# ОРФОГРАФИЯ В ИНТЕРНЕТЕ: АНАЛИЗ ОДНОЙ ОРФОГРАФИЧЕСКОЙ ОШИБКИ ORTHOGRAPHY IN THE INTERNET:

# ORTHOGRAPHY IN THE INTERNET THE ANALYSIS OF ONE MISSPELL

**Богданов А.В.** (bidon@inbox.ru) Московский государственный университет им. М.В. Ломоносова

В работе рассматривается орфография в сети Интернет как естественный языковой процесс. Делается анализ одной, отдельно взятой, орфографической ошибки, а именно ошибочного написания мягкого знака в формах третьего лица единственного числа возвратных глаголов (делаеться, придеться). Далее делается попытка доказать, что орфографические ошибки могут различаться по своей частотности, а значит и по своему праву на установление новой орфографической нормы.

#### 1. Введение

С развитием Интернета, в том числе и русскоязычной его части, у людей, активно пользующихся сетью, появилось новое языковое пространство — в этом пространстве, с одной стороны, приходится использовать письменный язык, но с другой — нет практически никаких ограничений на правильность его использования. В самом деле, до появления Интернета практически все сферы использования письменного языка были таковы, что в них не допускались или, по крайней мере, не приветствовались орфографические ошибки. Будь то диктант по русскому языку, заявление, письменная жалоба, отчет о проделанной работе, или же просто записка — любой документ, написанный от руки, предполагал грамотное написание текста. Ошибки могли повлиять на судьбу автора в разных случаях с разной степенью, но так или иначе ошибок пытались избегать.

В новом языковом пространстве – в Интернете – ситуация с правильностью использования письменного языка отличается от вышеописанной коренным образом. Поскольку Интернет предоставляет возможность не только формального общения, но и общения в высшей степени неформального, то регистры [Беликов & Крысин 2001] языка, используемые в сети, могут очень сильно варьировать. Ну а если человек общается неформально и с использованием такого регистра языка, который в принципе не предполагает существование некоторого правильного написания<sup>1</sup>, то тем самым складываются все условия для того, чтобы освободиться от гнета правил орфографии, и единственным ограничением, которое накладывается в этом случае на процесс записи слов, остается стремление быть понятым. В самом деле, зачем писать длинное и сложное слово сейчас и тратить на него целых шесть букв, если можно ограничиться простым и понятным коротким *щас*, сэкономив три буквы. Конечно, если бы слово сейчас не редуцировалось бы до одного слова в устной речи, оно, вероятно, не могло бы быть записано таким образом. Тем не менее в этом примере мы видим как отход от произносительной нормы, так и от орфографической.

Таким образом, можно заметить, что использование письменной формы языка в таком свободном языковом пространстве, каким является Интернет, становится все больше и больше похожим на использование устного языка. Например, как видно из примера со словом сейчас, в этом языковом пространстве начинают действовать такие типичные закономерности устной речи, как максимы Грайса [Grice 1969]. В использовании щас вместо сейчас проявляется баланс между двумя максимами – говори коротко и избегай непонятности. Это действительно минимальное количество символов, необходимое для того, чтобы идентифицировать данное слово. Заметим, что в обычном письменном дискурсе (в котором соблюдаются в том числе правила орфографии) максимы Грайса на уровне орфографии не действуют: мы все еще вынуждены писать по правилам, например, слово здравствуйте, хотя такое количество букв явно излишне, как бы это слово ни редуцировалось в речи.

Итак, мы можем наблюдать, по крайней мере в таком узком языковом пространстве, как Интернет, за новым языковым процессом – за превращением одного из жанров письменной формы языка в естественный

<sup>&</sup>lt;sup>1</sup> Например, такого правильного написания не предполагает редуцированная форма слова 4mo. Как правильно писать:  $4\ddot{e}$  или 4o? На этот вопрос, насколько нам известно, пока не может ответить ни один орфографический словарь.

# Орфография в Интернете: анализ одной орфографической ошибки

язык. То есть в такую систему, которая развивается по своим внутренним естественным законам и не терпит вмешательств извне. В таком случае разумно было бы предположить, что норма в этом новом естественном языке, то есть понятие о том, что правильно, а что неправильно, может начать быстро изменяться. Так же быстро, как это происходит в обычном естественном языке, например, с фонетической нормой.

#### 2. Описание ошибки

В этой работе мы попытались проследить за одной очень частотной орфографической ошибкой и взглянуть на нее с позиций естественности письменной формы языка и возможности становления в этой системе новой орфографической нормы.

Эта ошибка заключается в написании мягкого знака после m в формах единственного числа третьего лица возвратных глаголов. Несколько примеров из Интернета:

- (1) а. Мне свет солнца кажеться тусклым.
- б. Сегодня состоиться матч Лиги Чемпионов.
- в. Кто боиться правды?

Эта ошибка в настоящее время действительно является очень распространенной, и особенно в текстах в Интернете. Также мягкий знак может писаться и в форме множественного числа, но мы для простоты ограничимся здесь формами единственного числа.

Отдельную интересную проблему представляет собой задача выделить тот класс глаголов (по-видимому, на основе фонологических признаков основы), для которых такая ошибка типична. Ведь можно привести примеры таких возвратных глаголов, для которых доля случаев написания с ошибкой среди всех случаев написания в Интернете много меньше соответствующей доли для приведенных в (1) глаголов. Так, например, доля написаний *понижаеться* среди всех использований этой формы в Интернете составляет 0,4 %, тогда как соответствующая доля для написания *боиться* — 3,9 %. Мы, однако, в этой работе не пытались выделить те параметры, которые могут влиять на возможность такого типа ошибок для конкретной глагольной основы.

### 3. Анализ частотности ошибки

Наша задача состояла в том, чтобы проанализировать динамику частотности данной ошибки. Для этого нами была использована поисковая система Яндекс. В этой поисковой системе имеется возможность ограничивать область поиска временным интервалом. Собственно, анализ динамики состоял в том, что были взяты десять разных глаголов и для каждого из них произведены следующие вычисления: количество<sup>2</sup> употреблений данной словоформы с ошибкой в течение одного года (для всех годов с 2000 г. по 2007 г.) и количество употреблений данной словоформы без ошибки в течение одного года. Далее эти показатели суммировались и вычислялось отношение количества употреблений с ошибкой к сумме всех употреблений для каждого года. ВременнУю шкалу было решено начать с 2000 года, так как в поисковой системе Яндекс возможность поиска по дате появилась сравнительно недавно, и поэтому для более ранних периодов точность разметки ресурсов по дате уже заметно хуже, чем для периода начиная с 2000 года.

Приведем десять глагольных форм, которые были отобраны для этой работы, в написании с ошибкой:

- боиться
- делаеться
- захочеться
- кажеться
- называеться
- получаеться
- придеться
- состоиться
- считаеться
- являеться

Все эти словоформы были отобраны благодаря следующим свойствам. Ни одна из этих словоформ в написании с ошибкой не совпадает ни с одной реальной словоформой русского языка (в том числе она не совпадает и с инфинитивом своего глагола, что бывает довольно часто). Это свойство позволило осуществлять поиск по запросу состоящему только из самой этой словоформы (в кавычках), при этом была гарантия, что в

<sup>&</sup>lt;sup>2</sup> В качестве показателя количества употреблений бралось количество найденных в Интернете страниц.

## Богданов А.В.

выдаче по этому запросу попадутся именно и только данные словоформы в ошибочном написании, и никакие другие словоформы языка. Также эти слова обладают сравнительно высокой частотностью, что позволяет увеличить общую точность расчета частоты.

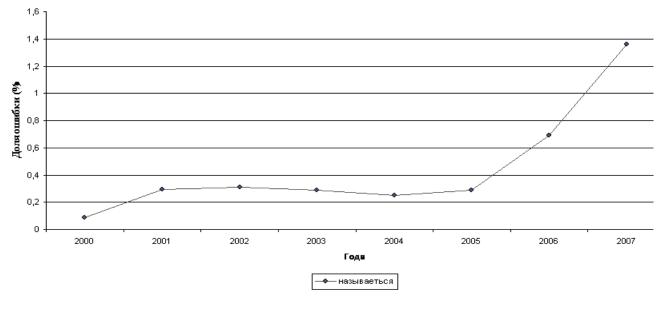
Приведем пример одного такого расчета для одной из этих словоформ.

Словоформа: называеться

Год	Количество употреблений с	Количество употреблений без ошибки	Доля употреблений с ошибкой
	ошибкой ( <i>называеться</i> )	(называется)	(округлено)
2000	22	25147	0,09 %
2001	121	40967	0,29 %
2002	224	71819	0,31 %
2003	410	140577	0,29 %
2004	337	135469	0,25 %
2005	629	218168	0,29 %
2006	3382	486633	0,69 %
2007	34123	2478693	1,36 %

Таблица 1.

На основе представленных в таблице 1 данных можно для наглядности изобразить такой график динамики доли ошибки для данной словоформы.



**График 1** Динамика доли ошибки для сповоформы называеться

Как видно уже только из этого графика, данные оказались весьма интересными, поскольку можно заметить, что начиная с 2005 года доля написаний с такой ошибкой начала резко возрастать.

Аналогичные вычисления были проведены для каждой словоформы из списка, представленного выше. На графике 2 показана динамика доли ошибки для всех десяти словоформ из нашего списка, а также прерывистой линией изображена средняя доля ошибки.

Итак, в графике 2 показано, что по данным, вычисленным для десяти разных словоформ, была построена кривая, показывающая среднюю долю ошибки. Средняя доля считалась просто: для каждого года бралось среднее арифметическое показателей доли ошибки для каждой словоформы.

Мы, однако, не ограничились констатацией того факта, что средняя доля ошибок такого типа начиная с 2005 года начала резко возрастать. Было решено сравнить динамику для этой ошибки с общей динамикой доли орфографических ошибок.

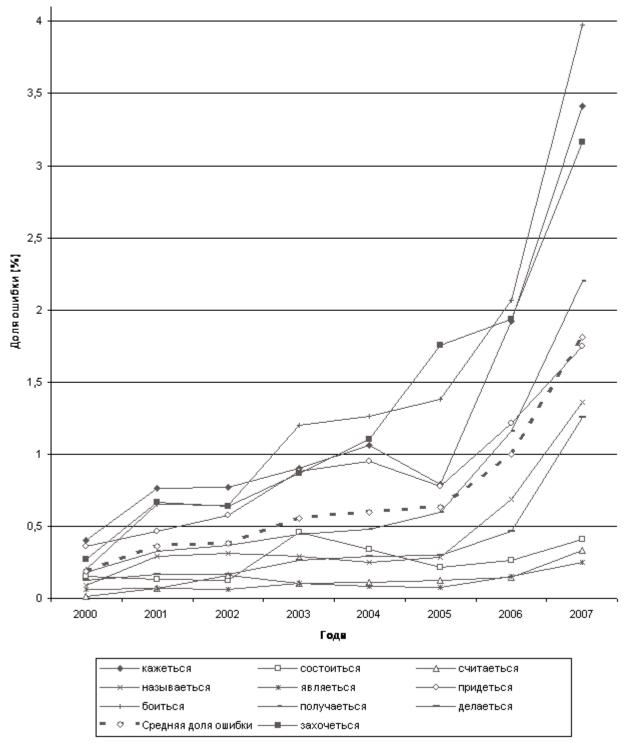


График 2 Динамика доли ошибки для всех десяти словоформ

4. Анализ общей динамики орфографических ошибок

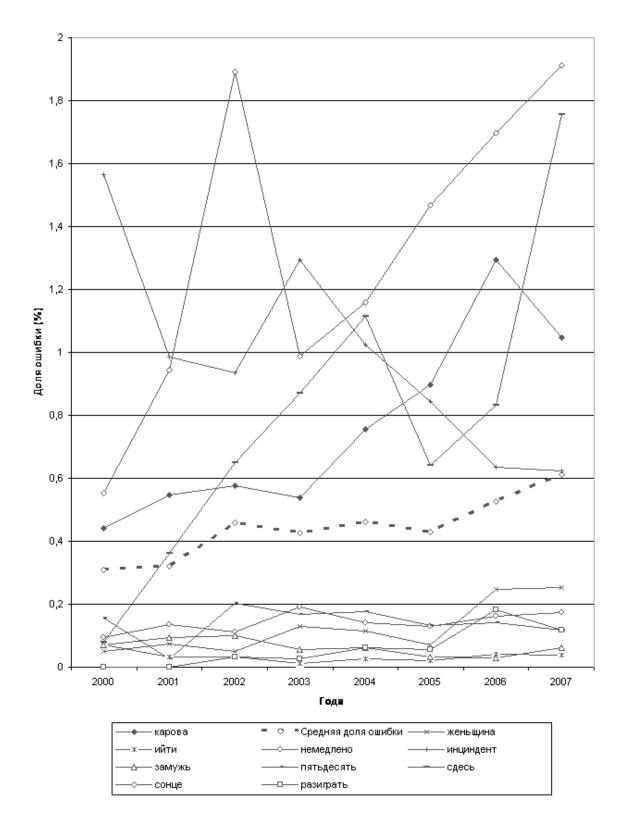
Для того, чтобы вычислить общую динамику орфографических ошибок пользователей Интернета, было взято десять типичных орфографических ошибок, не связанных с написанием мягкого знака в глагольных формах. Список состоял из следующих словоформ (приводится в записи с ошибкой).

- карова
- uŭmu
- немедлено

# Богданов А.В.

- женьщина
- инциндент
- замужь
- пятьдесять
- сдесь
- сонце
- разиграть

Грефик 3 Динамика других орфографических ошибок



# Орфография в Интернете: анализ одной орфографической ошибки

Все эти словоформы, как и глагольные словоформы из списка выше, характерны тем, что каждая из них в таком написании не является никакой реальной словоформой русского языка, что также помогло избежать шума в поисковой выдаче. К тому же они тоже являются довольно частотными словами. Также можно заметить, что все десять словоформ представляют собой примеры разных орфографических ошибок. При выборе этих словоформ мы также пытались избежать пересечения этих написаний с элементами современного сетевого жаргона (в соответствии с которым используются, например, такие написания: превед, креведко и т.п.). На наш взгляд, написания, представленные в списке выше, вряд ли могли бы быть основаны на использовании этого жаргона.

Для этих словоформ были проведены аналогичные вычисления. На графике 3 представлена динамика доли ошибки для всех десяти словоформ, а также прерывистой линией показана средняя доля ошибки.

Отметим, что при всем разнообразии поведения кривых на графике 3 значение образующих точек остается в пределах двух процентов, а среднее значение доли ошибки изменяется от 0,3 процента до 0,6 процента (округлено). Среднее значение на этом графике, как видно, тоже постепенно повышается но далеко не так резко, как среднее значение на графике 2.

#### 5. Сравнение динамики доли ошибок на -ться с общей динамикой орфографических ошибок

Теперь нам осталось лишь сравнить две кривые, каждая из которых представляет собой усредненное значение доли ошибок. Первая – доли ошибок на *тыся/тся*, вторая – доли других орфографических ошибок. На графике 4 эти две кривые представлены вместе.

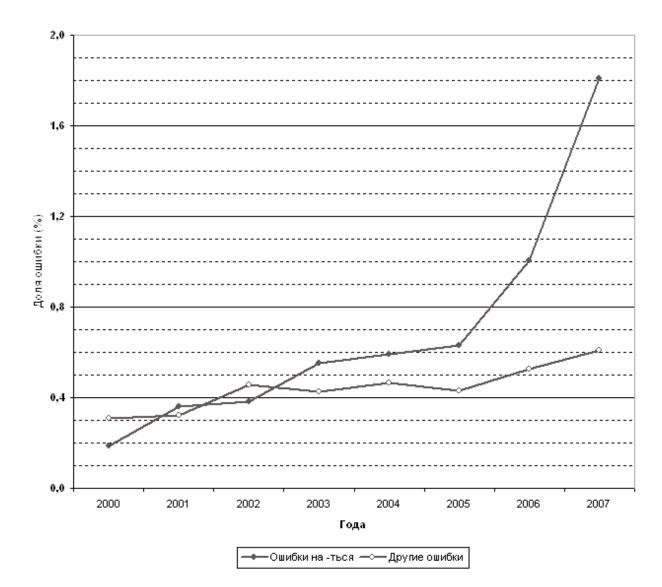


График 4 Сравнение динамики доли ошибок

### Богданов А.В.

Как видно из графика 4 средние значения для доли ошибок в обоих случаях увеличиваются, но начиная с 2005 года доля ошибок на -ться начинает возрастать гораздо более резко, чем доля других ошибок. Конечно, на таком маленьком количестве данных сложно сделать достаточно достоверный вывод, но, как нам кажется, тенденцию можно рассмотреть не прибегая к более обширным исследованиям.

Итак, результатом наших вычислений стал тот факт, что доля написаний глагольных форм третьего лица единственного числа возвратных глаголов с мягким знаком не вписывается по скорости возрастания в общую тенденцию увеличения количества орфографических ошибок. На основании этих данных можно сделать вывод о некоторой иной природе таких ошибок.

Если соображения, высказанные в введении к настоящей работе, имеют право на существование и письменная форма языка в Интернете действительно начинает вести себя так же, как естественный язык, то мы вынуждены констатировать, что в данном случае мы имеем дело с новой нормой написания таких глагольных форм. В самом деле, что еще может являться доказательством становления новой нормы как не аномально возрастающая доля употребления нового варианта на фоне стабильно низкого колебания вариативности в других областях?

Конечно, данное соображение не претендует на повод для изменений в современных правилах правописания, а лишь призвано обратить внимание специалистов на природу этого явления. Однако, если тенденция, показанная нами выше, верна, то при условии возможного влияния новых норм письменного языка Интернета на «офлайновый» письменный язык, изменение действующих правил правописания может стать неизбежным.

#### Список литературы

- 1. Grice H.P. Utterer's meaning and intentions // Philosophical Review, 1969. Vol.78. P. 54 70
- 2. Беликов В.И. & Крысин Л.П. Социолингвистика, М., 2001