

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2020»

Москва, 17–20 июня 2020 г.

ЧЕМУ НАС УЧАТ ОШИБКИ: ТРУДНОСТИ ПРИ ОБРАБОТКЕ СЛОВ С ЧАСТОТНЫМИ ОРФОГРАФИЧЕСКИМИ ОШИБКАМИ¹

Чернова Д. А. (chernovadasha@yandex.ru)

СПбГУ, Санкт-Петербург, Россия

Алексеева С. В. (mail@s-alexeeva.ru)

СПбГУ, Санкт-Петербург, Россия

Слюсарь Н. А. (slioussar@gmail.com)

НИУ ВШЭ, Москва, и СПбГУ, Санкт-Петербург, Россия

DOI: 10.28995/2075-7182-2020-19-147-159

WHAT DO WE LEARN FROM MISTAKES: PROCESSING DIFFICULTIES WITH FREQUENTLY MISPELLED WORDS

Chernova D. A. (chernovadasha@yandex.ru)

St. Petersburg State University, St. Petersburg, Russia

Alexeeva S. V. (mail@s-alexeeva.ru)

St. Petersburg State University, St. Petersburg, Russia

Slioussar N. A. (slioussar@gmail.com)

HSE, Moscow, and St. Petersburg State University,
St. Petersburg, Russia

¹ Исследование выполнено при поддержке РФФИ в рамках научного проекта № 18-00-0646 (18-00-00640) «Закономерности обработки лингвистической информации в ситуации многозначности: процессы активации и выбора вариантов значений».

Even if we know how to spell, we often see words misspelled by other people — especially nowadays when we constantly read unedited texts on social media and in personal messages. In this paper, we present two experiments showing that the incidence of orthographic errors reduces the quality of lexical representations in the mental lexicon—even if one knows how to spell a word, repeated exposure to incorrect spellings blurs its orthographical representation and weakens the connection between form and meaning. As a result, it is more difficult to judge whether the word is spelled correctly, and—more surprisingly—it takes more time to read the word even when there are no errors. We show that when all other factors are balanced the effect of misspellings is more pronounced for the words with lower frequency.

We compare our results with the only previous study addressing the problem of misspellings' influence on the processing of correctly spelled words—it was conducted on the English data. It may be interesting to explore this issue in a cross-linguistic perspective. In this study, we turn to Russian, which differs from English by a more transparent orthography. Much larger corpora of unedited texts are available for English than for Russian, but, using a different way to estimate the incidence of misspellings, we obtained similar results and could also make some novel generalizations.

In Experiment 1 we selected 44 words that are frequently misspelled and presented in two conditions (with or without spelling errors) and were distributed across two experimental lists. For every word, participants were asked to determine whether it is spelled correctly or not. The frequency of the word and the relative frequency of its misspelled occurrences significantly influenced the number of incorrect responses: not only it takes longer to read frequently misspelled words, it is also more difficult to decide whether they are spelled correctly.

In Experiment 2 we selected 30 words from the materials of Experiment 1 and for every selected word, we found a pair that is matched for length and frequency, but is rarely misspelled due to its orthographic transparency. We used a lexical decision task, presenting these 60 words in the correct spelling, as well as 60 nonwords. We used LMMs for statistics. Firstly, the word type factor was significant: it takes more time to recognize a frequently misspelled word, which replicates the results obtained for English. Secondly, the interaction between the word type factor and the frequency factor was significant: the effect of misspellings was more pronounced for the words of lower frequency. We can conclude that high frequency words have more robust representations that resist blurring more efficiently than low frequency ones.

Finally, we conducted a separate analysis showing that the number of incorrect responses in Experiment 1 correlates with RTs in Experiment 2. Thus, whether we consciously try to find an error or simply read words orthographic representations blurred due to exposure to frequent misspellings make the task more difficult.

Keywords: mental lexicon, orthographic errors, processing, Russian

1. Введение

Не все слова, которые мы знаем, мы знаем одинаково хорошо. В каких-то случаях мы не вполне уверены в значении слова, иногда у нас есть сомнения, как оно произносится или пишется. Согласно так называемой гипотезе

лексического качества, Lexical Quality Hypothesis [Perfetti 1985], [2007]; а также [Nelson Taylor, Perfetti 2016]; [Perfetti, Hart, 2001]; [2002]), репрезентации таких слов в ментальном лексиконе оказываются «менее качественными»: некоторые формальные или семантические характеристики более размыты и связи между ними слабее, чем у других слов. Недостаточно четкая информация о буквенном составе слова ведет к орфографическим ошибкам.

Орфографические ошибки — страшная напасть, но мы привыкли думать, что терпение и труд помогут с ней справиться. Однако, даже если мы сами научились писать правильно то или иное слово, никто не может уберечь нас от ошибок, сделанных другими. Эта проблема стала особенно актуальной сейчас, с появлением многочисленных электронных СМИ, чатов, социальных сетей, где колоссальные массивы текстов публикуются без корректорской проверки или не предполагают ее в принципе. Если мы регулярно видим какое-то слово с ошибкой, можно предположить, что неправильное написание фиксируется в ментальном лексиконе и получает подкрепление каждый раз, когда мы сталкиваемся с ним снова. Тогда ментальные репрезентации слов, в которых часто допускают ошибки, окажутся менее качественными даже у тех, кто сам таких ошибок не делает.

Более того, согласно различным теориям, описывающим процессы обучения в целом и применительно к языку (например, [Baayen et al. 2011]; [Ramscar, Dye, McCauley 2013]; [Rescorla, Wagner 1972]), в ситуации, когда два разных набора формальных признаков связаны с одним и тем же значением, формируются более слабые связи. Это мешает запоминанию информации и затрудняет ее последующее извлечение из памяти. Предположим, что с такой точки зрения можно посмотреть и на слова, в которых часто допускают орфографические ошибки. Тогда можно ожидать, что не только написание, но и обработка таких слов при чтении окажется затруднена — даже тогда, когда эти слова написаны правильно.

Подобная гипотеза впервые была сформулирована в работе С. Рахманьян и В. Купермана [Rahmanian, Kuperman 2019], которые протестировали ее на материале английского языка. Они обратились к семимиллиардному корпусу, содержащему неотредактированные сообщения из новостных групп на платформе USENET [Shaoul, Westbury, 2013], и отобрали слова, которые чаще или реже встречаются в неправильном написании. Неопределенность, создающаяся за счет конкуренции между различными вариантами, было решено описывать при помощи такой меры, как энтропия (H) — ее апробация для лингвистических исследований была проведена в статье [Milin et al. 2009] и др. Получилось, что, например, для слова *innocent* ‘невинный’ $H=0,89$: 141 960 вхождений с правильным написанием и 62 665 вхождений (31 %) с ошибкой: *inocent*. У слова *necessary* ‘необходимый’ $H=0,14$ — ошибочный вариант *necessary* составляет всего 2 % от общего числа вхождений.

Затем было проведено два эксперимента. В первом участникам было предложено читать предложения, содержащие отобранные слова и создающие для них нейтральный контекст. Велась запись движений глаз. Во втором эксперименте использовался метод лексического решения: участники видели на экране цепочки

букв и должны были как можно скорее ответить, являются ли они словами английского языка или псевдословами, нажав на одну из двух клавиш. И в первый, и во второй эксперимент целевые слова вошли в правильное написание (в качестве псевдослов использовались цепочки, не имеющие к ним отношения). В обоих экспериментах получилось, что фактор энтропии оказывает значимое влияние на скорость чтения слов — наравне с такими хорошо известными факторами, как частотность и длина. Таким образом, даже когда слово написано правильно, наличие частотных орфографических ошибок затрудняет его обработку.

С. Рахманьян и В. Куперман также предположили, что эффект энтропии будет более сильным для слов с более высокой частотностью и для более начитанных участников. Чем чаще человек видит то или иное слово (за счет частотности слова или за счет того, что этот человек больше читает), тем лучше должна закрепиться информация о вероятностном распределении различных вариантов. Такая логика не кажется нам бесспорной: ведь у более редких слов и менее опытных читателей должно быть больше проблем с качеством лексических репрезентаций, что должно способствовать раздуванию эффекта энтропии. В любом случае, подтвердить или опровергнуть эти гипотезы пока не удалось.

Участники экспериментов проходили несколько тестов, призванных оценить общую начитанность, орфографическую и орфоэпическую грамотность и словарный запас, включая тест ART/MRT на знание фамилий авторов и названий журналов [Stanovich, West 1989], произносительный тест TOWRE [Torgeson et al. 1999] и орфографические тесты из работы [Andrews, Hersch 2010]. Однако, хотя результаты тестов коррелировали с общей скоростью чтения, ни в одном случае не удалось выявить их взаимодействия с фактором энтропии. Что касается фактора частотности, результаты двух экспериментов были противоположными: в первом эффект энтропии был больше для более частотных слов, а во втором — наоборот, для менее частотных.

Так как пока было сделано всего одно исследование на материале одного языка и некоторые результаты оказались противоречивыми, мы провели два эксперимента на русском языке. К сожалению, у нас нет корпуса неотредактированных текстов, сравнимого по объему с корпусом USENET [Shaoul, Westbury, 2013], что не позволяет с высокой точностью оценить энтропию для разных орфографических вариантов. Поэтому мы начали с того, что, опираясь на корпусные и другие источники, составили список слов, в которых особенно часто допускают орфографические ошибки (подробнее об этом см. в разделе 2). В первом эксперименте мы отобрали часть этих слов и подобрали к ним пары, в которых ошибки встречаются очень редко, сбалансировав слова попарно по длине и частотности. Мы сравнили скорость чтения в двух группах, используя метод лексического решения.

Второй эксперимент изначально задумывался как независимое исследование. Дело в том, что для русского языка есть много упражнений и проверочных заданий на знание орфографии, но нет таких тестов, как упомянутые выше для английского языка. Результат в них имеет числовое выражение, которое затем можно использовать в исследованиях, где орфографическая грамотность является одним из факторов. Так как нам был необходим такой тест для большого

проекта, посвященного чтению на родном и иностранном языке, мы разработали его, используя упомянутый выше список «орфографически сложных» слов. Во время прохождения теста участник видит половину слов в правильном написании, половину — в частотном неправильном, его задача — отличить первые от вторых, нажимая на одну из двух клавиш.

Основная характеристика для тестов такого рода — достаточно дробные результаты, позволяющие эффективно ранжировать участников, и с этой точки зрения созданный нами тест оказался удачным. Однако, когда мы стали анализировать данные подробнее и задались вопросом, почему ошибки в одних словах замечаются намного чаще, чем в других, стало понятно, что результаты интересно сопоставить с первым экспериментом — особенно учитывая, что у С. Рахманьян и В. Купермана не было подобных данных для английского языка (в их работе целевые слова всегда предъявлялись в правильном написании). Ниже мы начинаем с анализа этих данных, а затем переходим к эксперименту с лексическим решением.

2. Эксперимент 1 с задачей определения ошибок

Устойчивость графического представления слов в памяти, или орфографическая зоркость, может быть оценена с помощью теста на узнавание правильно и неправильно написанных слов. Для английского языка такой инструмент, Spelling recognition test, был разработан С. Эндрюс и коллегами [Andrews, Hersch, 2010]. Мы задались целью создать аналогичный тест на материале русского языка. В этой работе собранные нами данные анализируются не с точки зрения информативности теста (разброс между участниками, корреляция результатов с различными социалингвистическими переменными и др.), а чтобы выявить влияние ряда факторов на успешность распознавания правильных и ошибочных написаний.

2.1. Материал

Для первичного отбора материала мы опирались на данные различных ресурсов для подготовки к экзаменам по русскому языку, где перечислены трудные для написания слова, например, <http://slova.textologia.ru/>. Затем мы отобрали 44 слова, которые часто встречаются в неверном написании — как по данным основного подкорпуса Национального корпуса русского языка (НКРЯ, www.ruscorpora.ru), куда входят в основном редактируемые тексты, так и по данным подкорпуса «Социальные сети» корпуса «Тайга» [Shavrina, Kurmachova, n.d.], куда входят тексты нередктируемые (этот корпус больше подходит нам по содержанию, но еще не отредактирован окончательно для удаления повторяющихся фрагментов и т.д.). Ниже при анализе результатов используется информация о частотности, извлеченная из корпуса «Тайга», так как там намного шире представлены неправильные написания. В числе отобранных оказались слова с непроизносимыми согласными (*агентство, сердце, соответственно*), заимствованные

слова с удвоенными согласными (*аппендицит, аккомпанемент, терраса*) и без таковых (*галерея, коридор*), слова, имеющие ошибочный вариант с эпентезой (*поскользнуться, педиатр, эспрессо, прецедент*), слова с орфограммами в приставках (*предыстория, расчет, чересчур*), корнях (*просвещение, помощник, участвовать*) и суффиксах (*серебряный, ветреный, девчонка*). Было составлено два протокола: в первом одна случайно отобранная половина слов была предъявлена в верном написании, а другая — в частотном ошибочном, во втором протоколе — наоборот.

2.2. Участники

В исследовании на добровольной основе приняли участие 166 носителей русского языка: 136 женщин и 30 мужчин в возрасте от 18 до 42 лет.

2.3. Процедура

Участникам на экране компьютера предъявлялись слова и ставилась задача определить, написаны они верно или неверно. Слова были упорядочены по алфавиту. Участники не были ограничены во времени, скорость принятия решения не регистрировалась. Длительность эксперимента составила в среднем около 5 минут.

2.4. Анализ и результаты

Найти в стимуле определенную ошибку или удостовериться, что в нем нет ни одной ошибки, — две разные задачи, поэтому данные по правильным и неправильным написаниям надо анализировать отдельно. В силу ограниченного объема данной работы мы анализируем только данные по правильным написаниям — именно они важны в первую очередь для сравнения со вторым экспериментом и с предыдущими исследованиями. Был проведен статистический анализ с использованием смешанной логистической регрессии (GLMM). Зависимой переменной была точность ответа (верный / неверный), а независимыми — частотность ошибочных вариантов написания слова и общая его частотность в корпусе «Тайга» (фиксированные эффекты).

Был обнаружен эффект частотности слова ($b = 0,68$, $SE = 0,14$, $z = 4,68$, $p < 0,01$): чем частотнее слово, тем легче понять, что оно написано правильно (см. **Рис. 1**). Это ожидаемый результат. Более важно, что мы также обнаружили влияние частотности ошибочных вариантов ($b = -0,45$, $SE = 0,19$, $z = -2,42$, $p = 0,02$): чем частотнее у слова ошибочные написания, тем сложнее понять, что оно написано правильно (см. **Рис. 2**). Это можно объяснить изложенной выше теорией, согласно которой ошибочные варианты размывают репрезентацию слова в ментальном лексиконе. Важно подчеркнуть, что зависимость удалось установить несмотря на то, что информация об ошибочных написаниях, доступная для русского языка, намного беднее, чем та, что доступна для английского. Далее проверим, влияют ли они на *скорость* обработки правильно написанных слов, как это показали С. Рахманьян и В. Куперман для английского языка.

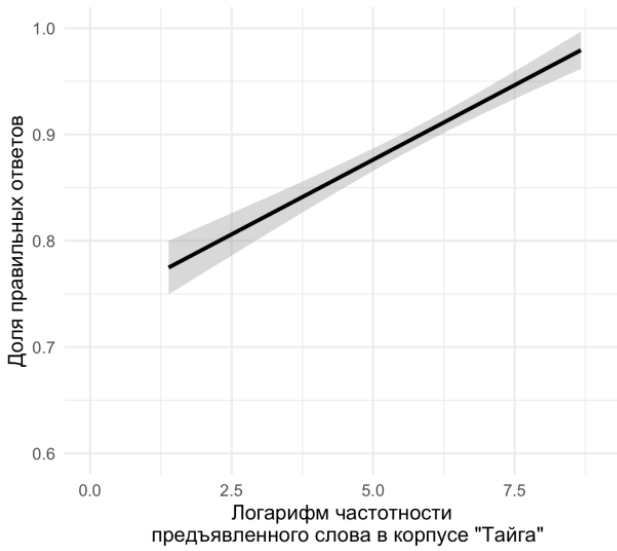


Рис. 1. Эффект частотности слова в Эксперименте 1

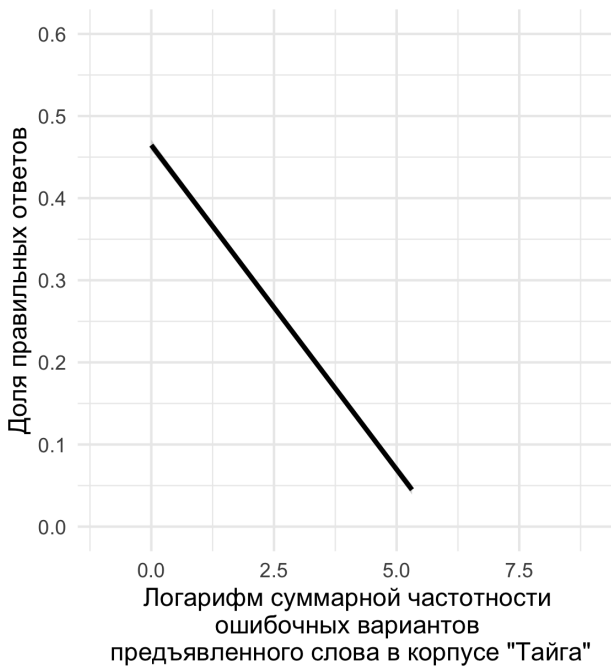


Рис. 2. Эффект частотности ошибочных вариантов в Эксперименте 1

3. Эксперимент 2 с задачей лексического решения

3.1. Материал

В качестве стимульного материала были отобраны 30 слов из числа использованных в предыдущем эксперименте. К каждому из этих слов, которые условно считались «сложными для написания», было подобрано в качестве пары «простое для написания» слово, то есть такое, где написание легко восстанавливается из звучания, а в НКРЯ и в корпусе «Тайга» не встречается ни одного неверного варианта. При этом «сложное» и «простое» слово совпадали по длине и имели минимальные отличия в частотности, например, *дилемма* (ipm 2,9) — *флагман* (ipm 2,1), *расчет* (ipm 108,9) — *клетка* (ipm 109,6)². В эксперименте все слова предъявлялись в правильном написании. Кроме этого, в стимульный материал было включено 60 псевдослов, на одну букву отличающихся от реально существующих.

3.2. Участники

В эксперименте на добровольной основе принял участие 41 носитель русского языка: 35 женщин и 6 мужчин в возрасте от 18 до 21 года, не участвовавшие в предыдущем исследовании.

3.3. Процедура

Была использована методика лексического решения: участник должен был определить, является ли предъявленная ему на экране последовательность букв словом русского языка. Эксперимент проводился на персональном компьютере с использованием программы PsychoPy [Peirce et al. 2019]. Сначала для фиксации взгляда в центре экрана на 500 мс предъявлялся символ «*», после чего также в центре экрана появлялась стимульная последовательность. Если испытуемый не принимал решения в течение 1 секунды, она сменялась символом «*» и следующим стимулом. Целевые слова и псевдослова предъявлялись в случайном порядке. Длительность эксперимента составляла около 10 минут.

3.4. Анализ данных и результаты

Никто из участников не ошибся более чем в 15% случаев, поэтому ничьи данные не были исключены из дальнейшего анализа. Мы анализировали время реакции испытуемых на целевые слова. При этом мы удалили те 149 случаев (6% данных), где испытуемые дали неверный ответ. Затем, пользуясь процедурой, описанной в работе [Ratcliff 1993], мы рассчитали для каждого

² Для подбора пары использовалась база данных StimulStat (http://stimul.cognitivestudies.ru/ru_stimul/, [Alexeeva et al. 2018]), данные о частотности лемм здесь и далее даны по словарию [Lyashevskaya, Sharov 2009].

экспериментального условия пороговые значения времени реакции (среднее время реакции плюс или минус 2,5 стандартных отклонения) и удалили все данные за пределами этих пороговых значений, а именно, 40 случаев, что составляет 1,7% от общего количества.

Для статистического анализа мы воспользовались смешанными линейными моделями (LMM)³. Значимым оказался эффект «тип слова» ($b = -0,07$, $SE = 0,03$, $t = -2,75$, $p = 0,01$): участникам исследования требовалось больше времени, чтобы распознать правильно написанное слово, имеющее частотные ошибочные варианты написания, чем правильно написанное слово, которое таких вариантов не имеет (см. **Рис. 3**). Это подтверждает общую гипотезу С. Рахманьян и В. Купермана [Rahmanian, Kuperman 2019] на новом материале.

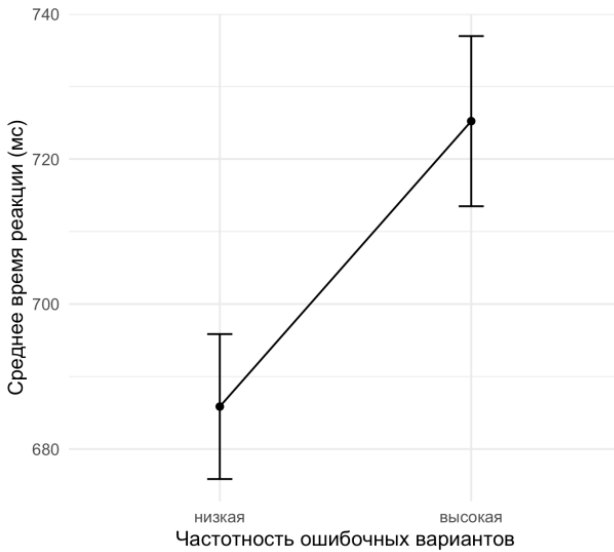


Рис. 3. Эффект наличия ошибочных вариантов в Эксперименте 2

Затем мы провели дополнительный анализ: помимо фактора «тип слова» (сложное / простое) модель включала факторы «длина слова» (количество букв) и «частотность слова» (высокая / низкая), а также взаимодействия между типом и частотностью/длиной. Разделение на высоко- и низкочастотные слова осуществлялось на основе медианного значения (средняя частотность в первой группе равнялась 88,5 ipm, а во второй — 9,5 ipm). Были обнаружены эффекты длины ($b = 0,04$, $SE = 0,006$, $t = 4,07$, $p < 0,01$) и частотности слова ($b = 0,12$, $SE = 0,026$, $t = 4,79$, $p < 0,01$): высокочастотные слова обрабатываются быстрее низкочастотных, а короткие — быстрее длинных. Так же, как и в предыдущем

³ Во всех анализах с использованием LMM зависимая переменная была введена в модель в трансформированном виде (использовалась обратная трансформация). Это было необходимо для того, чтобы так называемые остатки модели (та вариативность, которую не удалось предсказать) были распределены нормально.

анализе, был значим фактор «тип слова», т. е. наличие частотных ошибочных вариантов ($b = -0,07$, $SE = 0,026$, $t = -2,70$, $p < 0,01$). Однако более важно, что мы получили значимое взаимодействие между частотностью и типом слова ($b = -0,11$, $SE = 0,052$, $t = -2,16$, $p = 0,03$): репрезентация слова в ментальном лексиконе страдает от ошибочных вариантов сильнее, если слово низкочастотное (см. **Рис. 4**). Это противоречит гипотезе С. Рахманьян и В. Купермана [Rahmanian, Kuperman 2019] о роли частотности, но согласуется с данными, которые они получили в аналогичном эксперименте. Взаимодействие между длиной и типом слова не было значимым.

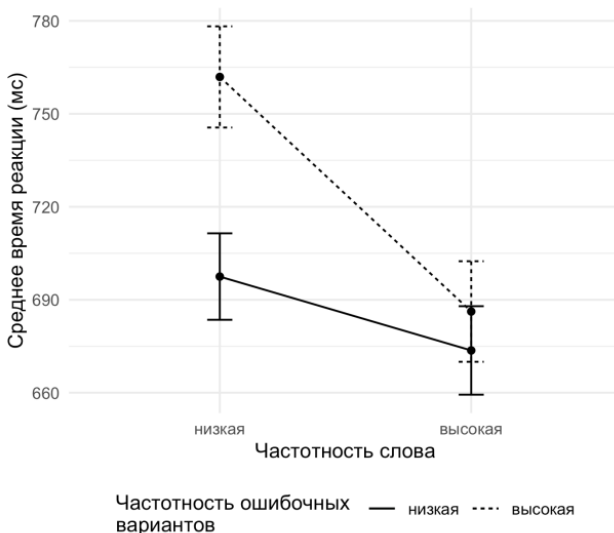


Рис. 4. Эффект частотности слова и наличия ошибочных вариантов в Эксперименте 2

Наконец мы задались вопросом, связана ли легкость распознавания правильного написания слова в первом эксперименте со временем его прочтения во втором. Для этого мы отобрали времена реакции на слова, которые имеют частотные ошибочные варианты — только они были представлены в первом эксперименте. Статистический анализ зафиксировал значимое влияние успешности определения правильного написания на скорость обработки ($b = -0,05$, $SE = 0,19$, $t = -2,82$, $p = 0,01$): чем больше правильных ответов слово получает в тесте на орфографическую грамотность, тем быстрее оно обрабатывается в эксперименте на лексическое решение (см. **Рис. 5**).

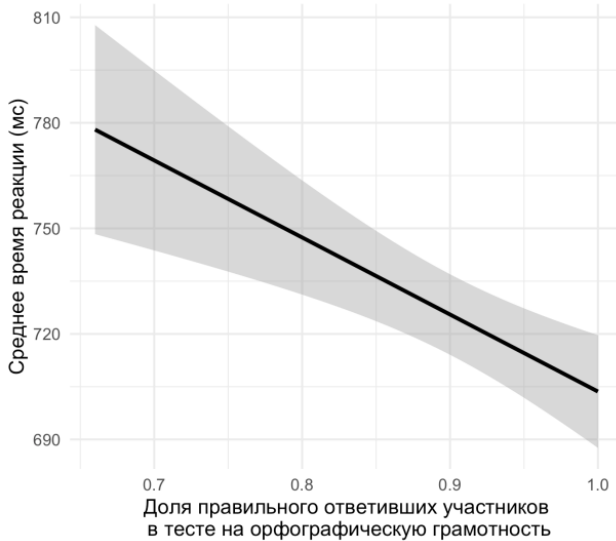


Рис. 5. Влияние успешности распознавания правильного написания в Эксперименте 1 на время реакции в Эксперименте 2

4. Заключение

Орфографическим ошибкам посвящено много исследований, но до последнего времени изучались только закономерности их возникновения при порождении. Между тем, из гипотезы лексического качества ([Perfetti 1985], [2007] и др.), согласно которой частые ошибки свидетельствуют о более размытой репрезентации слова в ментальном лексиконе, и из общих моделей обучения (например, [Baayen et al. 2011]; [Ramscar, Dye, McCauley 2013]; [Rescorla, Wagner 1972]) следует, что слова, в которых чаще допускают ошибки, должны вызывать повышенные трудности и при восприятии. Эта гипотеза впервые была протестирована С. Рахманьян и В. Куперманом [Rahmanian, Kuperman 2019] в двух экспериментах на материале английского языка. В целом она подтвердилась, однако некоторые результаты оказались противоречивыми.

Чтобы проверить эту гипотезу на другом материале и разрешить возникшие противоречия, мы провели два эксперимента на материале русского языка. Основная сложность заключалась в том, что только для английского языка есть корпуса, содержащие огромные массивы неотредактированных текстов. Тем не менее, частотности правильных и неправильных написаний, извлеченные из подкорпуса «Социальные сети» корпуса «Тайга» [Shavrina, Kurmachova, n. d.], показали значимую корреляцию с результатами первого эксперимента, в котором участникам надо было определять слова, написанные правильно и с орфографическими ошибками, и позволили эффективно разделить стимулы во втором эксперименте на две группы: орфографически сложные и простые.

Оба эксперимента подтверждают гипотезу о том, что слова, в которых чаще допускают орфографические ошибки, оказываются более сложными и при восприятии — даже когда они написаны правильно. Такие слова медленнее читаются — результат, аналогичный полученному С. Рахманьян и В. Куперманом на материале английского языка, — а оценка правильности их написания вызывает больше сложностей. Результаты двух экспериментов показали значимую корреляцию друг с другом.

С. Рахманьян и В. Куперман предположили, что этот эффект будет сильнее для высокочастотных слов, которые мы видим чаще, чем для низкочастотных, однако их эксперименты дали противоречивые результаты. Проведенный нами эксперимент свидетельствует в пользу обратной картины: эффект значимо сильнее для слов с более низкой частотностью. Это можно объяснить тем, что графический облик высокочастотных слов более устойчив в ментальном лексиконе и менее подвержен «размыванию» ошибочными вариантами написания.

Литература

1. Alexeeva S., Slioussar N., Chernova D. (2018), StimulStat: a lexical database for Russian. *Behavior Research Methods*, Vol. 50, pp. 2305–2315.
2. Andrews S., Hersch J. (2010), Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, Vol. 139, p. 299.
3. Baayen R. H., Milin P., Durdevic D. F., Hendrix P., Marelli M. (2011), An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, Vol. 118, pp. 438–481.
4. Lyashevskaya O., Sharov S. (2009), The frequency dictionary of modern Russian language based on Russian National Corpus [Chastotnyj slovar' sovremennogo russkogo jazyka (na materialakh Nacional'nogo korpusa russkogo jazyka)]. Moscow: Azbukovnik.
5. Milin P., Kuperman, V., Kostic A., & Baayen R. H. (2009), Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. *Analogy in grammar: Form and acquisition*, pp. 214–252.
6. Nelson Taylor J., Perfetti C. A. (2016), Eye movements reveal readers' lexical quality and reading experience. *Reading and Writing*, Vol. 29, pp. 1069–1103.
7. Peirce J. W., Gray J. R., Simpson S., MacAskill M. R., Höchenberger R., Sogo H., Kastman E., Lindeløv J. (2019), [PsychoPy2: experiments in behavior made easy](#). *Behavior Research Methods*, Vol. 51, pp. 195–203.
8. Perfetti C. A. (1985), *Reading ability*. Oxford, Oxford University Press.
9. Perfetti C. A. (2007), Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, Vol. 11, pp. 357–383.
10. Perfetti C. A., Hart L. (2001), The lexical basis of comprehension skill. On the consequences of meaning selection: Perspectives on resolving lexical ambiguity, American Psychological Association, Washington, DC, pp. 67–86.

11. *Perfetti C. A., Hart L.* (2002), The lexical quality hypothesis. Precursors of functional literacy, Vol. 11, pp. 67–86.
12. *Rahmanian S., Kuperman V.* (2019), Spelling errors impede recognition of correctly spelled word forms. *Scientific Studies of Reading*, Vol. 23, pp. 24–36.
13. *Ramscar M., Dye M., McCauley S. M.* (2013), Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, Vol. 89, pp. 760–793.
14. *Ratcliff R.* (1993), Methods for dealing with reaction time outliers. *Psychological Bulletin*, Vol. 114, pp. 510–532.
15. *Rescorla R. A., Wagner A. R.* (1972), A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, Vol. 2, pp. 64–99.
16. *Shaoul C., Westbury C.* (2013), A reduced redundancy USENET corpus (2005–2011). University of Alberta, 39, pp. 850–863.
17. *Shavrina T., Kurmachova Y.* (n. d.), Taiga Corpus. An open-source corpus for machine learning, available at: https://tatianashavrina.github.io/taiga_site/.
18. *Stanovich K. E., West R. F.* (1989), Exposure to print and orthographic processing. *Reading Research Quarterly*, Vol. 24, pp. 402–433.
19. *Torgeson J. K., Wagner R. K., Rashotte C. A.* (1999), Test of word reading efficiency (TOWRE). Austin, TX, ProEd.