

Some errors are more harmful than others: the role of type and frequency of orthographic errors in word processing

Natalia Slioussar
HSE, St. Petersburg
State University,
St. Petersburg, Russia
slioussar@gmail.com

Ivan Gurkov
St. Petersburg
State University,
St. Petersburg, Russia
anhryrs@gmail.com

Daria Chernova
St. Petersburg
State University,
St. Petersburg, Russia
d.chernova@spbu.ru

Abstract

In the modern world of social media, we constantly read texts that were not subject to professional proof-reading and editing. As a result, we see misspelled words more often than the previous generations. The effects are interesting for several scientific disciplines, including psycholinguistics. Several experiments on different languages have recently demonstrated that the incidence of orthographic errors for a particular word reduces the quality of its lexical representation in the mental lexicon. 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.

In the present study, our goal was to find out whether the type of orthographic errors (the orthogram) plays a role in addition to their incidence. We selected six orthograms forming two groups: more and less difficult ones. Two experiments were conducted. In Experiment 1, participants were asked to judge whether stimulus words are spelled correctly. In Experiment 2, all stimulus words were presented in the correct spelling in a lexical decision task, i.e. we measured how fast they are processed. In Experiment 1, the type of the orthogram played a significant role, being more important than any other factor. In Experiment 2, the influence of this factor could be reduced to the incidence of incorrect spellings.

In other words, when we consciously decide how to spell, the type of orthogram plays a crucial role. As a result, some errors are more frequent than others (although it is obvious that the incidence of errors can be only partially predicted by the type of orthogram). However, when we simply read words, only the incidence of errors matters, i.e. the type of orthogram affects the reading speed only indirectly.

Keywords: mental lexicon, orthographic errors, word processing, Russian

DOI: 10.28995/2075-7182-2022-21-1149-1157

Одни ошибки вреднее других: роль типа и частотности орфографических ошибок в обработке слов

Слюсарь Н. А.
НИУ ВШЭ, СПбГУ,
Санкт-Петербург, Россия
slioussar@gmail.com

Гурков И. Е.
СПбГУ,
Санкт-Петербург, Россия
anhryrs@gmail.com

Чернова Д. А.
СПбГУ,
Санкт-Петербург, Россия
d.chernova@spbu.ru

Ключевые слова: ментальный лексикон, орфографические ошибки, обработка слов, русский язык

1 Введение

В последние годы в сети Интернет многократно возросло количество текстов, не подвергающихся редактуре и корректуре. Это записи в социальных сетях, чатах, форумах, статьи ряда электронных СМИ. В них часто попадаются слова с орфографическими ошибками. Соответственно, общее количество ошибочных написаний, с которыми нам приходится сталкиваться, также возрастает.

Как показывают многочисленные исследования, орфографические нормы осваиваются и закрепляются в процессе чтения (например, Andrews, Hersch, 2010; Hersch, Andrews, 2012). Следовательно, часто встречающиеся ошибки оказывают непосредственное влияние на репрезентацию слов в ментальном лексиконе носителя языка. Ошибочные написания фиксируются в памяти и

получают подкрепление каждый раз, когда мы сталкиваемся с ними снова, доля встретившихся нам верных написаний при этом снижается, и неправильный вариант закрепляется в ментальном лексиконе наряду с верным (Perfetti, Hart, 2002).

В соответствии с моделями дискриминативного научения (Baayen et al., 2011; Ramscar et al., 2010; Ramscar et al., 2013; Rescorla, Wagner, 1972), связь между сигналом и его значением ослабевает, если на то же значение указывает также и другой сигнал. Это в свою очередь снижает так называемое лексическое качество слова (Perfetti, 1985; 2007; Nelson et al., 2016; Perfetti, Hart, 2001). Таким образом гипотеза лексического качества объясняет, как неправильные варианты написания слова могут существенно повлиять на его репрезентацию в ментальном лексиконе, даже если они заметно менее частотны, чем правильное.

Эта гипотеза изначально использовалась для того, чтобы объяснить, в каких словах мы чаще допускаем орфографические ошибки. Однако С. Рахманьян и В. Куперман предположили, что более низкое лексическое качество будет влиять и на обработку (Rahmanian, Kuperman, 2019). Они показали на материале английского языка, что слова, в которых часто встречаются ошибки, при прочих равных читаются медленнее, чем слова, в которых ошибки редки (в исследовании все слова фигурировали в правильном написании). Аналогичные результаты были затем получены на материале языков с различными типами орфографии: греческого, финского, иврита, китайского (Kuperman et al. 2021). В отдельном исследовании эта гипотеза была подтверждена на материале русского языка (Чернова и др., 2020).

В этой статье также было показано, что частотность неправильных написаний влияет на скорость и точность в задании на определение орфографической правильности (где половина слов предъявляется в неправильном написании). Этот результат можно считать более ожидаемым, чем выводы экспериментов о скорости прочтения слов в правильном написании, но он окажется важным для дальнейшей дискуссии. Кроме того, Чернова с коллегами показали интересные корреляции между результатами экспериментов, использующих разные методики.

Однако в этих работах не ставился вопрос о том, влияет ли на эти процессы тип орфограммы. Русский язык оптимально подходит для исследования этой проблемы, так как носители русского регулярно допускают ошибки разных типов: выбирают не ту гласную в безударном слоге, сомневаются, писать одну согласную или две и так далее. Можно ли утверждать, что все эти ошибки одинаково легко — или одинаково трудно — заметить, и важна только относительная частотность неправильных написаний, исследованная в статье (Чернова и др., 2020)? Или же тип орфограммы — отдельный фактор, возможно, даже более сильный, чем частотность? Ответу на этот вопрос посвящен наш первый эксперимент, в котором мы анализировали скорость и точность в задании на определение орфографической правильности. Во втором эксперименте исследовалось, влияет ли тип ошибки на скорость чтения слов в правильном написании.

2 Эксперимент 1

2.1 Участники

Участники — 31 носитель русского языка (23 женщины, 8 мужчин) в возрасте от 19 до 28 лет (средний возраст — 21 год). Все испытуемые приняли участие в эксперименте на добровольной основе.

2.2 Материалы

При подборе стимульного материала мы опирались на данные Генерального Интернет-корпуса Русского Языка (ГИКРЯ) (Piperski et al., 2013; Belikov et al., 2013). Данный корпус включает в себя материалы крупнейших ресурсов Рунета и содержит 19 801 миллионов слов. Он позволяет получить статистические данные о частотности употребления орфографически правильных и ошибочных вариантов слов в неотредактируемых текстах (в отличие от НКРЯ, где представлены преимущественно тексты, прошедшие редактуру и корректуру). Важным параметром для нашего исследования была относительная частотность, или доля, неправильных написаний в ГИКРЯ. Так как мы опирались на данные ГИКРЯ для оценки этого параметра, общую частотность слова (включающую как правильные, так и неправильные варианты написания) мы также рассчитывали по ГИКРЯ.

Для исследования были подобраны слова, содержащие орфограммы следующих типов: двойная согласная (ДС, *кристалл, аммиак*); ложная двойная согласная (ЛДС, *калории, продюсер*); безударная непроверяемая гласная (НБГ, *нотариус, винегрет*); гласная после шипящего (ГПШ, *парашиют, трущобы*); безударная проверяемая гласная (БПГ, *храбрец, тишина*) и глухая согласная на конце слова (ГС, *шкаф, синтез*). Опираясь на различные ресурсы, где приводятся частотные ошибки, а также на собственную интуицию, мы отобрали для первых четырех типов по десять слов с достаточно высокой частотностью неправильных написаний. Последние два типа рассматривались нами как контрольные условия: взрослые носители русского языка редко делают ошибки такого типа, и процент неправильных написаний у отобранных нами слов был низким ($\leq 0,6\%$).

Далее мы будем называть эти две группы орфограмм *сложными* и *простыми*. Отметим, что в случае т.н. простых орфограмм можно проверить, какое написание является правильным, поставив слово в форму косвенного падежа или подобрав морфологически родственное слово. В случае сложных орфограмм надо либо знать более сложное общее правило (правописание гласных после шипящих), либо просто помнить правильное написание наизусть. Для слов, имеющих несколько вариантов ошибочных написаний (*порашиют, парошиют, порошиют, парашиут*), выбирался наиболее частотный вариант и в соответствии с ним слово попадало в ту или иную категорию орфограмм.

2.3 Процедура

По принципу латинского квадрата было составлено два экспериментальных листа: в первом половина слов была предъявлена в верном написании, а другая — в ошибочном; во втором листе — наоборот. Эксперимент проводился с помощью интернет-платформы IxexFarm (Drummond et al., 2016). На экране компьютера испытуемого в случайном порядке предъявлялись слова и ставилась задача определить, верно ли они написаны (первые три пробы были тренировочными). На принятие решения отводилось максимум 5 сек, межстимульный интервал составлял 1 сек. Скорость и точность ответа испытуемого регистрировалась.

2.4 Результаты и обсуждение

Средняя скорость и точность ответов для слов в ошибочном и в правильном написании в зависимости от типа орфограммы приведена в Таблице 1. Также там приводится средний процент неправильных вариантов. Из таблицы явствует, что он сильно отличается для «простых» и «сложных» орфограмм, но и внутри двух этих групп есть существенные отличия. Тем не менее, видно, что различия в точности явно определяются не только или даже не столько этим фактором: в группе НБГ больше неправильных ответов, чем в группах ДС и ГПШ.

Орфограмма	Средний % неправильных вариантов	Написание	% неверных ответов	Среднее время ответа (в мс)
Ложная двойная согласная (ЛДС)	8,1%	правильное	28%	1742,5
		неправильное	59%	1812,2
Двойная согласная (ДС)	7,0%	правильное	8%	1676,9
		неправильное	20%	1598,0
Гласная после шипящего (ГПШ)	6,5%	правильное	5%	1603,3
		неправильное	30%	1849,0
Непроверяемая безударная гласная (НБГ)	3,6%	правильное	16%	1834,0
		неправильное	40%	1895,7
Глухая согласная на конце слова (ГС)	<0,1%	правильное	5%	1236,6
		неправильное	0%	1181,9
Проверяемая безударная гласная (БПГ)	<0,1%	правильное	1%	1090,2
		неправильное	1%	1101,9
Всего		правильное	10%	1517,3
		неправильное	25%	1557,6

Таблица 1. Результаты эксперимента 1

Для статистического анализа скорости и точности в зависимости от различных факторов использовался метод смешанной логистической регрессии (GLMM) и смешанной линейной регрессии (LMEM). Для попарных сравнений *post-hoc* использовался тест Тьюки. Весь статистический анализ проводился с использованием языка программирования R (R Core Team, 2021). Использовались библиотеки *lme4* (Bates et al., 2015) и *lmerTest* (Kuznetsova et al., 2017).

Таблица 1 подтверждает реальность разделения орфограмм на простую и сложную группы: в первой в среднем ниже время реакции и выше точность. Также можно отметить, что точность для правильных написаний значимо выше, чем для неправильных ($b = 1,42$, $SE = 0,16$, $z = 9,05$, $p < 0,001$). Это может объясняться тем, что правильные написания всегда частотнее неправильных, поэтому уверенность испытуемых выше. Кроме того, неправильные написания следует отвергнуть несмотря на то, что многие из них регулярно встречаются носителю русского языка и потому достаточно хорошо зафиксированы в памяти. При этом для времени реакции значимых отличий нет.

Доля неправильных написаний оказывает значимое влияние на точность ответа ($b = 0,13$, $SE = 0,04$, $z = 3,58$, $p < 0,001$): чем чаще слово встречается с ошибкой, тем больше ошибок испытуемые допускают в тесте. Аналогичная картина наблюдается со скоростью ответа ($b = 25,29$, $SE = 7,28$, $z = 3,48$, $p < 0,001$). Общая частотность слова влияет на точность ответов, но в меньшей степени, чем доля неправильных написаний ($b = -0,004$, $SE = 0,001$, $z = -2,42$, $p < 0,001$): чем частотней слово, тем проще определить, правильно ли оно написано. Влияние этого фактора на скорость ответов не достигло значимости. Длина стимульных слов не оказалась значимым фактором. Теперь перейдем к сравнению различных орфограмм.

Точность ответа. На первом этапе мы включили в модель два фактора: написание и тип орфограммы (за базовый уровень были приняты правильное написание и орфограмма ГС). Как показывает Таблица 2, оба фактора оказались значимы, после чего для типа орфограмм были проведены попарные сравнения (Таблица 4). Затем мы добавили в модель относительную частотность неправильных написаний (Таблица 3). Этот фактор коррелирует с типом орфограммы, и нам было важно определить, есть ли у типа орфограммы независимая роль, не сводящаяся к этому фактору. При этом коэффициенты для разных значений фактора «тип орфограммы» увеличились, однако сам фактор относительной частотности неправильных написаний на значимость не вышел (Таблица 4). Это указывает на то, что тип орфограммы влияет на результаты больше, чем доля неправильных написаний.

	b	SE	z	p
Ложная двойная согласная (ЛДС)	3,90	0,53	7,35	<0,001
Двойная согласная (ДС)	1,97	0,56	3,53	<0,001
Гласная после шипящего (ГПШ)	2,29	0,54	4,21	<0,001
Непроверяемая безударная (НБГ)	3,00	0,55	5,51	<0,001
Проверяемая безударная (ПБГ)	-1,14	0,78	-1,46	0,144
Написание	1,42	0,16	8,80	<0,001

Таблица 2. Эксперимент 1: регрессионная модель по точности ответа

	b	SE	z	p
Ложная двойная согласная (ЛДС)	3,67	0,54	6,79	<0,001
Двойная согласная (ДС)	1,88	0,54	3,47	<0,001
Гласная после шипящего (ГПШ)	2,17	0,54	4,04	<0,001
Непроверяемая безударная (НБГ)	2,85	0,53	5,41	<0,001
Проверяемая безударная (ПБГ)	-1,12	0,76	-1,47	0,14
Доля неправильных вариантов	-0,0004	0,02	-0,02	0,98

Таблица 3. Эксперимент 1: регрессионная модель по точности ответа (с учетом фактора частотности неправильных вариантов)

	b	SE	z	p
ЛДС — ГС	3,67	0,54	6,79	<0,001
ДС — ГС	1,89	0,54	3,47	0,003
ГПШ — ГС	2,17	0,54	4,04	<0,001
ПБГ — ГС	-1,12	0,76	-1,47	0,284
НБГ — ГС	2,85	0,53	5,41	<0,001
ДС — ЛДС	-1,79	0,40	-4,43	<0,001
ГПШ — ЛДС	-1,5	0,38	-3,97	<0,001
ПБГ — ЛДС	-4,79	0,7	-6,8	<0,001
НБГ — ЛДС	-0,83	0,39	-2,13	0,133
ГПШ — ДС	0,29	0,42	0,69	0,49
ПБГ — ДС	-3,002	0,71	-4,25	<0,001
НБГ — ДС	0,96	0,42	2,31	0,104
ПБГ — ГПШ	-3,29	0,7	-4,69	<0,001
НБГ — ГПШ	0,67	0,4	1,69	0,274
НБГ — ПБГ	3,96	0,69	5,72	<0,001

Таблица 4. Эксперимент 1: попарные сравнения по точности ответа

Рассмотрим данные из Таблицы 4. Простые орфограммы ГС и ПБГ не отличаются друг от друга, но значимо отличаются от всех сложных. Среди сложных мы не видим значимых различий между ДС, ГПШ и НБГ, но они значимо отличаются от ЛДС — это самая сложная орфограмма.

Скорость ответа. Мы включили в модель два фактора: относительную частотность неправильных написаний и тип орфограммы. Как показывает Таблица 5, хотя фактор частотности неправильных написаний значим, когда анализируется отдельно, при анализе в сочетании с типом орфограммы только последний оказался значимым. Попарные сравнения разных типов орфограмм представлены в Таблице 6.

	b	SE	t	p
Ложная двойная согласная (ЛДС)	551,05	116,58	4,73	<0,001
Двойная согласная (ДС)	421,12	112,67	3,74	<0,001
Гласная после шипящего (ГПШ)	507,00	111,93	4,53	<0,001
Непроверяемая безударная (НБГ)	651,76	110,49	5,91	<0,001
Проверяемая безударная (ПБГ)	-113,46	102,25	-1,11	0,27
Доля неправильных ответов	2,43	6,02	0,42	0,69

Таблица 5. Эксперимент 1: регрессионная модель по скорости ответа

	b	SE	z	p
ЛДС — ГС	573,45	101,76	5,636	<0,001
ДС — ГС	434,37	106,95	4,061	<0,001
ГПШ — ГС	522,79	104,03	5,025	<0,001
ПБГ — ГС	-113,34	101,46	-1,112	1
НБГ — ГС	661,77	106,83	6,194	<0,001
ДС — ЛДС	-139,08	104,94	-1,325	1
ГПШ — ЛДС	-50,66	101,96	-0,497	1
ПБГ — ЛДС	-686,79	99,34	-6,913	<0,001
НБГ — ЛДС	88,32	104,82	0,843	1
ГПШ — ДС	88,42	107,15	0,825	1
ПБГ — ДС	-547,70	104,66	-5,233	<0,001
НБГ — ДС	227,40	109,87	2,070	0,27
ПБГ — ГПШ	-636,12	101,67	-6,256	<0,001
НБГ — ГПШ	138,98	107,03	1,299	1
НБГ — ПБГ	775,10	104,54	7,415	<0,001

Таблица 6. Эксперимент 1: попарные сравнения по скорости ответа

Как показывает Таблица 6, не все условия удалось сравнить попарно одновременно ($p=1$ указывает на то, что модель не сходится). Тем не менее, как и в случае с точностью ответов, мы видим значимые различия во всех попарных сравнениях простых и сложных орфограмм. Однако значимых различий между разными сложными орфограммами выявлено не было.

3 Эксперимент 2

3.1 Участники

Участники — 44 носителя русского языка (28 женщин) в возрасте от 19 до 52 лет (средний возраст — 29 лет). Все испытуемые приняли участие в эксперименте на добровольной основе.

3.2 Материалы

В качестве стимульного материала были взяты все слова из Эксперимента 1, только в правильном их написании. К каждому из этих слов было сгенерировано по псевдослову (например, *таридор* для слова *коридор*).

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 с использованием интернет-платформы Open Lab (Shevchenko 2022). Сначала для фиксации взгляда в центре экрана предъявлялся символ «+», время показа варьировалось в диапазоне от 1000 до 1595 мс, с шагом в 5 мс. После фиксации в центре экрана появлялась стимульная последовательность. Если испытуемый не принимал решения в течение 0,6 секунды, она сменялась белым экраном, который предъявлялся максимум 2 секунды, во время которых испытуемый все еще мог дать ответ. Целевые слова и псевдослова предъявлялись в случайном порядке.

3.4 Результаты и обсуждение

Мы анализировали время реакции испытуемых на целевые слова. При этом мы удалили те 106 случаев (4% данных), где испытуемые дали неверный ответ, а также ответы, превышающие для данной выборки пороговые значения: среднее время реакции плюс или минус 2,5 стандартных отклонения (Ratcliff 1993). Их было 5%. Средняя скорость ответов в зависимости от типа орфограммы приведена в Таблице 7. Статистический анализ проводился так же, как для Эксперимента 1.

	Средний % неправильных вариантов	Среднее время ответа (в мс)
Ложная двойная согласная (ЛДС)	8,1%	731,8
Двойная согласная (ДС)	7,0%	725,7
Гласная после шипящего (ГПШ)	6,5%	733,0
Непроверяемая безударная (НБГ)	3,6%	728,0
Проверяемая безударная (ПБГ)	<0,1%	688,0
Глухая согласная на конце слова (ГС)	<0,1%	680,0

Таблица 7. Результаты эксперимента 2

На первый взгляд, результаты Эксперимента 2 аналогичны результатам Эксперимента 1: в Таблице 7 очень заметна разница между простыми и сложными орфограммами. Однако нам предстоит проверить, является ли тип орфограммы независимым фактором, который не сводится к относительной частотности неправильных вариантов. В Эксперименте 1 мы получили на этот вопрос положительный ответ и даже показали, что тип орфограммы играет более важную роль.

Чтобы ответить на этот вопрос для Эксперимента 2, проанализируем сперва влияние различных факторов по отдельности. На время ответа значимо влияет общая частотность стимула ($b = -0,03$, $SE = 0,01$, $z = -2,85$, $p = 0,006$) и доля неправильных написаний ($b = 3,05$, $SE = 1,12$,

$z = 2,71, p = 0,009$), при этом надо заметить, что роль первого фактора больше, чем второго. В Эксперименте 1 роль общей частотности оказалась значимой только для точности ответов и была меньше, чем у доли неправильных написаний. Когда эти два отчасти взаимосвязанных фактора анализировались в рамках одной модели, оба достигли значимости, но взаимодействие между ними на значимость не вышло. Также значимое влияние на скорость ответов оказала длина стимула ($b = 8,25, SE = 3,70, z = 2,23, p = 0,030$).

Теперь перейдем к типу орфограммы. Если анализировать этот фактор изолированно, он оказывается значимым на уровне тенденции (Таблица 8). Как и в Эксперименте 1, за базовый уровень была принята орфограмма ГС. Однако, если в модель добавляется доля неправильных написаний, значимой оказывается только она (Таблица 9). Таким образом, мы наблюдаем картину, противоположную той, что была выявлена в Эксперименте 1: фактор частотности неправильных написаний оказывается более важным и влияние типа орфограммы, можно, по сути, свести к этому фактору.

	b	SE	z	p
Ложная двойная согласная (ЛДС)	48,12	24,38	1,97	0,054
Двойная согласная (ДС)	46,61	24,41	1,94	0,058
Гласная после шипящего (ГПШ)	54,95	24,31	2,26	0,028
Непроверяемая безударная (НБГ)	47,57	24,30	1,96	0,056
Проверяемая безударная (ПБГ)	7,34	24,25	0,30	0,763

Таблица 8. Эксперимент 2: регрессионная модель по скорости ответа.

	b	SE	z	p
Ложная двойная согласная (ЛДС)	32,05	26,46	1,21	0,231
Двойная согласная (ДС)	30,15	26,83	1,18	0,238
Гласная после шипящего (ГПШ)	41,99	25,60	1,64	0,107
Непроверяемая безударная (НБГ)	40,46	24,52	1,65	1,105
Проверяемая безударная (ПБГ)	7,28	23,98	0,30	0,763
Доля неправильных вариантов	2,41	1,12	2,16	0,035

Таблица 9. Эксперимент 2: регрессионная модель по скорости ответа (с учетом фактора частотности неправильных вариантов).

4 Заключение

В предыдущих работах на материале разных языков было показано, что относительная частотность неправильных написаний слова влияет и на то, насколько быстро и точно мы можем определить, правильно ли оно написано, и, что менее очевидно, на то, с какой скоростью мы его читаем, когда оно написано без ошибок, даже если задание никак не фокусирует нашего внимания на проблеме орфографической правильности (Rahmanian, Kuperman, 2019; Kuperman et al. 2021; Чернова и др., 2020). В данной работе мы задались вопросом, влияет ли на эти процессы такой фактор, как тип орфограммы. Русский — очень удачный язык для таких исследований, так как можно подобрать разнообразные орфограммы и разделить их на более простые и более сложные.

Мы провели два эксперимента, которые позволяют прийти к следующим выводам. Когда мы осознанно обращаемся к проблеме орфографической правильности, определяя, правильно ли написано слово, фактор типа орфограммы оказывается ключевым. Он играет большую роль, чем относительная частотность неправильных написаний. Когда мы просто читаем слова, не заостряя свое внимание на проблеме орфографической правильности, фактора относительной частотности неправильных написаний оказывается достаточно, чтобы объяснить наблюдаемую вариацию в скорости чтения (наряду с такими хорошо изученными в этой области факторами, как общая частотность слова и его длина).

Конечно, взрослые носители русского языка делают меньше ошибок в словах с простыми орфограммами, поэтому в каком-то смысле тип орфограммы всё равно влияет на скорость чтения. Однако для нас важно то, что это влияние имеет сугубо косвенный характер — через посредство другого фактора, доли неправильных написаний. Исходя из этого мы предполагаем, что только доля неправильных написаний, с которой сталкивается читающий, непосредственно влияет на качество лексической репрезентации слов. Фактор типа орфограммы, вероятно, действует на уровне орфографической системы языка в целом. Мы были бы склонны предположить, что его влияние будет принципиально зависеть от выполняемой задачи и от индивидуальных характеристик читающего (например, от его читательского опыта и орфографической зоркости). Мы планируем проверить эти гипотезы в будущих исследованиях.

Acknowledgements

Поддержано грантом РФФ 21-18-00429

References

- [1] Andrews S., Hersch J. Lexical precision in skilled readers: Individual differences in masked neighbor priming // *Journal of Experimental Psychology: General*. — 2010. — Т. 139. — №. 2. — С. 299.
- [2] Baayen R. H. et al. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning // *Psychological review*. — 2011. — Т. 118. — №. 3. — С. 438.
- [3] Bates D., Maechler M., Bolker B., Walker S. lme4: Linear mixed-effects models using Eigen and S4 // R package version 1. — 2015. — P. 1–8.
- [4] Belikov V. et al. Corpus as language: from scalability to register variation // *Komp'juternaja lingvistika i intellektual'nye tehnologii*. — 2013. — №. 12. — С. 84-95.
- [5] Drummond A., von der Malsburg T., Erlewine M.Y., Vafaie M. Ibex farm. // GitHub. — 2016. — URL: <https://github.com/addrummond/ibex>
- [6] Hersch J., Andrews S. Lexical quality and reading skill: Bottom-up and top-down contributions to sentence processing // *Scientific Studies of Reading*. — 2012. — Т. 16. — №. 3. — С. 240-262.
- [7] Kuznetsova A., Brockhoff P.B., Christensen R.H.B. lmerTest Package: Tests in Linear Mixed Effects Models // *Journal of Statistical Software*. — 2017. — №. 82. — P. 1–26.
- [8] Perfetti C. A. Reading ability. — Oxford university Press, 1985.
- [9] Perfetti C. A. The psycholinguistics of spelling and reading. — 1997.
- [10] Perfetti C. A., Hart L. The lexical basis of comprehension skill. — 2001.
- [11] Perfetti C. A., Hart L. The lexical quality hypothesis // *Precursors of functional literacy*. — 2002. — Т. 11. — С. 67-86.
- [12] Perfetti C. Reading ability: Lexical quality to comprehension // *Scientific studies of reading*. — 2007. — Т. 11. — №. 4. — С. 357-383.
- [13] Piperski A. et al. Big and diverse is beautiful: A large corpus of Russian to study linguistic variation // *Proc. 8th Web as Corpus Workshop (WAC-8)*. — 2013. — С. 24-29.
- [14] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. — 2020. — Vienna, Austria. URL <https://www.R-project.org/>.
- [15] Rahmanian S., Kuperman V. Spelling errors impede recognition of correctly spelled word forms // *Scientific Studies of Reading*. — 2019. — Т. 23. — №. 1. — С. 24-36.
- [16] Ramscar M. et al. The effects of feature-label-order and their implications for symbolic learning // *Cognitive science*. — 2010. — Т. 34. — №. 6. — С. 909-957.
- [17] Ramscar M., Dye M., McCauley S. M. Error and expectation in language learning: The curious absence of "mouses" in adult speech // *Language*. — 2013. — С. 760-793.

- [18] Ratcliff R. Methods for dealing with reaction time outliers // *Psychological Bulletin*. — 1993. — Vol. 114. — Pp. 510–532.
- [19] Rescorla R. A., Wagner, A. R. et al. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement // *Current research and theory*. — 1972. — С. 64-99.
- [20] Shahar-Yames D., Share D. L. Spelling as a self-teaching mechanism in orthographic learning // *Journal of research in reading*. — 2008. — Т. 31. — №. 1. — С. 22-39.
- [21] Shaoul C., Westbury C. A reduced redundancy USENET corpus (2005–2011) // *University of Alberta* — 2013 — Т. 39. — №. 4. — С. 850–863.
- [22] Shevchenko Yu. Open Lab: A web application for running and sharing online experiments. // *Behavior Research Methods*. — 2022. — 10.3758/s13428-021-01776-2.
- [23] Taylor J. N., Perfetti C. A. Eye movements reveal readers' lexical quality and reading experience // *Reading and Writing*. — 2016. — Т. 29. — №. 6. — С. 1069-1103.
- [24] Чернова Д. А., Алексеева С. В., Слюсарь Н. А. Чему нас учат ошибки: трудности при обработке слов с частотными орфографическими ошибками // *Комп'ютернаја Lingvistika i Intellektual'nye Tehnologii*. — 2020. — №. 19. — С. 147-159.