

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ РУССКИХ ПОТЕНЦИАЛЬНЫХ ГЛАГОЛОВ С ПРИСТАВКАМИ *О-* И *У-**

© 2018

Анна Александровна Эндресен[®],
Лора Алексис ЯндаУниверситет Тромсё — Норвежский арктический университет, Тромсё, Норвегия;
[®]anna.endresen@gmail.com

В статье представлено экспериментальное исследование потенциальных слов русского языка, образованных по двум продуктивным моделям, а именно деадъективных глаголов с приставками *о-* и *у-* типа *овнешнить* и *уконкретить*, *обытовить* и *украсивить*, и др. В частности, проверяются четыре гипотезы о том, определяются ли суждения о приемлемости потенциальных глаголов а) частотностью и статусом слова в языке, б) продуктивностью словообразовательной модели, в) возрастом и г) полом носителя языка. В проведенном эксперименте собраны суждения 121 носителя русского языка о приемлемости таких низкочастотных маргинальных слов в сравнении с высокочастотными глаголами, с одной стороны, и квазиглаголами, с другой стороны. Результаты эксперимента говорят о том, что частота использования слова является не менее важным фактором в оценках потенциальных слов, чем способность распознавать деривационные структуры.

Ключевые слова: глаголы, маргинальные слова, потенциальные слова, приставки, русский язык, суждения о грамматической приемлемости, эксперимент

AN EXPERIMENTAL STUDY OF RUSSIAN POSSIBLE VERBS PREFIXED IN *O-* AND *U-*

Anna Endresen[®], Laura A. JandaUniversitetet i Tromsø: Norges arktiske universitet, Tromsø, Norway; [®]anna.endresen@gmail.com

This article reports on an experimental study of Russian possible words formed via two productive derivational patterns, namely deadjectival verbs prefixed in *o-* and *u-* like *ovnešnit* 'externalize' and *ukonkretit* 'concretize', *obytovit* 'vulgarize' and *ukrasivit* 'make prettier'. We test four hypotheses as to whether acceptability scores of these verbs correlate with a) corpus frequency and word status; b) productivity of derivational pattern; c) speaker's age, and d) speaker's gender. We collected acceptability judgments from 121 native Russian speakers. The study compared low frequency marginal verbs to high frequency verbs on the one hand and nonce verbs on the other hand. Our findings suggest that in perception of possible words, frequency of use is a stronger factor than speakers' ability to unpack morphological patterns.

Keywords: acceptability judgments, experiment, marginal words, possible words, prefixes, Russian, verbs

* Наше исследование проведено при поддержке гранта № 222506 Норвежского совета по научным исследованиям и Университета Тромсё — Норвежского арктического университета. За помощь в проведении эксперимента авторы глубоко признательны учителям МБОУ школы-лицея № 41 г. Ижевска Ксении Александровне Беляковой и Виктории Евгеньевне Нетребиной. Отдельное спасибо членам исследовательской группы CLEAR университета Тромсё за критические замечания к предшествующим версиям статьи.

Введение

В письменной и устной речи носители русского языка нередко образуют новые слова с опорой на существующие в языке продуктивные модели. В русистике подобные слова принято называть потенциальными словами, поскольку возможность их возникновения заложена в языковой системе [Земская 1981: 230; Петрухина 2014: 421]. В качестве примеров часто приводят сингулятивы типа *сахаринка*, образованные по модели *сор* > *соринка*, *пыль* > *пылинка* [Мусатов 2010: 193], или наименования деятеля типа *спрашиватель* и *возражатель*, образованные по аналогии с уже существующими в языке *читатель*, *писатель* и др. [Розенталь, Теленкова 2008: 314]. Подобным же образом по образцу таких привычных глаголов как *одомашнить*, *ускорить*, *объяснить*, *уменьшить* носители русского языка образуют глаголы *обытовить*, *уконкретить*, *оприличить*, *упрозрачить* и мн. др. Так, в [НКРЯ] можно найти следующие употребления потенциальных глаголов *уконкретить* и *овнешнить*:

- (1) *Но тем не менее **уконкретим** технические параметры — для понимающих читателей...* [Компьютеры будут новые (2003)].
- (2) *Фильмы о чудовищах, мутантах и маньяках выполняют <...> и определенную терапевтическую миссию: они позволяют зрителю <...> экстериоризировать, «**овнешнить**» проблемы в форме кинообразов, перенести их из себя в безопасное пространство — на экран...* [А. Каменецкий. США как объект психотерапии (2003)].

Потенциальные слова не фиксируются в словарях. Они одновременно существуют и не существуют. Существуют, потому что их кто-то когда-то (хоть однажды) образовал и использовал в устной речи или письменном тексте. Не существуют, потому что большинство говорящих на русском языке, возможно, никогда не сталкивались с этими конкретными словами, никогда их не слышали и не образовывали сами.

Поскольку потенциальные слова не принадлежат «языковому стандарту», то есть закрепленным в языке словам с устоявшейся традицией употребления, они маргинальны по определению и находятся в промежутке между общепринятым и несуществующим в языке. Несмотря на это, как правило, потенциальные слова оказываются понятными даже тем, кто видит их впервые. Например, знание слова *серьезный*, с одной стороны, а с другой стороны, слов *обогащать*, *округлить*, *освежить*, *осложнить*, мотивированных прилагательными *богатый*, *круглый*, *свежий*, *сложный*, позволяет с легкостью по принципу аналогии определить и значение потенциального глагола *осерьезнить*: ‘сделать серьезным’.

Потенциальные слова важно отличать от неологизмов, то есть слов, которые появились в языке недавно, вошли в основной лексический фонд, стали узуальными и продолжают употребляться [Фельдман 1957: 64; Лопатин 1973; Петрухина 2014: 421], например *перестройка*, *тусовка*, *топ-менеджер*, *ретрит*, *вебинар*, *лайкнуть*, *форвардировать* и т. д. С другой стороны, потенциальные слова отличаются от авторских окказионализмов, которые создаются поэтами и писателями вопреки существующим в языке продуктивным моделям и потому могут обладать затемненной, непонятной без контекста семантикой [Фельдман 1957; Земская 1972; Ханпира 1972: 248; Петрухина 2014: 422], например *магазинус*, *звень*, *стрекозел* и пр. (подробно о вопросах терминологии см. [Захарова 2013]).

Мы предлагаем следующее определение потенциального слова. Это слово, которое засвидетельствовано хотя бы один раз, не имеет устоявшейся традиции употребления и, таким образом, не закреплено в языковой системе. Как правило, это слово, образованное в устной или письменной речи спонтанно, на ходу, намеренно, и потому уместное в рамках определенного контекста. Потенциальное слово создается на основе продуктивного словообразовательного типа, его структура прозрачна, а семантика композициональна, то есть значение целого легко выводится из суммы значений составляющих частей [Земская 1981: 230].

В связи с этими особенностями возникает ряд интересных вопросов. Как потенциальные слова воспринимаются носителями языка: как естественные и «нормальные» русские слова или, скорее, как странные и не существующие в языке? Как носители языка оценивают потенциальные слова? Принимают ли их за «свои» или за «чужие»? Меняется ли отношение к таким словам с возрастом? Играет ли роль в восприятии таких слов гендерный фактор? И, наконец, зависит ли принятие или непринятие потенциальных слов от степени продуктивности словообразовательной модели? Ответы на эти вопросы далеко не тривиальны и до сих пор не найдены. В настоящей статье мы описываем экспериментальное исследование, которое позволило дать предварительные ответы на эти вопросы на материале двух продуктивных деривационных моделей русского языка.

В рамках эксперимента при помощи шкалы Ликерта (см. раздел 2.1) были собраны суждения 121 носителя русского языка о приемлемости¹ 20 потенциальных деадъективных глаголов, зафиксированных в НКРЯ и образованных по двум продуктивным моделям с приставками *о-* и *у-* (например, *овнешнить* и *уонкретить*). Помимо таких низкочастотных маргинальных слов в эксперименте также использовались для сравнения и контроля 20 высокочастотных глаголов и 20 квазиглаголов. Высокочастотные слова прочно закреплены в языке, как правило, зафиксированы в словарях и понятны большинству носителей языка. В этой статье мы будем называть такие слова стандартными. Квазислова соответствуют правилам русской фонотактики, но не имеют никакого закрепленного за ними значения. Примерами квазислов могут быть элементы знаменитой фразы Л. В. Щербы *Глокая кудра штеко будланула бокра и кудрячит бокренка* [Лыков 1976: 21].

Структура статьи следующая. В разделе 1 мы формулируем гипотезы, протестированные в эксперименте, а в разделе 2 опишем его дизайн. Далее мы обратимся к полученным результатам в разделе 3 и подведем итоги в Заключение.

Собранные данные, список предложений со стимульными глаголами, анкета и программа статистического анализа доступны в архиве TROLLing (<http://opendata.uit.no/>) по ссылке <https://opendata.uit.no/dataset.xhtml?persistentId=hdl:10037.1/10256>.

1. Тестируемые гипотезы

Главной целью эксперимента была проверка зависимости суждений информантов о приемлемости потенциальных слов от четырех факторов: а) частотности и статуса слова в языке, б) продуктивности словообразовательной модели, в) возраста и г) пола носителя языка. В связи с этим были сформулированы гипотезы, которые проверялись в эксперименте.

1.1. Частотность и статус слова

В настоящем исследовании нас интересовали три категории слов, которые можно описать в терминах статуса и частотности: стандартные, потенциальные и квазислова.

Центральный вопрос, интересующий нас в данном случае, — как носители языка воспринимают потенциальные слова: как существующие или как не существующие в языке, как слова или как не-слова? Сближаются ли потенциальные слова в восприятии носителей языка со стандартными словами или с квазисловами? Воспринимаются ли три категории слов отлично друг от друга или образуют единый непрерывный континуум? Поэтому первую гипотезу можно сформулировать так:

¹ Хотя метод извлечения суждений о грамматической приемлемости традиционно ассоциируется с генеративным изучением синтаксиса (см. обзор работ в [Федорова 2013]), сегодня он существует и развивается и в других направлениях лингвистики (ср. когнитивные исследования [Dąbrowska 2010; Bermel, Knittl 2012]), и на других уровнях языка (например, морфология, словообразование), о чем скромно свидетельствует наша статья.

Гипотеза 1:

Существует статистически значимое соотношение между статусом слова в языке и его восприятием носителями языка.

1.2. Продуктивность словообразовательной модели

Играет ли роль в восприятии потенциальных слов продуктивность лежащей в их основе словообразовательной модели, а точнее бóльшая или меньшая степень продуктивности? Мы обратились к данному вопросу на примере глаголов с приставками *о-* и *у-*².

Продуктивность комбинирования приставки и суффикса при образовании новых отыменных и деадъективных глаголов отмечалась в русском языке уже в 1970-е гг. Так, В. В. Виноградов [1972: 348] описывал как инновации слова *углубить*, *разбазарить*, *приземлиться*, *уплотнить*, *раскулачить*, *скособочить*, которые на сегодняшний день уже являются частью общего лексического фонда. Другие глаголы, образованные по данной модели (например, *упрозрачить* [Земская 1992: 86], *приохотить*, *распестрить* [Улуханов 1996: 141—145]), так и остались маргинальными. Для нас важно, что из числа всех русских приставок именно *о-* и *у-* наиболее продуктивно используются для образования подобных глагольных дериватов [РГ 1980: § 891, § 902; Мельчук 1998: 384; Townsend 1968: 143].

В исследовании [Endresen 2014: 272—273] на материале НКРЯ было показано, что в образовании новых глаголов со значением ‘придать X-у признак Y’ на основе имен прилагательных приставка *о-* (глаголы типа *опохабить*) является в два раза более частотной и продуктивной³, чем приставка *у-* (глаголы типа *усовременить*). Можно предположить, что потенциальные глаголы, образованные по более продуктивной словообразовательной модели, имеют больше шансов быть воспринятыми как более естественные и приемлемые, чем глаголы, образованные по менее продуктивной модели. В связи с этим можно сформулировать следующую гипотезу:

Гипотеза 2:

Деадъективные потенциальные глаголы с приставкой *о-* воспринимаются носителями языка как более естественные и приемлемые, чем деадъективные потенциальные глаголы с приставкой *у-*.

1.3. Возраст

Третий вопрос, на который мы хотели получить ответ при проведении эксперимента: зависят ли оценки потенциальных слов от возраста информанта? При этом нас интересовали две возрастные группы: школьники в возрасте 14—17 лет и взрослые от 25 лет и старше. Мы ожидали, что разделение информантов на две указанные возрастные категории позволит определить, изменяются ли предпочтения носителей языка в отношении исследуемых глаголов между периодом старшего школьного возраста и периодом профессиональной деятельности.

Поскольку в подростковом возрасте активно используется молодежный сленг, а процесс усвоения литературных норм еще полностью не завершен [Доценко 2000: 79; Eisikovits 1998: 42], мы предположили, что старшие школьники могут быть более открыты для незнакомых

² Для деадъективных глаголов с общей семантикой ‘придать X-у признак Y’ в русскоязычных работах до сих пор нет общепринятого термина. Следуя англоязычной литературе, такие глаголы называют «фактивными» [Townsend 1968: 143; Эндресен 2013: 38, 47] либо «глаголами изменения состояния» [Levin 1993: 240; Endresen, Janda 2016].

³ Продуктивность здесь и далее понимается как способность приставки быть использованной для образования новых слов [Langacker 1999: 114]. Обзор работ по продуктивности см. в [Barddal 2008], а о способах измерения продуктивности см. [Baayen, Lieber 1991].

новых и потенциальных слов и, возможно, склонны давать им более либеральные оценки, чем взрослые, чьи языковые предпочтения уже сформировались. Приблизительным порогом взрослости в этом отношении был выбран возраст 25 лет⁴, к которому в России большинство людей уже имеет законченное образование и выходит на рынок труда, а сленговые слова перестают определять языковую самоидентификацию. В связи с этим можно сформулировать гипотезу 3.

Гипотеза 3:

Взрослые склонны демонстрировать более консервативные оценки в отношении потенциальных глаголов, чем школьники.

1.4. Пол

Задания, предложенные в эксперименте, апеллировали к языковой интуиции носителей языка, которая может опираться как на представления о языковом стандарте, так и на индивидуальную свободу в использовании языковых средств. Поскольку предпочтения стандартных либо нестандартных форм связаны с половыми различиями⁵, пол может быть важным фактором при оценке потенциальных слов. Поэтому в эксперименте мы проверяли гипотезу о роли пола, сформулированную ниже.

Гипотеза 4:

Пол носителя языка играет существенную роль в распределении оценок приемлемости. Информанты женского пола дают более консервативные (низкие) оценки потенциальным словам, чем информанты мужского пола.

2. Дизайн эксперимента

2.1. Шкала Ликерта

Эксперимент состоял из 60 заданий, в каждом из которых участнику предлагалось прочитать предложение и оценить выделенный в нем глагол по пятибалльной шкале, выбрав одну из рубрик. Экспериментальное задание выглядело следующим образом:

Давно пора как-то оприличить наше общение более мягкими выражениями.

- 5 баллов: Это совершенно нормальное слово русского языка.
- 4 балла: Это слово нормальное, но его мало используют.
- 3 балла: Это слово звучит странно, но, может быть, его кто-то использует.
- 2 балла: Это слово звучит странно, и его вряд ли кто-то использует.
- 1 балл: Этого слова в русском языке нет.

⁴ Порог 25 лет для разграничения консервативных и либеральных языковых приоритетов носителей языка отмечается и в других исследованиях, см., например, [Nilsen Márkus 2015: 46, 231] в отношении венгерского языка. На материале русского языка порог 25 лет использовался при разделении возрастных групп в психолингвистических экспериментах, описанных в диссертации [Горошко 1996].

⁵ В классической работе В. Лабова [Labov 1972: 301] на материале английского языка было показано, что в официальных ситуациях женщины стараются придерживаться нормативных, социально более престижных стандартов и поэтому более консервативны при выборе языковых средств, чем мужчины (см. аналогичные выводы также в [Wolfram 1969; Trudgill 1972]). На материале русского языка речевая дифференциация мужчин и женщин была описана в работах [Земская и др. 1993; Горошко 1996; Yokoyma 1999; Andrews 2003]. Вопрос о том, насколько половая дифференциация актуальна для речи подростков, изучен мало. В этом отношении особый интерес представляют исследования [Доценко 2000; Eisikovits 1998].

Предложенная шкала оценок построена по принципу шкалы Ликерта⁶ [Likert 1932], но совмещает количественное измерение (оценки в баллах) с качественными характеристиками (оценочные суждения). Количество баллов и сопровождающее его суждение отражает степень приемлемости интересующего нас глагола для носителя русского языка. Выбранный метод имеет важное преимущество: сопровождение баллов оценочными суждениями позволяет обеспечить условия для единообразной интерпретации баллов испытуемыми.

Пятибалльная шкала оценок также выбрана неслучайно: она широко используется в российской системе образования и хорошо знакома всем россиянам. Один балл представляет собой наихудшую оценку, тогда как пять баллов соответствуют наивысшей оценке. Поэтому в нашем эксперименте один балл получает слово, которого «нет в русском языке», а пятью баллами оценивается «совершенно нормальное слово русского языка». Промежуточные деления на шкале сформулированы таким образом, чтобы по возможности создать относительно равные интервалы между 5 и 4, 4 и 3, 3 и 2, 2 и 1 баллами. Кроме того, предложенные формулировки отсылают участников к использованию русского языка в целом, а не к языковой практике отдельных индивидов (ср. неопределенно-личные конструкции и выбор местоимений). Вертикальная шкала подчеркивает возрастание баллов и улучшение оценок и основана на когнитивном иконическом соответствии «выше = лучше» [Lakoff, Johnson 1980: 16].

2.2. Стимулы

В эксперимент были включены 20 стандартных слов русского языка, 20 потенциальных и 20 квазислов. В каждой двадцатке 10 слов имели приставку *о-* и 10 слов — приставку *у-*. Равное количество стимулов в каждой группе предлагалось с целью сбалансировать состав стимулов в соответствии с требованиями к подобного рода исследованиям [Schütze 1996: 184]. Стандартные единицы и квазислова выполняли роль контрольных стимулов, позволяющих определить, понял ли информант задание, а потенциальные слова-стимулы были тестируемыми единицами.

Все стандартные и потенциальные стимулы были деадъективными глаголами, не имеющими параллельных глагольных мотивирующих основ. Например, глагол *объяснить*⁷, образованный от прилагательного *ясный* при отсутствии глагола **яснить*, или дериват *усложнить*, образованный от прилагательного *сложный* при отсутствии бесприставочного глагола **сложнить*. Семантика каждого стимула была типичной для глаголов изменения состояния и может быть в обобщенном виде представлена как ‘придать X-у признак Y’. Каждый из выбранных глаголов образован при помощи приставочно-суффиксального способа [РГ 1980 § 202], имея комбинацию приставки *о-* или *у-* и суффикса *-и-*. Квазислова по форме (начальные *о-* и *у-*, конечное *-ить*) были похожи на существующие в русском языке слова. Они были образованы в соответствии с особенностями русской фонотактики и ранее уже использовались в эксперименте, описанном в работах [Эндресен 2013; Vaayen et al. 2013].

⁶ Шкала Ликерта и подобные ей шкалы широко применяются не только в лингвистических экспериментах [Глазанова 2000; 2001; Dąbrowska 2010; Bermel, Knittl 2012], но и в социологических опросах, психологических исследованиях и др. [Lavrakas 2008: 429]. В нашем случае методика шкалирования позволяет обратиться к градуальной природе языковой интуиции, отмечаемой во многих современных работах [Bard et al. 1996; Keller 2000].

⁷ Приставки *о-* и *об-* мы считаем вариантами единой морфемы в соответствии с исследованиями [Эндресен 2013; Endresen 2014: 102—151]. Заметим, что алломорфия *о-* и *об-* в рамках словообразовательного типа деадъективных глаголов с семантикой изменения состояния не подвергалась сомнению даже сторонниками идеи семантического расхождения *о-* и *об-* [Кронгауз 1998: 131—148]. В частности, вариант *об-*, как правило, появляется (хотя и не без исключений) перед начальным сонорным согласным основы [Кронгауз 1998: 141—143; Эндресен 2013: 39], отсюда *об-* в глаголах *объяснить* и *облегчить*. Эта тенденция учтена нами и в квазиглаголах *обномить* и *обмомлить*.

Чтобы исключить возможность влияния дополнительных факторов, все стимульные глаголы были даны в форме инфинитивов совершенного вида. Контексты для стандартных и потенциальных глаголов были подобраны из НКРЯ. В редких случаях более ясный контекст употребления потенциального глагола был найден в поисковых системах www.yandex.ru и www.google.ru. Ниже мы приводим стандартные слова в использованных в эксперименте предложениях (3) и (4), потенциальные слова в примерах (5) и (6) и квазислова в примерах (7) и (8).

Общепринятые стандартные слова:

- (3) *С помощью этой очищающей маски можно легко увлажнить кожу и **освежить** цвет лица.*
 (4) *Новая система позволяет **ускорить** доставку грузов и **повысить** безопасность персонала.*

Потенциальные слова:

- (5) *Голливуд ухитрился **опохобить** почти все шедевры литературы.*
 (6) ***Усовершенить** архитектуру в городах России можно было бы путем сноса ветхих домов в центре многих городов.*

Квазислова:

- (7) *Новые компьютерные игры могут **отовить** любого: на это работает и графика, и цветовая гамма, и сюжет.*
 (8) *Если вы возьмете с собой собаку, придется **укампить** стоимость экскурсии на 40 рублей.*

Все стандартные и потенциальные глаголы-стимулы засвидетельствованы в НКРЯ. Все стандартные глаголы, использованные в эксперименте, являются высокочастотными по данным НКРЯ и приведены в таблице 1. Глаголы приведены в порядке убывания количества примеров в корпусе, частотности даны по современному подкорпусу НКРЯ, т. е. текстам, созданным в 1950—2012 гг. (дата обращения: сентябрь 2012 г.).

Таблица 1

Высокочастотные в НКРЯ русские деадъективные глаголы
(группа стимулов I, стандартные слова)

Глаголы с приставкой <i>о-</i>	Примеры	Глаголы с приставкой <i>у-</i>	Примеры
<i>объяснить</i>	18 149	<i>уточнить</i>	2860
<i>облегчить</i>	1 802	<i>уменьшить</i>	2010
<i>ослабить</i>	1 401	<i>ускорить</i>	2008
<i>округлить</i>	939	<i>улучшить</i>	1899
<i>обогащать</i>	800	<i>упростить</i>	1350
<i>ожесточить</i>	686	<i>укоротить</i>	787
<i>осложнить</i>	410	<i>усложнить</i>	311
<i>оголеть</i>	387	<i>утеплить</i> ⁸	205
<i>осчастливить</i>	343	<i>уплотнить</i>	201
<i>освежить</i>	280	<i>ухудшить</i> ⁹	199

⁸ Бесприставочный глагол *теплить* фиксируется некоторыми словарями, но, судя по современному подкорпусу НКРЯ, употребляется крайне редко (10 вхождений), поэтому мы не считаем его главной мотивирующей основой для глагола *утеплить*.

⁹ Несмотря на то, что глагол *ухудшить* имел частотность чуть ниже 200 употреблений в современном подкорпусе НКРЯ, в рамках эксперимента он был включен в состав высокочастотных глагольных стимулов.

В таблице 2 приведены все потенциальные глаголы-стимулы в порядке возрастания количества вхождений от 1 до 8 по данным НКРЯ. Несмотря на низкую частотность, и структура, и семантика каждого потенциального глагола абсолютно прозрачна, а связь с мотивирующим прилагательным очевидна: *омеждународить* < *международный*, *увкуснить* < *вкусный* и т. д.

Таблица 2

**Низкочастотные в НКРЯ русские деадъективные глаголы
(группа стимулов II, потенциальные слова)**

Глаголы с приставкой <i>о-</i>	Примеры	Глаголы с приставкой <i>у-</i>	Примеры
<i>омеждународить</i>	1	<i>увкуснить</i>	1
<i>опохабить</i>	1	<i>умедлить</i>	1
<i>оприличить</i>	1	<i>украсивить</i>	1
<i>осерьезнить</i>	1	<i>усерьезнить</i>	1
<i>остеклянить</i>	1	<i>уконкретить</i>	1
<i>оржавить</i>	2	<i>усовременить</i>	1
<i>осуровить</i>	2	<i>устрожить</i>	3
<i>обытовить</i>	3	<i>уцеломудрить</i>	3
<i>овнешнить</i>	4	<i>упрозрачить</i>	4
<i>омузыкалить</i>	4	<i>удорожить</i>	8

В таблице 3 приводится список всех использованных квазислов.

Таблица 3

**Не существующие в русском языке глаголы
(группа стимулов III, квазислова)**

О-глагол	У-глагол
<i>осурить, отовить, одуктить, огабить, окончить, ошаклить, очавить, облусить, обномить, обмомлить</i>	<i>усаглить, утулить, удамлить, угузвить, укампить, ушадрить, учопить, улоприть, унокрить, умарвить</i>

Таким образом, три категории глаголов, представленные в эксперименте, различаются их статусом в русском языке и частотой употребления.

2.3. Процедура проведения эксперимента

При выполнении заданий участники эксперимента не были ограничены во времени. В среднем на заполнение анкеты затрачивалось 20 минут. Дети заполняли анкету на бумаге, а взрослые — электронную анкету, которая была создана в программе <http://www.questionpro.com> и повторяла дизайн бумажной анкеты.

Перед заполнением анкеты каждый участник должен был ознакомиться с инструкцией, в которой объяснялась суть заданий, шкала оценок, и, кроме прочего, сообщалось, что в эксперименте представлены существующие и не существующие в русском языке слова.

2.4. Участники

Мы проанализировали ответы 121 участника, исключив из рассмотрения информантов «промежуточного» возраста (от 18 до 24 лет) и информантов, проживающих за пределами России. Все оставшиеся участники эксперимента — носители русского языка, выросшие, получившие образование и проживающие в России. В их числе 70 школьников в возрасте 14—17 лет и 51 взрослый в возрасте 25—62 лет. Среди участников 47 информантов мужского пола и 74 информанта женского пола, равномерно распределенных по двум возрастным группам.

3. Результаты эксперимента

3.1. Общая картина распределения данных

В ходе эксперимента было собрано 7260 ответов. При обработке результатов мы подсчитывали количество различных типов ответов, а также общее количество баллов, набранных каждым словом.

Диаграммы 1—4 позволяют наглядно представить распределение зависимой переменной (оценок приемлемости стимульных глаголов) в соответствии с четырьмя независимыми переменными, взятыми по отдельности, а именно статусом слова-стимула в языке, приставкой, возрастом информанта и полом информанта.

Distribution of scores across word categories

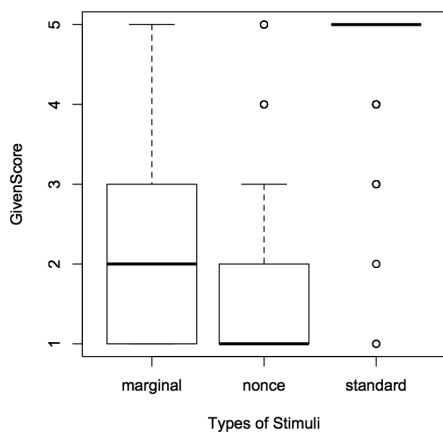


Диаграмма 1. Влияние статуса слова.

Distribution of scores across prefixes

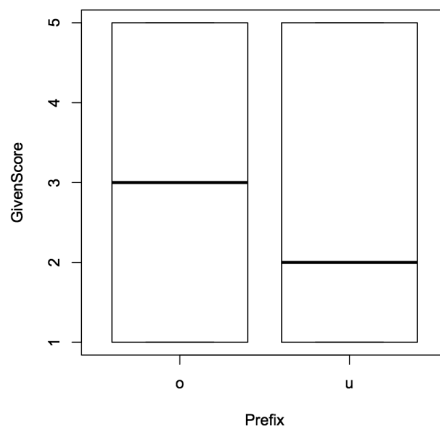


Диаграмма 2. Влияние приставки.

Distribution of scores across age groups

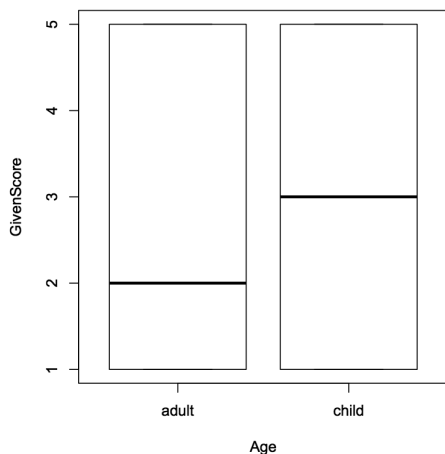


Диаграмма 3. Влияние возраста.

Distribution of scores across genders

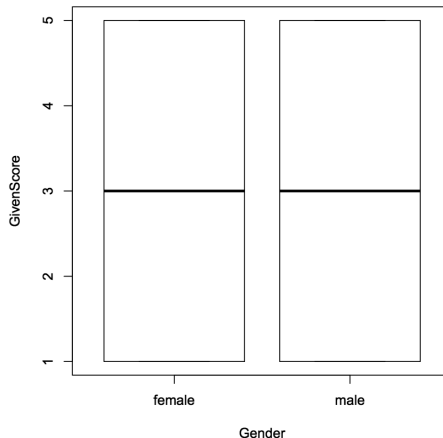


Диаграмма 4. Влияние пола.

На каждой диаграмме вся совокупность собранных данных расположена вдоль вертикальной оси, где жирная линия показывает медиану (срединное значение). Ровно половина данных лежит выше значения медианы, и ровно половина данных располагается ниже значения медианы. Так, сравнивая медианы распределения оценок, можно составить общее представление о соотношении каждого из четырех факторов с собранными оценками стимульных глаголов. Границы прямоугольников определяются по значениям 1-го и 3-го квартилей, при этом нижняя граница прямоугольника показывает 25-й процентиль, а верхняя граница прямоугольника соответствует 75-му. На диаграммах 2, 3 и 4 вся совокупность данных представлена внутри прямоугольников, поскольку дистрибуция данных по этим факторам очень широкая, или рассеянная (от 1 до 5). Диаграмма 1 отличается тем, что дистрибуция данных относительно типов стимулов здесь более узкая, или плотная (для стандартных глаголов она практически сжата в одну линию).

Диаграмма 1 говорит о том, что все три типа слов (стандартные, потенциальные и квазислова) демонстрируют отличные друг от друга типы поведения. В целом потенциальные глаголы получили неожиданно низкие оценки: половина потенциальных глаголов получили самые низкие оценки в 1 и 2 балла. Кроме жирных линий медиан на диаграмме 1 мы также видим пунктирные «усы» и концы усов, обозначенные короткими горизонтальными линиями. Концы усов показывают нам те данные (ответы информантов), которые отклоняются от 1-го и 3-го квартилей, обозначенных прямоугольниками, не больше, чем на полтора межквартильных размаха ($1,5 * IQR$) [Г. Фалин, А. Фалин 2011: 9]. Для потенциальных глаголов таким значением является оценка в 5 баллов, а для квазислов — оценка в 3 балла. Кроме того, маленькие кружки на диаграмме 1 обозначают так называемые «выбросы» (аномальные значения) — данные, выходящие за границы усов. Выбросы показывают, что в оценках некоторых информантов некоторые квазислова получили 4 и даже 5 баллов, а некоторые стандартные слова, наоборот, получили всего 1, 2, 3 или 4 балла. Однако такие оценки являются редкими исключениями, тогда как основной состав данных (50% данных, а именно 2-й и 3-й квартили) обозначен прямоугольниками, а центральные тенденции — медианами.

Диаграмма 2 показывает, что глаголы с приставкой *o-* в целом имеют тенденцию получать больше баллов, чем глаголы с приставкой *y-*: половина *o-*глаголов получили оценки выше 3 баллов, тогда как половина *y-*глаголов получили оценки выше 2 баллов. Согласно диаграмме 3, школьники в целом дают стимульным глаголам более высокие оценки, чем взрослые. Как видно на диаграмме 4, пол информанта не играет никакой роли в распределении оценок приемлемости слов.

3.2. Вариативность и противоречивость в оценках потенциальных глаголов

В предыдущем разделе мы анализировали всю совокупность данных целиком (диаграммы 1—4). Теперь рассмотрим три тестируемые группы слов по отдельности.

На диаграммах 5—7 видна степень вариативности в оценках стандартных, потенциальных и квазиглаголов, различная для этих групп слов. На каждой диаграмме вертикальная ось показывает общее число баллов, полученных каждым стимульным глаголом в ходе эксперимента со взрослыми и детьми.

Чтобы сопоставить «успех» различных глаголов среди взрослых и детей, а также сравнить «успех» стимулов из разных групп, общее число баллов по каждому глаголу дано в процентном соотношении от максимально возможного количества баллов, которое одинаково для всех стимульных глаголов. Любой стимульный глагол мог получить максимум 350 баллов среди детей ($= 5 \text{ баллов} * 70 \text{ школьников}$) и 255 баллов среди взрослых ($= 5 \text{ баллов} * 51 \text{ взрослый}$). Многие стандартные глаголы получили максимальный процент баллов, как показывает диаграмма 5, в то время как многие потенциальные глаголы получили менее 50% от максимального числа баллов (см. диаграмму 6). Более того, столбцы черного и серого цвета на диаграммах показывают, что оценки, данные одним

и тем же словам взрослыми и детьми, далеко не всегда одинаковы. Особенно различия между двумя возрастными группами дают о себе знать в оценивании потенциальных (диаграмма 6) и квазиглаголов (диаграмма 7).

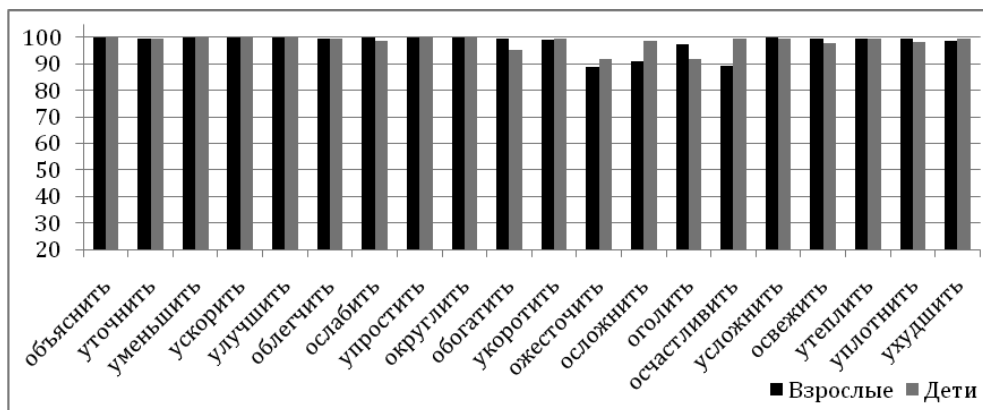


Диаграмма 5. Вариативность в оценках стандартных глаголов.

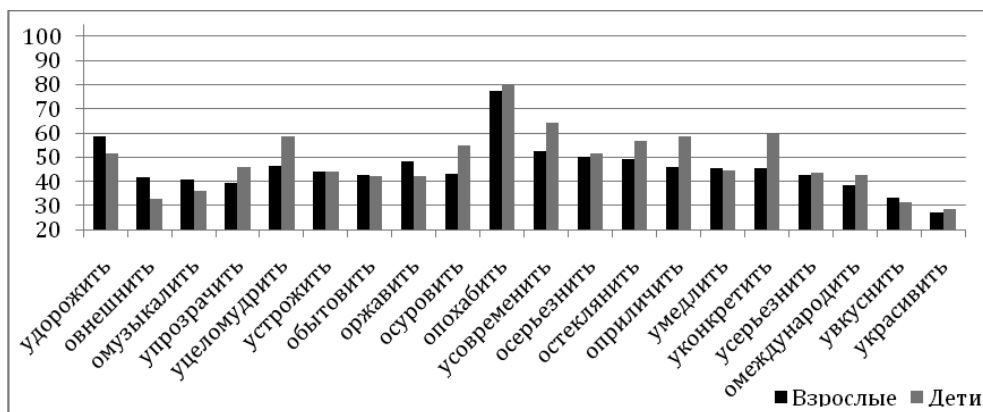


Диаграмма 6. Вариативность в оценках потенциальных глаголов.

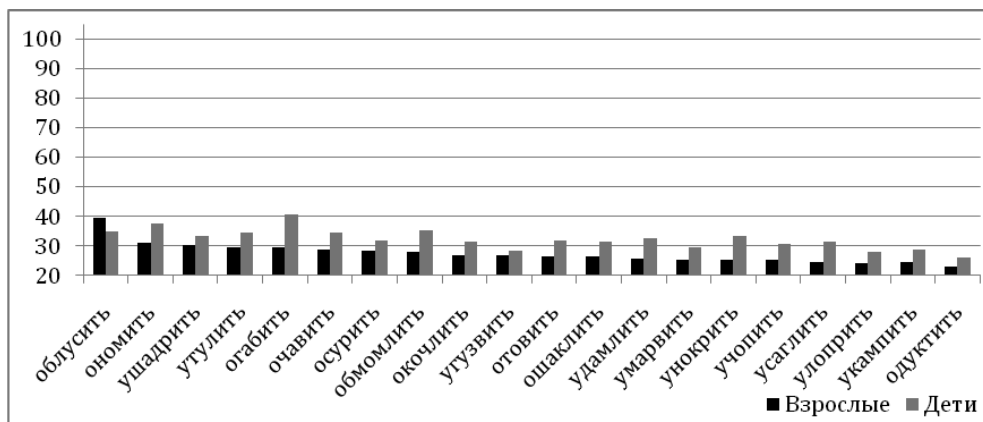


Диаграмма 7. Вариативность в оценках квазиглаголов.

В целом можно заметить, что диапазон, в котором варьируются баллы, полученные стимулами из трех групп, максимален для потенциальных глаголов (диаграмма 6), и это является особенностью этой группы стимулов.

Благодаря данным, приведенным в таблице 4, можно сравнить общее число баллов, полученное максимально «успешным» глаголом (MAX) и минимально «успешным» глаголом (MIN) в каждой группе стимулов. Это позволяет подсчитать, насколько различны диапазоны вариативности внутри каждой группы стимулов для двух возрастных групп информантов.

Таблица 4

**Диапазон вариативности общего числа баллов в трех группах стимулов
для двух возрастных групп информантов**

Информанты	Стандартные глаголы	Потенциальные глаголы	Квазиглаголы
Взрослые (N = 51)	MAX 100 % (<i>объяснить</i>) – MIN 89 % (<i>ожесточить</i>) = 11 %	MAX 77 % (<i>опохабить</i>) – MIN 27 % (<i>украсивить</i>) = 50 %	MAX 39 % (<i>облусить</i>) – MIN 23 % (<i>одуктить</i>) = 16 %
Дети (N = 70)	MAX 100 % (<i>объяснить</i>) – MIN 92 % (<i>оголить</i>) = 8 %	MAX 80 % (<i>опохабить</i>) – MIN 29 % (<i>украсивить</i>) = 51 %	MAX 41 % (<i>огабить</i>) – MIN 26 % (<i>одуктить</i>) = 15 %

Таблица 4 показывает, что диапазон вариативности равен 8—11 % для стандартных глаголов, 15—16 % для квазиглаголов и 50—51 % для потенциальных глаголов. Иными словами, для обеих возрастных категорий информантов потенциальные глаголы образуют наиболее неоднородную группу стимулов, в которой отдельные глаголы сильно отличаются друг от друга по общему количеству полученных баллов.

Более детально вариативность оценок, типичную для потенциальных глаголов, можно проследить на уровне разных типов ответов. В частности, если в восприятии стандартных и квазислов информанты в целом давали схожие суждения, то в восприятии отдельных потенциальных слов мнения сильно расходились. Так, например, как показано в таблице 5, глагол *усовременить* 28 человек посчитали несуществующим в русском языке (1 балл), а для 22 испытуемых он оказался совершенно нормальным словом русского языка (5 баллов). Кроме того, многие оценили этот глагол в промежутке между этими полярными суждениями: 18 человек решили, что глагол *усовременить* звучит странно и его вряд ли используют (2 балла); 27 человек оценили этот глагол как странный, но возможный в использовании (3 балла); а 26 участников назвали его нормальным, но малоиспользуемым (4 балла). Похожая ситуация с сильным расхождением во мнениях наблюдается и в отношении потенциального глагола *оприличить*, приведенного в таблице 5, а также многих других потенциальных глаголов.

Таблица 5

**Вариативность в оценках потенциальных глаголов на уровне
отдельных информантов: противоречивые суждения**

Маргинальные глаголы	Количество информантов, оценивших глагол				
	5 баллов	4 балла	3 балла	2 балла	1 балл
<i>усовременить</i>	22	26	27	18	28
<i>оприличить</i>	9	25	33	22	31

Можно заключить, что вариативность в оценках потенциальных глаголов существует как на уровне индивидуальных лексем, так и на уровне отдельных информантов. В целом, судя по оценкам информантов, потенциальные глаголы составляют наиболее разнородную группу стимулов (несмотря на почти одинаково низкую частотность в НКРЯ), сильно отличаясь в этом отношении и от стандартных слов, и от квазислов.

3.3. Статистический анализ взаимодействия факторов

Чтобы определить, какие факторы играют первостепенную роль в распределении суждений информантов, мы проанализировали собранные данные при помощи метода классификационных деревьев¹⁰.

Деревья классификации относятся к непараметрическим нелинейным статистическим методам¹¹ и позволяют учитывать сложные многоуровневые взаимодействия многочисленных факторов [Ваауен 2008: 154]. В нашем случае использование именно этого метода мотивировано характером зависимой переменной (суждений информантов о потенциальных глаголах), категории которой упорядочены от меньшего к большему, но не обязательно образуют шкалу с равными интервалами¹². Результатом анализа собранных данных является диаграмма 8, на которой представлено дерево классификации.

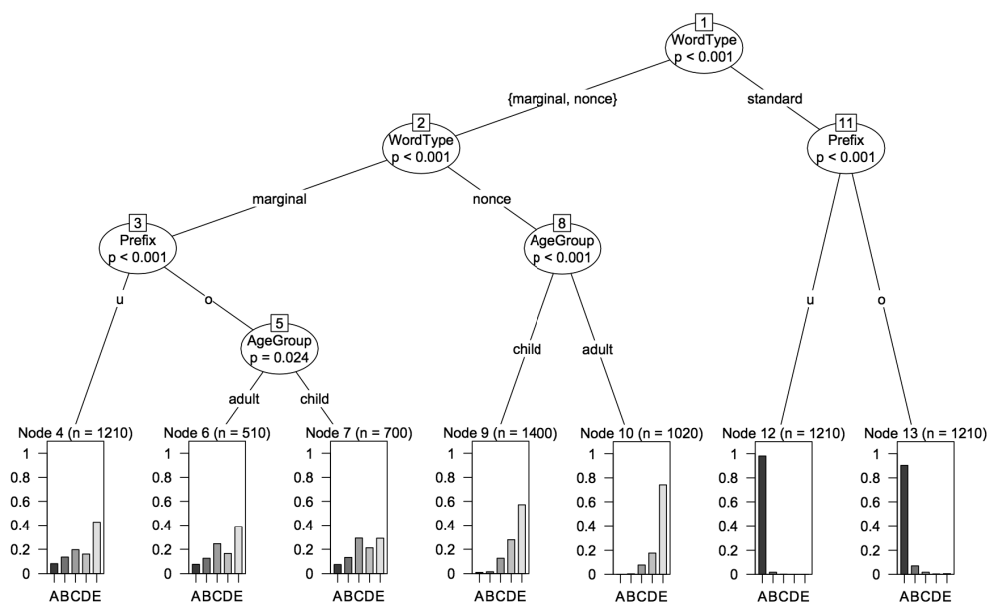


Диаграмма 8. Дерево классификации ответов информантов.

Мы включили в статистический анализ все четыре фактора, используя формулу $Score \sim WordType + Prefix + AgeGroup + Gender$, где слева от тильды располагается зависимая переменная $Score$ (оценка потенциального глагола в баллах), а справа следуют независимые переменные (предикторы): $WordType$ (статус слова), $Prefix$ (приставка), $AgeGroup$ (возраст информанта) и $Gender$ (пол информанта). Последний фактор, пол информанта, оказался статистически незначимым и поэтому не представлен на дереве классификации.

Дерево классификации на диаграмме 8 является графическим представлением наиболее оптимального алгоритма сортировки данных, который учитывает заданные тестируемые

¹⁰ Этот метод сравнительно новый, но быстро приобретает широкую популярность в различных научных дисциплинах — генетике, медицине, социологии, лингвистике — благодаря своим многочисленным достоинствам [Strobl et al. 2009: 324; Tagliamonte, Waayen 2012; Waayen et al. 2013].

¹¹ Непараметрические методы не требуют нормального распределения данных. Нелинейные методы не исходят из того, что зависимость между переменными является линейной.

¹² Подробнее о применимости и результатах других статистических методов в данном исследовании см. [Endresen, Janda 2016]. Приведенный здесь анализ выполнен в программе R версии 2.15.0 [R Core Team 2015].

факторы (предикторные переменные: приставка, статус слова, пол, возраст) и позволяет предсказать ответ информанта (зависимую переменную) в каждом конкретном случае.

Рассмотрим подробнее, как устроено дерево на диаграмме 8. Графическое изображение дерева позволяет пройти по его ветвям от корневого узла (1) к семи терминальным вершинам, или «листьям», — узлам 4, 6, 7, 9, 10, 12, 13. В листьях дерева сосредоточены категории зависимой переменной, т. е. суждения информантов. При этом в каждом узле указывается количество ответов (наблюдений) — величина *n*. Так, например, узел 4 группирует 1210 наблюдений, а узел 6 представляет 510 ответов и т. д. Сами суждения информантов, данные ими в соответствии с пятибалльной шкалой оценок, мы обозначили с помощью букв, чтобы отразить их категориальный статус: А — 5 баллов, В — 4 балла, С — 3 балла, D — 2 балла и Е — 1 балл. Вертикальная шкала, приведенная для каждой терминальной вершины, в виде столбцов показывает процентное соотношение ответов каждого типа. Так, например, в узле 12 ответы с оценками А (5 баллов) составляют почти 100% из 1210 наблюдений, а в узле 13 преобладание оценок А несколько менее выражено, поскольку около 10% ответов — это оценки В и С (т. е. 4 и 3 балла соответственно).

К каждому листу (терминальной вершине) классификационного дерева ведет последовательность ветвей. Дерево классификации оптимально в каждом своем разветвлении, поскольку в ходе построения дерева из всех возможных разветвлений программа выбирает такое разбиение данных, которое позволяет максимально предсказать характер зависимой переменной. Все разветвления, включенные в структуру дерева, отражают статистически значимые различия в распределении данных. Это видно из значений величины *p*, приведенных в нелистовых узлах.

На диаграмме 8 видно, что полученное дерево классификации имеет шесть разветвлений в узлах 1, 2, 3, 5, 8, 11. Первое разветвление в узле 1 разбивает весь состав данных на две группы по параметру статуса стимульного глагола (WordType). Левая ветвь выделяет в одну группу потенциальные глаголы и квазиглаголы, тогда как правая ветвь выделяет стандартные глаголы. Это разделение показывает, что в суждениях информантов потенциальные глаголы сближаются с квазиглаголами — результат, который мы уже наблюдали на диаграмме 4. Второе разбиение данных происходит в узле 2 в соответствии с тем же самым параметром — статусом стимульного глагола (WordType). Таким образом, суждения информантов относительно потенциальных глаголов выделяются в отдельный подкласс (узлы 3—7), подтверждающая наш вывод о том, что поведение этих стимулов имеет природу, отличную и от квазиглаголов, и от стандартных глаголов (см. раздел 3.1). Листовые узлы 4, 6 и 7 ясно показывают, что отличие потенциальных глаголов состоит прежде всего в том, что их оценки ранжируются на всем спектре возможностей от 1 до 5 баллов (см. столбцы графиков для А, В, С, D, Е), тогда как квазиглаголы имеют другой профиль оценок (от 1 до 3 баллов: С, D, Е на графиках узлов 9 и 10). Подводя итог, можно сказать, что первые два разветвления дерева поддерживают гипотезу 1 о соотношении между статусом слова в языке и его восприятием.

Наибольший интерес для нашего исследования представляют два разбиения данных в узлах 3 и 5: это те факторы, которые определяют суждения информантов относительно различных слов в рамках группы потенциальных глаголов. Мы видим, что здесь на первое место выходит фактор приставки (Prefix), выделяя в отдельную группу суждения о глаголах с приставкой *y-*. Далее следует разбиение суждений о глаголах с приставкой *o-* в узле 5 в соответствии с фактором возраста информанта (AgeGroup): суждения взрослых о потенциальных глаголах (узел 6) распределяются по оценкам А, В, С, D, Е неидентично суждениям школьников (узел 7). В частности, отрицание существования слова посредством оценки в 1 балл («Этого слова в русском языке нет») чаще наблюдается среди взрослых, чем среди детей (ср. столбцы Е в графиках узлов 6 и 7). Именно это отличает и потенциальные глаголы с приставкой *y-* (узел 4): отрицание их «нормальности», или наличия в языке, посредством оценки в 1 балл здесь встречается несколько чаще, чем в отношении глаголов с приставкой *o-* (узлы 6 и 7).

На основании этих наблюдений можно сделать три важных вывода. Во-первых, существует статистически значимое соотношение суждений информантов о потенциальных глаголах, с одной стороны, и типом словообразовательной модели (приставка *о-* или *у-*) и возрастом информанта, с другой стороны. Роль словообразовательной модели и возраста можно осторожно интерпретировать как факторы, влияющие на оценку потенциального слова. При этом фактор словообразовательной модели имеет больший вес, чем фактор возраста, как показано ниже на диаграмме 9.

Во-вторых, более частое неприятие глаголов с приставкой *у-* подтверждает гипотезу 2 о том, что деадъективные потенциальные глаголы с приставкой *о-* воспринимаются носителями языка как более естественные и приемлемые, чем деадъективные потенциальные глаголы с приставкой *у-*. Различие статистически значимо, но невелико (коэффициент Крамера $V = 0,1$ показывает, что сила связи переменных невелика).

В-третьих, в отношении потенциальных глаголов с приставкой *о-* подтверждается гипотеза 3, так как взрослые информанты в целом дают более консервативные оценки в отношении потенциальных глаголов, чем школьники. Это проявляется в более частом отрицании «нормальности» стимульного *о-*глагола взрослыми участниками. В данном случае отличие взрослых от школьников невелико (сила связи переменных небольшая: коэффициент Крамера $V = 0,1$), но все же статистически значимо.

Остается рассмотреть два разветвления дерева в узлах 8 и 11, которые очень показательны для локальной роли уже отмеченных факторов возраста и словообразовательной модели. В узле 8 все собранные суждения о квазиглаголах разбиваются по параметру возраста информанта. Школьники дают меньше оценок в 1 балл, чем взрослые (см. столбцы графиков для категории E), что поддерживает гипотезу 3, хотя она сформулирована, прежде всего, для потенциальных глаголов. Разветвление в узле 11 демонстрирует роль параметра приставки для класса стандартных глаголов: количество оценок ниже 5 баллов несколько больше для глаголов с приставкой *о-*. Заметим, что это не противоречит гипотезе 2, поскольку она сформулирована для потенциальных слов.

Важной особенностью деревьев классификации является их иерархическое устройство: разбиение данных по параметрам происходит последовательно, выделяя факторы, которые оказываются определяющими для общего состава данных (в нашем случае это фактор статуса слова), и факторы, роль которых более локальна (здесь — факторы приставки и возраста). В целом, в отличие от диаграмм 1—4, классификационное дерево показывает сложное взаимодействие трех факторов. Как уже было сказано, четвертый фактор — пол информанта — не представлен на дереве классификации. Это говорит о том, что его роль в распределении данных статистически не значима, что противоречит гипотезе 4 о воздействии фактора пола (ср. аналогичный результат на диаграмме 4).

Чтобы оценить значимость протестированных факторов предикторных переменных относительно друг друга, мы дополнительно применили метод случайных лесов. В основе этого метода лежит сопоставление целого ансамбля деревьев принятия решений с целью определения иерархии факторов в общем распределении данных. В рамках этого анализа статистическая программа в случайном порядке устраняет из рассмотрения различные порции данных и различные факторы, чтобы выявить, какой из факторов может лучше предсказать дистрибуцию данных (подробнее см. [Вауен et al. 2013: 265—267]). Результатом анализа является диаграмма 9 (с. 60), на которой представлена единая шкала значимости, располагающая четыре фактора относительно друг друга в порядке убывания их значимости.

Диаграмма 9 показывает, что наибольшую значимость в распределении всех суждений информантов имеет статус стимульного слова. Напомним, что именно этот фактор определяет ключевое разбиение данных в корневом узле (1) дерева классификации. Далее диаграмма 9 показывает, что по сравнению с фактором статуса слова значимость приставки и возраста информанта в общем распределении данных очень мала. Как мы видели на диаграмме дерева, эти факторы играют локальную роль в рамках определенных подгрупп ответов. Наконец, диаграмма 9 показывает, что роль пола информанта равна нулю.

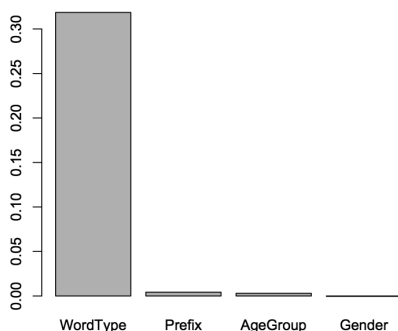


Диаграмма 9. Шкала значимости предикторных переменных¹³.

Заключение

Тестирование четырех гипотез на собранном нами материале позволяет сделать вывод о факторах, определяющих существование потенциальной лексики в языке, то есть охарактеризовать потенциальность лексики как феномен, непосредственно связанный и с системой языка (отсюда внимание к типу словообразовательной модели, по которой образованы глаголы, а также их статусу в системе), и с носителями языка (отсюда проверка социолингвистических факторов, таких как возраст и пол информантов).

Собранные данные подтверждают гипотезу о том, что статус слова оказывается ключевым фактором, определяющим его восприятие носителями языка. Мы показали, что слова каждой группы стимулов, а именно стандартные, потенциальные и квазислова, оцениваются носителями русского языка по-разному. В частности, потенциальные слова демонстрируют поведение, отличное и от стандартных, и от квазислов: разброс оценок и степень вариативности суждений в отношении потенциальных слов оказывается гораздо выше.

Интересным результатом эксперимента стали в целом неожиданно низкие оценки потенциальных слов. Несмотря на то, что слова типа *омузыкалить* семантически прозрачны (в отличие от квазислов типа *отовить*), по оценкам носителей русского языка потенциальные слова оказываются гораздо ближе к квазисловам, чем к стандартным словам. Возможно, это объясняется специфической для русских языковой культурой, заботой о правильности речи, прескриптивной природой нормы литературного языка [Крысин 2007]. С другой стороны, низкие оценки потенциальных слов говорят о том, что для носителей русского языка частотность слова важнее, чем его понятность, семантическая прозрачность. Поэтому, возможно, память носителя языка, фиксирующая все случаи использования языкового элемента в ментальном лексиконе [Taylor 2012], — это более мощный фактор в языке, чем применение продуктивных правил.

Тип словообразовательной модели и возраст информанта тоже играют роль в суждениях о потенциальных словах. В частности, деадъективные потенциальные глаголы с более

¹³ Вертикальная шкала на диаграмме 9 является математическим построением, основанным на поведении предикторов в процессе построения случайного леса. При построении шкалы используется индекс Джини (Gini, см. [Strobl et al. 2009: 335]), который в данном случае оказывается мерой значимости предикторных переменных в распределении данных. Индекс Джини представляет собой среднее значение увеличения «чистоты» данных в разветвлениях дерева, достигаемой каждой из предикторных переменных.

продуктивной приставкой *о-* воспринимаются носителями языка как более естественные и приемлемые, чем деадъективные потенциальные глаголы с менее продуктивной приставкой *у-* (гипотеза 2), и эта тенденция более характерна для школьников, чем для взрослых. В целом взрослые склонны давать более консервативные оценки в отношении потенциальных слов, чем школьники (гипотеза 3). Таким образом, можно предположить, что с возрастом отношение к потенциальным словам меняется. При этом пол говорящего, в отличие от возраста, не играет существенной роли в распределении оценок приемлемости (что противоречит гипотезе 4).

Наше исследование ограничено одним типом словообразовательной продуктивности в одном языке, но имеет значение для целой области изучения продуктивных моделей словообразования в других языках, а также может в перспективе изменить наше теоретическое представление о влиянии продуктивности и частотности на говорящего.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Виноградов 1972 — Виноградов В. В. Русский язык. М.: Высшая школа, 1972. [Vinogradov V. V. *Russkii yazyk* [The Russian language]. Moscow: Vysshaya Shkola, 1972.]
- Глазанова 2000 — Глазанова Е. В. Методика лексико-семантического исследования с использованием субъективного шкалирования и ее надежность // Вестник молодых ученых. 2000. № 2. С. 44—50. [Glazanova E. V. The procedure of lexical-semantic study using the method of magnitude estimation and its reliability. *Vestnik molodykh uchenykh*. 2000. No. 2. Pp. 44—50.]
- Глазанова 2001 — Глазанова Е. В. Типы связей в ментальном лексиконе и экспериментальные методы их исследования. Дис. ... канд. филол. наук. СПбГУ, 2001. [Glazanova E. V. *Tipy svyazei v mental'nom leksikone i eksperimental'nye metody ikh issledovaniya*. Kand. diss. [Types of connections in the mental lexicon and experimental methods of their investigation. Cand. diss.]. St. Petersburg: St. Petersburg State Univ., 2001.]
- Горошко 1996 — Горошко Е. И. Особенности мужского и женского вербального поведения (психолингвистический анализ). Дис. ... канд. филол. наук. М.: Институт языкознания РАН, 1996. [Goroshko E. I. *Osobennosti muzhskogo i zhenskogo verbal'nogo povedeniya (psikholingvisticheskii analiz)*. Kand. diss. [Special aspects of masculine and feminine verbal behavior (a psycholinguistic study). Cand. diss.]. Moscow: Institute of Linguistics, Russian Academy of Sciences, 1996.]
- Доценко 2000 — Доценко Т. И. Влияние фактора «пол» на внутренний лексикон подростка // Крысин Л. П. (ред.). Русский язык сегодня. Вып. 1. М.: Азбуковник, 2000. С. 78—85. [Dotsenko T. I. The impact of gender on the teenager's inner lexicon. *Russkii yazyk segodnya*. No. 1. Krysin L.P. (ed). Moscow: Azbukovnik, 2000. Pp. 78—85.]
- Захарова 2013 — Захарова О. С. К проблеме соотношения понятий «окказионализм» и «потенциальное слово» // Вестник Российского университета дружбы народов. Сер. Лингвистика. 2013. № 3. С. 14—26. [Zakharova O. S. The relation between the terms “occasional word” and “potential word”. *Vestnik Rossiiskogo universiteta druzhby narodov. Ser. Lingvistika*. 2013. No. 3. Pp. 14—26.]
- Земская 1972 — Земская Е. А. Окказиональные и потенциальные слова в русском словообразовании // Актуальные проблемы русского словообразования. Самарканд: Изд-во Самаркандского гос. ун-та, 1972. С. 19—28. [Zemskaya E. A. Occasional and potential words in the Russian word-formation. *Aktual'nye problemy russkogo slovoobrazovaniya*. Samarkand: Samarkand State Univ. Publ., 1972. Pp. 19—28.]
- Земская 1981 — Земская Е. А. Словообразование // Белошاپкова В. А. (ред.). Современный русский язык. М.: Наука, 1981. [Zemskaya E. A. Word-formation. *Sovremennyi russkii yazyk*. Beloshapkova V. A. (ed.). Moscow: Nauka, 1981.]
- Земская 1992 — Земская Е. А. Словообразование как деятельность. М.: Наука, 1992. [Zemskaya E. A. *Slovoobrazovanie kak deyatel'nost'* [Word-formation as an activity]. Moscow: Nauka, 1992.]
- Земская и др. 1993 — Земская Е. А., Китайгородская М. А., Розанова Н. Н. Особенности мужской и женской речи // Русский язык в его функционировании. Коммуникативно-прагматический аспект. М.: Наука, 1993. С. 90—137. [Zemskaya E. A., Kitaigorodskaya M. A., Rozanova N. N. Distinctive characteristics of men's and women's speech. *Russkii yazyk v ego funktsionirovanii. Kommunikativno-pragmaticheskii aspekt*. Moscow: Nauka, 1993. Pp. 90—137.]
- Кронгауз 1998 — Кронгауз М. А. Приставки и глаголы в русском языке: семантическая грамматика. М.: Языки русской культуры, 1998. [Krongauz M. A. *Pristavki i glagoly v russkom yazyke*:

- semanticheskaya grammatika* [Prefixes and verbs in Russian: Semantic grammar]. Moscow: Yazyki Russkoi Kul'tury, 1998.]
- Крысин 2007 — Крысин Л. П. Русская литературная норма и современная речевая практика // Русский язык в научном освещении. 2007. № 2(14). С. 5—17. [Krysin L. P. Norms of the Russian literary language and actual use in spoken discourse. *Russkii yazyk v nauchnom osveshchenii*. 2007. No. 2(14). Pp. 5—17.]
- Лопатин 1973 — Лопатин В. В. Рождение слова. Неологизмы и окказиональные слова. М.: Наука, 1973. [Lopatin V. V. *Rozhdenie slova. Neologizmy i okkazional'nye slova* [How words are born. Neologisms and occasional words]. Moscow: Nauka, 1973.]
- Лыков 1976 — Лыков А. Г. Современная русская лексикология (русское окказиональное слово). М.: Высшая школа, 1976. [Lykov A. G. *Sovremennaya russkaya leksikologiya (russkoe okkazional'noe slovo)* [Modern Russian lexicology (the Russian occasional word)]. Moscow: Vysshaya Shkola, 1976.]
- Мельчук 1998 — Мельчук И. А. Курс общей морфологии. Т. 2. М.; Вена: Языки русской культуры, 1998. [Mel'chuk I. A. *Kurs obshchei morfologii* [A course of general morphology]. Vol. 2. Moscow, Vienna: Yazyki Russkoi Kul'tury, 1998.]
- Мусатов 2010 — Мусатов В. Н. Русский язык. Морфемика. Морфонология. Словообразование. М.: Флинта, 2010. [Musatov V. N. *Russkii yazyk. Morfemika. Morfonologiya. Slovoobrazovanie* [The Russian language. Morphemics. Morphology. Word formation]. Moscow: Flinta, 2010.]
- НКРЯ — Национальный корпус русского языка // <http://www.ruscorpora.ru>. [Natsional'nyi korpus russkogo yazyka [Russian National Corpus]. Available at: <http://www.ruscorpora.ru>.]
- Петрухина 2014 — Петрухина Е. В. Образование новых слов в русском языке: теоретические аспекты и когнитивно-дискурсивный анализ // Mengel S. (ed.). *Slavische Wortbildung im Vergleich. Theoretische und pragmatische Aspekte*. Slavica Varia Halensia. Bd. 12. Berlin: Lit Verlag, 2014. С. 417—435. [Petrukhina E. V. Formation of new words in Russian: Theoretical aspects and cognitive discourse analysis. *Slavische Wortbildung im Vergleich. Theoretische und pragmatische Aspekte. Slavica Varia Halensia*. Bd. 12. Mengel S. (ed.). Berlin: Lit Verlag, 2014. Pp. 417—435.]
- РГ 1980 — Русская грамматика: В 2 т. Т. I / Гл. ред. Шведова Н. Ю. М.: Наука, 1980. [*Russkaya grammatika* [Russian grammar]: In 2 vol. Vol. I. Shvedova N. Yu. (ed.). Moscow: Nauka, 1980.]
- Розенталь, Теленкова 2008 — Розенталь Д. Э., Теленкова М. А. Словарь лингвистических терминов. Минск: Онискс, 2008. [Rozenal' D. E., Telenkova M. A. *Slovar' lingvisticheskikh terminov* [A dictionary of linguistic terms]. Minsk: Oniks, 2008.]
- Улуханов 1996 — Улуханов И. С. Единицы словообразовательной системы русского языка и их лексическая реализация. М.: Русские словари, 1996. [Ulukhanov I. S. *Edinitsy slovoobrazovatel'noi sistemy russkogo yazyka i ikh leksicheskaya realizatsiya* [Units of the Russian word-formation system and their lexical realization]. Moscow: Russkie Slovari, 1996.]
- Г. Фалин, А. Фалин 2011 — Фалин Г. И., Фалин А. И. Квартили в описательной статистике // Математика. 2011. № 15. С. 8—14. [Falın G. I., Falın A. I. Quartiles in descriptive statistics. *Matematika*. 2011. No. 15. Pp. 8—14.]
- Федорова 2013 — Федорова О. В. Об экспериментальном синтаксисе и о синтаксическом эксперименте в языкознании // Вопросы языкознания. 2013. № 1. С. 3—21. [Fedorova O. V. Experimental syntax and a syntactic experiment in linguistics. *Voprosy jazykoznanija*. 2013. No. 1. Pp. 3—21.]
- Фельдман 1957 — Фельдман Н. И. Окказиональные слова и лексикография // Вопросы языкознания. 1957. № 4. С. 64—73. [Fel'dman N. I. Occasional words and lexicography. *Voprosy jazykoznanija*. 1957. No. 4. Pp. 64—73.]
- Ханпира 1972 — Ханпира Э. Окказиональные элементы в современной речи // Стилистические исследования. Сб. статей. М.: Изд-во МГУ, 1972. С. 245—318. [Khanpira E. Occasional elements in modern spoken language. *Stilisticheskie issledovaniya. A collection of articles*. Moscow: Moscow State Univ. Publ., 1972. Pp. 245—318.]
- Эндресен 2013 — Эндресен А. А. Самостоятельные морфемы или позиционные варианты? Морфологический статус приставок О- и ОБ- в свете новых данных: корпус и эксперимент // Вопросы языкознания. 2013. № 6. С. 33—69. [Endresen A. A. Distinct morphemes or allomorphs? Morphological status of the Russian prefixes о- and об- in the light of the new data: Corpus and experiment. *Voprosy jazykoznanija*. 2013. No. 6. Pp. 33—69.]
- Andrews 2003 — Andrews D. R. Gender effects in a Russian and American matched-guise study: A sociolinguistic comparison. *Russian Linguistics*. 2003. Vol. 27. Pp. 287—311.
- Baayen 2008 — Baayen R. H. *Analysing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge Univ. Press, 2008.

- Baayen, Lieber 1991 — Baayen R. H., Lieber R. Productivity and English derivation: A corpus-based study. *Linguistics*. 1991. Vol. 29. No. 5. Pp. 801—844.
- Baayen et al. 2013 — Baayen R. H., Janda L. A., Nessel T., Endresen A., Makarova A. Making choices in Slavic: Pros and cons of statistical methods for rival forms. *Russian Linguistics*. 2013. Vol. 37 (Space and time in Russian temporal expressions). Pp. 253—291.
- Bard et al. 1996 — Bard E. G., Robertson D., Sorace A. Magnitude Estimation of linguistic acceptability. *Language*. 1996. Vol. 72. No. 1. Pp. 32—68.
- Barddal 2008 — Barddal J. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam: John Benjamins. 2008.
- Bermel, Knittl 2012 — Bermel N., Knittl L. Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory*. 2012. Vol. 8. No. 2. Pp. 241—275.
- Dąbrowska 2010 — Dąbrowska E. Naive vs. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review*. 2010. Vol. 27. Pp. 1—23.
- Eisikovits 1998 — Eisikovits E. Girl-talk / Boy-talk: Sex differences in adolescent speech. *Language and gender: A reader*. Coates J. (ed.). Oxford: Blackwell, 1998. Pp. 42—55.
- Endresen 2014 — Endresen A. *Non-standard allomorphy in Russian prefixes: Corpus, experimental, and statistical exploration*. Doct. diss. Tromsø: University of Tromsø: The Arctic University of Norway, 2014. Available at: <http://hdl.handle.net/10037/7098>.
- Endresen, Janda 2016 — Endresen A., Janda L. A. Five statistical models for Likert-type experimental data on acceptability judgments. *Journal of Research Design and Statistics in Linguistics and Communication Science*. 2016. Vol. 3. No. 2. Pp. 217—250.
- Keller 2000 — Keller F. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Doct. diss. Univ. of Edinburgh, 2000.
- Labov 1972 — Labov W. *Sociolinguistic patterns*. Philadelphia: Univ. of Pennsylvania Press, 1972.
- Lakoff, Johnson 1980 — Lakoff G., Johnson M. *Metaphors we live by*. Chicago: Univ. of Chicago Press, 1980.
- Langacker 1999 — Langacker R. W. A dynamic usage-based model. *Usage-based models of language*. Barlow M., Kemmer S. (eds.). Stanford (CA): CSLI Publications, 1999. Pp. 1—64.
- Lavrakas 2008 — Lavrakas P. J. *Encyclopedia of survey research methods*. Thousand Oaks (CA): Sage Publications, 2008.
- Levin 1993 — Levin B. *English verb classes and alternations: A preliminary investigation*. Chicago: Univ. of Chicago Press, 1993.
- Likert 1932 — Likert R. *A technique for the measurement of attitudes*. Doct. diss. Columbia University. Series “Archives of Psychology”. 1932. Vol. 140. New York: The Science Press. Pp. 1—55.
- Nilsen Márkus 2015 — Nilsen Márkus A. *Taming the Hungarian (in)transitivity zoo. Undiagnosed species and a complete derivation of the morphological patterns*. Doct. diss. Tromsø: University of Tromsø: The Arctic University of Norway, 2015. Available at: <http://munin.uit.no/handle/10037/8274>
- R Core Team 2015 — R: *A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Available at: <http://www.R-project.org/>.
- Schütze 1996 — Schütze C. T. *The empirical base of linguistics: Grammaticality judgements and linguistic methodology*. Chicago: Univ. of Chicago Press, 1996.
- Strobl et al. 2009 — Strobl C., Tutz G., Malley J. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*. 2009. Vol. 14. No. 4. Pp. 323—348.
- Tagliamonte, Baayen 2012 — Tagliamonte S. A., Baayen R. H. Models, forests and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change*. 2012. Vol. 24. No. 2. Pp. 135—178.
- Taylor 2012 — Taylor J. R. *The mental corpus: How language is represented in the mind*. Oxford: Oxford Linguistics, 2012.
- Townsend 1968 — Townsend C. E. *Russian word-formation*. New York: McGraw-Hill, 1968.
- Trudgill 1972 — Trudgill P. Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*. 1972. Vols. 1, 2. Pp. 179—195.
- Wolfram 1969 — Wollfram W. A sociolinguistic description of Detroit Negro Speech. *Urban Language Series*. Vol. 5. Washington (DC): Center for applied linguistics, 1969.
- Yokoyama 1999 — Yokoyama O. Russian genderlects and referential expressions. *Language in Society*. 1999. Vol. 28. Pp. 401—429.