

Санкт-Петербургский государственный университет

Максимова Полина Дмитриевна

Выпускная квалификационная работа

**Орфографическая репрезентация слов в ментальном лексиконе: данные
русского языка**

Уровень образования: магистратура

Направление 45.04.02 «Лингвистика»

Основная образовательная программа ВМ.5624 «Теория и история языка и
языки народов Европы»

Профиль «Психолингвистика»

Научный руководитель:

Старший научный сотрудник
Института когнитивных исследований,
кандидат филологических наук,
Чернова Дарья Алексеевна

Рецензент:

Доцент, кандидат филологических наук,
Краснощекова Софья Викторовна

Санкт-Петербург
2022

Оглавление

1.	Введение.....	3
2.	Глава 1. Теоретические основы лексического качества и его связь с читательским опытом.....	7
1.1	Лексические репрезентации и гипотеза лексического качества.....	7
1.2	Орфографические репрезентации, правописание и чтение.....	19
1.3	Читательский опыт и способы его оценки.....	28
1.4	Особенности русской орфографии.....	36
3.	Глава 2. Анализ орфографических репрезентаций в ментальном лексиконе.....	42
2.1	Постановка задачи.....	42
2.2	Материал исследования.....	42
2.3	Проведение эксперимента.....	46
2.4	Обработка результатов.....	50
2.5	Обсуждение результатов.....	52
4.	Заключение.....	55
5.	Список литературы.....	58
6.	Приложение 1.....	66
7.	Приложение 2.....	68

Введение

В настоящее время значительную часть психолингвистических исследований занимают исследования организации ментального лексикона, посвященные тому, как в долговременной памяти носителя языка хранятся слова, их звуковые, графические, семантические репрезентации.

В современных исследованиях ментальный лексикон, то есть внутренний словарь человека, понимается как сложноорганизованная сеть [Белов, 2020]. Существует множество моделей, как ментальный лексикон может быть организован, например, на основе принципов: фонетических, семантических и/или ассоциативных [Aitchison, 2012]. В области изучения ментального лексикона большинство работ посвящено восприятию речи, так как этот процесс проще отследить и затем проанализировать, нежели порождение речи, где мы можем видеть только конечный результат, а процесс – скрыт от глаз наблюдателя. Основной метод, который используется для изучения ментального лексикона – задание на принятие лексического решения, где участник эксперимента видит перед собой комбинацию букв и должен ответить – слово перед ним или нет. Или это могут быть слова с ошибками и без, как в нашем исследовании.

В данной работе мы хотим исследовать устойчивость графического облика слов в ментальном лексиконе носителей русского языка.

В недавнем исследовании на материале английского языка и иврита (Rahmanian, Kuperman, 2018) было показано, что с увеличением количества не редактируемых текстов (различные социальные сети, электронные СМИ и др.) растет частотность неверных написаний слов, которые могут зафиксироваться в ментальном лексиконе читающего и тем самым ослабить взаимосвязи между формальными и семантическими репрезентациями и вызвать трудности с последующим извлечением информации о слове из памяти.

Вслед за ним, в 2021 было опубликовано кросс-лингвистическое исследование Купермана и коллег [Kuperman, Bar-On, Bertram, Boshra, Deutsch, Kyröläinen, Mathiopoulou, Oralova, Protopapas, 2021] в котором представлены новые данные по английскому языку и ивриту и данные по еще 3 языкам: китайскому, финскому и греческому. Там они также говорят о том, что большое количество различных вариантов слова, которые существуют (правильные и неправильные) ведет к более медленному распознаванию данного слова, даже если оно было написано верно. Это было верно для всех языков, несмотря на то, что они все разные по своей орфографической и лексической системам.

Актуальность нашего исследования состоит в том, что оно позволит получить новые данные о том, как происходит формирование и изменение структуры графического компонента ментального лексикона на материале языка с другим типом орфографии.

Учитывая то, что именно люди порождают слова (в их правильных и неправильных вариациях), стоит обратить внимание на индивидуальные различия между людьми, тем откуда они берутся и какой вклад они вносят в то, каким образом мы воспринимаем данные слова. Поэтому в своем исследовании мы хотим посмотреть, существует ли зависимость грамотности от начитанности и как они влияют друг на друга в контексте индивидуума.

Объектом исследования является восприятие слов, содержащих и не содержащих орфографические ошибки.

Предметом исследования является орфографическая репрезентация слов в ментальном лексиконе носителя русского языка.

Целью исследования является выявление механизмов восприятия слов, имеющих конкурирующие варианты написания.

В связи с поставленной целью, можно вывести следующие задачи исследования:

1. проанализировать существующую научную литературу по данному вопросу и обозначить основные теоретические концепции для исследования;
2. подобрать материал для исследования на материале русского языка, используя корпус не редактируемых текстов;
3. провести экспериментов, направленный на оценку устойчивости графических репрезентаций слов в ментальном лексиконе носителей русского языка с различным читательским опытом
4. обработать результаты эксперимента и сделать обобщающие выводы.

Научная новизна нашего исследования состоит в том, что ранее подобные исследования проводились в основном на материале английского языка, тогда как мы хотим посмотреть на те же процессы уже на материале русского языка, который значительно отличается по типу орфографии от английского языка.

Наше исследование проводилось на основе данных из Генерального Интернет-корпуса Русского Языка (ГИКРЯ). Корпус включает в себя текстовые материалы из блогосферы, социальных сетей, крупнейших новостных ресурсов и литературных журналов. Данный корпус содержит большое количество не отредактированных материалов и является одним из немногих мега-корпусов, достигающих объемом нескольких миллиардов слов.

Результаты данного исследования могут быть использованы при подготовке методических материалов для преподавания русского языка как родного и как иностранного, что обуславливает практическую значимость работы. Методика исследования была апробирована во время сессии по лингвистике в Государственном бюджетном учреждении дополнительного образования «Ленинградский областной центр развития творчества одаренных детей и юношества “Интеллект”». Данная работа послужила основой для одного из

проектов, в котором школьники проверяли, зависит ли грамотность их сверстников от читательского опыта.

Структура работы: работа состоит из введения, двух глав, заключения, списка литературы и трех приложений.

Глава 1. Теоретические основы лексического качества и его связь с читательским опытом.

1.1 Лексические репрезентации и гипотеза лексического качества.

Лексические репрезентации – это образы слова в ментальном лексиконе носителя языка. Новые слова усваиваются и добавляются в ментальный лексикон, когда мы выучиваем их через устную или письменную речь. Когда мы слышим, видим и используем эти слова много раз и в разных контекстах, их значения конкретизируются, а их позиция в ментальном лексиконе укрепляется. Сила и стабильность репрезентаций в ментальном лексиконе определяет так называемое лексическое качество слова.

Гипотеза лексического качества впервые была подробно описана в работах Перфетти [Perfetti 1985; 2007]. Она говорит о том, что для правильного распознавания слова в нашем ментальном лексиконе должна храниться качественная лексическая репрезентация, то есть облик слова должен быть закреплён фонологически, орфографически и семантически. Три составляющие лексической репрезентации подробно описаны в статьях Перфетти и Харта [Perfetti, Hart 2001; 2008]. Когда мы говорим о фонологии, мы имеем в виду произношение слова. Когда мы говорим об орфографии, мы имеем в виду правописание конкретных слов. Когда мы говорим о семантике, мы имеем в виду свойства значения слова, которые помещают слово на его место в лексиконе среди других слов.

Таким образом, у нас получается комплекс из трех составляющих.

Возьмем, к примеру, слово *gate* (см. рис.1). У нас есть орфографическая составляющая – то, как это слово правильно пишется (*gate*). Фонологическая составляющая – то, как это слово правильно произносится (*[geɪt]*). Семантическая составляющая – то, какое точное значение имеет это слово. Таким образом, мы видим перед собой (рис.1) репрезентацию высокого качества, где составляющие слова связаны между собой тесными связями.

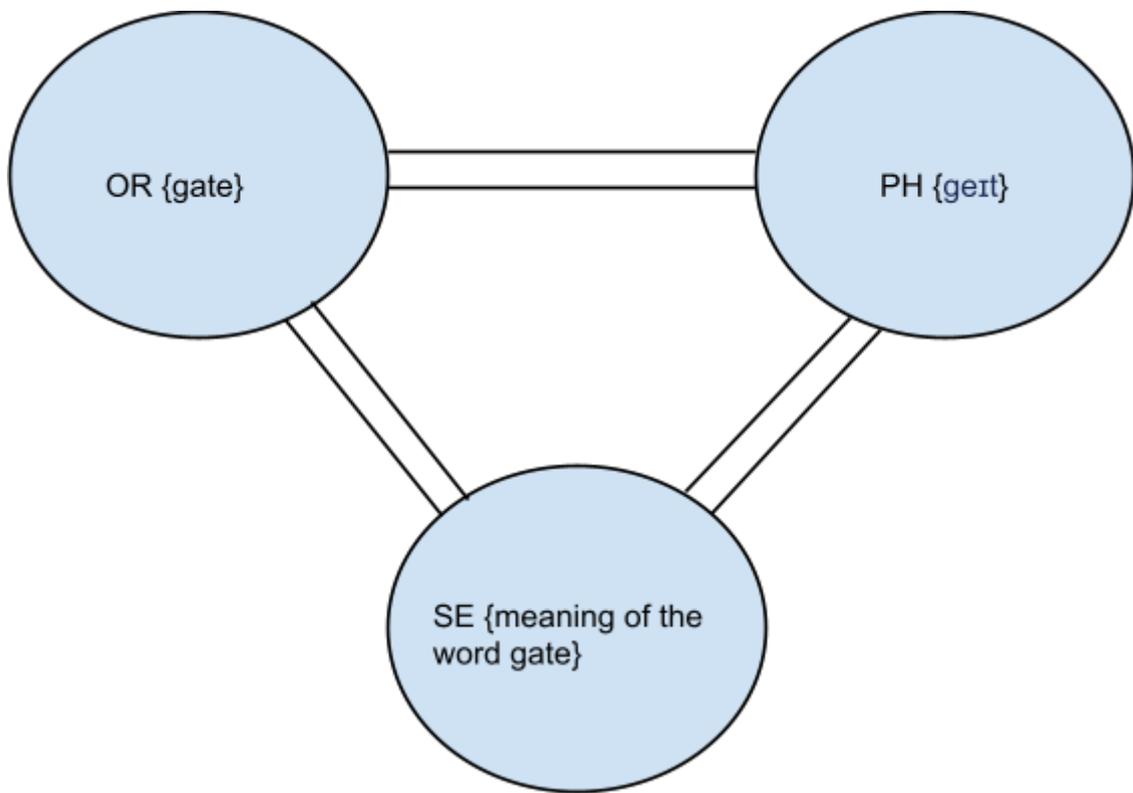


рис.1 Лексическая репрезентация слова «gate» [Perfetti, Hart 2001; 2008]

Когда орфографическая, фонологическая и семантическая информация о слове активируется одновременно, связи между ними укрепляются и формируется единая лексическая репрезентация. Чем сильнее связь, тем больше вероятность, что именно эта комбинация лексических составляющих активирует одну и только одну конкретную триаду. Таким образом, четкая и связанная лексическая репрезентация позволяет читающему извлечь из ментального лексикона именно то слово, которое он перед собой видит, а не фрагменты, которые могут ассоциироваться и с другими словами.

Качественные лексические репрезентации быстро реактивируются в ментальном лексиконе, когда мы сталкиваемся с этим словом на письме или в устной речи. Это делает лексическую информацию доступной для понимания.

Про раздельное формирование графического и звукового облика также пишет А. А. Залевская: «ментальная репрезентация звучания слова отделена от его написания, и между ними существуют более или менее сложные взаимоотношения: в долговременной памяти могут иметься отдельные репрезентации для графических форм слов, противопоставляемых фонологическим формам тех же слов. Первыми усваиваются фонологические формы слов, по меньшей мере для той части лексикона, которая усваивается до овладения письмом. Отсюда следует, что начальные шаги в овладении чтением и письмом могут опосредоваться уже имеющимися фонологическими формами слов, но далее обособляются самостоятельные, хотя и связанные с исходными, письменные формы слов (организованные как структуры букв)». [Залевская, 2005]

Лексические репрезентации высокого качества должны быть конкретными и избыточными. То есть, такая лексическая репрезентация имеет точно определенную орфографическую репрезентацию (правильное написание слова), а так же избыточные фонологические репрезентации (одну, полученную из устной речи, и одну, полученную при прочтении этого же слова, то есть, когда мы соотносим звуки в речи и буквы на письме). Такая лексическая репрезентация будет связной и надежной. Под связной здесь подразумевается, что все составляющие (орфографическая, фонологическая и семантическая) синхронизированы и доступны для извлечения в процессе порождения как единое целое. Под надежной подразумевается то, что когда мы много раз встречаем одно и то же слово, у нас активизируется одна и та же лексическая репрезентация с одним и тем же набором составляющих.

В свою очередь, ментальные репрезентации низкого качества подразумевают под собой, то, что какая-то составляющая слова была недостаточно хорошо закреплена или же связи между этими составляющими слишком слабые. Соответственно, нестабильность графического или звукового облика, или же

значения слова, препятствует развитию связной и надежной лексической репрезентации слова в ментальном лексиконе человека.

Есть такие слова, про которые можно сказать, что они изначально имеют лексическую репрезентацию низкого качества. Это омофоны, омографы и неоднозначные слова.

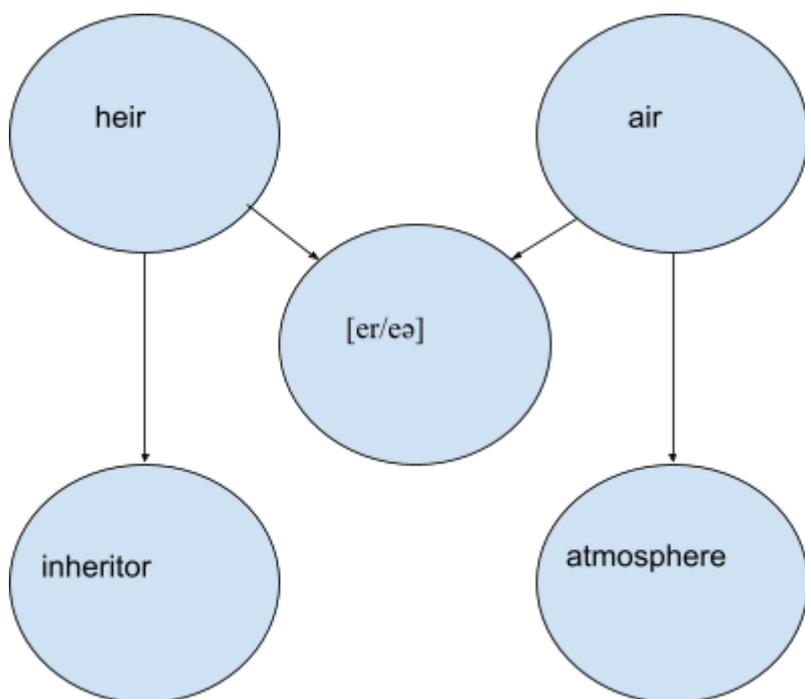


рис. 2 Лексические репрезентации слов «heir» и «air» [Perfetti, Hart 2001; 2008].

У омофонов одна фонологическая репрезентация и две орфографические репрезентации, каждая из которых имеет свою отдельную семантическую репрезентацию, например английские слова *heir* и *air* (см. рис.2) имеют одну и ту же фонологическую репрезентацию [er/eə], но первая орфографическая репрезентация означает «наследник», а вторая – «воздух». Проблема омофонов заключается в том, что в конкуренцию вступают две орфографические и две семантические репрезентации соответственно.

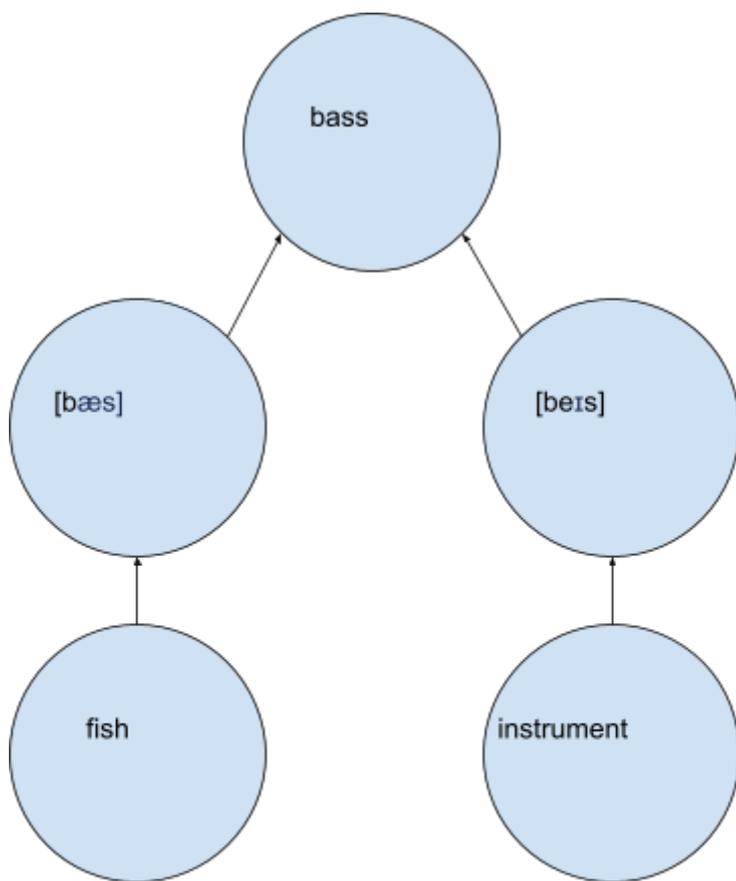


рис.3 Лексическая репрезентация слова «bass» [Perfetti, Hart 2001; 2008]

У омографов одна орфографическая репрезентация, но две фонологические репрезентации, каждая из которых имеет свою отдельную семантическую репрезентацию, например, если английское слово bass (см. рис.3) произнести с коротким звуком [a], то оно будет обозначать рыбу (окунь), а если с длинным – музыкальный инструмент (бас). То есть, при виде одной орфографической репрезентации, у нас не активируется одна конкретная триада, а две, из которых нам нужно выбирать. Контекст, в котором было использовано слово, необходим для того, чтобы разрешить неоднозначность.

У неоднозначных слов: одна орфографическая и одна фонологическая репрезентации соотносится с двумя или более семантическими репрезентациями, например, английское слово bank (см. рис.4) – многозначное; два его основных значения: банк и берег реки. Это тоже не конкретная лексическая репрезентация, так как когда у нас активируются

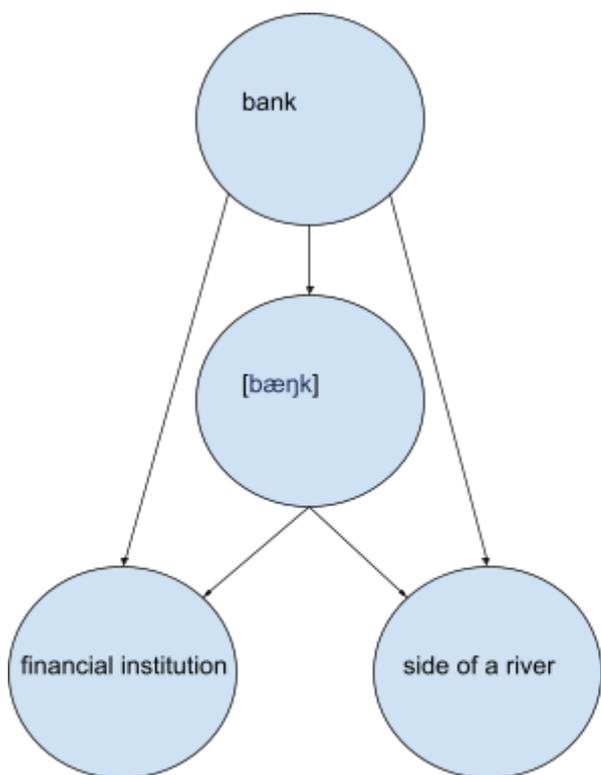


рис.4 Лексическая репрезентация слова «bank» [Perfetti, Hart 2001; 2008]

одна фонологическая и одна орфографическая составляющие, у нас активируются две лексические единицы с двумя разными значениями. Контекст, в котором появляется это слово, опять играет большую роль для того, чтобы снять неоднозначность. Также, частотность использования слова в каждом из значений, поможет выбрать нужную среди нескольких лексических единиц [Collins, 2002].

А теперь обратимся к примеру плохого лексического качества, связанного именно с недостаточной языковой компетенцией носителя и плохого закрепления слова в ментальном лексиконе человека. Рассмотрим следующие случаи:

1. Человек сталкивается со словом «грейпфрут», он может правильно его произнести и знает, что он обозначает какой-то фрукт, но, допустим, никогда с ним не сталкивался и точно не знает, что это и как оно выглядит.

2. Человек сталкивается со словом «грейпфрут» и испытывает проблемы с его произношением, порождая, например, “грейфрукт”.
3. Человек сталкивается со словом «грейпфрут», он может его произнести и знает, что это цитрус с красной мякотью, похожий на апельсин, но крупнее. Когда он использует это слово в речи и рассказывает кому-то про то, как он купил этот фрукт в магазине, он произносит его как “грейпфрут”, а иногда как “грейпфрукт”.
4. Человек сталкивается со словом «грейпфрут» и может правильно произнести это слово, знает его значение и может верно употребить его в предложении, но пишет его правильно только в некоторых случаях.

Таким образом, мы видимо перед собой несколько случаев слабой лексической репрезентации слова грейпфрут [grɛɪp'frut] – плод субтропического вечнозеленого дерева рода цитрус. Такие лексические репрезентации будут иметь разные последствия для человека и предоставлять разные трудности в процессе порождения речи и в процессе понимания. Например, случай номер 4 знаком многим грамотным людям: есть семантическая и фонологическая компетенция в сочетании с орфографическими трудностями. Первый случай может представлять начитанного человека с не очень большим опытом, который может декодировать слово, которое он не знает, но слышал его в речи и таким образом имеет фонологическую репрезентацию.

На этом примере предлагается следующая довольно общая концепция, которая применима как к опытным читателям, так и к менее опытным читателям. У опытных читателей в ментальном лексиконе могут быть лексические репрезентации низкого качества для многих слов, как низкочастотных слов из общего словаря, так и слов из специализированных словарей. То, что опытные читатели имеют в качестве преимущества, – это основополагающие ресурсы, которые позволяют им извлекать максимум информации из некачественных лексических репрезентаций. Опытный

читатель имеет больше шансов добавить новую информацию (об орфографии, произношении или значении слова) к лексической репрезентации низкого качества. Однако наш анализ больше фокусируется на словах, чем на отдельных людях. Не отдельные люди, а лексические репрезентации различаются по своему качеству. Опытный читатель, в дополнение к базовым ресурсам (навыкам расшифровки, правописания и грамматики), – это тот, кто обладает множеством качественных лексических репрезентаций слов в своем ментальном лексиконе. Менее квалифицированный читатель – это тот, у кого меньше лексических репрезентаций слов высокого качества.

Происхождение большого количества качественных лексических репрезентаций слова – это вопрос опыта, который порождает соответствующие знания. Конечно, базовые фонологические знания способствуют раннему прогрессу в чтении, а последующая практика чтения, аудирования, говорения и письма способствует формированию более качественных лексических репрезентаций слов. Серьезные различия в том, сколько читает тот или иной индивидуум, достаточны для того, чтобы вызвать степень различий в качестве лексических репрезентаций, которые можно наблюдать в любых возрастных группах людей. Такой опыт влияет на качество слова и количество слов с лексическими репрезентациями высокого качества. По сути, существует континуум лексических знаний и именно этот континуум разделяется пополам, когда исследователи обращаются к опытным и менее опытным читателям.

В своей работе, посвященной орфографическим ошибкам в английском языке, Рахманьян и Куперман [Rahmanian, Kuperman, 2017] выдвигают следующую гипотезу – орфографические ошибки возникают не только как следствие того, что орфографическая репрезентация слова в ментальном лексиконе слабая, так как слово было плохо выучено и его орфографический облик не закрепился. Они считают, что само существование

орфографических ошибок в свою очередь становится причиной возникновения различных сложностей при освоении орфографии и распознавании слов. То есть, получается, что не очень грамотные люди пишут в соответствии со своими лексическими репрезентациями низкого качества некие тексты. Другие люди, возможно даже более грамотные, читают данные тексты (которые не были подвергнуты никакой редакции – различные посты в социальных сетях, независимые электронные СМИ, самиздат и так далее). Соответственно, в их ментальные лексикон попадают и неправильные варианты написания из этих не подверженных редактуре источников (правильные варианты к ним попадают, допустим, из книг и газет, которые были проверены редакторами). И таким образом, мы получаем несколько вариантов написания слова в ментальном лексиконе человека – правильные и неправильные. В результате у слов появляются различные псевдоомофоны, из-за чего даже уже хорошо сформированные репрезентации начинают размываться. Вариантов написания слова в ментальном лексиконе человека становится все больше и больше и становится все сложнее и сложнее распознавать, и порождать слова в письменной форме.

Куперман и Бар-Он [Bar-On, Kuperman, 2019] в своей статье пишут о таком явлении как *spelling entropy* “орфографическая энтропия”, которая количественно определяет степень неопределенности при выборе одного из доступных вариантов с учетом их распределения вероятностей. Иными словами, степень конкуренции между вариантами написания слова.

Исследования речепорождения [Ernestus, 2014] показывают, что альтернативные формы слова (в нашем случае, слова с орфографическими ошибками) создают свои собственные лексические репрезентации, которые хранятся в ментальном лексиконе одновременно с правильными вариантами и конкурируют друг с другом в процессе порождения речи. То есть, каждый раз, когда мы встречаем на письме неверно написанный вариант, он все прочнее укрепляется в нашем ментальном лексиконе и велика вероятность,

что в следующий раз, при написании, мы выберем именно вариант с ошибкой, который встречался нам чаще.

Согласно этому, Рахманьян и Куперман [Rahmanian, Kuperman, 2017] сделали предположение, что люди будут испытывать наибольшие трудности при усвоении и распознавании слов, когда неопределенность в выборе между альтернативными вариантами написания является наиболее сильной, например, когда существует много альтернативных вариантов написания и/или когда их относительная частота встречаемости более равномерна, например, такие слова как «ноль» и «нуль» (оба варианта написания считаются правильными) и такие варианты как «апелляция», «апелляция» и «апелляция» (первый вариант верный). Ожидается, что слово, которое пишется единообразно всегда (то есть, имеет один вариант написания), не приведет к каким-либо усилиям со стороны человека при распознавании или во время усвоения слова.

Возвращаясь снова к показателю энтропии. Показатель энтропии количественно определяет неоднозначность преобразования графем в фонемы и наоборот [Borgwaldt, Hellwig, De Groot, 2005], [Protopapas, Vlahou, 2009].

Общая идея такого подхода заключается в следующем: если данная графема (или фонема) полностью соответствует одной конкретной фонеме (или графеме), то преобразование одного в другое абсолютно предсказуемо и показатель энтропии равен нулю. Чем больше альтернативных вариантов произношения (или написания) имеет графема (или фонема), тем менее предсказуемым становится преобразование одного в другое и тем выше будет значение энтропии.

Выражение (не) однозначности в этих преобразованиях в виде значений энтропии, следовательно, приведет к непрерывным переменным, начинающимся с нуля для полностью однозначных преобразований и будут постепенно увеличиваться по мере увеличения степени неопределенности. В

дополнение к количеству различных произношений (или вариантов написания), относительная частота этих альтернативных вариантов вносит свой вклад в величину энтропии. Если существует одно доминирующее соответствие графема-фонема, и некоторые из этих альтернативных вариантов встречаются очень редко, значение энтропии будет ниже, чем когда все произношения происходят примерно с одинаковой частотой, что приводит к довольно минимальному влиянию исключительных альтернативных вариантов.

Энтропия H по распределению вероятностей определяется следующим образом:

$$H = - \sum_i^n p_i \log p_i,$$

где p - относительная частота (вероятность) каждого варианта написания i в суммарной частоте всех n вариантов слова.

В своем исследовании 2005 года Боргвальдт и коллеги [Borgwaldt et al., 2005] исследовали относительный вклад гласных и согласных в общую орфографическую прозрачность и проанализировали влияние значений энтропии как предикторов времени реакции в заданиях на называние.

Ни в одной из изученных орфографий (исследование проводилось на материале английского, французского, португальского, немецкого, голландского, итальянского и венгерского языков) не было обнаружено полностью однозначных соответствий между буквами и звуками. С точки зрения общего соответствия орфографии фонетике, проанализированного на уровне буква-фонема, английский язык имел самую неоднозначную орфографию, за ним следовали французский, немецкий, португальский, голландский, итальянский и венгерский. Когда согласные и гласные буквы анализировались отдельно, картина немного изменилась: с учетом гласных английский оставался языком с наиболее неоднозначными соответствиями букв и фонем, за ним следовали немецкий, голландский, французский, португальский, итальянский и венгерский, причем последний язык демонстрирует совершенно однозначные сопоставления гласных

букв/гласных фонем. Рассматривая согласные, наиболее неоднозначные соответствия букв и звуков были обнаружены во французском языке, за которым следуют английский, немецкий, венгерский, итальянский, голландский и португальский. Авторы утверждают, что вычисления начальной энтропии не только информируют нас об общей орфографической прозрачности отдельного языка, но также позволяют нам ранжировать отдельные слова в соответствии со степенью неоднозначности их написания. Эта переменная значительно коррелирует с задержками в скорости присвоения имен. Как показал их анализ, семь языков демонстрировали разные характеристики с точки зрения неоднозначности согласных и гласных, что, в свою очередь, может объяснить специфичное для языка поведение фонологического кодирования в процессе чтения. Авторы заостряют внимание на том, что двусмысленность сопоставлений букв и фонем не следует игнорировать в пользу исключительного внимания к более крупным по размерам единицам, таким как морфемы.

В более поздней работе, Куперман и коллеги [Kuperman, Bar-On, Bertram, Boshra, Deutsch, Kyrgöläinen, Mathiopoulou, Oralova, Protopapas, 2021] говорят об обучении, основанном на ошибках, и снова возвращаются к своей гипотезе. Они говорят о том, что в психолингвистике орфографические ошибки долгое время считались значимым свидетельством того, по каким траекториям происходит усвоение грамоты у нормальных детей и у детей с расстройствами (дислексия, дисграфия и т.д.). Фактически, орфографические ошибки рассматриваются как несовершенные орфографические репрезентации или слабые ассоциативные связи между орфографическими и фонологическими формами слов или частей слов. То есть, орфографические ошибки это такой побочный продукт продолжающегося развития грамотности человека.

В связи с этим они говорят о том, что связь между альтернативными вариантами слов и качеством лексической репрезентации является взаимной,

а не однонаправленной. То есть, орфографические ошибки могут как отражать, так и порождать не успешное усвоение грамоты.

1.2 Орфографические репрезентации, правописание и чтение.

Как мы уже говорили в предыдущем разделе, когда слово хранится в ментальном лексиконе, мы можем получить доступ к его орфографической, фонологической и семантической составляющим. Однако, мы также говорили о том, что звуки в произносимом слове не всегда указывают на то, какие буквы содержатся в его написании, и часто лексический облик слова не имеет прямого сходства с его семантическим значением. Таким образом, изучение слов – это процесс создания и совершенствования этих отдельных компонентов словесной идентичности, а также связей между ними.

Согласно гипотезе лексического качества, о которой мы подробно говорим в разделе 1.1, выдвигается гипотеза, что фонология, орфография и семантика являются необходимыми факторами, влияющими на качество лексики. Соответственно предполагается, что недостаток любого из этих компонентов затруднит лексический доступ к слову и приведет к сложностям при чтении слов. Например, доступ к словам с лексическими репрезентациями более низкого качества будет менее эффективен и, следовательно, читать такие слова мы будем медленнее, чем слова с более качественными репрезентациями. Одним из способов измерения орфографического качества является проверка правописания; постоянство правильных написаний будет отражать наличие качественных орфографических репрезентаций. Тем не менее, влияет ли качество орфографических репрезентаций непосредственно на скорость чтения остается неясным. Именно этот вопрос послужил мотивом для исследования Сандры Мартин-Чанг, Мэлани Мэддэн и Джина Уэллетта [Martin-Chang, Ouellette, Madden, 2014].

Целью их исследования было изучение взаимосвязи между качеством (точностью и стабильностью) орфографических репрезентаций в ментальном лексиконе, индексируемых точностью правописания при повторных испытаниях, и непосредственно измеренной скоростью чтения слов. Гипотеза лексического качества утверждает, что порождаемые слова отражают качество орфографических репрезентаций, и доступ к словам с репрезентациями высокого качества должен быть более эффективным, чем к словам с более низким качеством. Эта гипотеза непосредственно проверяется с использованием орфографической точности и стабильности в качестве независимых переменных и скоростью чтения в качестве зависимой переменной. Учитывая, что правильное написание слова в различных ситуациях рассматривается в качестве главного показателя лексического качества, они определили "идеализированное орфографическое представление" как орфографическую репрезентацию, которая была точной и стабильной в течение пяти проб. Предполагалось, что слова, которые всегда были написаны правильно (т.е. во всех пяти пробах), имеют самую качественную орфографическую репрезентацию в наборе слов. Слова, которые никогда не были написаны правильно, следовательно, имели самую низкую по качеству орфографическую репрезентацию. Беглость чтения одного слова была реализована как точное и быстрое распознавание слова.

В их исследовании оценивалась связь между основными орфографическими репрезентациями, отраженными в точности и стабильности правописания, и скоростью чтения. Согласно предпосылке гипотезы лексического качества, неточные или нестабильные лексические репрезентации должны приводить к более медленной активации слова в ментальном лексиконе и, следовательно, к более медленному прочтению слов. Данные в исследовании Сандры Мартин-Чанг и коллег в значительной степени подтвердили эту гипотезу: слова, которые не могли быть написаны точно, читались медленнее. Слова с более устойчивыми репрезентациями прочитывались быстрее, чем слова с менее устойчивыми репрезентациями.

Способность участников читать слова, которые они не могли правильно написать, что произошло со всеми словами из категории "никогда", то есть словами, которые в течение эксперимента никогда не были записаны правильно участниками, подтверждает идею о том, что чтение слов с частичными или ошибочными орфографическими репрезентациями, то есть, орфографическими репрезентациями низкого качества, все еще может происходить. Как предположили Берт и Фьюри [Burt, Fury, 2000], возможно, что орфографические репрезентации при чтении должны быть достаточно точными только для того, чтобы отличить один элемент от другого. Однако для точного написания необходимо точная и четкая орфографическая репрезентация более высокого качества. Данное исследование показывает, что, хотя репрезентаций низкого качества может быть достаточно для чтения, они не способствуют скорости чтения в той же степени, что и полностью стабильные, точные репрезентации высокого качества.

Исследования показали, что люди, которые хорошо владеют орфографией, также, как правило, быстрее читают [Holmes, Castles, 2001] и более способны к так называемой "bottom-up processing", то есть способности к подбору слов и звуков исходя из услышанных звуков и своего словарного запаса [Andrews, Bond, 2009] [Hersch, Andrews, 2012]. Результаты исследования Сандры Мартин-Чанг и коллег [Martin-Chang, Ouellette, Madden, 2014] также показывают корреляцию между способностью к правописанию и скоростью чтения, предполагая, что у тех, кто лучше владеет орфографией, активация слов при чтении происходит быстрее.

Эти результаты подтверждают и дополняют предыдущие работы, показывающие умеренную или высокую положительную корреляцию между способностью читать и грамотным письмом (например, [Burt, Fury, 2000]).

Хотя исследования, перечисленные выше, согласуются с гипотезой лексического качества, они не исключают возможность того, что грамотно пишущий человек может отличаться от неграмотно пишущего человека по

другим параметрам, например, более быстрое автоматическое запоминание, более высокий IQ и так далее. Однако общепризнано, что лексическое качество варьируется не только ‘у разных людей для данного слова’, но и ‘у разных слов для данного человека’. Этот второй аспект гипотезы лексического качества предполагает, что даже в целом грамотно пишущие люди могут иметь в ментальном лексиконе слова, которые имеют более низкие по качеству орфографические репрезентации, а в целом неграмотно пишущие люди могут иметь в ментальном лексиконе некоторые слова, которые имеют более высокие по качеству орфографические репрезентации.

В данном исследовании участники не определялись по общим способностям, а скорее сами слова определялись тем, насколько точно они были написаны каждым из участников. Использование одних и тех же слов как в заданиях по правописанию, так и в заданиях по чтению дает представление о том, как лексическое качество каждого слова соотносится со скоростью чтения отдельных слов. Действительно, анализ для каждого из участников показал, что связь между лексическим качеством, индексируемым точностью правописания, и скоростью чтения сохраняется внутри индивида: участники читают слова, которые они лично могли правильно написать, быстрее, чем слова, которые они не могли написать.

В своей статье “Does a Reading Lexicon Provide Orthographic Representations for Spelling?” Дженнифер Берт и Хелена Тейт [Burt, Tate, 2002] поднимают вопрос организации ментального лексикона для чтения и правописания: один это лексикон или же их два? Несмотря на значительный интерес к правописанию со стороны исследователей в последнее время, мало что известно об организации орфографических знаний, используемых для правописания, или о том, как именно правописание связано с чтением. Хотя визуальное распознавание слов также должно основываться на изучении орфографии отдельных слов, неизвестно, представлена ли орфографическая информация, относящаяся к конкретному слову, полученная для чтения, и

хранится ли она в форме, доступной для правописания. То есть, может быть существует два ментальных лексикона, один для приобретения и один для вывода информации, с не обязательно одинаковой информацией в каждом. Альтернативная точка зрения, которой Берг и Тейт отдают предпочтение, заключается в том, что единый или орфографический лексикон служит и для чтения и для правописания.

Вопросы об организации орфографических знаний поднимают основные вопросы о том, как орфографическая информация представлена в памяти. В настоящее время развитие теории правописания ограничено недостатком информации о природе и приобретении орфографических репрезентаций, а также отсутствием четко определенной функциональной модели правописания и ее взаимосвязи с чтением. Эти вопросы важны не только для фундаментальных исследований в области грамотности, но и для педагогов, которые хотят способствовать оптимальному овладению навыками грамотности, необходимыми для современной жизни.

Среди грамотных взрослых, которые находятся в центре внимания настоящих исследований, превосходство в правописании ассоциируется с превосходством в фонологических навыках и знанием орфографических условностей, а также словообразовательной и флективной лексической морфологии. Как правило, грамотно пишущие люди также являются высококвалифицированными специалистами в визуальной идентификации слов.

Гипотеза о том, что для чтения и правописания используется единый орфографический словарь, послужила мотивом для исследования Берг и Тейт. Эта точка зрения с одним лексиконом согласуется с умеренной или высокой корреляцией между тестами на чтение и правописание, обычно наблюдаемыми у детей [Frith, 1980], и она обычно выдвигается в теориях усвоения чтения [Ehri, 1986], [Perfetti, 1992] и процессов правописания [Simon, Simon, 1973]. Дети и взрослые, которые читают часто, имеют много

возможностей научиться писать знакомые и незнакомые слова, и важность чтения для достижения навыков правописания была предложена многими исследователями в области развития [Bradley, Bryant, 1980], [Bradley, Bryant 1985], [Ehri, 1986],[Frith, 1980], [Perfetti, 1992], [Smith, 1982]. В соответствии с этой точкой зрения, показатели воздействия читательского опыта имеют умеренную корреляцию с точностью правописания у взрослых [Stanovich, Cunningham, 1992], [Stanovich, West, 1989].

Доказательства, подтверждающие альтернативную позицию двойного лексикона, то есть того, что существуют отдельные орфографические лексиконы для чтения и для правописания, получены из исследований о диссоциациях между чтением и правописанием в исследованиях развития [Bruck, Waters, 1988, 1990], [Burden, 1992], [Frith, 1980] и из нейропсихологической литературы.

В исследовании Дженнифер Берт и Хелены Тейт [Burt, Tate, 2002] изучалась поэтапная взаимосвязь между чтением и правописанием у студентов университетов, от которых можно ожидать, что они будут обладать, по крайней мере, адекватными навыками чтения. Цель исследования состояла в том, чтобы представить доказательства, которые имеют отношение к наборам элементов и участникам, о влиянии точности написания элементов на эффективность распознавания слов.

Вердикт по вопросу о том, используется ли при чтении и письме один орфографический лексикон или два разных, имеет важные теоретические и образовательные последствия. Это исследование показало, что, когда задание на распознавание слов выполняется в условиях, ограничивающих возможности участников проводить проверку правописания, существует соответствие между точностью написания элементов слов и эффективностью распознавания слов, что непосредственно вытекает из теории одного ментального лексикона для чтения и правописания.

Понятие 'лексическое качество' широко используется в литературе по распознаванию слов и чтению для обозначения систематического измерения индивидуальных различий, которое предсказывает выполнение целого ряда задач по распознаванию слов и чтению как у развивающихся читателей, так и у квалифицированных взрослых людей. Для оценки лексического качества использовалось множество различных подходов, но лишь немногие из них учитывали орфографическую точность, которая занимает центральное место в конструкции. В статье Салли Эндрюс и коллег [Andrews, Veldre, Clarke, 2020] описываются и оцениваются орфографические диктанты и тесты на распознавание слов, которые были разработаны для обеспечения точных измерений компонента лексического качества у студентов колледжей, которые являлись опытными читателями, которые и составили основную часть испытуемых в данном исследовании. Результаты анализа представлены для 785 учащихся, которые выполнили тесты по правописанию в сочетании со стандартизированными показателями понимания прочитанного, словарного запаса и скорости чтения, из которых 107 также прошли тесты на читательский опыт и фонематическую расшифровку. Анализ внутренней согласованности показал, что оба теста на правописание были относительно одномерными и демонстрировали хорошую внутреннюю согласованность, хотя тест на распознавание содержал слишком много простых элементов. Анализ на уровне элементов был включен, чтобы обеспечить основу для дальнейшего совершенствования этих инструментов. Тесты на правописание умеренно коррелировали с другими показателями владения письменным языком, но факторный анализ показал, что они последовательно определяют отдельный компонент, демонстрируя измерение вариативности, которое частично не зависит от различий в понимании прочитанного, скорости и словарном запасе. Эти компоненты, по-видимому, соответствуют параметрам точности и согласованности лексического качества.

Последовательные доказательства того, что тесты на орфографические способности выявляют частично независимое измерение вариативности

навыков чтения, согласуются с компонентом точности конструкции лексического качества [Perfetti, 2007]. Это также подтверждает предположение [Andrews, 2008, 2012, 2015] о том, что показатели орфографической способности могут использоваться для определения лексической точности и учета уникальных различий, которые не учитываются показателями понимания прочитанного и словарного запаса, которые обычно используются для оценки индивидуальных различий между опытными читателями.

Лексическое качество определяется как точностью орфографических представлений, так и последовательной, синхронной активацией связанных с ними фонологических и семантических кодов [Andrews, 2015], [Perfetti, 2007]. Второй компонент, последовательно идентифицируемый в анализе основных компонентов, определялся показателями знаний и процессов более высокого уровня: семантические знания, индексируемые словарным запасом, показатели понимания и скорости чтения на уровне текста, а также тест на читательский опыт. Таким образом, он, по-видимому, задействует процессы, связанные с эффективным, синхронным поиском и использованием комплекса лексических кодов, связанных с конкретной словоформой, – процессы, которые относятся к измерениям "связности" лексического качества [Perfetti, 2007] (подробнее об этом см. раздел 1.1, с. 8). Это согласуется с данными о том, что люди с более высокими баллами по компоненту связности, а не по компоненту точности, демонстрируют более сильное семантическое влияние праймов как в задании на лексическое решение [Andrews, Lo, 2013], так и в задачах семантической категоризации [Andrews, Lo, Xia, 2017]. Имеющиеся данные о том, что лексика теста Нельсона – Денни (Данный тест включает два субтеста. Первый подтест, «Словарь», состоит из 80 заданий с множественным выбором, каждый из которых имеет пять вариантов ответа. Слова взяты из учебников средней школы и колледжа и различаются по сложности. Вторым подтестом, «Понимание», требует от испытуемых прочитать пять коротких отрывков

(также взятых из учебников для средней школы и колледжа) и ответить на 38 вопросов с несколькими вариантами ответов о содержании этих отрывков. Примерно половина этих элементов относится к конкретному фактическому содержанию, тогда как другая половина носит более выводный характер. Общий балл по чтению получается путем суммирования баллов за словарный запас с баллами за понимание. (Последний балл удваивается, чтобы компенсировать большее количество заданий в первом субтесте.) Во время прохождения первого отрывка в субтесте на понимание также оценивается скорость чтения. В основном тест Нельсона – Денни используется в качестве скринингового теста на проблемы чтения и в качестве предиктора академической успеваемости. Однако, он не подходит для клинической оценки нарушений чтения [Brown, Bennett, Hanna, 1981], [Brown, Fishco, Hanna, 1993].) одинаково нагружена как по факторам точности, так и по факторам связности для выборки с половинным временем только тогда, когда в качестве предиктора выступал показатель скорости чтения, предполагают, что знание значений слов сочетается с высокой орфографической точностью для облегчения быстрого и последовательного поиска комплекса лексических кодов, определяющих слово. Это также подразумевает, что вклад связности между орфографическими и семантическими кодами в ранний поиск лексики может быть более эффективно использован, когда словарный запас индивидуума оценивается под давлением показателя скорости [Andrews, Veldre, Clarke, 2020].

Многомерный характер лексического качества, выявленный в результате этих анализов, подчеркивает ценность включения нескольких показателей лексического мастерства в исследования, направленные на изучение индивидуальных различий среди опытных читателей.

В этом разделе мы подробно рассмотрели то, как связаны между собой лексическое качество слова и наша способность к более продуктивному чтению, а также взаимосвязь начитанности и правописания и способы

качественного измерения показателя правописания. Остается вопрос, какие тесты можно применить для измерения показателя начитанности и как это можно сделать наиболее объективно? Об этом мы и расскажем в разделе 1.3, где мы рассмотрим важность показателя начитанности и то, какими способами мы можем его измерить, чтобы добиться наибольшей объективности.

1.3 Измерение показателя начитанности и тест на распознавание авторов (Author Recognition Test).

Широко распространено мнение, что большой читательский опыт делает нас умнее и способствует успеху в жизни. Однако подтверждается ли это предположение научными данными? Сюзанна Мол и Адриана Бас [Mol, Bus, 2011] отмечают, что ранее не предпринималось попыток ответить на эти вопросы путем обобщения фактических данных, о детях, находящихся на разных уровнях развития. Поэтому в своей работе они в деталях рассматривают данный вопрос.

В начальных классах дети в основном знакомятся с повествовательными текстами, в то время как начиная с четвертого класса их взаимодействие с текстами смещается в сторону описательных и технических текстов, поскольку они должны читать, чтобы получить знания в различных специальных областях. Задания по чтению для студентов колледжей и университетов также включают в себя больше научно-популярных текстов, чем повествовательных. Таким образом, чтение художественной литературы все чаще становится добровольным занятием, которое влечет за собой дополнительную и независимую практику чтения и, следовательно, вероятно, будет отличать более мотивированных читателей от менее мотивированных. Кроме того, поскольку когнитивная обработка текста улучшается в зависимости от вовлеченности читателя, а также потому, что повествовательные тексты больше стимулируют воображение и являются

лично значимыми и / или эмоционально привлекательными, чем описательные, чтение художественной литературы может особенно способствовать консолидации и расширению знаний о формах слов и значениях слов. Таким образом, чтение повествовательных текстов в качестве досуга может по-разному влиять на навыки чтения в разных возрастах и на разных уровнях образования.

По сути, чтение – это когнитивный процесс понимания записанной речи. Маленькие дети формируют основные понятия о связях между произносимыми и написанными словами, что приводит к распознаванию слов и знакомству с написанием слов. Первоначально дети узнают алфавит, приобретают навыки фонологической и орфографической обработки слов.

Эти базовые навыки чтения считаются наиболее ограниченными по времени навыками: после периода быстрого роста в начальных классах достигается потолок. Аналогичным образом, навыки техники чтения и правописания могут развиваться по аналогичной траектории, ограниченной временем, хотя для достижения мастерства в точности и беглости чтения слов, а также в правильном написании слов требуется больше времени. С самого раннего возраста способность к чтению слов может зависеть не только от базовых навыков чтения, но и от навыков устной речи, таких как словарный запас.

Проведя серию экспериментов, они выяснили, что книги обеспечивают значимый контекст для обучения чтению, не только как способ стимулирования понимания прочитанного, но и как средство развития технических навыков чтения даже в раннем детстве. У менее опытных читателей они обнаружили, что воздействие литературы было умеренно связано с устной речью и базовыми знаниями о чтении. Чтение книг оставалось важным для детей в школе, которые были более мотивированными читателями в детстве. Их анализ показал, что занятия чтением, которые являются частью досуга ребенка, дают ему существенные преимущества при развитии навыка устной речи. Интересно, что чтение книг

книг вне школьной программы, также позволяет развивать конкретные знания о словоформах и лучше развивать навык правописания. Наконец, студенты колледжей и университетов, которые читают для удовольствия, также могут быть более успешными в учебе.

То есть, было показано, что чтение влияет на правописание. Для того, чтобы провести более конкретный анализ корреляции начитанности и грамотности нам нужно как-то измерить читательский опыт человека.

Показателем читательской компетенции является высокая интеллектуальная активность при чтении, способность глубоко и критически осмыслить текст. Предполагается, что отличительная особенность восприятия художественного текста читателями с менее обширным читательским опытом – сниженная потребность в анализе прочитанного.

Одним из показателей компетенции читателя является предрасположенность к самостоятельному чтению (*free voluntary reading*) – устойчивая потребность и способность читать книги по осознанному выбору в свободное время.

Существуют следующие способы измерения читательского опыта:

1. Оценивание скорости понимания текста (*Reading comprehension speed test, RCS*). Оценивает скорость чтения в соотношении с тестом на понимание прочитанного [Steiner, 1978].
2. Опросник читательских привычек может содержать в себе примерно следующие вопросы:
 - а. Испытуемому предлагается оценить количество книг, а также журналов и газет, прочитываемых в месяц. Также необходимо указать приблизительное время, затраченное на их прочтение. Кроме того, испытуемому предлагается оценить следующую ситуацию: «Представьте, что вы остановились в незнакомом отеле и обнаружили, что вам совершенно нечего читать. Что вы почувствуете? Ответы ранжировались от 1 (отсутствие всякого

беспокойства) до 4 (крайняя степень беспокойства). Итоги служили непрямым признаком того, насколько велика мотивация человека к занятию чтением.

- b. Похожим образом критерием мотивации может служить «интервал чтения», измеряющийся оценкой следующей воображаемой ситуацией: «Представьте, что вы можете читать книгу, которая вам очень нравится, без вынужденных перерывов такое количество времени, сколько захотите. В течение какого времени вы бы продолжали?»
- c. Время, требующееся на выбор книги (Book Choice Decision Time) было измерено, исходя из ответов на следующий вопрос «Вы ищете в библиотеке книгу для чтения в свободное время. Вы выбираете одну, просматриваете её и решаете, взять ли или поставить обратно на полку. Сколько времени в среднем у вас занимает подобное решение?»
- d. Подобным измерением является время сортировки. Испытуемых просят без ограничения по времени просмотреть 30 кратких отрывков без указания авторства из различных книг и отнести каждую в одну из четырёх категорий: очень понравилось, понравилось, не очень понравилось, очень не понравилось. Такая процедура в целом является аналогичным повседневному выбору книг для чтения: как правило, выбирая книгу, читатель быстро пробегает глазами одну-две случайно выбранные страницы.
- e. Наконец, последний вопрос опросника читательских привычек звучит так: «Подумайте о книге, которую вы прочитали недавно и которая доставила вам удовольствие. Какого рода удовольствие вы обычно получаете от таких книг?»

3. Международное мониторинговое исследование PIRLS (Progress in International Reading Literacy Study) направлено на изучение читательской грамотности четвероклассников. Исследование

организовано Международной ассоциацией по оценке учебных достижений IEA (International Association for the Evaluation of Educational Achievement). В этом исследовании читательская грамотность определяется как «способность понимать и использовать письменную речь во всем разнообразии ее форм для целей, определяемых обществом и/или ценных для индивида. Младшие школьники читают, чтобы учиться, чтобы участвовать в школьных и внешкольных читательских сообществах, и для удовольствия». Международное мониторинговое исследование читательской грамотности PIRLS позволяет сравнить уровень и качество чтения и понимания текста учащимися начальной школы в различных странах мира, а также выявить различия в системах образования разных стран. В исследовании PIRLS оцениваются два вида чтения, которые чаще других используются учащимися во время уроков и вне школы: – чтение с целью приобретения опыта чтения художественной литературы; – чтение с целью освоения и использования информации [Центр оценки качества образования, 2016].

С помощью указанных методик можно изучить, есть ли связь между временем, проводимым за чтением, и предрасположенностью к чтению; что мотивирует людей больше времени посвящать чтению; и другое.

В своих статьях, Станович и коллеги [Stanovich, 1986], [Stanovich, West, 1989], [Stanovich, West, Harrison, 1995] говорят о следующих видах тестов на измерение начитанности:

1. Activity Preference Questionnaire – опросник на предпочитаемый вид деятельности в свободное время: дается следующая инструкция: вам на выбор будут даны два вида деятельности и вам нужно выбрать только одну, которой вам хотелось бы заниматься в свободное время. Пожалуйста, отметьте только один вариант. То есть, даже если вам нравятся оба вида деятельности, пожалуйста, отметьте только тот,

который вам нравится больше всего. Аналогичным образом, даже если вам не нравятся оба вида деятельности, отметьте тот, который вы предпочли бы выполнять. Далее идут 12 предложений по типу “Я предпочел бы послушать музыку (по моему выбору) или я предпочел бы посмотреть телевизионную передачу (по моему выбору). Шесть вопросов касались чтения, а остальные шесть служили филлера, чтобы отвлечь внимание от вопросов про чтение. В этих шести пунктах "прочитать книгу по моему выбору" было дополнено "посмотреть телевизионную программу по моему выбору", "заняться спортом на открытом воздухе по моему выбору", "послушать музыку по моему выбору", "поговорить с друзьями по моему выбору", "сходить в кино по моему выбору," и "потратить время на мое хобби". Количество раз, когда чтение предпочитали какому-то другому виду деятельности, и являлось оценкой участника за это задание. Таким образом, баллы варьировались от 0 до 6.

2. Reading habits composite – опросник читательских привычек: Наряду с серией демографических вопросов участники ответили на пять вопросов об их читательских привычках. Вопросы представляли собой вопросы с несколькими вариантами ответов, которые проверяли, читают ли они для удовольствия (почти никогда, пару раз в год и т. д.), регулярно покупают или подписываются на журналы, читают газеты (более одной в день, по одной каждый день, иногда и т. д.), сколько книг они читали в течение года (0, 1-2, 3-10 и т.д.) и насколько им нравилось читать (не очень сильно, немного, очень сильно).
3. Author Recognition Test (ART) – Тест на распознавание авторов: этот тест должен позволить измерить начитанность и был разработан так, чтобы обойти проблему субъективности (человек может завышать количество прочитанных книг и говорить, что очень любит читать, хотя на самом деле это не так), а так же проблему искажения

результатов опроса из-за желания соответствовать ожиданиям общества (ведь чтение это одно из самых одобряемых обществом занятий). ART представляет собой контрольный список, в котором участники указывают, знакомы ли они с именем конкретного популярного автора, поставив галочку рядом с именем. Участник не может просто отметить все имена подряд из-за наличия филлеров – имен людей, которые не являются популярными писателями или авторами. Таким образом, при проверке теста из количества правильно отмеченных реальных авторов будет отнято количество напрасно отмеченных имен-филлеров. ART содержит 40 имен авторов и 40 филлеров. Многие авторы книг регулярно появляются в списках бестселлеров. Хотя статистическая выборка авторов не проводилась, была предпринята попытка объединить авторов из самых разных жанров. Таким образом, в тесте представлены большинство основных категорий научной литературы (например, наука, политика, юмор, религия, история, бизнес, финансы, путешествия) и художественной литературы (например, мистика, детектив, романы, исторические романы, вестерны, рассказы, научная фантастика). При составлении списка были отобраны авторы, с которыми мы вероятнее всего встретились бы за пределами школы. Это было сделано для того, чтобы ART было косвенным показателем внешкольного читательского опыта. Таким образом, была предпринята попытка избежать авторов, которые включены в школьную программу. 40 филлеров в ART были именами, взятыми из 22 выпуска журнала Reading Research Quarterly за 1987 год. Очевидно, что ART не измеряет абсолютные уровни воздействия литературы с точки зрения времени, затраченного на чтение, или количества прочитанных слов. Чтобы получить такие оценки, необходимо использовать другие методы, такие как сбор дневников видов деятельности. То, что эти показатели являются косвенными, конечно, проблематично в некоторых контекстах, но, с

другой стороны, иногда это может быть и сильной стороной. Очевидно, что услышать об авторе по телевидению, не ознакомившись с реальной письменной работой, будет проблематично. Возникновение такого рода ситуаций, конечно, снижает валидность теста. Однако рассмотрим постэкспериментальный комментарий, который иногда дают участники: они знали, что определенное имя принадлежит какому-то автору, но никогда не читали ничего из того, что он написал. На вопрос о том, откуда они узнали, что этот человек писатель, участники неизменно отвечали, что видели одну из книг автора в книжном магазине, видели книгу автора в разделе "новая художественная литература" в библиотеке, читали рецензию на работу автора в Newsweek, увидел объявление в газете и так далее. Таким образом, знание имени этого автора было косвенным показателем чтения, даже несмотря на то, что конкретный автор на самом деле не был прочитан. Таким образом, хотя очевидно, что существуют способы ознакомления с именами, которые снижают достоверность (телевидение, радио), многие модели поведения, ведущие к знакомству с именами авторов, сами по себе являются свидетельствами читательского опыта.

4. Для того, чтобы сбалансировать тест на распознавание авторов, были разработаны Magazine Recognition Test (MRT) – тест на распознавание журналов и Newspaper Recognition Checklist – тест на распознавание газет.

Как нами уже было сказано в пункте 3 перечня тестов на начитанность про Author Recognition Test, что он обладает следующими характеристиками, которые делают его лучше других тестов на определение читательского опыта; во-первых, он наиболее объективен, так как человек, который его проходит не может зависить количество прочитанных им книг или указать, что любит читать, даже если это не так, и тем самым исказить результаты теста; во-вторых, участник не может просто отметить галочкой все имена и

это будет считаться как стопроцентный результат в тесте, так как половина имен – это не писатели и соответственно, баллы за выбор филлеров будут вычтены из итогового результата; в-третьих, тест на распознавание авторов содержит имена авторов не столько из школьной программы, сколько имена авторов бестселлеров, с которыми человек скорее всего должен был бы познакомиться вне школы, то есть самостоятельно, что также косвенно свидетельствует о его тяге к чтению и о том, что чтение – это то, чем человек любит заниматься в свободное время.

Естественно, чтобы использовать ART в экспериментах на других языках требуется адаптация теста и полная его переработка, учитывая специфику конкретного языка и культуры, которая с ним связана. ART был адаптирован для корейского [Lee, Seong, Choi, Lowder, 2019], нидерландского [Brysbart, Sui, Dirix, Hintz, 2020], китайского [Chen, Fang, 2013] и, конечно, для русского [Чернова, Бахтурина, 2021] – именно эта русскоязычная версия используется в данной работе.

1.4 Особенности русской орфографии.

В своем исследовании мы много говорим о грамотности и начитанности, поэтому стоит принять во внимание такой важный фактор, как строение орфографической системы языка, ведь он несомненно будет влиять на оба показателя.

Элизабет Борлеффс и коллеги [Borleffs, Maassen, Lyytinen, Zwarts, 2017] пишут о различных орфографических системах с алфавитным письмом.

Они говорят о том, что независимо от того, какая буквенная орфография изучается, начинающему читателю, по сути, необходимо научиться связывать буквы со звуками, чтобы получить доступ к фонологическим представлениям известных слов.

После целенаправленной практики и после того, как лексические представления слов закрепились в памяти читателя, опытному читателю

больше не нужно полагаться на фонетику, когда он снова сталкивается с тем же словом; чтение стало быстрым и высокоэффективным процессом распознавания слов [Sprenger-Charolles, Cole, 2003]. Та специфическая орфография, которую осваивает ребенок, была определена как центральный фактор, относящийся к окружающей нас обстановке, влияющий на овладение чтением и дислексию [Ziegler, Goswami, 2005]. Характеристики конкретной орфографии, которую необходимо выучить, формируют стратегии фонологического перекодирования и чтения.

Исследования показали, что прозрачные орфографии с очень регулярными графемно-фонемными соответствиями легче усваиваются, чем сложные и непрозрачные орфографии с высокой долей неправильных и непоследовательных вариантов написания (например, [Aro, Wimmer, 2003]; [Seymour, Aro, Erskine, 2003]). Было даже высказано предположение, что дети, находящиеся на нижнем конце спектра способностей к чтению, проявляют менее серьезные симптомы на языках с прозрачной орфографией, по крайней мере, с точки зрения точности [Landerl, Wimmer, Frith, 1997]. В непрозрачных орфографиях владение алфавитным принципом обеспечивает только часть ключа для декодирования, и многие слова не могут быть точно озвучены без доступа к сохраненному фонологическому представлению всего слова. Это может привести к разработке нескольких стратегий перекодирования, которые позволяют учащемуся декодировать с несколькими различными единицами, дополняя соответствие графем и фонем распознаванием буквенных шаблонов для рифм и попыток распознавания целых слов [Ziegler, Goswami, 2005], что требует задействования более широкого спектра когнитивных навыков.

Так как наше исследование сосредоточено именно на русском языке, стоит отдельно поговорить о специфике именно русской орфографии. Известно, что устройству русского языка, главным образом отвечает фонематический (морфологический) принцип, согласно которому “одни и те же буквы алфавита обозначают фонему во всех ее видоизменениях, как бы она ни

звучала в том или ином фонетическом положении”. [Ильинская, Сидоров, 1953] Таким образом, при написании мы можем сделать ошибку, потому что записали слово так, как его услышали или произнесли бы сами. Некоторые фонемы в русском языке могут обозначаться различными графемами на письме, например: фонема [ɕ:] может обозначаться как [сч] в слове “счастье” [‘ɕ:æstʲiɐ], как [щ] в слове “щека” [ɕ:ɪ’ka] и как [жч] в слове “мужчина” [mʊ’ɕ:inə]. Получается, что можно сделать ошибку в написании слова без ущерба смыслу.

В своей работе “Орфографические ошибки и опечатки. Особенности и причины появления” Чернышов Николай Владимирович пишет следующее: Выбор знака в орфограмме строго регламентирован орфографическими правилами. Конкуренция знаков системна и возникает только в том случае, если буквы для обозначения звуков становятся либо однозвучными (первый тип орфограмм), либо становятся парными знаками одной фонемы (т.к. буквы для гласных функционально осложнены сигналом о твёрдости или мягкости предшествующего согласного, и эта функция нейтрализуется при обозначении гласных звуков после шипящих и «ц»), либо конкуренция возникает между удвоением букв для согласного в интервокальном положении, что свидетельствует о долготе, и неудвоением (третий тип орфограмм), либо конкурируют небуквенные знаки дефис (полуслитное написание) – пробел (раздельное написание) – контакт (слитное написание). Конкуренция знаков – буквенных и небуквенных – создаёт орфографическую ситуацию выбора одного из нескольких. Выбор может осуществляться принудительно, без вариантов, и избирательно, когда для обозначения необходимо выбрать только один знак. Выбор знака подконтролен тому или иному орфографическому принципу. [Чернышов, 2007]

Также, существует фонетический принцип: записываются те буквы, которые обозначают именно тот звук, который слышится. К счастью, этот принцип действует очень не часто, иначе разобрать что написано было бы практически невозможно, потому что каждый писал бы как хотел.

Тем не менее по фонетическому принципу организованы следующие орфографические правила:

- приставки на –З, –С (перед глухим пишется С, перед звонким – З; обычно так и слышится: растянуть – перед глухим слышится С; раздавить – перед звонким слышится З);
- Ы после приставок на согласный (но из правила много исключений: И сохраняется после всех иноязычных приставок, а также русских приставок сверх- и меж-);
- написание приставок раз- и рас- с буквой А, в то время как под ударением слышится О: россыпь, но рассыпать;
- Ы после Ц в суффиксах и окончаниях (в этом месте слышится Ы, потому что Ц всегда твердый);
- О в ударных суффиксах после шипящих (девчонка);
- и некоторые другие, более редкие правила (например, написание слова «свадьба», так как в этом месте слышатся Д, хотя родственные ему «свататься», «сватовство» содержит в этом месте Т).

Последний принцип – традиционный. Здесь все просто, это те написания, которые сейчас не объясняются какими-то конкретными правилами, но восходят к каким-либо историческим процессам в русской орфографии. Сейчас такие написания нужно просто запомнить (крокодил, цифра, тонна, прочь).

Керек и Ниemi [Kerek, Niemi, 2009] пишут в своей статье об орфографии русского языка с точки зрения того, как происходит обучение чтению на русском.

В первую очередь, они говорят о том, что исследований по изучению орфографии русского языка необычайно мало. Недостаток исследований особенно досаден, поскольку русская фонология и орфография обладают рядом уникальных особенностей, что делает русский язык интересным примером на котором можно было бы протестировать современные теории распознавания слов и усвоения чтения. Конечно, в настоящее время (статья

вышла в 2009), существует довольно много исследований на тему русской орфографии, но по сравнению с другими исследованиями, их количество все еще остается небольшим.

Русский язык имеет довольно элегантную орфографическую систему: с одной стороны, довольно сложную и иерархическую, а с другой стороны, организованную вокруг доминирующего принципа и, следовательно, достаточно регулярную и предсказуемую, несмотря на то, что количество исключений высокое. Это сочетание сложности и регулярности делает русскую орфографию интересной для сравнительного исследования чтения.

В этой статье система письма была рассмотрена с точки зрения возможных единиц, используемых как для начального, так и для квалифицированного чтения на русском языке. Этот подход был вызван недавними предположениями [Ziegler, Goswami, 2005; 2006] о том, что различия в орфографических системах приводят к различиям в стратегиях чтения, разработанных начинающими читателями. Обсуждение принципов написания на русском языке и споры, связанные с их определениями, подчеркивают возможность использования нескольких единиц не только при письме, но и при чтении.

Подробный анализ фонологической и морфологической структуры русских слов служит основой для прогнозирования возможных сложностей обучения чтению на русском языке. Такие специфические черты русской фонологии, как различие между мягкими и твердыми согласными, о чем в письменной форме сигнализирует следующая буква, требуют особого внимания и долгосрочной стратегической подготовки начинающих читателей. Тот факт, что русская фонотактика допускает сложные включения, ничуть не облегчает эту задачу.

Морфемная (фонемная, согласно московской школе) преемственность в письменном русском языке требует хорошо развитого морфологического знания. Однако разбор русского слова на составляющие его морфемы часто бывает затруднен из-за широко распространенного морфемного слияния.

Развитию морфологической осведомленности также препятствуют надсегментарные особенности, такие как крайняя изменчивость русского ударения и тот факт, что различимость в произношении разных частей русских слов зависит от их положения по отношению к ударному слогу. Фонетические изменения в качестве русских согласных также не позволяют носителю русского языка сохранить образ морфемы как отдельной единицы восприятия.

Во многих отношениях трудности, с которыми сталкиваются начинающие читатели на русском языке, могут быть похожи на трудности, описанные для изучающих датский язык Элбро [Elbro, 2006], например, буквенно-звуковые паттерны с условным произношением (слоги), варианты написания, основанные на знании морфемной орфографии, и орфографические паттерны, специфичные для слов (традиционные варианты написания). Однако сопоставимая степень сложности в русском языке не обязательно означает, что она принесет те же результаты, что и в датском. Причиной такого оптимизма является предсказуемость многих неоднозначностей в написании русских слов.

Глава 2. Анализ орфографических репрезентаций в ментальном лексиконе.

2.1 Постановка задачи.

В первой части мы говорили о том, что такое лексические репрезентации (образы слова в ментальном лексиконе носителя языка, состоящие из трех составляющих: орфографическая, фонетическая и семантическая) и о том, почему они должны быть качественными (качественные лексические репрезентации быстро реактивируются в ментальном лексиконе, когда мы сталкиваемся с этим словом на письме или в устной речи).

Так же мы говорили о том, что такие качественные репрезентации формируются путем чтения и у опытных читателей больше качественных репрезентаций слов в ментальном лексиконе или же их опыт помогает им восполнять пробелы в менее качественных репрезентациях.

Эксперименты на эту тему были проведены на материале английского языка, иврита и ряда других языков. Но на материале русского языка, так же как и других восточно-славянских языков, таких экспериментов еще не было.

В своем эксперименте мы планируем проверить связаны ли между собой чтение и правописание в русском языке или, возможно, задействуются какие-то другие механизмы.

2.2. Материал исследования.

В качестве материала нашего исследования мы взяли различные слова, в которых люди часто допускают ошибки.

В разделе 1.4 мы говорили об особенностях русской орфографии, поэтому, при выборе материала нам нужно было учесть принципы орфографии русского языка и постараться подобрать различные орфограммы для того, чтобы наш материал достоверно отображал лексику русского языка и даже на небольшом количестве были представлены все ее особенности: безударная гласная (аккумулятор), удвоенная согласная или ее отсутствие (*комментарий, калории*), приставки пре-при (*непреренно*), разделительные мягкий и твердый знаки (*объект*), буквы и/ы после приставок (*предыстория*), гласные после

шипящих (*девчонка*), различные чередования (*экстремальный/экстрим*), слитное/раздельное написание (*в общем*).

Полный список слов с частотностями в корпусе представлен в приложении 1. Наше исследование проводилось на основе данных из Генерального Интернет-корпуса Русского Языка (ГИКРЯ) (URL: <http://www.webcorpora.ru/>). ГИКРЯ является мегакорпусом с более чем двадцатью миллиардами слов, созданный при помощи полностью автоматической технологии сбора и разметки текстов из Рунета и основанный на современных достижениях компьютерной лингвистики. Данный корпус был создан в 2013 году. Корпус включает в себя текстовые материалы из блогосферы, социальных сетей, с крупнейших новостных ресурсов и из литературных журналов. Из всех словоупотреблений 49 % приходится на «ВКонтакте», 40 % — на Живой Журнал, еще по 4 % — на Блоги@Mail.ru и Новости, и 2 % — на Журнальный зал. В сегменте Новостей собраны материалы из источников: РИА Новости, Regnum, Lenta.ru, Росбалт. Тексты снабжены метаразметкой (по дате создания текста, полу, месту и году рождения автора, интернет-жанру и так далее); все тексты снабжены автоматической морфологической разметкой и лемматизированы. Большую часть текстов создают записи за 2013—2014 годы, хотя на некоторых сегментах, например, в Журнальном зале, собраны тексты, начиная с 1994 года [Belikov et al., 2013] [Belikov et al., 2013], [Sharoff et al., 2015].

ГИКРЯ был выбран нами потому, что в этом корпусе представлены не редактируемые тексты из социальных сетей и блогов, которые как раз позволяют нам взглянуть на ошибочные варианты написания слов.

Нами было отобрано 66 трудных для написания слов (полный список слов с данными о них, выгруженными из корпуса, находится в приложении 1).

Затем, все эти слова мы внесли в ГИКРЯ и проверили их частотность в корпусе.

Введите ваш запрос	Запрос количества слов	Результатов:	IPM:
асимметрия 	<input type="text"/> 	2987	~0.152791
<input type="button" value="Добавить запрос"/>	<input type="button" value="Построить график"/>		
Искать среди слов: <input type="text"/>		или найти столько результатов: <input type="text"/>	
Имя запроса: ***	<input type="button" value="Сохранить"/>	<input type="button" value="Удалить"/>	<input type="button" value="Поделиться с"/>
<input type="checkbox"/> Стат. запрос <input type="checkbox"/> Тестовый режим <input type="checkbox"/> Базовый запрос			

рис. 5 результаты для слова асимметрия: количество данного слова в корпусе и IPM.

Затем, к каждому правильному написанию мы подобрали ошибочный вариант (к некоторым словам несколько вариантов). Если неправильных вариантов написания было больше, чем один, то для итогового эксперимента мы выбрали наиболее частотный вариант. Например, в случае со словом *асимметрия*, которое мы уже рассматривали выше, из всех возможных вариантов (*ассиметрия*, *ассиметрия*, *асиметрия*) мы взяли вариант *ассиметрия* как самый частотный:

Введите ваш запрос	Запрос количества слов	Результатов:	IPM:
ассиметрия 	<input type="text"/> 	1120	-0.058989
<input type="button" value="Добавить запрос"/>	<input type="button" value="Построить график"/>		
Искать среди слов: <input type="text"/>		или найти столько результатов: <input type="text"/>	
Имя запроса: ***	<input type="button" value="Сохранить"/>	<input type="button" value="Удалить"/>	<input type="button" value="Поделиться с"/>
<input type="checkbox"/> Стат. запрос <input type="checkbox"/> Тестовый режим <input type="checkbox"/> Базовый запрос			

рис. 6 результаты для варианта ассиметрия: количество данного варианта в корпусе и IPM

Введите ваш запрос	Запрос количества слов	Результатов:	IPM:
ассиметрия 	<input type="text"/> 	68	~0.004571
<input type="button" value="Добавить запрос"/>	<input type="button" value="Построить график"/>		
Искать среди слов: <input type="text"/>		или найти столько результатов: <input type="text"/>	
Имя запроса: ***	<input type="button" value="Сохранить"/>	<input type="button" value="Удалить"/>	<input type="button" value="Поделиться с"/>
<input type="checkbox"/> Стат. запрос <input type="checkbox"/> Тестовый режим <input type="checkbox"/> Базовый запрос			

рис. 7 результаты для варианта ассиметрия: количество данного варианта в корпусе и IPM

Введите ваш запрос  Запрос количества слов  Результаты: IPM:
60 ~0.004229

Искать среди слов: или найти столько результатов:

Имя запроса:

Стат. запрос Тестовый режим Базовый запрос

рис. 8 результаты для варианта асиметрия: количество данного варианта в корпусе и IPM

Для каждого слова мы рассчитали показатель энтропии по формуле, которая приведена в разделе 1.1.

После того, как мы проделали такую операцию со всеми нашими словами мы разделили слова на два протокола для эксперимента: в первом половина слов написана верно, а половина с ошибкой, во втором то же самое, но наоборот (те, что в первом комплекте были правильно написаны, будут с ошибкой, а те, что были в первом комплекте с ошибкой, будут написаны верно). Соответственно, половине наших участников достанется первый протокол, половине – второй.

Мы планируем проверить, как воспринимаются слова с искаженным написанием, когда они предъявлены изолированно: участник видит перед собой на экране слово и должен принять решение, правильно оно написано или нет.

В ходе эксперимента мы измеряем следующие показатели:

1. правильность ответа, данного участником (варианты: правильно ответил, что ошибки нет; правильно ответил, что ошибка есть; неправильно ответил, пропустив ошибку; неправильно ответил, заподозрив ошибку там, где ее не было);
2. скорость обдумывания ответа.

При этом нас интересует, как правильность и скорость ответа участника зависит от частотности правильного написания слова, частотности его неправильного написания, а также соотношения частотностей правильного и неправильного написания слова в корпусе.

В данной ситуации, наша гипотеза такова, что чем чаще слово встречается в неправильном написании, тем дольше над ним будут думать и тем больше неправильных ответов дадут. Причем неправильных не только в смысле «не заметил ошибку, которая была», но и наоборот «счел верное написание ошибочным». Потому что представление о том, как пишется слово, становится более размытым. С другой стороны, мы подозреваем, что это скорее всего будет происходить с низкочастотными словами, потому что лексические репрезентации для высокочастотных слов более устойчивы.

2.3 Проведение эксперимента.

В условиях пандемии коронавирусной инфекции (COVID-19) нам пришлось проводить наш эксперимент и собирать наши данные удаленно.

Соответственно, в связи с этим, наш эксперимент был выполнен на платформе PCIBex (PennController for Internet Based Experiments) [Zehr, Schwarz, 2018], позволяющей создать свой эксперимент и удаленно собирать данные.

На рис. 9 представлена стартовая страница эксперимента, где написана инструкция к эксперименту, краткая анкета для облегчения процесса сбора и дальнейшей обработки данных и информированное согласие на прохождение эксперимента:

Добрый день!

Пожалуйста, внимательно ознакомьтесь с инструкцией.

Ваша задача - определить, ПРАВИЛЬНО ли написано слово на экране.

Если слово написано ПРАВИЛЬНО - нажмите на клавиатуре клавишу "1".

Если слово написано С ОШИБКОЙ - нажмите на клавиатуре клавишу "0".

Первые три слова - тренировочные.

Пожалуйста, укажите Ваш ID или никнейм, возраст и пол. Эти сведения нужны только для удобства обработки данных и не будут использоваться с иными целями.

ID или никнейм:

Возраст:

Пол: Мужской Женский

Я согласен принять участие в исследовании

[→ Click here to continue](#)

рис. 9 стартовая страница эксперимента с инструкцией и согласием.

Сначала участникам предъявляются три слова для тренировки, они не входят в дальнейший анализ и нужны для того, чтобы участнику стала понятна процедура эксперимента. Затем начинается сам эксперимент. Слова предъявлялись изолированно и в случайном порядке для каждого участника. Стимулы располагались по центру экрана, 32 кегль, шрифты Verdana, Geneva (семейство шрифтов sans-serif, чтобы на разных устройствах текст выглядел одинаково).

На каждое слово давалось временное ограничение в 5 секунд. Если участник думал больше данного времени, слово автоматически пропускалось и осуществлялся переход к следующему стимулу. Время нажатия на кнопку регистрировалось и при обработке результатов мы оценивали скорость реакции.

Между стимулами участникам предъявлялся фиксационный крест в середине экрана на 1 секунду.

Когда участник прошел первую часть со словами – ему предлагается вторая часть – тест на распознавание авторов, оценивающий уровень читательского опыта (подробнее см. раздел 1.3), где нужно выбрать имена всех писателей, которые знакомы участнику эксперимента, если участник знал такого писателя – ему нужно было нажать клавишу 1, если нет – 0.

Всего в тесте на распознавание авторов было 75 имен писателей, плюс – 75 филлеров (не писателей). Стимулы предъявлялись по одному в центре экрана на пять секунд, так же, как и в первой части эксперимента.

Между стимулами участникам также предъявлялся фиксационный крест.

После прохождения второй части эксперимента участникам предъявлялось сообщение о том, что эксперимент успешно завершен и его результаты записаны.

В зависимости от скорости прохождения эксперимент в среднем занимал от 15 до 20 минут.

Участники-носители русского языка приглашались с помощью объявления в социальных сетях:

Предлагаю Вам пройти один очень интересный эксперимент, который я сделала в рамках своей магистерской работы. Он связан с ментальным лексиконом – внутренним словарем, который мы используем, когда порождаем и воспринимаем речь. Для эксперимента обязательно нужен компьютер, а по времени он занимает 15 минут. Подробная инструкция есть по ссылкам:

- Если Ваш день рождения летом или зимой – перейдите по этой ссылке: <https://farm.pcibex.net/p/QdycUr/>
- Если весной или осенью - по этой: <https://farm.pcibex.net/p/jTeMxT/>

Спасибо большое всем заранее за участие! Если не можете пройти - сделайте, пожалуйста, репост. Может кому-то из Ваших знакомых это будет интересно. Спасибо за помощь!

рис. 10 пример объявления для прохождения эксперимента в социальных сетях.

Также, ссылки размещались на ресурсе Яндекс.Толока. Это краудсорсинговый проект, созданный в 2014 году для быстрой разметки большого количества данных, которые затем используются для машинного обучения и совершенствования поисковых алгоритмов. Как правило, размещаемые задания достаточно простые, для их исполнения не требуется специальная подготовка. В основном они связаны с анализом и оценкой контента. Поэтому любой желающий может выполнять задания и получать за них вознаграждение. Толока — это не только инструмент для решения прикладных задач, но и площадка для научных исследований в разных

предметных областях. В мае 2019 года команда сервиса начала публиковать наборы данных для некоммерческих, академических целей, чтобы поддержать научное сообщество и привлечь исследователей в Толоку. Датасеты могут быть интересны исследователям разных направлений: для лингвистических исследований, для задач компьютерного зрения, для тестирования моделей агрегации вердиктов исполнителей и обучения чат-ботов.

Участниками нашего эксперимента являлись взрослые здоровые носители русского языка, так как наши стимулы подразумевают то, что эти слова будут знать те, кто уже окончил школу или является как минимум старшеклассником, так как эти словарные слова проходят в 9–11 классах.

Всего в нашем эксперименте приняло участие 75 человек, из них 36 мужчин и 39 женщин.

Распределение участников по возрасту представлено на линейчатой диаграмме ниже:



рис. 11 распределение участников по возрасту.

Средний возраст участников – 35 лет.

Во время обработки результатов нам пришлось исключить несколько участников, которые недобросовестно проходили эксперимент, то есть по результатам видно, например, что они нажимали на одну и ту же кнопку или просто пропускали все слова, не выбрав никакой вариант. Таким образом, число участников сократилось до 61.

В первом протоколе у нас получилось 33 участника, во втором – 28 (сводные таблицы с данными всех участников можно посмотреть по ссылке: <https://drive.google.com/drive/folders/14Wt6n7mnp31DDS9IsaTRsMZwdl0C2S0C?usp=sharing>).

Затем мы приступили к анализу наших результатов.

2.4 Обработка результатов.

При обработке результатов мы исключили все результаты, которые показали скорость реакции больше 5 секунд. Нами также был рассчитан порог выбросов – три стандартных отклонения от среднего значения, он составил 4700 мс, все данные (всего таких результатов было 35, то есть 0,9% от общего числа), превысившие этот порог были исключены из рассмотрения.

Для того, чтобы сделать выводы, нам нужно было сравнить следующие показатели и выявить, как на точность и скорость ответов в задании на правописание влияет частотность верных и неверных написаний слова в корпусе, а также уровень читательского опыта, оцененный при помощи теста на распознавание имен писателей.

Анализ данных мы проводили в программе Rstudio (URL: <https://www.rstudio.com/>). Это свободная среда разработки программного обеспечения с открытым исходным кодом для языка программирования R, который предназначен для статистической обработки данных и работы с графикой. RStudio написана на языке программирования C++ и использует фреймворк Qt для графического интерфейса пользователя [Van der Loo, Jonge, 2012].

Для анализа мы использовали метод смешанных линейных моделей. Сущность данного метода заключается в следующем. Предполагается, что эффекты (факторы), оказывающие влияние на зависимую переменную, можно условно разделить на два типа: фиксированные и случайные. Случайные те эффекты, которые случайным образом варьируются в исследовании и не отражают все возможные значения фактора в генеральной совокупности. Фиксированные эффекты, с другой стороны, — это то, что обычно является предметом интереса исследователя, то есть те независимые переменные, уровни которых он устанавливает или контролирует [Четвериков, 2015].

Анализ (полные данные по всем статистическим данным в виде таблиц можно найти в приложении 2) показал, что:

На правильность ответа влияет результат теста участника на начитанность:

$b = 0,020600$, $SE = 0,005694$, $z = 3,618$, $p = 0,000297$

и частотность слова в корпусе¹ (см. таблицу 1):

$b = 0,029800$, $SE = 0,004741$, $z = 6,285$, $p = 3,27e-10$

На скорость ответов влияет только частотность слов в корпусе (см. таблицу 2):

$b = -5,5321$, $SE = 1,8015$, $t = -3,071$, $p = 0,0032$;

Кроме того, на правильность ответа значимо влияет показатель энтропии [Var-On, Kuperman, 2019], то есть степень конкуренции между вариантами написания слова на правильность ответов, о котором мы подробно говорили в разделе 1.1 (см. таблицу 3):

$b = -2,514335$, $SE = 0,398166$, $z = -6,315$, $p = 2,71e-10$;

Также показатель энтропии значимо влияет и на скорость ответов (см. таблицу 4):

$b = 378,602$, $SE = 145,434$, $t = 2,603$, $p = 0,0116$.

¹ Данные о частотности вносились в модель в логарифмированном виде.

2.5 Обсуждение результатов

В первую очередь, мы проверили влияние читательского опыта и частотности верных и неверных написаний на правильность ответов. Чем больше у участника результат за тест на распознавание авторов, оценивающий уровень читательского опыта, тем выше у него будет результат в тесте на грамотность (он укажет правильно предъявленные слова как правильные, а неправильно предъявленные – как ошибку). Это подтверждает данные о том, что более опытные читатели являются и более грамотными людьми в целом, о чем мы говорили в связи с гипотезой лексического качества [Perfetti, 1985; 2007] в разделе 1.1.

У более квалифицированных читателей либо присутствует большое количество лексических репрезентаций высокого качества, либо если нет – то они обладают бóльшими ресурсами, чем менее квалифицированные читатели, для того, чтобы восстанавливать недостающие составляющие в лексической репрезентации меньшего качества.

Это согласуется с данными, которые были получены на других языках: английский, китайский, иврит, финский, греческий [Kuperman, Bar-On, Bertram, Boshra, Deutsch, Kyröläinen, Mathiopoulou, Oralova, Protopapas, 2021]. Также в ходе эксперимента выявлено, что частотность предъявленного слова очень сильно повлияла на точность ответов: чем частотнее слово, тем больше вероятность правильного ответа (то есть, правильное слово будет указано участником как правильное, а ошибочное как ошибочное). Это подтверждает данные о том, что более частотные слова имеют более устойчивые репрезентации и их сложнее размыть, чем менее частотные.

Кроме того, мы анализировали влияние различных факторов на скорость ответа.

Скорость ответа зависит от частотности правильного варианта написания слова: чем выше будет частотность слова в его правильном варианте написания, тем меньше времени участник затратит на то, чтобы дать ответ.

Это ожидаемый результат, так как широко известно, что частотность слова влияет на скорость его распознавания и чем частотнее слово – тем быстрее оно будет распознано.

Средние значения по скорости распознавания для тройки наших самых частотных слов составили: в общем (IPM 141,484495) – 1983 мс, пожалуйста (IPM 114,381608) – 1714 мс, будущее (IPM 48,076729) – 1589 мс.

Результаты теста на начитанность на скорость ответов не влияют. Это тоже довольно ожидаемо, так как такой показатель, как начитанность связан именно с лексическими репрезентациями и их качеством в ментальном лексиконе конкретного читателя и вряд ли может оказать значительное влияние на скорость распознавания слов.

Показатель энтропии очень сильно повлиял на правильность ответа участника: чем ниже будет показатель энтропии, то есть, чем ниже будет степень конкуренции между вариантами написания слова, тем больше вероятность, что участник даст правильный ответ. Это подтверждает данные, полученные на материале других языков: английский, китайский, иврит, финский, греческий [Kuperman, Bar-On, Bertram, Boshra, Deutsch, Kyröläinen, Mathiopoulos, Oralova, Protopapas, 2021], где было также показано, что чем выше показатель энтропии для слова, тем сложнее будет распознать его, так как альтернативные варианты будут конкурировать между собой и затруднять принятие решения.

Кроме того, было выявлено, что показатель энтропии влияет на скорость ответов: чем выше будет степень конкуренции между вариантами написания слова, тем больше времени участнику потребуется, чтобы дать правильный ответ. Это опять же подтверждает те данные, которые мы приводили в первой главе, чем больше альтернативных вариантов написания имеет то или иное слово, тем более размытой становится лексическая репрезентация (ее орфографическая составляющая) в ментальном лексиконе человека и тем больше времени человек затрачивает на то, чтобы распознать данное слово,

потому что в процессе распознавания, начинается конкуренция между этими альтернативными вариантами.

Таким образом, мы можем видеть, что, несмотря на то, что в русском языке иное устройство орфографической системы, нежели чем, например, в английском, данные о корреляции читательского опыта и правописания совпадают и говорят нам о том, что чтение (только тех источников, которые были предварительно отредактированы) действительно влияет на правописание и помогает нам создавать более качественные лексические репрезентации слов в ментальном лексиконе, а значит, помогает нам качественнее порождать письменную речь.

Несмотря на разность орфографических систем, совпадают и данные о том, что ошибочные варианты написания действительно могут вступать в конкуренцию с правильными вариантами и мешать нам распознавать корректно написанные слова, неважно, являемся ли мы сами грамотными людьми или нет.

Заключение

Таким образом, данное исследование было направлено на изучение устойчивости графического облика слов в ментальном лексиконе носителей русского языка. Мы поставили перед собой цель – выявить механизмы восприятия слов, которые имеют конкурирующие варианты написания.

Результаты, полученные в ходе исследования, позволяют сделать следующие выводы:

1. В ходе исследования теоретической стороны вопроса, мы:
 - a. изучили такие понятия как:
 - i. гипотеза лексического качества, которая говорит о том, что для правильного распознавания слова в нашем ментальном лексиконе должна храниться качественная лексическая репрезентация, то есть облик слова должен быть закреплён фонологически, орфографически и семантически.
 - ii. показатель энтропии, который количественно определяет степень неопределенности при выборе одного из доступных вариантов с учетом их распределения вероятностей. Иными словами, степень конкуренции между вариантами написания слова.
 - b. подробно рассмотрели взаимосвязь читательского опыта и грамотности, описали, какие исследования уже были проведены на эту тему и рассказали о том, что читательский опыт положительно влияет на грамотность, так как при чтении создаются качественные лексические репрезентации в ментальном лексиконе читающего.
 - c. рассмотрели различные методики измерения читательского опыта: оценивание скорости понимания текста, различные опросники читательских привычек, тест на распознавание авторов (ART); и пришли к выводу, что именно тест на

распознавание является самым объективным, так как участнику нужно отметить только авторов (и не отмечать филлеры), если он отметит все стимулы – то в сумме получит 0 баллов.

2. В ходе практической части нашего исследования мы:

- a. рассказали о выборе материала для исследования, который должен был соответствовать принципам русской орфографии и в относительно небольшой выборке отражать многообразную лексику русского языка.
- b. рассказали о том, как проводился наш эксперимент.
- c. проанализировали результаты и сделали следующие выводы:
 - i. чем больше у участника результат за тест на распознавание авторов, оценивающий уровень читательского опыта, тем выше у него будет результат в тесте на грамотность;
 - ii. чем частотнее слово, тем больше вероятность правильного ответа;
 - iii. чем выше будет частотность слова в его правильном варианте написания, тем меньше времени участник затратит на то, чтобы дать ответ;
 - iv. чем ниже будет показатель энтропии, то есть, чем ниже будет степень конкуренции между вариантами написания слова, тем больше вероятность, что участник даст правильный ответ;
 - v. чем больше показатель энтропии, то есть чем выше будет степень конкуренции между вариантами написания слова, тем больше времени участнику потребуется, чтобы дать правильный ответ.

Если делать общий вывод, то получается, что читательский опыт (чтение именно тех источников, которые были подвержены редактуре) влияет на наше правописание и действительно позволил более опытным читателям получить более качественные лексические репрезентации. И, как мы и предполагали,

показатель энтропии влияет на правильность ответов и на скорость: более размытое слово, которое имеет много конкурирующих вариантов написания, потребует от человека больших усилий и замедлит его при распознавании. Таким образом, мы выполнили все задачи, которые поставили перед собой во введении к работе.

Список использованных источников:

1. Белов В. А. Семантические исследования организации и функционирования ментального лексикона / В. А. Белов // Научный диалог. — 2020. — № 8. — С. 29—51.
2. Залевская А. А. Психолингвистические исследования. Слово. Текст: Избранные труды. — М.: Гнозис, 2005. — 543 с.
3. Ильинская И. С., Сидоров В. Н. Современное русское правописание // Учебные записи Московского городского педагогического института им. В.П. Потемкина. Т. 22. Вып. 2. М., 1953. С. 3–40.
4. Чернова, Д. А., Бахтурина, П. В. Разработка русской версии теста на распознавание авторов: инструмент для оценки читательского опыта. // Когнитивная наука в Москве: новые исследования. Материалы конференции 23–24 июня 2021. Под ред. Е.В. Печенковой, М.В. Фаликман, А.Я. Койфман. – М.: ООО «Буки Веди», ИППиП. 2021 г. — 556 стр. (С. 469–473).
5. Чернова Д. А., Алексеева С. В., Слюсарь Н. А. Чему нас учат ошибки: трудности при обработке слов с частотными орфографическими ошибками. // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (Москва, 17–20 июня 2020 г.). Вып. 19 (26), 2020., С. 147–159.
6. Чернышов Н. В. Орфографические ошибки и опечатки. Особенности и причины появления. // По материалам конференции “Современные научные достижения. Филология”. Вып. 2, 2007 г., С. 32–35.
7. Четвериков А.А. Линейные модели со смешанными эффектами в когнитивных исследованиях // Российский журнал когнитивной науки. 2015. Том 2 (1), С. 41–51.
8. Шаров С. А., Беликов В. И., Копылов Н. Ю., Сорокин А. А., Шаврина Т. О. (2015) Корпус с автоматически снятой морфологической неоднозначностью: К методике лингвистических исследований.

Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной конференции Диалог, Бекасово.

9. Центр оценки качества образования (2016) Основная информация об исследовании PIRLS // Международное исследование “Изучение качества чтения и понимания текста”, 35 стр.
10. Aitchison J. *Words in the Mind. An Introduction to the Mental Lexicon* / J. Aitchison. — Oxford : Blackwell, 2012. — 307 p.
11. Andrews S. (2008). Lexical expertise and reading skill. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 49, pp. 247–281.
12. Andrews S., Bond R. (2009). Lexical expertise and reading skill: Bottom-up and top-down processing of lexical ambiguity. *Reading and Writing*, Vol. 22, pp. 687–711.
13. Andrews S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In J. S. Adelman (Ed.), *Visual Word Recognition*, Vol. 2, pp. 151–172.
14. Andrews S., Lo S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, Vol. 68, pp. 279–296.
15. Andrews S. (2015). Individual differences among skilled readers: The role of lexical quality. In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading*, pp. 129–148.
16. Andrews S., Lo S., Xia, V. (2017). Individual differences in automatic semantic priming. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 43, pp. 1025–1039.
17. Andrews S., Veldre A., Clarke I.E. (2020). Measuring Lexical Quality: The Role of Spelling Ability. *Behavior Research Methods*, Vol. 52, pp. 2257–2282.

18. Aro M., Wimmer H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, Vol. 24, pp. 621–635.
19. Bar-On A., Kuperman V. (2019). Spelling errors respect morphology: a corpus study of Hebrew orthography. *Reading and Writing* Vol. 32, pp. 1107–1128.
20. Beech J. R. (2002). Individual differences in mature readers in reading, spelling, and grapheme-phoneme conversion. *Current Psychology*, Vol. 21, pp. 121–132.
21. Belikov V., Kopylov N., Piperski A., Selegey V., Sharoff S. (2013). Corpus as language: from scalability to register variation. In *Dialogue, Russian International Conference on Computational Linguistics*, Bekasovo.
22. Belikov V., Kopylov N., Piperski A., Selegey V., Sharoff S. (2013). Big and diverse is beautiful: A large corpus of Russian to study linguistic variation. In *Web as Corpus Workshop (WAC-8)*.
23. Borgwaldt S. R., Hellwig F. M., De Groot A. M. B. (2005). Onset entropy matters — Letter to phoneme mappings in seven languages. *Reading and Writing: An Interdisciplinary Journal*, Vol. 18, pp. 211–229.
24. Borleffs E., Maassen B., Lyytinen H., Zwarts F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: a narrative review. *Reading and writing*, Vol. 30(8), pp. 1617–1638.
25. Brown J., Bennett J., Hanna G. (1981). *The Nelson-Denny Reading Test*. Lombard, IL: Riverside.
26. Brown, J.A., Fishco, V.V., & Hanna, G.S. (1993). *Nelson–Denny Reading Test: Manual for Scoring and Interpretation, Forms G & H*. Rolling Meadows, IL: Riverside Publishing.
27. Bruck M., Waters G. (1988). An analysis of the spelling errors of children who differ in their reading and spelling skills. *Applied Psycholinguistics*, Vol. 9, pp. 77–92.

28. Bruck M., Waters G. (1990). Effects of reading skill on component spelling skills. *Applied Psycholinguistics*, Vol. 11, pp. 425–437.
29. Bryant P., Bradley L. (1980). Why children sometimes write words which they do not read. In U. Frith (Ed.), *Cognitive processes in spelling*, pp. 355–370. London, UK: Academic Press.
30. Bryant P., Bradley L. (1985). *Rhyme and Reason in Reading and Spelling*. Ann Arbor: University of Michigan Press, p. 144.
31. Brysbaert M., Sui L., Dirix N., Hintz F. (2020). Dutch Author Recognition Test. *Journal of Cognition*, Vol. 3(1), pp.6.
32. Burden, V. (1992). Why are some “normal” readers such poor spellers? In C. M. Sterling & C. Robson (Eds.), *Psychology, spelling and education: Multilingual matters*, pp. 200–213. Witchford, UK: Multilingual Matters, Ltd
33. Burt J.S., Fury M. (2000). Spelling in adults: The role of reading skills and experience. *Reading and Writing*, Vol. 13, pp. 1-30.
34. Burt J.S., Tate H. (2002). Does a reading lexicon provide orthographic representations for spelling. *Journal of Memory and Language*, Vol. 46, pp. 518–543.
35. Chen S.-Y., Fang S.-P. (2013). Developing a Chinese version of an Author Recognition Test for college students in Taiwan. *Journal of Research in Reading*, Vol. 38(4), pp. 344–360.
36. Collins M. (2002). Interhemispheric communication via direct connections for alternative meanings of ambiguous words. *Brain and Language*, Vol. 80(1), pp. 77–96.
37. Ehri L. C. (1986). Sources of difficulty in learning to spell and read. In M. L. Wolraich, & D. Routh (Eds.), *Advancements in developmental and behavioral pediatrics*, pp. 121-195. Greenwich, CT: JAI Press.
38. Elbro, C. (2006). Literacy acquisition in Danish: A deep orthography in cross-linguistic light. In M. Malatesha Joshi & P. G. Aaron (Eds.),

- Handbook of orthography and literacy (pp. 31–45). Mahwah, NJ: Lawrence Erlbaum Associates.
39. Ernestus M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, Vol. 142, pp. 27–41.
 40. Frith U. (1980). Unexpected spelling problems. In U. Frith (Ed.), *Cognitive processes in spelling*, pp. 495-515. London: Academic Press.
 41. Harris L. N., Perfetti C. A. (2017). Individual Differences in Phonological Feedback Effects: Evidence for the Orthographic Recoding Hypothesis of Orthographic Learning. *Scientific studies of reading : the official journal of the Society for the Scientific Study of Reading*, Vol. 21(1), pp. 31–45.
 42. Harris L. N., Perfetti C. A. (2016). Lexical Stress and Linguistic Predictability Influence Proofreading Behavior. *Frontiers in Psychology*, Vol. 7.
 43. Hart L., Perfetti C. A. (2008). Learning words in Zekkish: Implications for understanding lexical representation. In E. L. Grigorenko & A. J. Naples (Eds.), *Single word reading: Behavioral and biological perspectives* (pp. 107–128). New York: Taylor & Francis.
 44. Hersch J., Andrews S. (2012). Lexical quality and reading skill: Bottom-up and top-down contributions to sentence processing. *Scientific Studies of Reading*, Vol. 16, pp. 240–262.
 45. Holmes V. M., Castles A. E. (2001). Unexpectedly poor spelling in university students. *Scientific Studies of Reading*, Vol. 5, pp. 319–350.
 46. Kerek, Eugenia & Niemi, Pekka. (2009). Russian orthography and learning to read. *Reading in a Foreign Language*, Vol. 21 (1), pp. 1–21.
 47. Kuperman V., Bar-On A., Bertram R., Boshra R., Deutsch A., Kyröläinen A., Mathiopolou B., Oralova G., Protopapas A. (2021). Prevalence of spelling errors affects reading behavior across languages. *Journal of experimental psychology. General*, Vol 150(10), pp. 1974–1993.

- 48.Landerl K., Wimmer H., Frith U. (1997). The impact of orthographic consistency on dyslexia: A German–English comparison. *Cognition*, Vol. 63, pp. 315–334.
- 49.Lee H., Seong E., Choi W., Lowder M. W. (2019). Development and assessment of the Korean Author Recognition Test. *Quarterly journal of experimental psychology*, Vol. 72(7), pp. 1837–1846.
- 50.Mark P.J. Van der Loo, Edwin de Jonge. *Learning RStudio for R Statistical Computing*. — Packt Publishing, 2012. — 126 c.
- 51.Martin-Chang S.L., Ouellette G.P., Madden M.S. (2014). Does poor spelling equate to slow reading? The relationship between reading, spelling, and orthographic quality. *Reading and Writing*, Vol. 27, pp. 1485–1505.
- 52.Mol S.E., Bus A.G. (2011). To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. *Psychological bulletin*, Vol. 137(2), pp. 267–96.
- 53.Perfetti C. A. (1985). *Reading ability*. Oxford, Oxford University Press, pp. 282.
- 54.Perfetti C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri,& R. Treiman (Eds.), *Reading acquisition*, pp. 145–174. Hillsdale, NJ: Erlbaum
- 55.Perfetti C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, Vol. 11, pp. 357–383.
- 56.Perfetti C. A., Hart L. (2001). The lexical basis of comprehension skill. In D. S. Gorfien (Ed.) *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity*, Washington, DC: American Psychological Assosiation, pp. 67–86.
- 57.Protopapas A., Vlahou E. L. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods*, Vol. 41, pp. 991–1008.

58. Rahmanian S., Kuperman V. (2017), Spelling errors impede recognition of correctly spelled word forms. *Scientific Studies of Reading*, Vol. 23, pp. 24–36.
59. Seymour P. H. K., Aro M., Erskine J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, Vol. 94, pp. 143–174.
60. Simon D. P., Simon H. A. (1973). Alternative uses of phonemic information in spelling. *Review of Educational Research*, Vol. 43, pp. 115–137.
61. Smith F. (1982). *Writing and the writer*. New York: Holt, Rinehart, & Winston, p. 257.
62. Sprenger-Charolles, L., & Cole, P. (2003). *Lecture et Dyslexie*. Paris: Dunod.
63. Stanovich K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of education*, Vol. 189(1–2), pp. 23–55.
64. Stanovich K. E., Cunningham A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, Vol. 20, pp. 51–68.
65. Stanovich K. E., West R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, Vol. 24(4), pp. 402–433.
66. Stanovich K. E., West R. F., Harrison M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, Vol. 31(5), pp. 811–826.
67. Steiner, K. (1978). ERIC/RCS: Speed Reading Revisited. *Journal of Reading*, 22(2), 172–176.
68. Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).
69. Ziegler J. C., Goswami U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, Vol. 131, pp. 3–29.

70. Ziegler J. C., Goswami U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science*, Vol. 9, pp. 429–453.

Приложение 1.

правильное написание	рез-ты в корпусе	IPM	слово с ошибкой	рез-ты в корпусе	IPM
аббревиатура	11139	0,565779	абревиатура	338	0,018672
агентство	325866	16,539541	агенство	27161	1,379765
аккаунт	124520	6,321597	акаунт	7338	0,392193
аккомпанемент	13907	0,706406	аккомпанимент	1987	0,104208
аккумулятор	47379	2,405113	аккамулятор	974	0,053364
апелляция	5298	0,26961	аппелляция	521	0,028981
аппендицит	6899	0,351003	апендицит	630	0,034352
асимметрия	2987	0,152791	ассиметрия	1120	0,058989
баннер	34478	1,751806	банер	6103	0,316404
будущее	8938	48,076729	будующее	4834	0,250754
в общем	2787987	141,484495	вообщем	317292	16,199203
ветренный	4307	0,219878	ветреный	1656	0,088103
винегрет	3863	0,635673	винигрет	790	0,042965
впоследствии	185626	9,420622	впоследствие	2909	0,14842
галерея	75276	3,821111	галлеря	2342	0,12681
гостиная	20742	1,053497	гостинная	3077	0,162189
гостиница	80999	4,110977	гостинница	1562	0,082834
девчонка	64756	3,292226	девченка	10908	0,581047
джентльмен	17054	0,866416	джентельмен	7266	0,378944
дилемма	18791	0,954109	диллема	1207	0,063881
достойн	78169	3,967528	достоен	5332	0,2722
естественно	726046	36,845993	естественено	2576	0,136706
извините	299337	15,19358	извените	8458	0,447348
издавна	22490	1,141861	издавно	229	0,012367
иммунитет	56341	2,859795	имунитет	988	0,053804
ингредиент	8438	0,430154	ингридиент	1455	0,076012
искусственный	40300	2,04564	искуственный	4111	0,218978
искусство	333061	16,902622	искуство	3798	0,199472
каллиграфия	3188	0,163064	калиграфия	163	0,009757
калории	19605	0,996862	калории	1228	0,065926
комментарий	267383	13,569584	коментарий	12440	0,654482
коридор	92393	4,689051	корридор	2619	0,134884
кроссовки	73717	3,742795	кросовки	4886	0,258742
линолеум	15050	0,767456	линолиум	1291	0,069066
массаж	193120	9,804579	масаж	3675	0,198524
мировоззрение	47752	2,424191	мировозрение	5032	0,262745
мороженое	109381	5,551944	мороженное	29499	1,504352
непрерывно	271646	13,786102	Неприменно	269	0,015239
обаяние	25753	1,307871	обояние	1149	0,061377

объект	200281	10,164713	объект	3939	0,209174
парадокс	78291	3,973597	пародокс	727	0,039131
параллельно	156257	7,930363	паралельно	10386	0,531831
паталогия	10794	0,54832	паталогия	978	0,052662
педиатр	12018	0,610383	педиатор	367	0,020622
периферия	3332	0,170137	переферия	356	0,01943
пирожное	3586	1,016283	пирожное	1471	0,078239
пожалуйста	2253750	114,381608	пожалуста	14931	0,789195
помощник	111825	5,675115	помошник	4535	0,232254
попробовать	485530	24,640399	попробывать	13436	0,697476
предыстория	21792	1,107951	предистория	3084	0,160953
прецедент	5460	1,469891	прецидент	148	0,008446
программный	10105	0,513498	програмный	411	0,023376
просвещение	17727	0,900282	просвящение	350	0,018813
рассчитать	22763	1,155716	расчитать	4782	0,251449
расчет	91587	4,648216	расчет	2913	0,150372
серебряный	45875	2,328565	серебрянный	4209	0,215487
симпатичный	72751	3,693811	симпотичный	3306	0,177454
соответственно	501911	25,471445	соответсвенно	12602	0,645001
терраса	3763	0,622396	терасса	1365	0,071444
трафик	42231	2,144825	траффик	10845	0,55342
удлинить	3366	0,171639	удлиннить	396	0,02245
участвовать	360644	18,302356	учавствовать	10334	0,544139
чересчур	72044	3,65664	черезчур	10201	0,524903
чрезвычайно	153657	7,798088	черезвычайно	3377	0,175234
чувствовать	389113	19,747183	чувствовать	3668	0,192047
экстремальный	3384	0,522303	экстримальный	1052	0,057477

таблица 1. расчет влияния читательского опыта и частотности верных и неверных написаний на правильность ответов.

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9463	-0.7878	0.3959	0.6153	4.2127

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	0.3557	0.5964
subject	(Intercept)	0.4570	0.6760

Number of obs: 3709, groups: word, 64; subject, 60

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.535910	0.256046	-2.093	0.036347 *
art	0.020600	0.005694	3.618	0.000297 ***
lg_cor	0.029800	0.004741	6.285	3.27e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)	art	
art	-0.647	
lg_cor	-0.600	0.013

таблица 2. расчет влияния читательского опыта и частотности верных и неверных написаний на скорость ответа.

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.1247	-0.7133	-0.2103	0.5400	3.6322

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	44791	211.6
subject	(Intercept)	154209	392.7
Residual		593865	770.6

Number of obs: 3709, groups: word, 64; subject, 60

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	2145.9344	146.9447	115.8393	14.604	<2e-16 ***
art	-0.7949	3.0475	57.1973	-0.261	0.7952
lg_cor	-5.5321	1.8015	60.4716	-3.071	0.0032 **
lg_incor	1.5004	1.8035	60.1000	0.832	0.4087

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	art	lg_cor
art	-0.607		
lg_cor	-0.581	0.000	
lg_incor	-0.581	0.000	0.436

таблица 3. расчет влияния читательского опыта и энтропии на правильность
 ОТВЕТОВ.

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9756	-0.7961	0.3982	0.6094	4.6760

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	0.3508	0.5923
subject	(Intercept)	0.4578	0.6766

Number of obs: 3709, groups: word, 64; subject, 60

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.267997	0.244088	5.195	2.05e-07 ***

art	0.020660	0.005699	3.625	0.000289 ***
entropy	-2.514335	0.398166	-6.315	2.71e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)	art	
art	-0.663	
entropy	-0.544	-0.015

таблица 4. расчет влияния читательского опыта и энтропии на скорость
ответов.

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.1197	-0.7157	-0.2111	0.5360	3.6187

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	50815	225.4
subject	(Intercept)	154288	392.8
Residual		593867	770.6

Number of obs: 3709, groups: word, 64; subject, 60

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	1888.754	117.590	87.486	16.062	<2e-16 ***
art	-0.800	3.048	57.223	-0.262	0.7939
entropy	378.602	145.434	61.517	2.603	0.0116 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)	art	
art	-0.759	
entropy	-0.407	-0.001