

ТЕОРЕТИЧЕСКИЕ И ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

УДК 81'23

ОРФОГРАФИЧЕСКИЕ СОСЕДИ В РУССКОМ ЯЗЫКЕ: БАЗА ДАННЫХ И ЭКСПЕРИМЕНТ, НАПРАВЛЕННЫЙ НА ИЗУЧЕНИЕ МОРФОЛОГИЧЕСКОЙ ДЕКОМПОЗИЦИИ

(Исследование выполнено при поддержке гранта РГНФ №14-04-12034
«База данных и веб-интерфейс, охватывающие важнейшие психолингвистические
характеристики для основного лексического фонда русского языка».)

Алексеева Светлана Владимировна

м. н. с. Лаборатории когнитивных исследований СПбГУ
190000, Санкт-Петербург, ул. Галерная, 58/60
mail@s-alexeeva.ru

Слюсарь Наталия Анатольевна

доцент Школы лингвистики НИУ ВШЭ
105066 г. Москва, ул. Старая Басманная, 21/4
доцент факультета свободных искусств и наук СПбГУ
190000, Санкт-Петербург, ул. Галерная, 58/60
slioussar@gmail.com

В целом ряде психолингвистических экспериментов исследуется влияние на восприятие слова близких по написанию слов, или орфографических соседей (например, *крот – корт*). Обращение к такому материалу позволяет выявить некоторые закономерности восприятия графического облика слов при чтении и поиска слов в ментальном лексиконе, а также пролить свет на ряд других актуальных для психолингвистики проблем. Однако на материале русского языка такого рода экспериментов практически нет – возможно, в силу отсутствия ресурса для автоматизированного подбора соседей, сбалансированных по разным параметрам. В данной работе мы представляем такой ресурс – базу данных StimulStat, а также результаты эксперимента, в котором она была использована.

В ряде экспериментов, изучающих слова-соседи с перестановкой соседних букв, ставился вопрос, какую роль играет то, происходит ли перестановка внутри одной морфемы или на стыке морфем. Это позволяет понять, происходит ли при восприятии многоморфемных слов морфологическая декомпозиция, и если происходит, то на каком этапе. В отличие от большинства более ранних исследований, в проведенном нами эксперименте (с применением методики лексического решения с праймингом) изучаются не производные слова, а словоформы. На основании полученных результатов можно сделать вывод о том, что формы существительных в косвенном падеже не хранятся в памяти целиком и подвергаются декомпозиции (во всяком случае, если речь идет не о самых высокочастотных формах), в то время как начальная форма обладает особым статусом.

Ключевые слова: ментальный лексикон, морфологический анализ, орфографические соседи, словоизменительная морфология, русский язык.

Введение

Изучение восприятия слов при чтении – одно из основных направлений современной психолингвистики (см., например, [Леонтьев 1979; Залевская 2001]). Важнейшая часть этого процесса – анализ буквенного состава слова. В этой области выделяется целая группа исследований, посвященных влиянию на восприятие слова так называемых орфографических соседей (например, [Acha, Carneiras 2014; Davis et al. 2009; Grainger 1992, 2008; Huntsman, Lima 2002]). Этот термин используется для близких по написанию слов, в частности для слов, отличающихся друг от друга на одну букву (например, *крот* – *кот*, *крот* – *кров*) или за счет перестановки соседних букв (например, *крот* – *корт*).

Исследуется, как на доступ к слову в ментальном лексиконе (в частности, на скорость его прочтения) влияет число слов-соседей, их частотность, как активация одного из слов-соседей влияет на доступ к другому, а также ряд других проблем. На основании полученных результатов можно, во-первых, делать выводы о принципах хранения слов в ментальном лексиконе и о механизмах доступа к ним. Во-вторых, слова-соседи используются в экспериментах, направленных на изучение различных вопросов, которые напрямую не связаны с анализом буквенного состава слова. В данной работе мы обратимся к исследованиям, посвященным проблеме морфологического анализа слов при восприятии. В-третьих, учитывая, что значимость факторов, связанных с орфографическим сходством, была подтверждена в целом ряде работ, эти факторы следует принимать во внимание в самых разных психолингвистических экспериментах. Например, число слов-соседей часто контролируется при подборе экспериментальных стимулов наравне с такими факторами, как длина и частотность [White et al. 2008; Whitney, Lavidor 2004 и др.].

В связи с этим для нескольких языков были созданы базы данных, позволяющие подбирать стимулы с определенными характеристиками, например, пары слов-соседей определенного типа, слова, у которых много или мало соседей и т.д. [Boudelaa, Marslen-Wilson 2010; Davis 2005; Davis, Perea 2005; Duchon et al. 2013; New et al. 2004; и др.]. Автоматизация поиска является в данном случае принципиально важной, так как информация такого рода не представлена в каких-либо словарях и корпусах. В данной работе мы представим базу *StimulStat* – первую подобную базу данных, созданную на материале русского языка.

Один из ключевых вопросов, обсуждающихся в психолингвистической литературе, заключается в том, происходит ли при восприятии морфологический анализ многоморфемных слов, и если он имеет место, то на каком этапе обработки и какие факторы могут оказывать на него влияние. Эксперименты на материале орфографических соседей вносят свою лепту в решение этого вопроса. Обращает на себя внимание то, что перестановкам на стыке основы и окончания посвящено всего две опубликованные работы [Luke, Christianson 2013; Zargar, Witzel 2016].

Вероятно, это связано с тем, что германские и романские языки, на материале которых проводится большая часть исследований, не обладают достаточно развитой словоизменительной морфологией, что заметно усложняет подбор стимулов. Между тем, учитывая, что многие авторы по-разному смотрят на морфологический анализ в случае словоизменительной и словообразовательной морфологии, такого рода исследования представляют большой интерес. В данной работе представлен первый эксперимент такого рода, проведенный нами на материале русского языка.

Обзор исследований на материале орфографических соседей, посвященных морфологическому анализу

В настоящее время выделяют четыре основных типа орфографических соседей: соседи с заменой одной буквы (*крот – кров*) [например, Coltheart et al. 1977]; соседи с удалением одной буквы (*крот – рот*) [например, Davis et al. 2009]; соседи со вставкой одной буквы (*кара – карат*) [например, De Moor, Brysbaert 2000]; соседи с перестановкой двух соседних (*баня – баян*) или несоседних (*кот – ток*) букв [например, Andrews 1996]. Слова, являющиеся соседями того или иного типа, образуют орфографическое соседство (пример соседства слов с перестановками: *кот, ток, кто*).

Большая часть экспериментов в этой области использует методику принятия лексического решения с праймингом. Испытуемый должен как можно быстрее определить, является ли предъявленная ему на экране компьютера цепочка букв реальным словом или нет. При этом перед стимульной последовательностью на небольшой промежуток времени показывается другая последовательность (прайм). Например, если прайм совпадает с целевым словом, последнее распознается значительно быстрее, чем в контрольном условии (где стимул и прайм никак не связаны). Это называется эффектом прайминга. Он связан с тем, что ключевое слово уже было активировано в ментальном лексиконе, и чтобы активировать его повторно, требуется меньше времени.

Эксперименты на материале соседей разных типов используются при изучении различных проблем, в частности морфологического анализа слов при восприятии. Существующие модели можно разделить на те, которые рассматривают словоформы [например, McClelland, Patterson 2002; Marslen-Wilson, Tyler 1997; Pinker 1991, 1999; Rumelhart, McClelland 1986; Ullman 2004] и производные слова [Baayen et al. 1997; Pollatsek et al. 2000; Seidenberg, McClelland 1989; Taft, Forster 1975; Taft 2004 и др.]. В каждом случае ставится вопрос о том, хранится ли в ментальном лексиконе сложное слово / форма целиком или же содержащиеся в ней морфемы (в некоторых моделях реализованы обе опции) (см. также [Черниговская 2010]). Одни модели опираются на морфологическую декомпозицию и поморфемный поиск, другие – на поиск целого слова или формы, а третьи предполагают, что могут быть задействованы оба пути, один из которых окажется более эффективным (в зависимости от частотности, семантической прозрачности, морфологической регулярности и ряда других факторов).

Существует несколько способов проверить, происходит ли во время лексического доступа морфологическая декомпозиция. Один из них – сравнить время опознания многоморфемных слов после праймов-псевдослов, полученных из целевых стимулов перестановкой двух букв либо на стыке морфем, либо внутри одной морфемы. Если морфемный анализ является обязательным, то прайминг-эффекта, характерного для слов с перестановкой, в первом случае не будет. Это будет также означать, что информация о морфемных границах становится доступной перед информацией о порядке следования букв в слове или одновременно с ней (см., например, [Zargar, Witzel, 2016]).

Предыдущие исследования с применением данной методики были сфокусированы на словообразовании: на сложных словах (состоящих из двух корней)

[Christianson et al. 2005; Perea, Carreiras 2006], суффиксальных и префиксальных производных [Beysersmann et al. 2013; Duñabeitia et al. 2007; Masserang, Pollatsek 2012; Rueckl, Rimzhim 2011; Sánchez-Gutiérrez, Rastle 2013]. Результаты этих исследований не согласованы. В работах [Christianson et al. 2005; Duñabeitia et al. 2007; Rueckl, Rimzhim 2011] не было обнаружено прайминг-эффекта в условии с перестановкой букв на стыке морфем, а в остальных исследованиях прайминг-эффект наблюдался независимо от условия.

Многие исследователи пытались понять, чем обусловлена разница в полученных результатах. Тестировались разные языки (английский и бакский) [Sánchez-Gutiérrez, Rastle 2013], проверялись влияние семантического сходства между праймом и таргетом [Beysersmann et al. 2013] и индивидуальные особенности читателей [Duñabeitia et al. 2014]. Только в последнем случае была выявлена разница между условиями с перестановкой на стыке и внутри морфемы: у читателей с высокой скоростью чтения, в отличие от «медленных» читателей, снижается размер прайминг-эффекта при перестановке на стыке морфем.

Наше исследование посвящено хранению флективных форм существительных. Насколько нам известно, рассматриваемый нами метод применялся для изучения словоформ только в двух работах на материале английского языка. В первом случае [Luke, Christianson 2013] исследовались регулярные формы прошедшего времени (например, *enjoyed* «наслаждался»), во втором случае [Zargar, Witzel, 2016] – как словоформы, так и производные слова: формы с аффиксами *-ing* (например, *focusing* «фокусирующийся»), *-ed* (например, *stopped* «остановился»), *-est* (например, *largest* «самый большой») и слова на *-er* (например, *cleaner* «очиститель»).

В первом исследовании прайминг-эффект был зафиксирован только при внутриморфемных перестановках в праймах, что свидетельствует в пользу морфологической декомпозиции. Во втором исследовании, напротив, где бы ни располагалась перестановка внутри прайма, прайминг-эффект был значим. Это говорит о том, что морфологическая декомпозиция, если и происходит, то уже после того, как становится доступной информация о порядке следования букв. Объяснить, чем обусловлены различия в полученных результатах, пока сложно – для этого нужны дополнительные исследования. Один эксперимент такого рода представлен в нашей работе.

База данных StimulStat

Как мы видели выше, слова-соседи используются в качестве материала в разнообразных психолингвистических экспериментах. Кроме того, во многих работах число соседей и их частотность контролируется при подборе стимулов. Однако исследований такого типа для русского языка практически нет. Мы полагаем, что во многом это связано с отсутствием ресурса, который позволял бы отчасти автоматизировать трудоемкую задачу поиска соседей, сбалансированных по различным параметрам.

Эта опция была реализована в базе данных *StimulStat*, которую мы разработали в соавторстве с Д.А. Черновой для подбора стимульного материала для экспериментов на русском языке [Алексеева и др. 2015]. Ресурс доступен по адресу *stimul.cognitivestuides.ru*. Он включает более 50 000 лексем и более более 1,9 миллионов образованных от них форм. База содержит параметры, связанные с частотностью словоформ и лемм, буквенным составом (длина в слогах и символах, первая

и последняя буква, обратная запись, позиция однозначной идентификации и др.), просодическими особенностями (место ударения в символах и слогах, слоговая структура, наличие сдвига ударения в парадигме и др.), полисемией и омонимией, а также различными грамматическими характеристиками. Кроме того, в базе представлена информация, связанная с орфографическими соседями разных типов.

Некоторые параметры были взяты нами из различных источников (среди них «Частотный словарь современного русского языка» [Ляшевская, Шаров 2009], проект [Ляшевская 2013], морфологический анализатор *rumorphy2* (<https://rumorphy2.readthedocs.org/en/latest/>), включающий словарь проекта OpenCorpora (opencorpora.org), «Грамматический словарь русского языка» [Зализняк 1987] и др.). В этом случае преимущество базы заключается в возможности автоматического поиска и одновременного учета. Другие параметры были рассчитаны нами при создании базы.

Среди параметров, которые мы рассчитали сами, остановимся подробнее на соседях различных типов. Мы включили в базу (отдельно для лемм и для форм) следующие типы соседей: соседи с заменой (*ток – сок*); соседи с перестановкой (*баян – баня*); соседи с удалением одной буквы (*крот – кот*) и со вставкой одной буквы (*кот – крот*); соседи, у которых второе слово в паре полностью включено в первое (*абориген – бор, ген*), и соседи, у которых первое слово в паре полностью включено во второе (*бор – абориген, забор*); соседи, имеющие в той или иной позиции общую биграмму (т.е. сочетание из двух букв) (*ток – тонка*) или триграмму (*порвать – поручень*). Для каждого отдельного соседства мы подсчитали количество входящих в него слов или форм, их суммарную частотность, натуральный и десятичный логарифм частотности, минимальную и максимальную частотность слов или форм. Также для каждого слова в соседстве известно, какое количество слов превышает по частотности текущее.

База может использоваться в экспериментальных исследованиях русского языка, а также в любых других работах, где требуется подбор примеров с определенными характеристиками.

Экспериментальное исследование

В этом разделе мы представим проведенное нами экспериментальное исследование. Мы использовали методику принятия лексического решения с праймингом, описанную во втором разделе. В качестве стимулов выступают пары словоформ-соседей, которые можно разделить на несколько групп:

два существительных в начальной форме (в именительном падеже единственного числа), перестановка затрагивает только основы (например, *крот – кот*);

два существительных в начальной форме, перестановка на стыке основы и окончания (например, *баян – баня*);

два существительных, одно из которых стоит не в начальной форме, а другое – в начальной (например, *клоун – клоун*).

Из предыдущих исследований известно, что при восприятии слов-соседей наблюдается эффект прайминга: если сперва предъявляется одно слово, а за ним второе, то второе воспринимается значительно быстрее, чем в контрольном условии. Наша гипотеза заключается в том, что степень выраженности этого эффекта будет зависеть от того, затрагивает ли перестановка в словах-соседях только основы или также окончание в одном из слов, а также являются ли оба соседа начальными фор-

мами слова или нет. На основании полученных результатов можно будет сделать выводы о том, осуществляется ли при восприятии форм морфологическая декомпозиция и на каком этапе. В проведении эксперимента принимала участие студентка СПбГУ Анастасия Петрова.

Участники

В эксперименте приняли участие 26 носителей русского языка в возрасте от 18 до 24 лет (15 женщин и 11 мужчин).

Материалы

Стимульный материал был подобран следующим образом. Используя базу данных *StimulStat*, представленную в предыдущем разделе, мы отобрали все пары соседей с перестановкой букв среди словоформ существительных в определенном диапазоне длины и частотности (известно, что эти два фактора играют в распознавании слов при чтении ключевую роль). Длина словоформ варьировала от 4 до 7 букв, а частотность лемм составляла от 0,4 до 111,3 на миллион согласно «Частотному словарю современного русского языка» [Ляшевская, Шаров 2009].

Отобранные пары словоформ мы разделили на две группы, каждая из которых делится на две подгруппы. В первой группе («начальной») обе словоформы-соседа – начальные формы, т.е. стоят в именительном падеже единственного числа. Для нее нам удалось набрать 18 пар слов¹. Эту группу мы разделили на подгруппы 1а («основа-окончание») и 1б («основа-основа»), в которых содержится по девять слов. В подгруппе 1а перестановка букв затрагивает окончание (*волна-волан*), а в подгруппе 1б не затрагивает (*гроб-горб*). Во всех парах прайм частотней целевого слова.

Во второй группе («косвенной») одно из слов-соседей стоит в форме косвенного падежа (эти формы использовались в качестве праймов), а второе – в начальной форме. Нам удалось набрать 36 таких пар. Во всех парах перестановка затрагивает стык основы и окончания. Мы разделили эту группу на подгруппы 2а («прайм>цель») и 2б («цель>прайм»), в которых содержится по 18 слов. В подгруппе 2а прайм частотней целевого слова² (*призам* (22,4) – *призма* (3,7)), а в подгруппе 2б целевое слово частотней прайма (*рифам* (3,5) – *рифма* (8,5)).

Далее следовало подобрать праймы для контрольного условия, а также квазислова (в задании на принятие лексического решения половина стимульных последовательной всегда не является реальными словами) и праймы для них. Поэтому для каждого уже отобранного нами прайма мы нашли по две словоформы, совпадающих с ним по длине и слоговой структуре и близкие по частотности, но не обладающие орфографическим сходством с ним. Пример приведен в Таблице №1. Эти словоформы использовались в качестве праймов в контрольном условии, а также перед квазисловами. Такого рода отбор гарантировал, что различные экспериментальные условия будут отличаться только с точки зрения интересующих нас факторов, связанных с орфографическим соседством.

¹ В целом следует отметить, что словоформ-соседей оказалось меньше, чем мы предполагали изначально.

² Здесь и далее частотность указывается в скобках в количестве упоминаний на миллион согласно словарю [Ляшевская, Шаров 2009].

Таблица №1

Пример набора целевых стимулов и праймов, использованных в эксперименте

Целевое слово	Прайм 1 (эксп. условие для целевого слова)	Прайм 2 (контр. условие для целевого слова и квазислова)	Прайм 3 (эксп. условие для квазислова)	Квази-слово
<i>призма</i> (3,7)	<i>призам</i> <i>приз</i> (22,4)	<i>блинам</i> <i>блин</i> (22,9)	<i>кранам</i> <i>кран</i> (23,3)	<i>кранма</i>

Квазислова были созданы на основании одного из праймов путем перестановки букв, как показано в Таблице №1. В результате пара «прайм 3 – квазислово» была аналогична паре «прайм 1 – реальное слово». Иначе, если бы испытуемые заметили, что орфографическое сходство наблюдается только в парах с реальными словами, это могло бы непредсказуемым образом отразиться на времени реакции. Кроме того, выбранная нами методика гарантировала, что целевые стимулы, являющиеся реальными словами и квазисловами, похожи друг на друга по таким ключевым характеристикам, как длина и слоговой состав.

Таким образом, у нас получилось 54 набора, включающих три различных прайма, целевое слово и квазислово. В Таблице №2 приведены средние значения частотности для целевых слов и праймов в различных группах. Средние значения длины внутри групп совпадают.

Таблица №2

Средние значения частотности в различных группах стимулов

Группа	Пример	Целевое слово	Прайм 1	Прайм 2	Прайм 3
1а «начальная, основа-окончание»	<i>волна – волан</i>	3,8	33,8	32,3	31,8
1б «начальная, основа-основа»	<i>гроб – горб</i>	3,1	13,5	12,9	12,7
2а «косвенная, прайм>цель»	<i>призам – призма</i>	2,9	17,4	16,9	17,2
2б «косвенная, цель>прайм»	<i>рифам – рифма</i>	22,4	2,7	2,7	2,6

Затем мы распределили стимулы по двум экспериментальным протоколам следующим образом. В первом протоколе целевое слово из первого набора предъявлялось после прайма 1 (т.е. в условии соседства), а квазислово – после прайма 2 (т.е. в контрольном условии). Целевое слово из второго набора предъявлялось после прайма 2 (в контрольном условии), а квазислово – после прайма 3 (в условии соседства) и т.д. Во втором протоколе условия были распределены зеркально.

Таким образом, внутри одного протокола праймы не повторялись, а также не было случаев, когда испытуемый видел, например, пару *блинам – призма* и прайм *призам* (который является соседом целевого слова) перед каким-то другим стимулом. Именно поэтому нам требовалось по три прайма в каждом наборе. Каждый испытуемый проходил только один протокол.

После получения экспериментальных данных мы планировали провести следующие сравнения:

- между группой 1 «начальной» (в целом) и 2а «косвенной, прайм>цель» (в обеих группах по 18 слов, прайм в обоих случаях частотнее, чем целевое слово, средняя частотность целевых слов близкая, можно сравнить начальные и неначальные словоформы в роли праймов);
- между группами 2а «косвенной, прайм>цель» и 2б «косвенной, цель>прайм» (обе группы по 18 слов, в обоих случаях прайм – неначальная форма, можно сравнить праймы, которые более частотны или менее частотны, чем целевые слова);
- между группами 1а «начальной, основа-окончание» и 1б «начальной, основа-основа» (обе группы по 9 слов³, средняя частотность целевых слов близкая, в обоих случаях прайм – начальная форма, можно сделать выводы по поводу роли того, затрагивает ли перестановка букв окончание или нет).

Процедура

Эксперимент проводился на персональном компьютере *MacBook* с использованием программы *PsyScope* [Cohen et al. 1993]. Все испытуемые тестировались индивидуально в тихом помещении. Никто из испытуемых не был осведомлен об истинной цели эксперимента. Сперва на экране предьявлялась следующая инструкция:

«Уважаемый участник! Положите указательный палец правой руки на клавишу «J» и указательный палец левой руки на клавишу «F». Перед Вами на экране будет появляться знак «*», цепочка букв и затем другая цепочка букв. Ваша задача: как можно быстрее определить, является ли вторая цепочка букв словом русского языка. Если это слово, нажмите правую кнопку. Если это не слово, нажмите левую кнопку»⁴.

Затем испытуемым предьявлялись восемь тренировочных пар стимулов. После этого у них была возможность задать уточняющие вопросы, если что-либо осталось непонятным. Далее следовал основной блок стимулов, включавший в себя 108 пар «прайм – целевое слово или квазислово».

Сперва испытуемый видел в центре экрана звездочку (точку фиксации). Временной интервал случайным образом выбирался из следующего списка: 1000, 1100, 1200, 1300, 1400 и 1500 мс. Это не позволяло испытуемым сформировать опреде-

³ Из-за того, что количество таких стимулов невелико, мы ожидали получить скорее предварительное представление, на основании которого можно создавать дальнейшие эксперименты, чем статистически значимые результаты.

⁴ Половина испытуемых, проходивших каждый протокол, видела инструкцию в таком виде, другая половина должна была нажимать на левую кнопку, если видит слово.

ленные ожидания относительно того, когда появится прайм. Затем в центре экрана (на том же месте, что и звездочка) предъявлялся прайм на 200 мс. Он на 200 мс сменялся пустым экраном, после чего появлялся целевой стимул. Если испытуемый не нажимал на клавишу в течение 500 мс, целевой стимул сменялся пустым экраном. Это было сделано для того, чтобы испытуемые отвечали как можно быстрее. Если испытуемый не отвечал в течение последующих 1500 мс, на экране снова проявлялась звездочка, а затем – следующие прайм и стимул.

Результаты

Как принято в такого рода исследованиях, мы анализировали количество правильных ответов и время реакции испытуемых. Прежде всего мы удалили данные двух испытуемых, которые допустили ошибки в 20% и 25% экспериментальных заданий. Все остальные участники ошибались менее чем в 15% случаев.

Таким образом, у нас осталось 2592 измерения времени реакции. Из них мы удалили те 243 случая, где испытуемые дали неверный ответ (это составляет 9,4%). Затем, пользуясь процедурой, описанной в работе [Ratcliff 1993], мы рассчитали для каждого экспериментального условия пороговые значения времени реакции (среднее время реакции плюс или минус 2,5 стандартных отклонения) и удалили все данные за пределами этих пороговых значений. В данном эксперименте их оказалось совсем немного – 10 случаев, что составляет 0,4% от общего количества.

Средние значения времени реакции для всех экспериментальных условий в различных группах, а также размер эффекта прайминга (разница между условием соседства и контрольным условием) представлены в Таблице №3. Сначала рассмотрим реальные слова. Можно заметить, что внутри «косвенной» группы (в подгруппах 2а «прайм>цель» и 2b «цель>прайм») эффект прайминга оказался небольшим, в то время как внутри «начальной» группы он выражен довольно ярко. Что касается квазислов, то для них сколько-нибудь заметная разница не наблюдается ни в одной из групп.

Таблица №3

Среднее время реакции испытуемых и размер эффекта прайминга в различных экспериментальных условиях (в мс)

Группа	Стимул / филлер	Контрольное условие	Условие соседства	Размер эффекта прайминга
1а, «начальная, основа-окончание»	реальное слово	755,9	693,6	62,3
1b «начальная, основа-основа»	реальное слово	758,4	683,2	75,2
2а «косвенная, прайм>цель»	реальное слово	774,2	750,7	23,4
2b «косвенная, цель>прайм»	реальное слово	719,4	698,4	21,0
1а «начальная, основа-окончание»	квазислово	780,4	787,8	-7,5
1b «начальная, основа-основа»	квазислово	795,6	787,1	8,5
2а «косвенная, прайм>цель»	квазислово	780,4	784,4	-4,0
2b «косвенная, цель>прайм»	квазислово	819,8	804,5	15,3

Чтобы установить, какие различия между условиями являются статистически значимыми, в программном пакете SPSS был проведен дисперсионный анализ с повторными измерениями (RM ANOVA). Сначала мы сравнили ответы в контрольном условии и в условии соседства внутри каждой группы. Значимая разница была обнаружена для групп 1а «начальная, основа-окончание» ($F(1, 175) = 4,06, p = 0,046$) и 1б «начальная, основа-основа» ($F(1, 173) = 6,33, p = 0,013$). Различия в остальных группах оказались незначимыми (для всех сравнений здесь и далее $p > 0,14$). Таким образом, значимый эффект прайминга в условии соседства наблюдается только в тех группах, где праймом является слово в начальной форме.

Чтобы непосредственно оценить различия между праймами в начальной форме и в формах косвенного падежа, мы сравнили группу 1 «начальную» (в целом) с группой 2а «косвенной, прайм>цель». В обеих группах прайм частотнее целевого слова. Был использован двухфакторный дисперсионный анализ, в качестве факторов рассматривалось наличие прайма-соседа (т.е. условие соседства или контрольное условие) и наличие прайма в начальной форме (т.е. есть группа 1 или 2а). Статистически значимым оказалось влияние первого фактора ($F(1, 690) = 9,31, p = 0,002$), а также взаимодействие между факторами ($F(1, 690) = 6,90, p = 0,009$). Это говорит о том, что эффект прайминга в условии соседства выражен в «начальной» группе значительно сильнее, чем в «косвенной».

Затем мы сравнили между собой «начальные» группы 1а и 1б, чтобы оценить, окажется ли эффект прайминга в условии соседства более значительным, если перестановка букв в словах-соседях затрагивает только основу (как в группе 1б) или основу и окончание (как в группе 1а). Был использован двухфакторный дисперсионный анализ, в качестве факторов рассматривалось наличие прайма-соседа (т.е. условие соседства или контрольное условие) и тип этого соседа (т.е. группа 1а или 1б). Значимым оказался только фактор соседства ($F(1, 348) = 10,21, p = 0,002$), второй фактор и взаимодействие факторов на значимость не вышли. Как мы отмечали выше, вероятно, это связано с небольшим количеством стимулов в этих группах.

Наконец, мы сравнили «косвенные» группы 2а и 2б. Эти две группы были введены в эксперимент, чтобы, если в них будет наблюдаться эффект прайминга в условии соседства, оценить, как на этот эффект влияет соотношение частотности прайма и целевого слова (в группе 2а частотней прайм, в группе 2б – целевое слово). Был использован двухфакторный дисперсионный анализ, в качестве факторов рассматривалось наличие прайма-соседа (т.е. условие соседства или контрольное условие) и соотношение частотностей прайма и целевого слова (т.е. группа 2а «прайм>цель» или 2б «цель>прайм»).

Второй фактор оказался статистически значимым ($F(1, 739) = 12,64, p < 0,001$). Это ожидаемый результат, учитывая, что целевые слова в группе 2б намного частотней, чем в группе 2а. Первый фактор, а также взаимодействие факторов значимости не достигли. Это хорошо соотносится с результатами первых внутригрупповых сравнений, показавших, что эффект прайминга в условии соседства статистически незначим в обеих «косвенных» группах.

Обсуждение полученных результатов и заключение

Как мы кратко продемонстрировали в обзоре литературы, эксперименты на материале орфографических соседей используются при решении целого ряда актуальных для психолингвистики проблем. Чтобы такого рода исследования было легче планировать на материале русского языка, мы включили информацию о соседях разных типов в созданную нами базу данных StimulStat (stimul.cognitivestuides.ru). Проведенный нами эксперимент также планировался с использованием этой базы.

В эксперименте мы ставили своей задачей проверить, хранятся ли падежные формы существительных в ментальном лексиконе целиком или подвергаются при восприятии морфологической декомпозиции. Была использована методика лексического решения с праймингом, в качестве материала были взяты пары словоформ с перестановкой двух соседних букв. Целевое существительное всегда стояло в начальной форме, а прайм – в начальной (*крот – корт*) или в косвенной (*призам – призма*). Сравнив эти две группы, «начальную» и «косвенную», мы выявили, что эффект прайминга в условии соседства (относительно контрольного условия) выражен в первой группе значительно сильнее, чем во второй.

Остановимся сначала на косвенных формах. Отсутствие прайминг-эффекта свидетельствует о том, что они подвергаются морфологической декомпозиции, причем на раннем этапе: либо до определения порядка следования букв в слове, либо – вероятнее – одновременно с этим процессом (считается, что именно в этот момент слово активирует своих орфографических соседей (см., например, [Zargar, Witzel 2016])). Начальные формы, напротив, вызывают прайминг-эффект, причем вне зависимости от того, затрагивает ли перестановка окончание.

Предположить, что косвенные формы подвергаются декомпозиции, а начальные – не подвергаются, было бы странно. Так как последние никак не маркированы в тексте, получается, что нужно проверять каждую форму на предмет того, не является ли она начальной и не хранится ли целиком, и уже затем приступать к декомпозиции. Мы полагаем, что полученные результаты легче всего объяснить в рамках моделей, сочетающих хранение целых форм и декомпозицию: все формы подвергаются декомпозиции, однако некоторые также хранятся целиком (в нашем случае наличие такой репрезентации оказывается ключевым для прайминг-эффекта). К таким моделям можно отнести, например, модель Х. Баайена и соавторов [Baayen et al. 1997] и поздние версии двусистемного подхода [Pinker 1999] (для последнего цельное хранение морфологически регулярных форм допускается скорее в виде исключения). В рамках дальнейших исследований, вероятно, уже с использованием других методов и материалов, необходимо будет ответить на вопрос, с чем связан выявленный нами особый статус начальной формы: с тем, что она, как правило, частотнее всех прочих форм в парадигме или с другими ее характеристиками?

Обратимся теперь к результатам еще одной манипуляции, которая была осуществлена при создании эксперимента: половина праймов в «косвенной» группе была более частотна, чем целевое слово, другая половина – менее частотна. Мы установили, что испытуемым требуется гораздо больше времени, чтобы принять решение, в первом случае, чем во втором. Но соотношение частотностей никак не повлияло на размер прайминг-эффекта, связанного с перестановкой (т.е. взаимодействие фактора частотности и фактора соседства было незначимым, значимым

оказался только первый фактор).

Экспериментов, где в качестве прайма и целевого стимула использовались бы два реальных слова-соседа с перестановкой, немного, в роли прайма чаще выступает квазислово. Остановимся на работах [Andrews 1996] и [Duñabeitia et al. 2009]. С. Эндрюс на материале английского языка показала, что прайминг-эффект в парах соседей с перестановкой был более выражен для высокочастотных целевых слов при низкочастотном прайме, чем для низкочастотных целевых слов при высокочастотном прайме. Она объясняет это тем, что активация высокочастотного прайма происходит быстрее и, соответственно, обработка целевого слова занимает меньше времени. Вероятно, в нашем случае получить похожие результаты помешала морфологическая декомпозиция. Кроме того, Эндрюс использовала задачу называния, а мы – задачу лексического решения. Наконец, важно отметить, что в исследовании Эндрюс прайминг-эффект был отрицательным, то есть сосед с перестановкой замедлял прочтение целевого слова. Исследовательница предположила, что при предъявлении прайма активируется и его сосед, они конкурируют между собой, и это замедляет время обработки (причем более сильное замедление наблюдается, если праймом является менее частотный сосед).

В работе [Duñabeitia et al. 2009] описана серия экспериментов на материале испанского языка. Использовалась методика принятия лексического решения с праймингом. Для соседей с заменой был выявлен отрицательный прайминг-эффект, как и в работе С. Эндрюс, а в случае соседей с перестановкой прайминг не наблюдался (при любом соотношении частотностей прайма и целевого слова). Исследователи полагают, что, с одной стороны, конкуренция на уровне слов должна замедлить восприятие целевого слова после прайма-соседа. С другой стороны, если у прайма и целевого слова одинаковый набор букв, это должно привести к повышенной активации на орфографическом уровне и ускорить восприятие целевого слова. Эти два эффекта уравнивают друг друга, и в результате ни отрицательный, ни положительный прайминг не наблюдается.

Таким образом, результаты нашего эксперимента не совпадают с полученными С. Эндрюс [Andrews 1996], а с результатами работы [Duñabeitia et al. 2009] совпадают лишь частично. Для «начальной» группы нами был зафиксирован положительный прайминг-эффект, для «косвенной» – нулевой эффект, при этом соотношение частотностей прайма и целевого слова не влияет на наличие и размер эффекта. К сожалению, сравнение трех работ затруднено тем, что в них использовалось разное время предъявления праймов и целевых стимулов, разные задания (в случае сопоставления со статьей С. Эндрюс), разные языки. Поэтому для объяснения полученных различий потребуются дополнительные исследования.

Литература

Алексеева С.В., Слюсарь Н.А., Чернова Д.А. StimulStat: база данных, охватывающая различные характеристики слов русского языка, важные для лингвистических и психологических исследований // Материалы 21-ой Международной конференции по компьютерной лингвистике «Диалог». 2015. [Электронный ресурс]. URL: <http://www.dialog-21.ru/digests/dialog2015/materials/pdf/AlexeevaSVSloussarNACHernovaDA.pdf> (дата обращения: 21.07.2016).

Залевская А.А. Текст и его понимание. Тверь: Тверской государственный

университет, 2001. 177 с.

Зализняк А.А. Грамматический словарь русского языка. Словоизменение. 3-е изд. М.: Русский язык, 1987. 880 с.

Леонтьев А.А. Восприятие текста как психологический процесс // Психолингвистическая и лингвистическая природа текста и особенности его восприятия. Киев: Киевский государственный университет, 1979. С. 18–30.

Ляшевская О.Н. Частотный лексико-грамматический словарь: проспект проекта // Компьютерная лингвистика и интеллектуальные технологии. 2013. №. 12. С. 478–489.

Ляшевская О.Н., Шаров С.А. Частотный словарь современного русского языка. М.: Азбуковник, 2009. 1090 с.

Черниговская Т.В. Мозг и язык: врожденные модули или обучающаяся сеть? // Мозг. Фундаментальные и прикладные проблемы. М.: Наука, 2010. С. 117–127.

Acha J., Carreiras M. (2014). Exploring the mental lexicon: A methodological approach to understanding how printed words are represented in our minds. *Mental Lexicon*, vol. 9, 196–231.

Andrews S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language*, vol. 35, 775–800.

Baayen R.H., Dijkstra T., Schreuder R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, vol. 37, 94–117.

Beyersmann E., Duñabeitia J.A., Carreiras M., Coltheart M., Castles A. (2013). Early morphological decomposition of suffixed words: Masked priming evidence with transposed-letter nonword primes. *Applied Psycholinguistics*, vol. 34, 869–892.

Boudelaa S., Marslen-Wilson W.D. (2010). Aralex: A lexical database for Modern Standard Arabic. *Behaviour Research Methods*, vol. 42, 481–487.

Christianson K., Johnson R.L., Rayner K. (2005). Letter transpositions within and across morphemes. *Journal of Experimental Psychology*, vol. 31, 1327–1339.

Cohen J.D., MacWhinney B., Flatt M.R., Provost J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods*, vol. 25, 257–271.

Coltheart M., Davelaar E., Jonasson T., Besner D. (1977). Access to the internal lexicon. In: *Attention and Performance VI*. New York: Academic Press. Pp. 535–555.

Davis C.J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, vol. 37, 65–70.

Davis C.J., Perea M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, vol. 37, 665–671.

Davis C.J., Perea M., Acha J. (2009). Re(de)fining the orthographic neighborhood: The role of addition and deletion neighbors in lexical decision and reading. *Journal of Experimental Psychology*, vol. 35, 1550–1570.

De Moor W., Brysbaert M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research*, vol. 63, 159–162.

Duchon A., Perea M., Sebastián-Gallés N., Martí A., Carreiras M. (2013). EsPal:

one-stop shopping for Spanish word properties, *Behavior Research Methods*, vol. 45, 1246–1258.

Duñabeitia J.A., Perea M., Carreiras M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition*, vol. 105, 691–703.

Duñabeitia J.A., Perea M., Carreiras M. (2009). There is no clam with coats in the calm coast: Delimiting the transposed-letter priming effect. *The Quarterly Journal of Experimental Psychology*, vol. 62, 1930–1947.

Duñabeitia J.A., Perea M., Carreiras M. (2014). Revisiting letter transpositions within and across morphemic boundaries. *Psychonomic Bulletin and Review*, vol. 21, 1557–1575.

Grainger J. (1992). Orthographic Neighborhoods and Visual Word Recognition. In: *Advances in Psychology*. Amsterdam: North-Holland. Pp. 131–146.

Grainger J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, vol. 23, 1–35.

Huntsman L.A., Lima S.D. (2002). Orthographic neighbors and visual word recognition. *Journal of Psycholinguistic Research*, vol. 31, 289–306.

Luke S.G., Christianson K. (2013). The influence of frequency across the time course of morphological processing: Evidence from the transposed-letter effect. *Journal of Cognitive Psychology*, vol. 25, 781–799.

Marslen-Wilson W.D., Tyler L.K. (1997). Dissociating types of mental computation. *Nature*, vol. 387, 592–594.

Masserang K.M., Pollatsek A. (2012). Transposed letter effects in prefixed words: Implications for morphological decomposition. *Journal of Cognitive Psychology*, vol. 24, 476–495.

McClelland J.L., Patterson K. (2002). Rules or connections in past-tense inflections: what does the evidence rule out? *Trends in Cognitive Sciences*, vol. 6, 465–472.

New B., Pallier C., Brysbaert M., Ferrand L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods*, vol. 36, 516–524.

Perea M., Carreiras M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychonomic Bulletin and Review*, vol. 13, 418–422.

Pinker S. (1991). Rules of language. *Science*, vol. 253, 530–535.

Pinker S. (1999). *Words and rules: the ingredients of language*. New York: Harper Collins. 352 p.

Pollatsek A., Hyönä J., Bertram R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology*, vol. 26, 820–833.

Ratcliff R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, vol. 114, 510–532.

Rueckl J.G., Rimzhim A. (2011). On the interaction of letter transpositions and morphemic boundaries. *Language and Cognitive Processes*, vol. 26, 482–508.

Rumelhart D., McClelland J. (1986). On learning the past tenses of English verbs. In: *Parallel distributed processing: explorations in the microstructure of cognition*. Vol. 2. Cambridge, MA: MIT Press. Pp. 216–271.

Sánchez-Gutiérrez C., Rastle K. (2013). Letter transpositions within and across

morphemic boundaries: Is there a cross-language difference? *Psychonomic Bulletin and Review*, vol. 20, 988–996.

Seidenberg M.S., McClelland J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, vol. 96, 523–568.

Taft M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, vol. 57A, 745–765.

Taft M., Forster K.I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, vol. 14, 638–647.

Ullman M.T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, vol. 92, 231–270.

White S.J., Johnson R.L., Liversedge S.P., Rayner K. (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology*, vol. 34, 1261–1276.

Whitney C., Lavidor M. (2004). Why word length only matters in the left visual field. *Neuropsychologia*, vol. 42, 1680–1688.

Zargar E.S., Witzel N. (2016). Transposed-letter priming across inflectional morpheme boundaries. *Journal of Psycholinguistic Research*, to appear.

**ORTHOGRAPHIC NEIGHBORS:
A DATABASE ON RUSSIAN LANGUAGE AND EXPERIMENTAL
STUDIES OF MORPHOLOGICAL DECOMPOSITION**

Svetlana V. Alexeeva

Junior research associate

Laboratory for Cognitive Studies

St. Petersburg State University (St. Petersburg, Russia),

190000, St.Petersburg, Galernaya Street, 58-60

mail@s-alexeeva.ru

Natalia A. Slioussar

Assistant professor

School of Linguistics, HSE (Moscow, Russia)

105066, Moscow, Staraya Basmanaya Street, 21/4

Assistant professor

Faculty of Liberal Arts and Sciences

St. Petersburg State University (St. Petersburg, Russia)

190000, St.Petersburg, Galernaya Street, 58-60

slioussar@gmail.com

The role of orthographic neighbors (e.g. *trail – trial*) in word processing has been discussed in many experimental studies. These studies address the problem of visual word recognition, lexical access in the mental lexicon and a number of other topics that are currently under debate in psycholinguistics. However, there are virtually no studies on

the material of the Russian language – potentially, due to the fact that there is no resource that would allow searching for such words balanced according to different parameters. In the present paper, we present *StimulStat*, a database that solves this problem, and an experimental study based on it.

A number of experiments studying transposition neighbors asked the following question: does it play a role whether the transposition takes place inside one morpheme or across the morphological boundary? This lets us understand whether multimorphemic words are decomposed in processing and, if yes, at what stage. Unlike the majority of earlier studies, our experiment (using lexical decision task with priming) studies inflected forms rather than derived words. The results let us conclude that noun forms in oblique cases are not stored as a whole and are decomposed (at least, when they are not highly frequent), while the initial (nominative singular) form has a special status.

Keywords: mental lexicon, morphological processing, orthographic neighbors, inflectional morphology, Russian.

References

Alexeeva S.V., Slioussar N.A., Chernova D.A. (2015). StimulStat: baza dannyx, oxvatyvajushhaja razlichnye xarakteristiki slov russkogo jazyka, vazhnye dlja lingvisticheskix i psihologicheskix issledovanij [StimulStat: A lexical database for linguistic and psychological research on Russian language] // Materialy 21-oj Mezhdunarodnoj konferencii po komp’juternoj lingvistike «Dialog» [Proceeding of the 21st International conference on computer science “Dialogue”]. 2015, URL: <http://www.dialog-21.ru/digests/dialog2015/materials/pdf/AlexeevaSVSlioussarNACHernovaDA.pdf>

Chernigovskaya T.V. Mozg i jazyk: vrozhdennye moduli ili obuchajushhajasja set’? [Brain and language: innate modules or a learning network?]. In: Mozg. Fundamental’nye i prikladnye problemy [Brain. Fundamental and applied issues]. Moscow: Nauka, 2010. pp. 117–127.

Leont’ev A.A. Vosprijatie teksta kak psixologicheskij process [Text processing as a psychological process]. In: Psixolingvisticheskaja i lingvisticheskaja priroda teksta i osobennosti ego vosprijatija [Psycholinguistic and linguistic nature of the text and characteristics of its processing]. Kiev: Kiev State University, 1979, pp. 18–30.

Lyashevskaya O.N. Chastotnyj leksiko-grammaticeskij slovar’: prospekt proekta [Lexico-grammatical frequency dictionary: a preliminary design]. Komp’juternaja lingvistika i intellektual’nye tehnologii [Computational Linguistics and Intellectual Technologies]. 2013, vol. 12, pp. 478–489.

Lyashevskaya O.N., Sharov S.A. Chastotnyj slovar’ sovremennogo russkogo jazyka [Frequency dictionary of modern Russian]. Moscow: Azbukovnik, 2009. 1090 p.

Zaljevskaja A.A. Tekst i ego ponimanie. [Text and its comprehension]. Tver: Tver State University, 2001. 177 p.

Zaliznjak A.A. Grammaticeskij slovar’ russkogo jazyka. Slovoizmenenie [Grammatical Dictionary of Russian Language. Inflection]. 3rd ed. Moscow: Russkij jazyk, 1987. 880 p.