), для которого подсказка предлагает исправление *бодяга* (РОС: «водоросль; разводить бодягу, кончай бодягу»). Можно сделать так, чтобы не подчеркивались оба этих написания – но не для других пользователей.

Нередко в текстах «по техническим причинам» происходит *склеивание* соседних слов. Спеллер текстового редактора WinWord не всегда замечает такие огрехи – в результате присутствия в его системных словарях некоторых малоупотребительных лексем. Например, пропускаются и не подчеркиваются глагольные формы *выбрасаете* и *вывершите*, причастие *наседававшие* – от глаголов *выбросать*, *вывершить* и *наседать*, которых нет даже в РОС [1], одном из самых представительных орфографических словарей.

Приведем довольно любопытный факт. В некоторых предыдущих версиях Word были заблокированы «нехорошие» слова: *голубой*, *розовый*, *залететь* (явно имелось в виду не первое значение, но глагол «реабилитирован» системой в 2018 г.), *негр*, *осёл* (написание *осел* тоже было убрано, причем вместе с глагольными формами *осела*, *осело*, *осели*) и др. После многочисленных критических замечаний некоторые слова все же были возвращены, но иногда без своей полной парадигмы: слова *дура* и *шлюха* уже есть, но словоформы *дур* и *шлюх* подчеркиваются красным; слово *подлец* в версии Word 2016 не склоняется (возможно, по ошибке оно приобрело статус междометия); слово *сволочь* спрягается в качестве глагола (*сволочу*, *сволочешь*), но косвенные падежи существительного отсутствуют (*сволочью*, *сволочам* и т.д.).

**5. Примеры специально сконструированных фраз с многочисленными ошибками, пропускаемых системой без подчеркиваний** синим или красным (см. ниже), демонстрируют некоторые недочеты в работе WinWord. В них мы видим склейки и разбиения слов, а также другие вероятные орфографические ошибки и опечатки, в результате которых получается *малоупотребительная* лексика, имеющаяся в системном словаре текстового редактора Word и пропускаемая по этой причине без красного подчеркивания. Замечаний к сочетаемости слов во фразах (подчеркиваний *синими линиями*) в этих примерах тоже, к сожалению, на данный момент нет.

1) *Не лъзя ли раз обратиться справами сroitъ трасу, кото рая не пода рок для эко логиш наши его кря.* В результате намеренных «опечаток» у нас появились следующие словоформы: отсутствующие в наиболее представительном из цитируемых нами в данной статье словарей РОС *лъзя* (!), *обратъся* и *трас* («горная вулканическая порода»); форма существительного *справа* (в толковом НСРЯ одно из значений – «одежда»); вместо *сroitъ* – *сroitъ* (НСРЯ: «соединить в один рой»); *кото* (РОС: «японская арфа»); родительный падеж существительного *под*; *эко*, которое в РОС

дано только прописными буквами; форма от сомнительного существительного *логиш* (в РОС отсутствует); междометие *кря*, получившееся в результате «пропавшей» буквы «а».

2) *Молод ежь брада пре доставленные отпуска, но по том буде за числена ректор ом.* Здесь «по ошибке» возникли такие словоформы: императив *ежь* от глагола *ежить* (*ёжить*); архаизм *брада* (борода) вместо *брала* – результат соседства на клавиатуре «л» и «д»; *пре* – дательный падеж архаизма *пря* (НСРЯ: «устар.: спор, борьба»); союз *буде* – при обрезании высокочастотной словоформы *будет*; краткая форма от *численный*; единица измерения *ом*.

3) *Сморите, как глад ко был изложин отче задругой пери од! Надо работать на советъ!* В результате наших «опечаток» получен маловероятный императив глагола *сморить*, а также формы следующих существительных: *глад* (РОС: «устар.; в поэзии»); низкочастотное *изложина* (НСРЯ: «местн.; лог, ложбина»); звательная форма *отче* от *отец* при отсутствии второй буквы «т» в слове *отчет*; *задруга* (НСРЯ: «патриархальная большая семья»); *перн* (НСРЯ: «перен. устар.»); *ода*. Во второй фразе после «на» получился инфинитив вместо существительного *совестъ*.

4) *Вы мня обиде ли. Я науцу на вас всю пресс-у. Я всегда миу за так ой дрему чий и лжи вый материл.* В первой фразе возникло деепричастие от глагола *мять* вместо *меня*. Во второй фразе получилась форма устаревшего глагола *науцитъ* (*науцать*) вместо *науцу* и опять разбиение словоформы дефисом. В третьей фразе: личная форма низкочастотного глагола *мишитъ* вместо *миу* («ш» и «щ» плохо различимы визуально и расположены на клавиатуре рядом); междометие *ой*; *чий* (РОС: «растение»); форма архаизма *выя*; форма глагола *материть* вместо *материал*.

5) *Все млеко питающие обо шлись без царап ин и уча щенного пульс а. После про верки до статочно дат-ь втору ю дозу вит аминов.* Из-за присутствия в системном словаре архаизма *млеко* (РОС: «устар. поэт. к молоко») без каких-либо предупреждений системой пропускается ошибочное разбиение пробелом. Далее в первой фразе: *царап* (БСЧ: «в знач. сказ.»); частица *ин*; форма мужского рода от *щённый* (ГСРЯ, но в БСЧ, РОС и НСРЯ слово дано только в женском роде, *щённая* – БСЧ: «о собаке, волчице, лисе»). Во второй фразе: *верки* (НСРЯ: «устар., оборотительные постройки»); форма от *статочный* («Статочное ли дело?»); мягкий знак через дефис; *втора* (НСРЯ: «второй голос в музыкальной партии»); буква «ю» – все без исключения *одиночные буквы* во фразах спеллером Word не подчеркиваются; краткая форма от *витый*; *амин* (ГСРЯ), но в РОС этого существительного нет.

6) *Он пой дет на нас воной и зало жит после рамена все наши ссуд ы!* Здесь появились: краткая форма страдательного причастия от *детъ*; *вона* (РОС: «ден. ед.»); *зало* (РОС: «устар. к зал»); форма родительного падежа множественного числа существительного *жито* (НСРЯ: «местн.»); архаизм *рамена* (РОС: «плечи»); одиночная буква «ы» в конце фразы.

7) *Мой ответ: это черти-что, что за дела в та ких фирм ах!* При «ошибочном» удвоении гласной

получилась личная форма глагола *отвелять* вместо существительного *ответ*. Нередко пользователи набирают «черти-что» (РОС: *чёрт-те что*), но Word на эту ошибку не реагирует. Далее: сомнительное деепричастие прошедшего времени *делав* в результате склейки *дела* с предлогом «в»; непонятное *ких* – в Интернете находим только аббревиатуру КИХ; междометие *ах* в конце фразы.

8) *Живи посредствам: при готовься к ужину уж е сутра!* При склейке с предлогом получилась форма множественного числа существительного *посредство*; *при* – предлог или императив глагола *переть*; при пропуске гласной «и» в словоформе *ужину* возникла личная форма глагола *ужинать* (с ударением на «а», от *жать*); при склейке с предлогом получилось слово *сутра* (ГСРЯ) – в приложении «Имена собственные» этого словаря имеется словарная статья *Камасутра* [3. С. 746].

Подобные фразы с наиболее вероятными ошибками и опечатками могут быть использованы при тестировании новых версий текстовых редакторов. Другие примеры некорректных фраз и синтагм со специально подобранными ошибками согласования, управления и примыкания – для тестирования автокорректоров – приведены, например, в [7. С. 121].

6. Опытным путем было обнаружено довольно много *малоупотребительных* слов, которые в текстовом редакторе Word 2016 (как и в предыдущих его версиях) пропускаются без подчеркивания, т.е. без указания на возможную опечатку, но могут появляться по ошибке: в результате *пропуска буквы* (например, при недостаточном сильном нажатии клавиши), при нажатии соседней с нужной клавиши и т.п.

**6.1. Примеры инфинитивов глаголов, отсутствующих в нормативном словаре БСЧ,** но не подчеркиваемых spellером системы Word, которые *с большей вероятностью* могут быть результатом *пропуска одной буквы*: *ботать* (*болтать*), *взмутить(ся)* (*возмутить(ся)*), *взнести(сь)* (*вознести(сь)*), *взрастать* и *взрасти(сь)* (*возрастать* и *возрасти(сь)*), *впивать* (*выпивать*), *вскальзывать* и *вскользнуть* (*выскальзывать* и *выскользнуть*), *вишбать(ся)* (*выишбать(ся)*), *вывить* (*выявить*), *высочить* (*высочить*), *вытраивать(ся)* и *вытроить(ся)* (*выстраивать(ся)* и *выстроить(ся)*), *густить(ся)* (*грустить*, *сгустить(ся)*), *завзаться* (*завезаться*), *запреть* (*запереть*), *захват* (*запахать*, *запахать*), *затесняться* (*застесняться*), *заушить* (*засушить*, *задушить*), *икриться* (*искриться*), *мрачить(ся)* (*омрачить(ся)*), *надвить* (*надавить*), *наторить* (*натворить*), *окорить* и *окорять(ся)* (*окормить*, *покорить* и *покорять(ся)*), *опороть* (*отпороть*), *оскользнуться* (*поскользнуться*), *оскрести(сь)* (*отскрести(сь)*, *поскрести(сь)*), *остареть* (*постареть*), *отелить* (*отделить*, *отселить*), *отоптать* (*оттоптать*, *потоптать*), *плотнить* (*уплотнить*), *повестить* (*оповестить*), *подирать(ся)* (*подбирать(ся)*), *подтирать(ся)*, *продирать(ся)*), *подлеть* (*подлезть*), *подсочить* (*подсочить*), *подустить* (*подпустить*, *подгустить*), *помститься* (*поместиться*), *попевать(ся)* (*подпевать(ся)*), *поспевать*, *поплевать(ся)*), *попеть* (*поспеть*, *пропеть*, *подпеть*), *постыть* (*простыть*, *поостыть*), *посунуться* (*просу-*

*нуться*, *подсунуться*), *потцититься* (*потацититься*), *престать* – в РОС отсутствует (*перестать*, *предстать*), *примститься* – в РОС нет (*примоститься*), *пробрести(сь)* (*приобрести(сь)*), *проглотать* (*проголотать*), *разоняться* – в РОС нет (*разгоняться*), *ристать* – в РОС нет, в словаре Даля ‘скакать, ездить’ (*пристать*), *сбирать(ся)* (*собирать(ся)*), *спахать* (*вспахать*), *спить* (*споить*, *испить*), *ставать* (*вставать*, *уставать*), *стаскать(ся)* (*истаскать(ся)*), *тихнуть* (*стихнуть*), *утихнуть*), *тлмить(ся)* (*томить(ся)*), *тропить(ся)* (*торопить(ся)*), *умесить* (*уместить*), *умирить(ся)* (*усмирить(ся)*), *усевать(ся)* (*успеивать*, *усеивать(ся)*), *утонить(ся)* и *утонять(ся)* (*уточнить(ся)* и *уточнять(ся)*), *хладеть* (*охладеть*).

**6.2. Примеры слов** (не только инфинитивов), **отсутствующих в нормативном словаре БСЧ,** не подчеркиваемых spellером системы Word, которые с большей вероятностью могут быть результатом *ошибки* или опечатки – *пропуск*, *непроизвольная замена*, *перестановка букв* и т.п. – в более употребительных словах: *вдёржка* и *вдержжка* (*выдержка*), *ветреть* (*вертеть*), *включатель* (*выключатель*), *возращенный* от *возрастить* (*возвращенный*), *впечь(ся)* (*в печь*, *выпечь(ся)*), *вымеищать* (*вымеищать*), *выпушенный* (*выпущенный*), *выскачу* от *выскакать* (*выскачу*), *достегать* и *застегать* от *стегать* (*достигать*, *застигать*), *завялю(сь)* от *завялеть(ся)* (*заявлю(сь)*), *зажуриуть(ся)* (*зажмурить(ся)*), *задуриуть* – «д» и «ж» на клавиатуре находятся рядом), *замета* (*заметка*), *зацепна* (*зацепка*), *казанное* (*оказанное*, *сказанное*, *указанное*), *купный* (*крупный*, *купанный*), *лыбиться* – в РОС отсутствует (*дыбиться* – при соседстве «л» и «д»), *млат* – в РОС нет, НСРЯ: «устар.», ‘молот’ (*мулат*), *нажевать(ся)*, *обжевать*, *отжевать*, *пережевать*, *сжевать(ся)* от *жевать* (*наживать(ся)*, *обживать*, *отживать*, *переживать*, *сживать(ся)*), *нервоз* – РОС: «нервное состояние» (*невроз* – РОС: «болезнь»), *огнь* – РОС: «устар., книжн. и поэтич.» (*огонь*), *отчина* – НСРЯ: «устар.; вотчина» (*отчизна*, *вотчина*), *падина* – НСРЯ: «падь» (*впадина*), *плотить(ся)* – РОС: «к плот» (*латить(ся)*), *поветь* – РОС: «помещение в крестьянском дворе, обл.» (*повесть*, *повеять*), *подог* – РОС: «палка, дубинка, трость, обл.» (*подлог*, *полог*, *поджог*), *подрожать* от *дрожать* (*подражать*, *подорожать*), *покланяться* (*поклоняться*), *полсть* – РОС: «устар. и обл. к полость – покрывало на ноги в экипаже» (*полость*), *попуск* – в РОС отсутствует (*пропуск*, *подпуск*), *постоять* – в РОС нет (*постоять*), *преставление* (*представление*), *придельный* – РОС: «от придел» (*предельный*), *приеденный* (*приведенный*), *сдой* – в РОС нет (*седой*, *содой*, *слои*), *севр* – РОС: «фарфор» (*север*), *середа* (*череда*: «ч» и «с» расположены рядом), *смериться* (*смириться*), *спряжённый* (*сопряжённый*), *средина* (*середина*), *сточка* (*строчка*), *уловный* (*условный*), *улыба* – в РОС нет, НСРЯ: «разг.; тот, кто много или часто улыбается» (*улыбка*), *умеищать* (*умеищать*), *утонение* (*уточнение*), *ухолить* (*уходить*), *четь* – архаизм со значением ‘четверть’ (*честь*, *сеть*).

7. Низкочастотные слова в словарях, как мы видим, иногда снабжены пометами «поэт.» или «поэтич.». Об

одной из особенностей практики стихосложения приведем цитату: «Стиховедами давно замечено, что стих фонетически плотнее, чем проза: в нем слова короче, и в равный слоговой объем их помещается больше» [8. С. 42]. Ниже даются примеры с подчеркиванием красной волнистой чертой при их цитировании из 6-й книги серии «Поэзия Московского университета: от Ломоносова и до...»<sup>6</sup>.

1) *Воздух под полдневных сил приливом...* [9. С. 58].

2) ...*Чтоб под шептанье вешних верб / Все люди обнялись как братья...* [Там же. С. 289]. Подсказка «Корректор» выдает: *шептунье, шептуньи, шептунья*, но только не *шептание*. Некоторые слова на -ье, представляющие собой варианты существительных с окончанием «-ие», в системный словарь WinWord не включены (пример: есть *создание*, но весьма употребительный вариант со стяжением *созданье* подчеркивается красным). Обратный случай: *веселье* и отсутствующий в цитируемых здесь словарях, но пропускаемый Word без замечаний архаичный вариант *веселие*. Пропускается также сомнительная форма *людьми* (наряду с нормативной формой *людьми*). Если имеется в виду старое название буквы (*люди*), то оно в словарях отмечено как несклоняемое.

3) ...*И упал, и задохся в крови императоров – зверь полудикий...* [Там же. С. 104]. Программа-подсказка слово *задохнулся* не предлагает (первый вариант – *задайся*).

4) ...*Поутру, помолясь на восток...* [Там же. С. 292]. «Корректор» деепричастия *помолившись* не выдает – замена -ясь на -ившись при поиске спеллером исправлений в системе пока не предусмотрена.

Приведем другие примеры кажущихся вполне нормальными деепричастий прошедшего времени от глаголов *совершенного вида*, часто встречающиеся (особенно в поэзии), но не признаваемые Word: *склонясь* и *склоня* (подсказка форму *склонив* не предлагает), *возвратясь*, *обратясь* (но в подсказке «Корректор» один из трех вариантов – *обретясь*), *остановясь* (в подсказке есть вариант «о становясь»), *превратясь*, *сократясь*, *утомясь* (префикс «у» спеллер предлагает отделить) и др. Нет в словаре системы Word формы *отстав* (есть только *отставши*), хотя *отстать* спрягается так же, как *достать* – с деепричастием *достав(ши)* [3. С. 131].

Сравнительная степень *нежней*, часто встречающаяся в стихотворных текстах, подчеркивается красным, но *нежнее* среди трех выдаваемых вариантов исправления не содержится.

В системном словаре Word нет некоторых вариантов форм существительных на -ою/-ею (ёю): *башкою*, *враждою*, *листвою*, *судьбою*, *тоскою*, а также *пищею*, *струею* (*струёю*) и др. В грамматическом словаре Зализняка (ГСРЯ), положенном в основу текстового редактора, читаем: «В данном подразделе все формы Т. ед. с окончанием **-ой**, **-ёй** или **-ей** имеют параллельный вариант с окончанием **-ою**, **-ёю**, **-ею** <...> В целях сокращения этот вариант в парадигмах не выписывается. В современном языке, в отличие от языка XIX в., вариант на **-ою**, **-ёю**, **-ею** употребляется очень редко

(преимущественно в поэзии)» [3. С. 45]. Однако в системе WinWord это пояснение учитывается не во всех случаях. Пример из стихов Николая Бухарина: *И с нами тихо говорит / Улыбкою Джоконды* [9. С. 98].

Другие примеры из антологии «Поэзия Московского университета: от Ломоносова и до...» на лексические и грамматические особенности поэтических произведений, во многих случаях диктуемые практической стихосложением, см. в статье [10].

8. В текстах пользователей нередко можно встретить **ненормативные** словоформы, которые Word обоснованно подчеркивает красным: *ботинков*, *грузинов*, *лезгинов*, *макаронов*, *румынов*, *сапогов*, *солдатов*, *турков*, *чулков*. Подсказка «Корректор» не предлагает нормативных форм родительного падежа множественного числа: *ботинок*, *грузин*, *лезгин*, *макарон*, *румын*, *сапог*, *солдат*, *турок*, *чулок*. Некоторые из этих подчеркнутых слов подсказка воспринимает как фамилии и предлагает набирать с прописной буквы. Для ненормативных, но не менее «популярных» словоформ *оладьев*, *соленьев* и *ясель* исправлений *оладий*, *солений* и *яслей* «Корректор» тоже не выдает.

Имеются случаи образования в системных словарях Word разных сомнительных словоформ. В ГСРЯ [3] особыми символами отмечены прилагательные, **краткие формы от которых почти никогда не употребляются** («предположительны» или «затруднительны»). Однако в лексической базе текстового редактора MS Word такие краткие формы образуются: *дальне*, *дочерня*, *легонька*, *миленька*, *остороженька*, *прежни*, *свеженьки*, *тихонька* и т.п. Намного более вероятным представляется их возникновение в текстах в результате случайного «обрезания» конечной буквы.

9. Елена Виленовна Бешенкова так описывает положение в современной орфографии русского языка: «Рекомендации кодификатора во многом определяются его целью – обеспечить письменную коммуникацию сегодня и понимание текстов какого-то числа предыдущих поколений. Теоретически существует несколько разных способов достичь этого: сохранить существующую систему, сознательно изменить ее или вести письмо к чисто традиционному. И хотя анализ орфографической кодификации последних десятилетий показал, что нормализаторы часто шли за узусом, закрепляя противоречащие системе варианты, все-таки их целью и сегодня остается сохранение и укрепление существующей системы» [11. С. 36].

Однако здесь представляется уместным привести цитату из статьи Владимира Андреевича Успенского: «Автор (Успенский. – Э.Л.) склоняется к тому, что понятие **нормы** имеет в своей основе статистику: если “так говорит” или “так понимает” абсолютное большинство носителей языка, то это и есть **норма**. С противопоставлением **‘правильно–неправильно’** дело обстоит сложнее: единого ответа на вопрос “Что есть истина?” дать, по всей видимости, невозможно. Тем не менее каждый отдельный пример устного или письменного словоупотребления допускает оценку (скорее всего, субъективную) по шкале **‘правильно–неправильно’**. Тогда, хотя бы теоретически, возникает возможность

**неправильной нормы**, когда нечто неправильно, но все так говорят» [12. С. 539].

В подтверждение этих строк можно заметить, что в узусе исключительно популярны формы *найм* и *займ* при нормативном написании *наём* и *заём* (но подсказка «Корректор» этой нормы «не знает»). Мы видим влияние более употребительных косвенных падежей и «выравнивание» парадигмы. Варианты *найм* и *займ* настолько часто в наше время встречаются в речи и на письме даже у высокообразованных членов общества, что уже почти превращаются в норму. Ненормативное *подскользнуться* в толковом словаре Ефремовой (НСРЯ) [4] имеет помету «разг.» и толкование: «То же, что: поскользнуться». Не исключено, что придется узаконить и вариант *подскользнуться*.

Еще примеры – уже из заимствованной лексики. Разработчики WinWord пошли «на поводу» у пользователей Интернета и включили в свой словарь *кэш*, а написание *кеш* признавать не посчитали нужным (но в РОС – только *кеш* через «е»). При этом наряду с написанием *кэширование* в системном словаре Word есть и *кеширование*. Не очень логичным выглядит подчеркивание красным слов *хеш* (РОС) и *хэштег* при одновременном присутствии во внутреннем словаре вариантов *хештег*, *хэи* и *тег* (последнее из этих слов имеется в РОС) – при этом вариант написания через «э» *тег* тоже не подчеркивается.

**10.** Даже авторитетные академические словари, по нашим наблюдениям, не во всех случаях снабжают устаревшие слова соответствующими пометами. Но компьютерные системы проверки правописания MS Word характеризуются полным отсутствием разделения на употребительную и низкочастотную (в современном русском языке) лексику. Нередко включение малоупотребительных и устаревших слов во внутренние словари spellера, как было показано выше, не только не способствует, но фактически во многих случаях препятствует эффективному выполнению задачи проверки и коррекции правописания в текстах пользователей.

Приведем мнение известного лексиколога и лексикографа Леонида Петровича Крысина об изъятии ряда лексем из словарей: «Устранение устаревших и периферийных для литературного языка (узкоспециальных, диалектных и т.п.) лексических значений, так же, как и исключение из словаря такого рода слов, должн<о> осуществляться с осторожностью, поскольку многие несомненные для современного состояния русского языка архаизмы и раритеты встречаются в русской классической литературе XIX – первой половины XX в., и, следовательно, словарь может служить важ-

ным подспорьем для читателя в понимании таких единиц. Тем не менее некоторые лексемы могут быть, кажется, устранены из МАСа (Малый академический словарь. – Э.Л.) без ущерба как для правильного представления о современном состоянии русской лексики, так и для интересов читателя» [13. С. 5].

**11. Выводы.** Стремительный прогресс в компьютерной сфере практически снимает ограничения, связанные с объемами баз данных. С одной стороны, желательнее, чтобы в каждой очередной версии системные словари текстовых редакторов как можно больше пополнялись. При этом могли бы более широко использоваться механизмы словообразования, создания словарных гнезд. С другой стороны, желательно выявлять низкочастотные словоформы, которые могут *совпасть с искажениями в результате наиболее вероятных ошибок и опечаток* достаточно употребительных словоформ. Такие «подводные камни» системных словарей способствуют пропуску разного рода огрехов (особенно при недостаточно разработанном синтаксическом анализе), а иногда мы их видим в списках рекомендуемых вариантов исправления. «Проблемные» низкочастотные слова и словоформы можно либо заблокировать до появления новых улучшенных версий системы, либо снабжать пометами – предупреждениями для пользователя (желательно с выделением их в текстах каким-то особым образом). Но на первых порах представляется целесообразным заблокировать в системных словарях хотя бы такие *короткие* устаревшие, узкоспециальные или просто низкочастотные слова, которые могут препятствовать выявлению ошибок и опечаток, тем более что опечатки или, наоборот, отсутствие ошибок в подчеркнутых spellером коротких словах визуально обнаружить легче, чем в длинных.

Мы предлагаем также дополнять системы коррекции правописания сведениями о наиболее вероятных и часто встречающихся искажениях некоторых слов – с их правильными вариантами, чтобы программа «подсказка» выдавала более адекватные рекомендации. Такие сведения о типичных ошибках будут способствовать также и минимизации числа отказов в выдаче вариантов исправления неопознанных слов при компьютерной коррекции текстов.

Справедливости ради следует заметить, что многие недочеты текстового редактора MS Word, о которых мы и другие пользователи Интернета писали, в последнее время уже устранены, но к практике наполнения системных (внутренних) компьютерных русских словарей по-прежнему остается много вопросов.

## ПРИМЕЧАНИЯ

<sup>1</sup> Например, все инфинитивы глаголов на -ть в словаре Зализняка собраны вместе (что удобно для лингвистов-исследователей), хотя и перемежаются немногочисленными другими вкраплениями: *пять*, *кровать*, *скатерть* и т.п.

<sup>2</sup> Автор книги Ирина Спира сетует: «Традиционная проверка правописания была реализована в Microsoft Word на высоком уровне. Программа замечала не только орфографические ошибки, но и “чувствовала” весьма тонкие грамматические и стилистические нюансы, решала даже непростые пунктуационные задачи. Но в Microsoft Word 2013 качество проверки правописания русского текста заметно ухудшилось».

<sup>3</sup> Примеры: для ошибочного *ассимметричный* с двумя «с» подсказка-2017 предлагает первым вариантом нормативное *асимметричный*, а вторым – ошибочное *ассиметричный*, которое и в текстах тоже не подчеркивается; слово *псевдоклассика* программа «Корректор» предлагает разделить на две части пробелом, хотя отдельного слова *псевдо* в РОС и других словарях нет.

<sup>4</sup> Ни подсказка-2016, ни новый «Корректор» для подчеркиваемого слова *халва* написания *халва* (в РОС это другое слово, с другим толкованием) не выдает, хотя слово *халва* включено во внутренний системный словарь и при поиске вариантов исправления достаточно поме-

нять всего одну букву. Некоторые рекомендации подсказки «Корректор» производят неожиданный эффект. Например, для просторечного *откуда* программа первым вариантом выдает «исправление» *оттуда*.

<sup>5</sup> Не все из этих и приводимых далее примеров слов имеются в цитируемых в предлагаемой статье словарях, но сложно требовать от бу- мажных изданий с их ограничениями на листаж порождения всей картины русского словообразования.

<sup>6</sup> 6-я и 7-я книги серии были удостоены премии «Лучшие книги и издательства – 2016».

## ЛИТЕРАТУРА

1. Русский орфографический словарь: около 200 000 слов / под ред. В.В. Лопатина, О.Е. Ивановой. 4-е изд., испр. и доп. М. : АСТ-Пресс книга, 2015. 896 с.
2. Букчина Б.З., Сазонова И.К., Чельцова Л.К. Орфографический словарь русского языка. 4-е изд., испр. М. : АСТ-Пресс книга, 2008. 1296 с.
3. Зализняк А.А. Грамматический словарь русского языка: Словоизменение. Ок. 110 000 слов. 6-е изд., стер. М. : АСТ-Пресс книга, 2010. 800 с.
4. Ефремова Т.Ф. Новый словарь русского языка. Толково-словообразовательный: в 2 т. 2-е изд. М. : Рус. яз., 2001.
5. Спира И. Microsoft Excel и Word 2013: Учиться никогда не поздно. СПб. : Питер, 2014. 256 с.
6. Лавошникова Э.К. Microsoft Word 2016 и синтаксический контроль // Современные информационные технологии и ИТ-образование. 2016. Т. 12, № 2. С. 205–210.
7. Лавошникова Э.К. «Проблемные» слова как причина пропуска ошибок при компьютерной проверке орфографии // Текст. Книга. Книгоиздание. 2017. № 15. С. 113–129. DOI: 10.17223/23062061/15/8
8. Гаспаров М.Л., Скулачев Т.В. Односложные слова в стихе: ритм и части речи // Русский язык в научном освещении. 2003. № 1 (5). С. 35–51.
9. Поэзия Московского университета: от Ломоносова и до... Книга 6: от Арсения Альвинга до Владислава Ходасевича, включая Глеба Анфилова, Николая Арсеньева, Николая Бухарина, Надежду Гиляровскую, Юрия Сидорова, Александра Тришатов. М. : НИВЦ МГУ – Бослен, 2011. 480 с.
10. Лавошникова Э.К. Вариативность в грамматике стихотворных текстов и их компьютерная коррекция (на материале антологии Поэзия Московского университета: от Ломоносова и до...) // Текст. Книга. Книгоиздание. 2017. № 14. С. 108–122. DOI: 10.17223/23062061/14/7
11. Бешенкова Е.В. Вариативность, узальная норма и политика нормализаторов // Сибирский филологический журнал. 2016. № 3. С. 35–42.
12. Успенский В.А. Субъективные заметки о неправильной норме // Русский язык сегодня. Вып. 4. Проблемы языковой нормы. М. : Ин-т рус. яз. им. В.В. Виноградова РАН, 2006. С. 537–571.
13. Крысин Л.П. Проблема обновления толковых словарей современного русского языка // Известия РАН. Сер. литературы и языка. 2011. Т. 70, № 1. С. 3–9.

Статья представлена научной редакцией «Филология» 20 августа 2018 г.

## WORD AND LOW FREQUENCY VOCABULARY IN DICTIONARIES OF THE TEXT EDITOR

*Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*, 2018, 435, 40–47.

DOI: 10.17223/15617793/435/5

**Elina K. Lavoshnikova**, Lomonosov Moscow State University (Moscow, Russian Federation). E-mail: elavoshnikova@mail.ru

**Keywords:** Microsoft Word; WinWord; computer spell checker; text editor; speller; obsolete vocabulary; spelling mistakes; regulatory writing; archaisms; Russian language.

This article describes problems of computerized spell checking of Russian-language texts. Microsoft Word® 2016™ (2018 modification) text editor's built-in spell checking engine is being investigated and evaluated. It is shown that the inclusion of some obsolete and low-frequency words into the internal (system) computer dictionaries sometimes not only does nothing to improve the work of the speller, but also leads to the skipping of errors and typos. It is worth mentioning that many flaws and gaps of previous MS Word versions have been patched in MS Word 2016. Nevertheless, computerized analysis of word concord – in phrases and in standalone word combinations – raises even more questions, especially when compared with previous Orfo™-based spellers. Even detection of spelling errors (as the most developed analysis area) and prompting of possible corrections are still far from being perfect. The suggesting program of the spell checker suggests possible correction variants of words underlined and unidentified by the system (no more than three options in 2017 edition) or revision of the phrase. The suggestion is not always able to propose the normative spelling of the word, especially if it differs from the underlined one by a few letters. In the list of options, words are often split by a space, without taking into account the coherence of the two resulting words. The article contains examples of words quite frequently used in modern phrases and not known by the WinWord system dictionary, which should not be detected as mistakes but should be skipped without remarks. At the same time, there is no reason to keep rare and low-frequent short lexical units which coincide with beginnings and endings of more commonly used words in the system dictionary, because they may appear when a word is unintentionally split by space. The article contains examples of *specialy constructed* phrases with errors: interchange of letters in a word, hyphaeresis or gemination, word split or concatenation. All such words resulting from errors are present or generated within the system dictionary. Word forms do not concord here; however, MS Word is unable to detect syntax errors of this type. Similar phrases can also be used for testing spell checkers of other MS Word versions, not only previous but also newer ones. The author provides a list of *rare words* considered by MS Word as correct despite of a significant chance of an error in writing more commonly used words. It is advisable *to remove* some 'specific' rare words *from internal system dictionaries or deactivate* them for the time being, until the spell checker is more informative about the contextual areas where the words can be used. Many of the flaws described by the author and by other Internet users have recently been eliminated from the MS Word text editor, but the content of its Russian system dictionary and the recommendations of the spell checker suggesting program leave a lot of questions.

## REFERENCES

1. Lopatin, V.V. & Ivanova, O.E. (eds) (2015) *Russkiy orfograficheskiy slovar': okolo 200 000 slov* [Russian spelling dictionary: c. 200,000 words]. 4th ed. Moscow: AST-Press kniga.
2. Bukchina, B.Z., Sazonova, I.K. & Chel'tsova, L.K. (2008) *Orfograficheskiy slovar' russkogo yazyka* [Spelling dictionary of the Russian language]. 4th ed. Moscow: AST-Press kniga.
3. Zaliznyak, A.A. (2010) *Grammaticheskiy slovar' russkogo yazyka: Slovoizmenenie. Ok. 110 000 slov* [Grammar Dictionary of the Russian Language: Word Change. C. 110,000 words]. 6th ed. Moscow: AST-Press kniga.

4. Efremova, T.F. (2001) *Novyy slovar' russkogo yazyka. Tolkovo-slovoobrazovatel'nyy: v 2 t.* [New dictionary of the Russian language. Interpretive word-formation: in 2 vols]. 2nd ed. Moscow: Rus. yaz.
5. Spira, I. (2014) *Microsoft Excel i Word 2013: Uchit'sya nikogda ne pozdno* [Microsoft Excel and Word 2013: It is never too late to learn]. St. Petersburg: Piter.
6. Lavoshnikova, E.K. (2016) Microsoft Word 2016 i sintaksicheskii kontrol' [Microsoft Word 2016 and syntactic control]. *Sovremennye informatsionnye tekhnologii i IT-obrazovanie – Modern Information Technology and IT-education*. 12(2), pp. 205–210.
7. Lavoshnikova, E.K. (2017) Specific words as a reason for not detecting errors by computerised spell check. *Tekst. Kniga. Knigoizdanie – Text. Book. Publishing*. 15. pp. 113–129. (In Russian). DOI: 10.17223/23062061/15/8
8. Gasparov, M.L. & Skulacheva, T.V. (2003) Odnoslozhnye slova v stikhe: ritm i chasti rechi [Monosyllabic words in verse: rhythm and parts of speech]. *Russkiy yazyk v nauchnom osveshchenii*. 1 (5). pp. 35–51.
9. Aydiyanyan, A. et al. (2011) *Poeziya Moskovskogo universiteta: ot Lomonosova i do...* [Poetry of Moscow University: from Lomonosov to . . .]. Book 6. Moscow: NIVTs MGU – Boslen.
10. Lavoshnikova, E.K. (2017) Variability in grammar of poetic texts and their computer correction (on the material of the anthology Poetry of Moscow University: from Lomonosov to...). *Tekst. Kniga. Knigoizdanie – Text. Book. Publishing*. 14. pp. 108–122. (In Russian). DOI: 10.17223/23062061/14/7
11. Beshenkova, E.V. (2016) Variants, usual way of writing, codification and policy of orthographers. *Sibirskiy filologicheskii zhurnal – Siberian Journal of Philology*. 3. pp. 35–42. (In Russian). DOI: 10.17223/18137083/56/3
12. Uspenskiy, V.A. (2006) Sub"ektivnye zametki o nepravil'noy norme [Subjective notes about the wrong norm]. In: Krysin, L.P. (ed.) *Russkiy yazyk segodnya* [Russian language today]. Is. 4. Moscow: Russian Language Institute, RAS. pp. 537–571.
13. Krysin, L.P. (2011) Problema obnovleniya tolkovykh slovarey sovremennoogo russkogo yazyka [The problem of updating the explanatory dictionaries of the modern Russian language]. *Izvestiya RAN. Ser. literatury i yazyka*. 70(1). pp. 3–9.

Received: 20 August 2018