

УДК 81'322.2

DOI: 10.17223/23062061/15/8

Э.К. Лавошникова

«ПРОБЛЕМНЫЕ» СЛОВА КАК ПРИЧИНА ПРОПУСКА ОШИБОК ПРИ КОМПЬЮТЕРНОЙ ПРОВЕРКЕ ОРФОГРАФИИ

Как уст румяных без улыбки,
Без грамматической ошибки
Я русской речи не люблю.

А.С. Пушкин. Евгений Онегин. Гл. 3, XXVIII

Аннотация. *Рассматривается проблематика компьютерных систем проверки правописания – для текстов, написанных на русском языке. Разбирается и тестируется с выявлением слабых мест работа наиболее актуального на данный момент автокорректора (спеллера), встроенного в текстовый редактор Microsoft Word (версия 2016 г.). На многочисленных примерах показывается, что перегруженность внутренних системных словарей устаревшей и низкочастотной лексикой приводит к пропуску ошибок и опечаток, особенно при слабом синтаксическом контроле. Предлагается дополнять системные словари текстового редактора списками словоформ и устойчивых словосочетаний с ошибками, не выявляемыми системой, и информацией об их правильном написании. Представлены примеры наполнения таких списков наиболее частыми ошибками пользователей Интернета – для выдачи рекомендаций по их исправлению.*

Ключевые слова: *Microsoft Word 2016; MS Word 2013; текстовый редактор; русский язык, спеллер; компьютерная проверка правописания; орфографические ошибки; нормативное написание.*

Введение. В книге, содержащей ровно 256 страниц (сакральное число для IT-специалистов), 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» [1. С. 59].

Справедливости ради следует отметить, что многие огрехи в системе Word 2016, отмечавшиеся нами и другими пользователя-

ми, уже устранены, но синтаксический анализ по-прежнему оставляет желать лучшего. И даже в наиболее разработанном направлении – выявлении орфографических ошибок – остаётся немало разного рода недочётов, в том числе не всегда адекватные рекомендации программы-«подсказки» по исправлению неопознанных слов.

Заметим, что спеллер «Орфо» первых версий Word'овского текстового редактора базировался на 1-м издании грамматического словаря Зализняка (1977 г.), размеченном при скудном, в отличие от типографского, наборе символов на ЭВМ ЕС-1022 и перенесённом на машинные носители (бобины с магнитной пленкой) в 1980-х гг. коллективом Лаборатории автоматизированных лексикографических систем НИВЦ МГУ им. М.В. Ломоносова [2. С. 32]. Электронная версия этого словаря со своей системой грамматических помет и таблиц легла в основу большинства современных компьютерных программ, работающих с русской морфологией. Системный словарь Word'a, как и последующие издания словаря Зализняка (см. [3]), продолжает пополняться новой лексикой. Академические словари, к сожалению, не всегда успевают закреплять написание новых заимствований, и разработчикам текстового редактора приходится опираться на транслитерацию большинства пользователей Интернета. Возникают расхождения также и между издаваемыми словарями (см.: [4. С. 50]).

При проверке текстов (здесь мы рассматриваем написанные на русском языке) автокорректор текстового редактора MS Word *красной волнистой линией* подчёркивает отдельные слова (или даже части сложных слов при написании через дефис), отсутствующие или не порождаемые в Word'овских системных словарях. Тем самым пользователю предлагается обратить на них внимание – нет ли в них ошибки или опечатки. *Синей волнистой линией* в текстовом редакторе Word 2016, а также в Word 2013 (или *зелёной* в предыдущих версиях, разработанных на основе системы «Орфо») подчёркиваются вероятные синтаксические огрехи – ошибки в согласовании и сочетаемости словоформ.

Подчёркиваемые системой Word 2016 слова и словосочетания (*красной или синей* волнистой чертой) мы в наших примерах тоже будем подчёркивать.

Написание приводимых в качестве примеров слов сверялось с «Русским орфографическим словарем» (РОС) [5], а также с 6-м изданием «Грамматического словаря русского языка» А.А. Зализняка [3]. Следует отметить, что словарь академика РАН Андрея Анатольевича Зализняка входит в список четырёх словарей, грамматик и справочников, рекомендованных в 2009 г. Межведомственной комиссией по русскому языку при Минобрнауки и содержащих нормы современного русского литературного языка.

Особенности работы программы-«подсказки» в версии MS Word 2016. *Программа-подсказка* (при её вызове) после замены, вставок, удаления и перестановок букв в слове с красным подчёркиванием ищет в системных словарях полученные «похожие» буквосочетания. Если находит, то выдаёт список таких словоформ как возможные варианты исправления неопознанного слова. Однако сначала подсказка в большинстве случаев старается разбить слово пробелом, не проверяя, к сожалению, получившееся словосочетание на согласованность.

Складывается впечатление, что подсказка-2016 не стала работать лучше, чем в MS Word 2010 и предыдущих версиях системы. Например, в некоторых версиях Word'a (основанных на системе «Орфо») подсказка выдавала рекомендацию по нормативному слитному написанию слова *невтерпёж*. К сожалению, в Word 2016 и Word 2013 последовательность с пробелами «не в терпеж», иногда встречающаяся в текстах пользователей Интернета, пропускается без замечаний.

В качестве исправления местоимения «ихний» в прежних версиях текстового редактора MS Word подсказка выдавала термин *ионий*, теперь же пропускает этот вариант формы *их* без замечаний, как и многие другие просторечные слова. Интересно, что для просторечного *откудова* вордовская подсказка предлагает *оттудова* (sic!). Текстовый редактор MS Word 2016 написание *прокремлёвский* (через букву Ё) считает неправильным и предлагает в подсказке исправление *прокремлевский* – через букву Е. Для неопознанного слова *холява* (РОС: «заготовки стекольного производства; голенище») подсказка-2016 среди вариантов исправления

слово *халыва* по непонятной причине не выдаёт, хотя оно имеется в Word'овском системном словаре и достаточно поменять всего одну букву.

В текстах нередко можно встретить ненормативные формы, которые Word подчёркивает красным: «ботинков», «румынов», «сапогов», «солдатов», «турков» (три последних слова предлагается писать с прописной буквы, как фамилии, а для «турков», кроме того, подсказка добавляет написание *туркав* – деепричастие от *туркать*). Подсказка-2016 не даёт нормативных форм родительного падежа мн. числа *ботинок*, *румын*, *сапог*, *солдат*, *турок*.

Ещё примеры. Подсказка системы MS Word 2016 при её вызове предлагает разбивать пробелом слова микроЭВМ (в РОС дано слитное написание) и псевдонаучность, хотя отдельных слов «микро» и «псевдо» в словарях [3] и РОС нет. Word'овская подсказка в неопознанных словах бесприцельный, бессобытийность, бессуффиксальный и др. отделяет пробелом приставку «бес-» (для прилагательного бестаможженный получается, например, единственный выдаваемый вариант «бес таможенный»). Подсказка предлагает писать отдельно «не» в словах: неаппетитность, негуманность, неотрицательность и некоторых других. У подсказки-2016 нет вариантов исправления для подчёркиваемого красным слова невоспроизводимость. В этом слове программа-подсказка почему-то уже не отделяет отрицание «не», хотя *воспроизводимость* пропускается без замечаний.

Проблематика наполнения Word'овского системного словаря. Из внутреннего системного словаря некоторых предыдущих версий Word'a были исключены многие «нехорошие» слова: *голубой*, *розовый*, *дурак*, *негр*, *осёл* (написание *осел* тоже было убрано – вместе с глагольными формами *осела*, *осело*, *осели*), *сдохнуть* и т.п. После многочисленных критических замечаний подобные слова всё же были возвращены, но иногда без своей полной парадигмы: слово *подлец* есть, однако в версии Word 2016 оно не склоняется; слово *сволочь* спрягается в качестве глагола (*сволочу*, *сволочишь*), но формы существительного подчёркиваются как не по-рождаемые в системе (сволочью, сволочам и т.д.).

Во внутренний словарь текстового редактора MS Word 2016 уже внесены слова *бойфренд*, *гламур*, *гуманизация*, *капучино*, *коллайдер*, *комплексовать*, *конфискат*, *лермонтовский* (но отымённые прилагательные *грибоедовский*, *булгаковский* всё ещё не опознаются), *месседж* (РОС) и довольно популярный в Интернете вариант *мессидж*, *минимизирующий*, *мониторить*, *наркодилер*, *невыездной*, *обязаловка*, *перфекционизм*, *подредактировать*, *политкорректный*, *продвинутость*, *растамозжить*, *репутационный*, *соинвестор*, *спецсеминар*, *харизматичный* и др. О желательности включения в словарь системы некоторых из этих лексем мы писали ещё 15 лет назад [6. С. 157].

К сожалению, в Word'овском системном словаре по-прежнему отсутствуют слова *влекомый*, *исхалтуриться*, *кремлинолог* (РОС) и встречающийся вариант *кремленолог*, *культуроведческий*, *проверяемость*, *телегеничность*, *турбуленция*. Нет слов *сверхрадикальный*, *суперэфективность*, *экстракардинальный* и некоторых других с префиксоидами «сверх-», «супер-», «экстра-», которые система предлагает писать отдельно, и т.п. Если в тексте спеллер подчеркнёт красным сложносоставное или не так давно заимствованное из других языков слово, то пользователь может решить, что эта лексема просто пока ещё не внесена в системный словарь текстового редактора, и не заметит опечатки. Поэтому такие «непростые» слова разработчикам следует как можно более оперативно включать во внутренний системный словарь.

Не включены в системный словарь Word'а и некоторые уменьшительно-ласкательные варианты слов, например: *защитничек*, *кабинетик*, *походочка*, *треугольничек*, *ухмылочка* и др.

Некоторые слова на «-ье», представляющие собой варианты существительных с окончанием «-ие», в системный словарь Word'а не включены (пример: без замечаний пропускаются слова *создание* и *таскание*, но варианты со стяжением *созда^{нь}е* и *таска^{нь}е* подчёркиваются красным). Обратный случай: *веселье* и отсутствующий в РОС и словаре [3], но пропускаемый Word'ом без замечаний устаревший вариант *веселие*.

В системном вордовском словаре для некоторых существительных нет вариантов форм на «-ую/-ею»: кривизною (подсказкой предложено разбиение «криви зною»), листвою, судьбою, тоскою, пищею, струёю и др.

Словоформу брелков (РОС), не всеми лингвистами признаваемую вариантом нормы, Word-2016 подчёркивает красным. В его системном словаре имеется (или порождается) для родительного падежа множественного числа только гораздо менее популярный в Интернете вариант брелоков. При этом неграмотную форму людя-ми спеллер признаёт приемлемым вариантом словоформы людьми и пропускает без замечаний.

Нет в системном словаре деепричастия отстав (есть только отставили), хотя отстать спрягается так же, как достать – с деепричастием достав(или) [3. С. 131]. Парадигмы некоторых других слов тоже оказываются неполными. Есть слово корова, но нет (по какой-то технической ошибке) словоформы коров. Отсутствуют косвенные формы слова стеклышко.

При проверке правописания в текстовом редакторе Word 2016 по неизвестным причинам оказываются неопознанными, в частности, следующие формы причастий: дублирующаяся, пишущиеся, порождающиеся, произносящиеся, цитирующиеся (но формы именительного падежа единственного числа этих причастий пропускаются без замечаний). Интересно, что подчёркивается как неопознанная форма превосходной степени свежайший, но менее частотное образование наисвежайший пропускается без замечаний. В системный словарь не внесены и некоторые другие формы превосходной степени: аппетитнейший, бездарнейший (подсказка-2016 предлагает «исправление» бездомнейший), наивыгоднейший. Но как раз в подобных длинных словах труднее заметить опечатку.

«Проблемные» слова в Word'овском системном словаре. Справедливости ради следует отметить, что из системного словаря наконец убраны некоторые низкочастотные слова, которые мешают находить ошибки, о чём мы предупреждали в своих статьях, например в [6. С. 151]: переплотить (РОС: «к плот»); поленица

(РОС: «богатырь») – это слово без удвоения буквы Н тоже уже подчёркивается красным (в системе есть *поленница* от слова *поле-но*); *пощепать* от слова *щепка*; *рождество* (в РОС отсутствует, есть *Рождество*); *эстокада* (РОС: «в фехтовании»).

К сожалению, в Word'овском внутреннем системном словаре всё ещё остаются *малоупотребительные* слова, не способствующие эффективному выявлению *орфографических ошибок*: написание *достегать* (РОС: «к стегать») скорее может получиться в результате ошибки в слове *достигать*; *картеж* (или *картёж*) – игра в карты, но можно сделать ошибку в слове *кортеж*; *мыслете* (РОС: «название буквы»), однако при довольно несовершенном синтаксическом контроле (см. примеры в [7. С. 206]) спеллер может не заметить орфографической ошибки в глагольной форме *мыслите*; *нажеваться* (РОС: «к жевать»), хотя глагол *нажизаться* употребляется гораздо чаще; *парадировать* (РОС: «от парад») – вполне вероятное неправильное написание глагола *пародировать*; *подрожать* (РОС: «от дрожать»), однако такое слово может встретиться вместо *подражать*. Кроме того, в системном словаре имеются глаголы *препираться* и *припираться* (РОС: «к припереться»). Для некоторых пар похожих по написанию слов – *адаптация* и *адоптация* (РОС: «усыновление»), *комплимент* и термин *комплемент*, *постигать* и *постегать* (РОС: «к стегать»), *презрение* и *призрение* – следующие версии системы могли бы, что было бы желательно, выдавать краткое *толкование* менее употребительного слова, не убирая его из своего внутреннего словаря.

Во внутренних системных словарях Word'a содержится довольно много *низкочастотных* слов, которые с большей вероятностью могут появляться в текстах в результате *пропуска буквы* (например, при недостаточно сильном нажатии клавиши) в более употребительных словах. Примеры: *браный* (РОС: «вытканый узорами»), *взмутиться*, *вскользнуть*, *вывить* (ср. *выявить*), *вытраивать*, *вытроить*, *замет* (ср. *заметка*), *затесняться*, *иступленный* (РОС: «к тупой») – «иступленный карандаш», *надавить* (ср. *надавить*), *наустить* (ср. *напустить*), *поветь* (РОС: «помещение в крестьянском дворе, обл.»), *подсочить*, *подустить*,

поместиться, собирать, строить, уточнить, а также *утонить, утоняться* ('делаться тоньше') – последние два глагола могут получаться при нечаянном пропуске буквы Ч. Такие «проблемные» лексемы желательно убрать из внутренних компьютерных словарей или заблокировать до лучших времён, когда система научится выдавать информацию (пометы) о сферах их употребления.

Часто в тех случаях, когда в словаре А.А. Зализняка [3] даются указания, что какие-либо формы из парадигмы слова затруднены или избегаются, проверщик правописания Word, тем не менее, пропускает их без замечаний. Примеры: *кусток, роток, часок*. Написание «кустками», «ротком», «часки» скорее может быть результатом непреднамеренного пропуска буквы И в словоформах *куст^Иками, рот^Иком, час^Ики*. Маловероятно употребление форм множественного числа от некоторых существительных (*мглами*), а также отдельных глагольных форм (*удаюсь, окислишься, подобраем*). Ниже в примерах использованы подобные глагольные формы (*ёкну, смори*), не подчёркиваемые вордовским spellером.

Кроме того, автокорректором порождаются сомнительные в стилистическом плане, пусть даже теоретически возможные формы деепричастий от глаголов *несовершенного вида*. Примеры: *грев* (слово может «оторваться» при нечаянной вставке пробела: «на грев», «при грев»), *давав, делав, дув, ев* (ср. «обид ев», «замер ев»), *изготовляв, крав* (ср. пропускаемое без подчёркиваний «обо крав»), *лежав, стояв* (ср. «лежа в...», «стоя в...»).

Бряд ли стоит держать в системном словаре текстового редактора *короткие односложные* лексемы и словоформы (их правописание не вызывает затруднений), совпадающие с началами или концами более употребительных слов: *ал* (кр. форма от *алый*), *ба* (междометие), *вый* (от устаревшего *выя*), *жит* (от *жито*, но может получиться при незамеченной вставке пробела: «поло жит», «умно жит» и др.), *ин* (частица), *ких, ким* (?), *ков* (от *ковы*), *кс*, *ку* (название буквы), *ла* (нет в РОС), *ми* (нота), *ом* (РОС: «ед. измер.»), *ор* (РОС: «к орать»), *ре* (нота), *су, че* (название буквы), *чик, ют* (совпадает с глагольным окончанием). Некоторые из таких ко-

ротких слов мы включили в примеры фраз с ошибочными пробелами и другими огрехами (см. ниже).

Примеры отсутствия реакции Word'a на фразы с ошибками. Далее приводятся *специально сконструированные фразы* с ошибками и опечатками (перестановка букв, пропуск или удвоение буквы, разбиение слова пробелом или склеивание слов). Согласование словоформ в них нарушено, однако Word 2016 не замечает в этих законченных предложениях *синтаксических* ошибок. **Нет подчёркиваний ни красной, ни синей** волнистой линией. Подобные фразы могут быть использованы при тестировании других версий Word'a – как предыдущих, так и последующих.

1. *Но ссунуть та ких обоз ленных умни ков в этот промысл – доя отела кадров лишь поддела.* В этой фразе как бы по ошибке получились формы от следующих слов: малоупотребительный глагол *ссунуть*; *ких* (в Интернете *КИХ*); *ленный* (РОС: «от лен», причём в этом словаре представлены *лен* и *лён*, склоняющиеся по-разному: *лена* и *льна*); императив глагола *умять*; род. падеж мн. ч. от *ковы*; слово *промысл*, отсутствующее в РОС; *доя* от глагола *доить* вместо *для* (литеры О и Л расположены на клавиатуре рядом); *отёл*. При наборе слова *поддела* при нечаянной перестановке согласных может получиться форма от малоупотребительного глагола *подлеть*.

2. Ещё пример фразы с опечатками – ненужными пробелами. *Это удиви тельное существ о с года ми не стари лось, хот я по срав нению с друг ими было старик ом.* Здесь мы видим прилагательное *тельный*; ноту *ми*; форму существительного *старь* [3], отсутствующего в весьма представительном издании РОС; музыкальный термин *хот*; низкочастотное слово *нения* (похоронная песня или причитание у древних греков и римлян).

3. *Все обезумили, что же вы пишите, что вы ищите ин туристов: дайка я вам нам екну.* В этой фразе мы видим примеры весьма распространённых ошибок в глагольных формах: *обезумили* от *обезумить* (кого-то) вместо подразумеваемой здесь формы *обезумели* от *обезуметь*; очень часто у пользователей Интернета получается императив *пишите* вместо личной формы мн.ч. *пише-*

те; не менее «популярно» написание *ищите* для подразумеваемого *ищите*. Далее мы видим малоупотребительную частицу *ин*. Из-за отсутствия дефиса появился геологический термин *дайка*. Последнее слово в этой искусственно составленной фразе – маловероятная форма 1-го лица от глагола *ёкнуть*.

4. *Смори на все обь явленные про верки с улыбой!* При пропуске буквы Т (соответствующий звук в глаголе *смотреть* тоже не всегда нами произносится) получился вряд ли используемый императив от глагола *сморить*; в Word'овском системном словаре имеется слово *обь* с твердым знаком на конце (!); *верки* – отдельные части укрепления; одушевлённое *улыба* уже устарело (отсутствует в РОС).

5. *На фото графии та кая цац ка на булаве – как в прежде й жизни и в пре жней одеж де.* Здесь из-за вставки пробела получилась форма от «графия». Этого слова нет в РОС и [3]; словарные статьи «кая», «кай», «каить» или «каять» в этих словарях тоже отсутствуют. Далее следует: род. падеж мн. ч. слова *цаца*; *ка* – название буквы; форма слова *булавка* с пропущенной буквой К; неупотребительное «прежде» от *прежний*. Заметим, что любые **одинокые буквы** Word 2016 пропускает без подчёркивания, в том числе Й (в этой фразе – «прежде й»). Словоформа *пре* – от устаревшего существительного *пря*, отсутствующего в РОС; затем получилась форма существительного *жнея*. Просторечное слово *одежа* текстовый редактор мог бы пропускать без замечаний только с буквой Ё, так как написание *одежа* может получиться в результате недостаточно сильного нажатия на клавишу буквы Д.

6. *Эти рез аки – из с плава и с вычурой купной ручкой.* Здесь *рез* – «в Древней Руси прибыль, процент от денег, данных взаймы»; *аки* – союз. Два предлога идут подряд. Слово *плав* отсутствует в РОС (в [3] – с пояснением «на плаву»). Далее получилась форма слова *вычура* [Там же] (в РОС есть словарная статья *вычуры*); *купный* – результат пропуска буквы Р (в РОС и словаре [Там же] имеется только наречие *купно*).

7. *При этом, кс тати, приедена придельная длинна из, кото рая не бол ее полу метра.* В этой фразе из-за вставки пробела получилось включённое в вордовский системный словарь не суще-

ствующее в русском языке слово *кс* (из Википедии: Counter-Strike (от англ. – «Контрудар»; сокр. CS или КС). Далее форма устаревшего *тать* – вор; при пропуске буквы В получилась краткая форма причастия *приеденный* (РОС); *придельный* (вместо *предельный*) – от *придел* со значением ‘пристройка’. В нашей редакторской практике иногда встречается написание *длинна* (краткая форма прилагательного) в роли существительного *длина*. Далее мы видим «из» вместо «их» – клавиши З и Х расположены рядом, что может приводить к такой опечатке, при этом за предлогом «из» следует запятая, на что Word’овский синтаксический контроль не реагирует; *кото* (‘японская арфа’); *бол* (нет в РОС и [3]); форма от *пол* или *полá*.

Желательно было бы удалить или временно (до улучшения ситуации с проверкой синтаксиса) заблокировать в Word’овском системном словаре некоторые *малоупотребительные* слова и словоформы, такие как в приведённых выше примерах фраз (получившиеся в результате намеренно допущенных нами огрехов). Мы видим, что их присутствие во внутреннем компьютерном словаре приводит к пропуску разного рода ошибок и опечаток в проверяемых текстах.

Другие придуманные фразы со специально подобранными ошибками для тестирования работы автокорректоров приведены, например, в [7. С. 206].

Предложения по усовершенствованию компьютерной проверки правописания. Мы предлагаем дополнять внутренние компьютерные словари *перечнями наиболее вероятных искажений слов и синтагм* с их исправленным написанием, которое по разным причинам программой-подсказкой не выдаётся. Конечно, для *некоторых первых* компонентов этих пар можно *с натяжкой* придумать примеры фраз, где эти словосочетания будут грамматически обоснованными.

Списки для выдачи программой-подсказкой рекомендованных исправлений могли бы состоять из пар наподобие приведённых ниже. В каждой такой заключённой в фигурные скобки паре слов или словосочетаний слева находится ошибочное, как правило,

написание, отмеченное астерiskом (звёздочкой), справа – предлагаемое его исправление.

СПИСОК 1

В этом списке слова из *левого* варианта подчёркиваются *красной* волнистой линией как не найденные в системном словаре, но программа-подсказка Word'a 2016 *нормативного варианта не выдаёт*.

{**аэрозолью* – *аэрозодем*} – есть тенденция употреблять слова с концовкой «-ль» в женском роде (**тольюю* вместо *тодем*, **тюльюю* вместо *тюдем*), сюда же можно добавить, например, слово *рояль*;

{**без препонов* – *без препон*};

{**в драбадан* – *вдрабадан*} – здесь оба варианта подчёркиваются красным как отсутствующие в Word'овском системном словаре (в РОС даётся слитное написание);

{**вуалем* – *вуалью*};

{**для избежания* – *во избежание*} – подсказка-2016 выдаёт два варианта для подчёркиваемого красной волнистой линией слова: *избежание* и *избегания* (таким образом, предлагаются исправления «*для избежание*» и «*для избегания*»);

{**досвидание* – *до свидания*} – подсказка Word'a (2013-го и 2016-го) выдаёт столь же неправильное «*до свидание*», которое система сама же в тексте подчёркивает *синей* волнистой линией;

{**ехай* – *поезжай*};

{**из под тишка* – *исподтишка*} – в первом члене пары сначала *синее* подчёркивание (обоснованно предлагается дефисное написание), потом *красное* для «*тишка*»: подсказка-2016 предлагает гипокористическое имя Тишка, а также «тушка» и «тёшка»;

{**камфорка* – *конфорка*} – подсказка-2016 выдаёт формы слова *каморка*, сюда же можно добавить *склонение* по числам и падежам: {**камфорки* – *конфорки*}, {**камфорке* – *конфорке*} и т.д.;

{**ложат* – *кладут*};

{**под шафе* – *подшофе*};

{**простыню* – *простыньёй*};

{**пятидесятью* – *пятьюдесятью*} – но подсказка Word'a 2016 такого исправления не выдаёт;

{*совецкий – советский} – хотя может встретиться *намеренное* искажение написания этого прилагательного;

{*тапком – тапкой} – в узусе это слово довольно часто встречается в форме мужского рода;

{*хочем – хотим} – для просторечной формы «хотете» подсказка-2016 нормативного варианта *хотите* тоже не даёт;

{*шампунью – шампунем}.

СПИСОК 2

В этом списке слова из *левого* варианта пропускаются текстовым редактором MS Word 2016 без замечаний, т.е. *не подчёркиваются* ни красным, ни синим.

{**в виду того, что* – *ввиду того, что*}, но «иметь в виду»;

{**в дребезги* – *вдребезги*} – таких случаев ненормативного раздельного написания (отделения приставки «в»), пропускаемого Word'ом 2016, довольно много, например для слов: *вдвое*, *вполгоса*, *вприпрыжку*, *вприсядку*, *впритык*, *вразвалку*, *всерьёз*, *встарь* и др.¹;

{**в не платном долгу* – *в неоплатном долгу*};

{**в отличии от* – *в отличие от*} – слева довольно распространённая ошибка;

{**в отсутстви* – *в отсутствие*}, хотя можно придумать фразу, где левый вариант нормативен: *В отсутстви* ошибок нашей заслуги нет;

{**в пол-оборота* – *вполоборота*};

{**в последствие* – *впоследствии*} – иногда левый вариант у пишущих получается по аналогии с некоторыми из предыдущих примеров;

{**в родные пенаты* – *к родным пенатам*} – пенаты в древнеримской мифологии 'боги-хранители и покровители домашнего очага';

¹ Нормативное написание со временем может меняться. В изданном в 1935 г. 1-м томе словаря Д.Н. Ушакова [8] мы видим слитное написание «впику» (подсказка Word'a-2016 среди вариантов исправления «в пику» не выдаёт), «вохотку», «врозницу», «втищи», а также «во-всю», подчёркиваемое в Word'e-2016 синей волнистой линией.

{**до света* – *досвета*} – подсказка-2016 подчеркивает красное второе (нормативное) написание, представленное в РОС;

{**до упада* – *до упаду*};

{**ей богу* – *ей-богу*};

{**зав. лаборатории* – *зав. лабораторией*};

{**заграницей* – *за границей*}, но можно придумать пример слитного написания: *С этой заграницей одни проблемы*;

{**извека в век* – *из века в век*} – в РОС имеются наречие *извека* и выражение «из века в век»);

{**играет значение* – *имеет значение*};

{**как кур во щи* – *как кур в ощи*};

{**мало по малу* – *мало-помалу*};

{**навеки веков* – *на веки веков*};

{**на вроде* – *навроде*} – подчёркивается красной волнистой линией только второй, правильный (хоть и просторечный) вариант;

{**на вскидку* – *навскидку*};

{**наощупь* – *на оощупь*} – Word-2016 допускает оба варианта, но слитное написание было нормативным только до реформы 1956 г.;

{**на перебой* – *наперебой*};

{**на утек* – *наутек*};

{**на цело* – *нацело*};

{**не больше, не меньше* – *ни больше ни меньше*} – наблюдается вытеснение усилительной частицы *ни*, довольно часто в узусе она заменяется отрицательной частицей *не* (отметим, что запятая в этом устойчивом обороте – как и в других подобных – не ставится);

{**не бум-бум* – *ни бум-бум*};

{**не в мочь* – *невмочь*};

{**не взирая на* – *невзирая на*};

{**не впопад* – *невпопад*};

{**не можетя* – *неможетяся*};

{**не повадно* – *неповадно*};

{**не уклюжий* – *неуклюжий*} – в РОС слово *уклюжий* [3] отсутствует;

{**никто иной, как* – *не кто иной, как*};

{**ничто иное, как* – *не что иное, как*};

{*одно и *тоже* – одно и *то же*};

{*пол Америки – пол-Америки};

{*при *том*, что – *притом что*} (это устойчивое выражение пишется с предлогом слитно и *без запятой* – вопреки синему подчёркиванию и рекомендации текстового редактора Word 2016, см. РОС);

{**тоже* самое – *то же* самое};

{**тютелька-в-тютельку* – *тютелька в тютельку*} – см. РОС;

{*через *чур* – *чересчур*};

{**черти-что* – *черт-те что*};

{**шиворот на выворот* – *шиворот-навыворот*}.

Список с другими некорректными словосочетаниями и их исправлениями (не выдаваемыми программой-подсказкой), который тоже мог бы быть использован при обновлениях систем проверки правописания, приведён, в частности, в статье [7. С. 209].

Очевидно, что множество подобных пар (если брать наиболее частотные примеры) – не бесконечно, а вполне обозримо. Перечни «популярных» в текстах пользователей Интернета ошибочных фрагментов с их исправленным написанием, подобные приведённым выше двум спискам, должны способствовать более эффективной работе текстового редактора при компьютерной проверке правильности написания фразеологизмов и других словосочетаний.

Литература

1. *Стира И.И.* Microsoft Excel и Word 2013: Учиться никогда не поздно. СПб. : Питер, 2014. 256 с.
2. *Казакевич О.А., Членова С.Ф.* Полвека лаборатории автоматизированных лексикографических систем НИВЦ МГУ им. М.В. Ломоносова // Вестник Российского государственного гуманитарного университета. 2014. Т. 16, № 8. С. 28–39.
3. *Зализняк А.А.* Грамматический словарь русского языка: Словоизменение. Ок. 110 000 слов. 6-е изд., стер. М. : АКТ-ПРЕСС КНИГА, 2010. 800 с. (Фундаментальные словари).
4. *Лавошникова Э.К.* Компьютерная коррекция орфографии и разночтения в словарях: возможны варианты? // Филологос. 2015. № 1 (24). С. 49–54.
5. РОС – 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.

6. Лавошникова Э.К. О «подводных камнях» в компьютерных системах проверки правописания // Вестник Московского университета. Серия 9: Филология. 2002. № 6. С. 151–162.

7. Лавошникова Э.К. Microsoft Word и синтаксический контроль // Современные информационные технологии и ИТ-образование. 2016. Т. 12, № 2. С. 205–210.

8. Толковый словарь русского языка : в 4 т. / под ред. Д.Н. Ушакова. М., 1935–1940.

‘SPECIFIC’ WORDS AS A REASON OF NOT DETECTING ERRORS BY COMPUTERISED SPELL CHECK

Tekst. Kniga. Knigoizdanie – Text. Book. Publishing, 2017, 15, pp. 113–129

DOI: 10.17223/23062061/15/8

Elina K. Lavoshnikova, Lomonosov Moscow State University (Moscow, Russian Federation). E-mail: el.lavoshnikova@yandex.ru

Key words: Microsoft Word 2016; MS Word 2013; text editor; Russian language; spellers; computer spell checker; spelling mistakes; regulatory writing.

This article describes problems of computerised spell checking of Russian-language texts. Microsoft Word ® 2016™ text editor’s built-in spell checking engine is being investigated and evaluated.

It is worth mentioning that many flaws and gaps of previous MS Word versions have been patched in MS Word 2016. Nevertheless, computerised 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 a most developed analysis area) and prompting of possible corrections are still far from being perfect.

When speller underlines a word that is very recently borrowed from foreign languages (marking as not recognised), users may think this lexical item is just not present in the system vocabulary. Users may overlook their misprints in many cases, especially if there is a compound or just a long word.

The article contains multiple examples of words quite frequently used in modern phrases and not known by MS Word 2016 system vocabulary, which should not be detected as mistakes but 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 vocabulary, because they may appear when a word is unintentionally split by space. The author provides a list of such short lexical units in the article, as well as 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.

The article contains examples of *specially 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 vocabulary.

Word forms do not concord here; however, MS Word 2016 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.

It is advisable **to remove** some ‘specific’ rare words **from internal system vocabularies or deactivate** them for the time being, until a spell checker will be more informative about the contextual areas where the words can be used.

The author also proposes to enrich the spell checker’s system vocabulary with **lists of the most probable misspellings** of words and phrases along with correct options to be chosen. Evidently, the most popular spelling errors that are omitted by MS Word 2016 are not infinite and thus may be collected. A collection of such common errors paired with normative writing should contribute to more efficient work of computerised spell checking systems.

References

1. Spira, I.I. (2014) *Microsoft Excel i Word 2013: Uchit'sya nikogda ne pozdno* [Microsoft Excel and Word 2013: It Is Never Late to Learn]. St. Petersburg: Piter.
2. Kazakevich, O.A. & Chlenova, S.F. (2014) Polveka laboratorii avtomatizirovannykh leksikograficheskikh sistem NIVTs MGU im. M.V. Lomonosova [Fifty years of the Laboratory of Automated Lexicographic Systems of M.V. Lomonosov Moscow State University]. *Vestnik Rossiyskogo gosudarstvennogo gumanitarnogo universiteta – Russian State University for the Humanities Bulletin*. 16(8). pp. 28–39.
3. Zaliznyak, A.A. (2010) *Grammaticheskii slovar' russkogo yazyka: Slovoizmenenie* [Grammatical Dictionary of the Russian Language: Inflection]. 6th ed. Moscow: AKT-PRESS KNIGA..
4. Lavoshnikova, E.K. (2015) Computer correction of spelling and differences in dictionaries: possible options? *Filologos*. 1(24). pp. 49–54. (In Russian).
5. Lopatin, V.V. & Ivanova, O.E. (eds) (2015) *Russkiiy orfograficheskii slovar'* [Russian Orthographic Dictionary]. 4th ed. Moscow: AST-PRESS KNIGA.
6. Lavoshnikova, E.K. (2002) O “podvodnykh kamnyakh” v komp'yuternykh sistemakh proverki pravopisaniya [On the “pitfalls” in computer spelling checking systems]. *Vestnik Moskovskogo universiteta. Seriya 9: Filologiya – Moscow State University Bulletin. Series 9. Philology*. 6. pp. 151–162.
7. Lavoshnikova, E.K. (2016) Microsoft Word i sintaksicheskii kontrol' [Microsoft Word and syntactic control]. *Sovremennye informatsionnye tekhnologii i IT-obrazovanie – Modern Information Technology and IT-education*. 12(2). pp. 205–210.
8. Ushakov, D.N. (ed.) (1935–1940) *Tolkovyy slovar' russkogo yazyka* [Explanatory Dictionary of The Russian Language]. Moscow: Sovetskaya entsiklopediya.