

УДК 303.[448+714]+811.161.1

ИЗМЕРЕНИЕ ПАССИВНОГО СЛОВАРНОГО ЗАПАСА РУССКОГО ЯЗЫКА

Григорий Владимирович Головин

к. ф.-м. н., старший научный сотрудник

Университет Небраски-Линкольн, США

Behlen Laboratory, UNL city campus, Lincoln, NE 68588. gregorygolovin@gmail.com

В статье представлена методика первого адаптивного теста, разработанного для русского языка с целью измерения пассивного словарного запаса. Разобраны этапы подготовки теста: построение частотного словаря, группировка словарных семей, отбор тестовых слов, выбор вида опроса, введение в тест слов-ловушек для контроля честности и внимательности прохождения. Детально рассмотрена новая методика подсчета словарного запаса, основанная на измерении функции знания с ее последующим интегрированием. Валидация и исследование точности показали, что коэффициент надежности теста (0,95) достаточно высок для проведения исследований как на групповом, так и на индивидуальном уровне. Погрешность измерения словарного запаса лежит в пределах 7 % для взрослых (>18 лет) респондентов. Работу теста иллюстрируют предварительные данные о возрастной зависимости словарного запаса носителей русского языка, полученные по выборке из 123 тыс. респондентов.

Ключевые слова: словарный запас; тестирование; русский язык.

1. Введение

Словарный запас является одним из важнейших параметров, отражающих уровень владения языком. Его качественные и количественные характеристики коррелируют с уровнем лингвистической компетенции, а также многое говорят об интеллекте, психологическом портрете, роде деятельности и даже привычках человека. Изучение этих характеристик важно для задач педагогики (например, для определения уровня владения языком и качества усвоения предметной лексики), а также психологии и социологии. К сожалению, исследования словарного запаса носителей русского языка практически не ведутся. Одна из причин этого – отсутствие необходимого инструмента. Действительно, один из самых распространенных методов изучения словарного запаса – это измерение его объема. Теоретические основы таких измерений хорошо разработаны для английского языка, для которого существует несколько общепринятых тестов [Read 1993; Schmitt, Schmitt, Clapham 2001]. В то же время аналогичных тестов для русского языка до недавнего момента не существовало. Тест, представленный в настоящей работе, восполняет этот пробел. Его задача – дать исследователям (педагогам, психологам, социологам, лингвистам) удобный и надежный инструмент для количественной оценки объема пассивного словарного запаса испытуемого.

Методика теста опирается на общепринятый статистический подход [Read 2000]. Его суть состоит в предположении, что вероятность знания респондентом слов, используемых в языке (в книгах, телепередачах, речи) с одной и той же частотой, примерно одинакова. Это позволяет проверять знание респондентом не всех лексем языка, что заведомо невозможно, а только небольшого количества специально отобранных тестовых слов, каждое из которых представляет целую группу лексических единиц примерно одинаковой частотности.

Технически тест проводится в два этапа. На первом этапе респондент получает 40 тестовых слов. Задача респондента – отметить знакомые слова, т. е. те, для которых он может объяснить хотя бы по одному значению. По полученным данным делается приблизительная оценка словарного запаса. На втором этапе респондент получает 80 новых тестовых слов, подобранных с учетом полученной приблизительной оценки таким образом, чтобы исключить из них слишком простые или слишком сложные. Респондент опять отмечает знакомые лексемы. По полученным данным ($40 + 80 = 120$ тестовым единицам) производится уже более точная оценка словарного запаса. Метод, по которому рассчитывается словарный запас и на первом и на втором этапах теста, основан на оценке параметров некоторой функции знания респондента с последующим ее

интегрированием. Этот метод является новым и не применялся в тестах на словарный запас ранее.

Настоящий тест задумывался как основной инструмент онлайн-проекта по изучению словарного запаса людей, говорящих на русском языке [Головин: электр. ресурс]. Проект был инспирирован следующими вопросами: сколько слов знает среднестатистический носитель языка? Какова динамика (рост и уменьшение) словарного запаса родного языка с возрастом? Какие факторы на нее влияют? Проект стартовал в апреле 2014 г. и продолжается до сих пор. На настоящий момент в нем приняли участие уже более 800 тыс. человек, что дает основания считать его самым масштабным проектом подобного типа.

В основной части статьи будут детально рассмотрены методика теста, этапы его проектирования, валидация, оптимизация и исследование точности теста. В конце будут представлены некоторые предварительные результаты онлайн-проекта, а именно возрастная динамика размера пассивного словарного запаса носителей русского языка.

2. Методика

2.1. Частотный словарь

Существующие частотные словари русского языка ограничены по объему и не содержат редких низкочастотных слов, необходимых для тестирования людей с большим словарным запасом. Так, частотный словарь под редакцией Л.Н. Засориной [Засорина 1977] содержит около 40 тыс. лемм, «Частотный словарь современного русского языка» О.Н. Ляшевской и С.А. Шарова [Ляшевская, Шаров 2009] – 52 тыс. лемм. Поэтому первой задачей разработки теста стало составление более полного частотного словаря русского языка. Для этого был использован толково-словообразовательный словарь Т.В. Ефремовой [Ефремова: электр. ресурс], содержащий 136 тыс. лемм, и для каждой леммы найдена ее частотность по Национальному корпусу русского языка [Национальный корпус русского языка: электр. ресурс]. Корпус состоит из большого количества (86 тыс.) текстов разной тематики (художественная литература, публицистика, научные и научно-популярные, религиозные и философские тексты, личная переписка, дневники); общий объем текстов – 230 миллионов слов. За счет большого объема и широкого охвата этот корпус представляет собой слепок современного (54 % всех текстов были созданы после 1950-го года) русского языка.

2.2. Словарные семьи

Основная идея статистического подхода к оценке словарного запаса заключается в том, что вероятность знания слова испытуемым зависит от частотности этого слова. Это, однако, не со-

всем так. К примеру, слово *думать* встречается в 100 тысяч раз чаще, чем *думающий*, однако если испытуемый знает одно из них, то, скорее всего, знает и другое. Поэтому можно сказать, что существуют некоторые «словарные семьи». Зная любое слово из такой семьи и обладая некоторым лингвистическим чутьем, можно догадаться о значении всех остальных слов этой семьи. В тестах на словарный запас английского языка подобные семьи называются «word families», причем знание одной word family приравнивается к знанию одного слова. Таким образом, в оценке словарного запаса обычно учитывается знание только основных слов, но не их производных. Существуют общепринятые правила формирования word families [Bauer, Nation 1993]. К сожалению, для русского языка таких правил найти не удалось.

Задача объединения слов из собранного частотного словаря в словарные семьи была решена следующим образом. Слово добавлялось в соответствующую словарную семью, если по отношению к хотя бы одному слову из этой семьи оно являлось:

- 1) соотносящимся по значению (*абонентный* и *абонент*);
- 2) уменьшительным (*абрикосик* к *абрикос*);
- 3) ласкательным (*душенька* к *душа*);
- 4) усилительным (*духотища* к *духота*);
- 5) уничижительным (*романчик* к *роман*);
- 6) мужской/женской параллельной формой (*еретичка* и *еретик*);
- 7) отвлеченным существительным (*естественность* к *естественный*);
- 8) существительным, означающим процесс действия по значению глагола (*амнистирование* к *амнистировать*);
- 9) страдательным залогом (*идеализироваться* к *идеализировать*);
- 10) причастием по значению глагола (*обещающий* к *обещать*);
- 11) совершенным/несовершенным видом глагола (*исчезнуть* и *исчезать*);
- 12) множественным числом (*банкноты* к *банкнота*).

136 тыс. слов были распределены таким образом в 88 тыс. словарных семей. В качестве примера можно привести следующие словарные семьи:

- рука, ручка, ручной, ручонка, рученька;
- день, дневной, денёк, денёчек, денной;
- частый, часто, частота, частотный.

Необходимо заметить, что настоящий список правил представляет собой достаточно грубое приближение к решению проблемы составления словарных семей, если эта проблема вообще может быть решена. Действительно, лингвистическое чутье значительно отличается у разных людей и зависит от объема словарного запаса: у об-

разованных и начитанных носителей языка словарные семьи больше, чем у людей, только начинающих учить язык. Возможно, словарные семьи должны строиться экспериментально на основе тестирования большого количества респондентов с разным словарным запасом [Guy, Browne, Culligan: электр. ресурс]. Понимая приблизительность принятых нами правил построения словарных семей, мы решили подсчитывать словарный запас в словах, а не в словарных семьях. Словарные семьи использовались только для перегруппировки частотного словаря. Слова в итоговом частотном словаре были отсортированы по суммарной частотности соответствующих словарных семей.

Возникает также вопрос о правомерности включения производных слов (например, уменьшительно-ласкательных) в словарь в качестве самостоятельных лексем при оценке словарного запаса. Он осложняется тем, что статус «производности» неоднозначен и по-разному решается лингвистами: в зависимости от точки зрения авторов словаря одни и те же слова могут быть представлены либо в рамках одной словарной статьи, либо в разных. В этом вопросе мы ориентировались на словарь Т.В. Ефремовой [Ефремова: электр. ресурс] и для определения объема словарного запаса считали самостоятельным каждое слово, оформленное в этом словаре собственной словарной статьей.

2.3. Тестовые слова

Идеальный тест на словарный запас должен проверять знание всех слов языка, что не представляется возможным. В реальном тесте используется ограниченное количество тестовых слов, где одно тестовое слово представляет собой целую группу близких по частоте слов. Тестовые слова должны быть максимально «нейтральны» и общеупотребительны, чтобы никто из респондентов не получал преимущества перед другими за счет специфических знаний. В противном случае такое «специфичное» слово будет нерепрезентативно группе, которую оно представляет. При отборе тестовых слов исключались:

- 1) диалектизмы (*кочет*);
- 2) специальная лексика (*диссоциация*);
- 3) заимствованные слова, о значении которых можно догадаться, зная соответствующий иностранный язык (*жантильность*);
- 4) устаревшие слова (*гужеед*);
- 5) жаргонизмы (*лабать*);
- 6) части фразеологизмов (*зга*);
- 7) обценная лексика.

2.4. Опрос

Есть несколько общепринятых способов проведения тестов на словарный запас, отличающихся методами опроса тестируемых. Рассмотрим только два наиболее часто встречающихся

варианта. Первый – это тест с множественными вариантами ответа. Тестовое задание может выглядеть так:

«Бестия» – это

- 1) персонаж древнегерманской мифологии;
- 2) пройдоха, плут;
- 3) скандалист;
- 4) привидение, призрак.

Второй – это «знаю/не знаю» тест, где испытуемого просят отметить слова, которые он знает. На первый взгляд, тест с множественными вариантами ответа предпочтительнее, так как должен давать более точный результат. В этом варианте знание тестовых слов проверяется, тогда как в варианте «знаю/не знаю» честность ответов остается на совести испытуемого. В действительности оба варианта опроса имеют свои достоинства и недостатки [Meara, Buxton 1987]. Впрочем, эти недостатки могут быть минимизированы при аккуратном подходе к составлению теста. При правильном составлении тестовых заданий, а также достаточном количестве тестовых слов, оба типа тестов дают схожие результаты [Culligan 2015; Pellicer-Sanchez, Schmitt 2012]. Рассмотрим причины, по которым для разработки теста был выбран более простой «знаю/не знаю» вариант опроса.

Во-первых, на каждое тестовое задание у респондента уходит определенное время. Как будет показано дальше, для достаточно точного определения словарного запаса необходимо проверить знание около 120 тестовых слов. В случае теста «знаю/не знаю» это занимает около 5 минут. В случае теста с множественными вариантами ответа – в несколько раз больше, так как испытуемому требуется прочитать все варианты ответа и мысленно «взвесить» каждый из них. Для онлайн-теста время прохождения критично, потому что проходящие его люди в основном недостаточно мотивированы чтобы тратить много времени.

Во-вторых, составление тестовых заданий в случае теста с множественными вариантами требует серьезных усилий [Аванесов 2005: электр. ресурс]. Каждый вариант должен быть достаточно реалистичен, чтобы избежать угадывания методом исключения, а также достаточно прост, чтобы не ввести респондента в заблуждение, если он знает тестовое слово. В случае «знаю/не знаю» теста тестовые задания как таковые отсутствуют, что значительно экономит время подготовки теста и облегчает его составление.

У выбранного «знаю/не знаю» типа опроса есть и недостатки. Во-первых, «знание» слова может интерпретироваться разными опрашиваемыми по-разному. Во-вторых, такой тест не позволяет проверить знание различных значений одного и того же слова. Принимая во внимание

эти ограничения, в обращении к опрашиваемым перед тестом мы специально оговорили, что следует понимать под «знанием слова»: «Считайте, что вы знаете слово, если можете дать определение хотя бы одному его значению. Не отмечайте слова, которые вы видели или слышали, но в значении которых не уверены до конца».

Вид опроса существенно влияет на то, что в итоге измеряет тест. Действительно, «знание» слова – явление многокомпонентное. Так, Пол Нэйшн выделил восемь аспектов знания [Nation 2013]:

- 1) как произносится слово;
- 2) как пишется слово;
- 3) каковы различные грамматические формы слова;
- 4) с какими другими словами обычно употребляется слово (коллокации);
- 5) как часто употребляется слово (частотность слова);
- 6) в каких регистрах обычно употребляется слово (стилистические свойства слова);
- 7) каковы значения слова;
- 8) каковы ассоциации слова с другими словами.

Выбор вида опроса определяет, какой из аспектов знания проверяется тестом. Опрос вида «знаю/не знаю» предполагает проверку только пунктов 2 и 7 (написание слова и его значение). Другие аспекты знания также могут быть исследованы, для чего должны быть использованы другие виды опроса [Webb, Sasao 2013].

2.5. Слова-ловушки

Другая серьезная проблема «знаю/не знаю» тестов – невозможность проверки правдивости ответов. Опрашиваемый может отметить некоторые тестовые слова как знакомые по невнимательности или специально, чтобы увеличить результат. Для отслеживания таких некорректных ответов обычно вводят тестовые слова-ловушки. Эти слова похожи на настоящие, но не существуют в языке, например, **очудей*, **онторология*, **тибильга*. В зависимости от количества «сработавших» слов-ловушек, результат теста может быть скорректирован [Anderson, Freebody 1983] – чем больше несуществующих слов отметил респондент, тем существеннее уменьшается его результат. Конкретная реализация такой коррекции, однако, вызывает разногласия [Pellicer-Sanchez, Schmitt 2012]. Действительно, невозможно понять, отметил ли респондент слово-ловушку как знакомое из-за невнимательности (прочитав его быстро и приняв за другое) или специально, чтобы увеличить результат. Коррекция в каждом из этих случаев должна быть разной. Из-за этой неопределенности (которую в рамках только «знаю/не знаю»

теста вряд ли можно преодолеть) было решено не проводить коррекцию результата. Слова-ловушки, тем не менее, использовались (на 120 тестовых слов – 4 слова-ловушки). Результаты респондентов, отметивших хотя бы одно такое слово, исключались из проводимых исследований словарного запаса.

2.6. Оценка словарного запаса

Для оценки словарного запаса использовался не применявшийся ранее алгоритм. Чтобы пояснить, как он работает, необходимо сделать концептуальное отступление.

Основная идея всех тестов на оценку пассивного словарного запаса – это представление довольно большой группы слов, близких по частотности, одним тестовым словом. Если респондент знает тестовое слово – считается, что он знает и всю группу, так как слова в ней близки по частотности к тестовому. Эту же идею можно переформулировать, используя понятие ранга – порядкового номера слова в отсортированном по частотности словаре (ранг самого частотного слова, таким образом, равен единице). Если респондент знает тестовое слово с некоторым рангом n , логично предположить, что он также знает и другие слова, ранги которых близки к n . Используя язык физики, можно сказать, что, предъявляя тестовое слово с рангом n и получая ответ от респондента, мы измеряем вероятность знания им слов с рангами, близкими к n . Используя физическую аналогию дальше, можно предположить, что есть некоторая функция вероятности знания слов, которую мы пытаемся измерить, и которая зависит от n – ранга слов. От респондента к респонденту такая функция будет варьироваться, характерный же вид ее, однако, будет общий – от единицы (вероятность знания – 100 %) для маленьких n (простых слов) до нуля (вероятность знания – 0 %) для больших n (сложных, редких слов). Каждое индивидуальное измерение является довольно грубым, так как дает либо 0, либо 1 (если тестовое слово отмечено как незнакомое или знакомое соответственно). Если, однако, провести много таких измерений, оказывается возможным достаточно точно измерить искомую функцию знания. Если эту функцию затем проинтегрировать, то получится не что иное, как оценка словарного запаса.

Так как характерный вид функции словарного запаса одинаков для всех респондентов, можно попытаться описать ее некоторой аналитической формулой, имеющей желаемую форму – от единицы при малых значениях аргумента до нуля при больших через плавный переход. Удобно использовать следующую двухпараметрическую функцию:

$$f(n) = \frac{1}{1 + e^{w_n(n-n_0)}}, \quad (1)$$

где n_0 определяет положение перехода ($f(n_0) = 1/2$), w_n задает крутизну перехода ($f'(n)|_{n=n_0} = -w_n/4$). Как показали дополнительные исследования, ограничиться однопараметрической функцией нельзя: при одинаковом положении перехода значение крутизны (для разных респондентов) варьируется в несколько раз; использование трех и более параметров неоптимально и приводит к неоправданному увеличению погрешности оценки.

Оценка словарного запаса (определенный интеграл от функции вероятности знания слов), таким образом, может быть выражена следующей формулой:

$$\int_1^{n_{max}} f(n)dn = n_{max} + \frac{\ln(1 + e^{w_n(1-n_0)})}{w_n} - \frac{\ln(1 + e^{w_n(n_{max}-n_0)})}{w_n} - 1, \quad (2)$$

где n_{max} – размер частотного словаря.

Алгоритм оценки словарного запаса был реализован следующим образом. Каждое тестовое слово (ранга n_i) давало одно измерение функции словарного запаса: $f_i = 1$, если респондент помечил тестовое слово как знакомое, и $f_i = 0$, если как незнакомое. Набор таких измерений аппроксимировался функцией (1) по методу наименьших квадратов. Аппроксимация давала значения коэффициентов n_0 и w_n , которые подставлялись в формулу (2) для получения оценки словарного запаса.

Измерение словарного запаса описанным методом имеет несколько достоинств. Во-первых, алгоритм позволяет оценить погрешность измерения, а также его качество. Действительно, аппроксимация по методу наименьших квадратов (или любому другому) позволяет оценить не только значения неизвестных коэффициентов функции словарного запаса (n_0 и w_n), но и их дисперсию. Это, в свою очередь, позволяет рассчитать дисперсию оценки словарного запаса. Аппроксимация также позволяет рассчитать коэффициент детерминации (он же R -квадрат), который показывает, насколько хорошо выбранная модель (то есть вид функции словарного запаса) аппроксимирует полученные данные. Если дисперсия оценки велика, а коэффициент детерми-

нации мал – это показатель того, что либо выбранный вид функции неверен, либо данных для оценки недостаточно, либо данные содержат аномалии. Как будет показано в разделе «3.2. Проверка выбранной функции распределения», первый пункт можно исключить. Поэтому при достаточном количестве тестовых слов (что исключает второй пункт), алгоритм позволяет выявить аномалии в данных. Такие аномалии могут свидетельствовать о невнимательном прохождении теста (в крайнем своем проявлении – о выборе ответа случайным образом).

Во-вторых, данный алгоритм позволяет поэтапно увеличивать точность оценки словарного запаса. Так, на первом этапе теста респонденту дается 40 тестовых слов. По этим данным алгоритм определяет приблизительную оценку словарного запаса. На втором этапе респонденту дается уже 80 тестовых слов, ранги которых относительно близки к приблизительной оценке. Эти новые данные добавляются к старым, и алгоритм дает уточненную оценку по уже 120 (40+80) тестовым словам.

В-третьих, оценка словарного запаса получается не в абстрактных единицах (баллах), а в словах, что делает интерпретацию результатов проще.

3. Валидация

3.1. Методика валидации и описание респондентов

Валидация теста включала в себя несколько исследований. Во-первых, было проверено, насколько выбранный вид функции (формула (1)) соответствует реальной функции вероятности знания слов. Во-вторых, были обоснованы преимущества при использовании тестирования в два этапа (против тестирования в один). В-третьих, была оценена точность теста. Эти исследования опирались на данные индивидуальных словарных запасов семи респондентов, прошедших расширенный тест с 1200 тестовыми словами (то есть с полным количеством тестовых слов, использующихся в тесте). Информация о респондентах приведена в таблице 1. Также по методу расщепления из классической теории тестов был рассчитан коэффициент надежности теста.

Таблица 1

Респонденты, прошедшие расширенную версию (1200 слов) теста

Респондент	Возраст, лет	Образование
А	29	Высшее, физическое
В	29	Высшее, информационные технологии
С	27	Высшее, инженерное
Д	29	Кандидат философских наук
Е	30	Кандидат физико-математических наук
Ф	54	Высшее, педагогическое
Г	55	Высшее, инженерное

3.2. Проверка выбранной функции распределения

По методике разрабатываемого теста для получения оценки словарного запаса респондента необходимо проинтегрировать его функцию знания. В обычном варианте теста (со 120 тестовыми словами) измерить функцию знания с необходимой точностью невозможно – именно поэтому необходимо использовать аппроксимацию. Однако в расширенной версии теста (с 1200 тестовыми словами) неизвестная функция знания может быть измерена достаточно качественно. Такое измерение необходимо для валидации выбранного вида аппроксимирующей функции.

Результат трех респондентов, прошедших расширенную версию теста показан на рисун-

ке 1. Каждый красный крест соответствуют одному тестовому слову (1 – если респондент отметил слово как знакомое, 0 – если как незнакомое). Видно, что для малых рангов (высокочастотных тестовых слов) плотность красных крестов на значении «1» высока – респонденты знают эти слова. По мере движения от малых рангов к большим (в сторону низкочастотных слов) красные кресты появляются на значении «0» все чаще: респонденты знают все меньше тестовых слов. Черные квадраты показывают усредненную вероятность знания тестовых слов (по 120 на квадрат). Синие кривые соответствуют аппроксимации тестовых данных функцией (1). Выбранная функция хорошо визуальнo аппроксимирует полученные для всех 7 респондентов результаты.

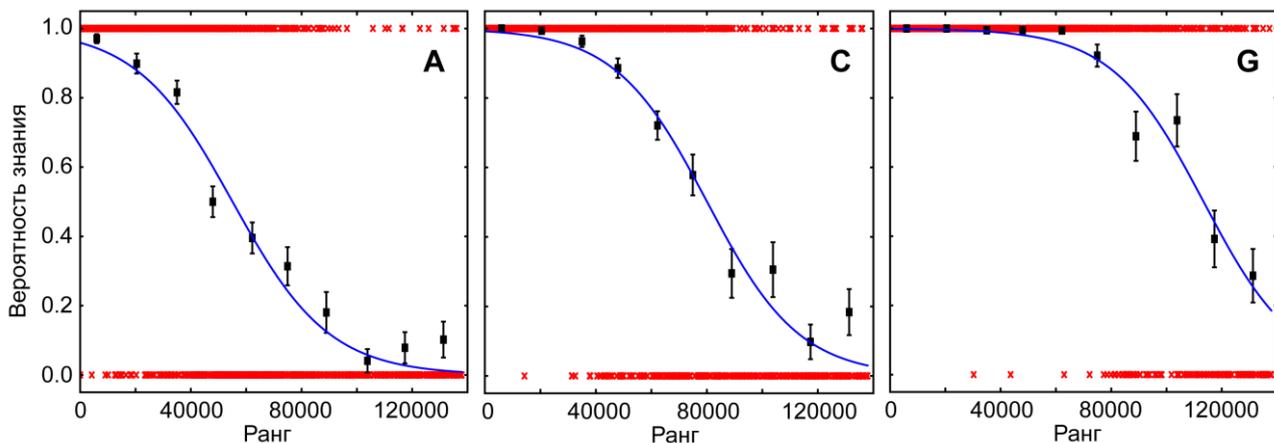


Рис. 1. Функции вероятности знания слов для трех респондентов

На рисунке представлены данные респондентов А, С, G (см. табл. 1). Красные кресты – индивидуальные тестовые слова (0 – ответ «не знаю», 1 – «знаю», всего 1200 тестовых слов). Черные квадраты – усредненные вероятности, каждый квадрат соответствует примерно 120 тестовым словам. Синие кривые – аппроксимации тестовых данных функцией, заданной формулой (1).

Для количественной оценки качества аппроксимации были сопоставлены две оценки словарного запаса. Одна была получена интегрированием измеренной функции знания (черные квадраты на рис. 1), другая – интегрированием ап-

проксимирующих функций. Результаты показаны в таблице 2. Разница оценок оказалась не больше 6 %, что может считаться хорошим результатом.

Таблица 2

Валидация выбора функции знания слов: сравнение оценок словарного запаса, полученных через аппроксимацию (по формуле (2)) и прямым интегрированием функции знания

РЕСПОНДЕНТ	А	В	С	Д	Е	F	G
Оценка словарного запаса через аппроксимацию (по формулам (1) и (2))	55600	67800	79800	86600	96000	97900	110100
Оценка словарного запаса (интегрирование без аппроксимаций)	58900	72200	82400	87300	97000	99000	110500
Разница оценок	5,6 %	6,1 %	3,2 %	0,8 %	1,1 %	1,1 %	0,4 %

3.3. Тест в один этап против теста в два этапа

Методика теста позволяет использовать любое количество тестовых слов. При этом очевидно, что чем больше слов используется, тем точнее получается оценка. Однако не все тестовые слова вносят одинаковый вклад в измерение. Например, для респондента с большим словарным запасом вероятность знания очень простых слов близка к единице, поэтому использование большого количества простых тестовых слов не увеличит точность теста. То же верно и наоборот – для респондента с маленьким словарным запасом вероятность знания очень сложных слов близка к нулю, то есть такому респонденту бессмысленно предъявлять сложные тестовые слова. Тест работает в два этапа. На первом этапе респонденту предъявляется 40 тестовых слов. По результатам этого этапа делается приблизительная оценка словарного запаса. Второй этап состоит из 80 слов, которые выбираются исходя из приблизительной оценки – слишком простые или сложные тестовые слова на этом этапе не присутствуют. Таким образом, тест в два этапа является адаптивным, так как тестовые слова второго этапа подбираются под каждого респондента индивидуально.

Для того чтобы проверить, действительно ли такой адаптивный подход более оптимален, чем неадаптивный (в один этап), было проведено следующее исследование. С помощью результа-

тов расширенной версии теста (1200 слов), пройденного респондентом Е (см. Таблица 1), были промоделированы различные варианты теста в два этапа. При этом варьировалось количество вопросов на первом и втором этапе. Результаты представлены на рисунке 2. Вертикальная ось показывает полное количество тестовых слов в двух этапах, горизонтальная – количество слов на втором этапе. Цвет кодирует точность теста (стандартное отклонение, полученное по 1000 реализаций теста со случайно выбранными тестовыми словами). Горизонтальная пунктирная линия соответствует фиксированному полному количеству тестовых слов (120). Самая левая точка на ней соответствует, по сути, тесту в один этап (120 тестовых слов на первом этапе, 0 – на втором), при движении по этой линии вправо количество тестовых слов на втором этапе увеличивается, на первом – уменьшается (сумма при этом остается постоянной). Точность теста вдоль этой линии непостоянна и имеет максимум в области около 80 тестовых слов на втором этапе. Действительно, для теста в один этап (120 слов) стандартное отклонение оценки словарного запаса – 4,8 %, для теста в два этапа (40+80 слов) – 4,1 %. Эту разницу можно интерпретировать и по-другому. Для того чтобы обеспечить точность теста в два этапа с оптимальным количеством слов (40+80=120), потребовался бы тест в один этап, состоящий из 150 тестовых слов, что на целых 25 % больше.

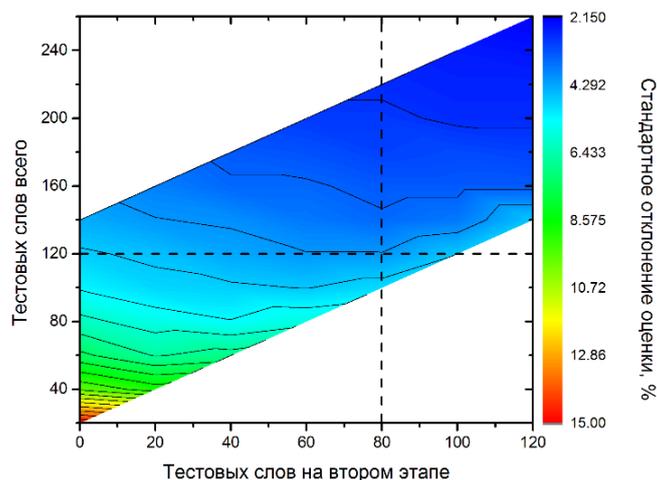


Рис. 2. Оптимизация количества тестовых слов на первом и втором этапах (на примере результатов респондента Е)

Каждое значение стандартного отклонения оценки словарного запаса получено по 1000 прохождениям теста со случайным образом выбранными тестовыми словами. Сплошные линии отмечают контуры одинаковых значений стандартного отклонения.

3.4. Точность теста

Оценка точности теста проводилась с помощью следующей процедуры. Для каждого респондента, прошедшего расширенную версию

теста, было промоделировано прохождение стандартной версии теста – 40 тестовых слов на первом этапе, 80 – на втором. Количество тестовых слов, используемых в стандартной версии

теста, в десять раз меньше полного их количества в базе теста, поэтому каждое прохождение – уникальное. Даже если один и тот же человек будет проходить тест раз за разом, тестовые слова будут повторяться довольно редко. Если бы такие прохождения приводили к значимо различным оценкам, это говорило бы либо о нероботоспособности метода, либо о плохом выборе тестовых слов. Было промоделировано 5000 таких прохождений каждым респондентом, чтобы

тестовые слова попали в тест в разных комбинациях. Полученные распределения оценок оказались нормальными (Гауссовыми), для них были рассчитаны средние значения и стандартные отклонения. Эти значения сравнивались с точными оценками словарного запаса каждого респондента, полученными по расширенной версии теста (1200 слов). Результаты сравнения приведены в таблице 3.

Таблица 3

Сравнение точности оценок словарного запаса, полученных по расширенной версии теста (1200 слов), с оценками, полученными по стандартной версии теста (два этапа, 40+80 слов)

РЕСПОНДЕНТ	A	B	C	D	E	F	G
Точная оценка словарного запаса (по 1200 словам)	55600	67800	79800	86600	96000	97900	110100
Оценка словарного запаса по первому этапу (40 слов)	57700	70100	80200	89700	96100	97300	110300
Стандартное отклонение оценки по первому этапу, %	14	13	10	10	10	8	7
Разница точной оценки и оценки по первому этапу, %	3,6	3,3	0,5	-0,8	0,1	-0,5	0,1
Оценка словарного запаса по двум этапам (40+80 слов)	56000	67800	79200	87600	92500	95600	108800
Стандартное отклонение оценки по двум этапам, %	7,0	6,1	4,2	4,2	4,1	3,1	2,7
Разница точной оценки и оценки по двум этапам, %	0,7	-0,1	-0,7	-3,0	-3,6	-2,2	-1,2

Можно сделать следующие выводы. Первый этап стандартной версии теста – 40 слов – приводит к несмещенной оценке словарного запаса. Разница между средним значением этой оценки и точным значением оказалась в пределах 4 %. Стандартное отклонение этой оценки довольно большое – до 14 %. Это значит, что, пройдя тест несколько раз, один и тот же респондент может получить значительно отличающиеся результаты. Другими словами, точность такой оценки

невысока. Это легко понять, так как число используемых тестовых слов невелико. Оценка, полученная по двум этапам, также является несмещенной (разница с точной оценкой – в пределах 4 %). Однако, стандартное отклонение этой оценки (оно же – точность теста) значительно лучше – в пределах 7 %. Полученные величины стандартных отклонений как функция величины словарного запаса приведены на рисунке 3.

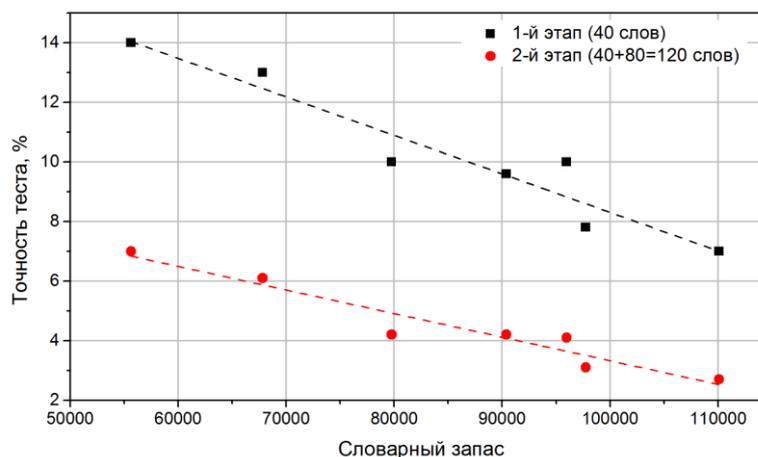


Рис. 3. Точность оценки (стандартное отклонение) словарного запаса

Каждая точка соответствует одному респонденту (A – E от меньших значений словарного запаса к большим; см. табл. 3).

Абсолютная точность теста, как оказалась, лежит в пределах от 3000 до 4000 слов. Это означает, что с вероятностью в 67 % оценка, полученная по тесту, будет лежать в пределах ± 4000 слов от реального значения словарного запаса; с вероятностью в 95 % – в пределах ± 8000 слов.

3.5. Надежность теста

Классическая теория тестов позволяет рассчитать коэффициент надежности теста – показатель, характеризующий как повторяемость, так и точность его результатов. На достаточно большой (19 тысяч) репрезентативной выборке респондентов, прошедших тест (и не отметивших ни одного слова-ловушки), был применен метод расщепления теста [Ким 2007]. Для каждого респондента был известен список из 120 тестовых слов, а также его ответы. Этот список был расщеплен на две равные половины, по которым независимо производилась оценка словарного запаса. Таким образом получилось 19 тысяч пар оценок. Коэффициент корреляции между этими оценками оказался равным 0,90. Так как оценка, полученная по половине (60) тестовых слов, имеет заниженную точность, коэффициент корреляции был скорректирован по методу Спирмена–Брауна. Итоговый коэффициент надежности оказался равным 0,95. Такое высокое значение надежности подтверждает то, что тест может ис-

пользоваться для измерений словарного запаса как на групповом, так и на индивидуальном уровне.

4. Предварительные результаты

Предварительные результаты были зафиксированы после прохождения интернет-теста 150 тысячами человек (все – носители русского языка). Из них 123 тысячи прошли тест аккуратно, не отметив ни одного слова-ловушки. Таким образом, было отсеяно 18 % неаккуратных прохождений.

На рисунке 4 показаны перцентили полученного распределения словарного запаса в зависимости от возраста респондентов. Например, самая нижняя кривая (10-ый перцентиль) для 20 лет дает 40 тысяч слов. Это означает, что 10 % респондентов этого возраста имеют словарный запас ниже этого значения, а 90 % – выше. Выделенная синим центральная кривая (медиана) соответствует такому словарному запасу, который больше показателей словарного запаса одной половины выборки и меньше показателей другой половины, т. е. половина респондентов соответствующего возраста продемонстрировала результат ниже данного числа, вторая половина – выше. Самая верхняя кривая – 90-й перцентиль – отсекает результат, выше которого показали только 10 % респондентов с максимальным словарным запасом.

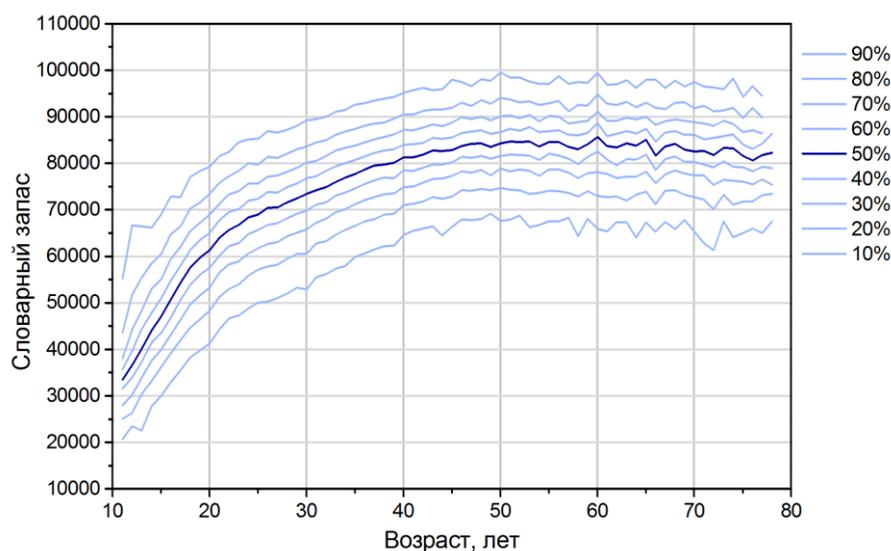


Рис. 4. Предварительные результаты теста (123 тысячи респондентов)

Сплошные кривые соответствуют перцентильям от 10 до 90 %, темно-синяя центральная кривая – медиана.

Полученные данные позволяют сделать следующие выводы.

1. Словарный запас растет с практически постоянной скоростью примерно до 20 лет, после чего скорость его набора уменьшается, сходя на

нет к 45 годам. После этого возраста словарный запас уже практически не меняется.

2. Во время обучения в школе подросток учит по 10 слов в день. Эта величина кажется неестественно большой, но объясняется тем, что в те-

сте производные слова учитывались отдельно, как самостоятельные.

3. К моменту выпуска из школы подросток в среднем знает 51 тысячу слов.

4. За время обучения в школе словарный запас увеличивается примерно в 2,5 раза.

5. После выпуска из школы и до достижения среднего возраста человек в среднем узнает 3 новых слова в день.

6. После достижения 55 лет словарный запас начинает несколько снижаться. Это может быть связано с забыванием слов, которые не используются достаточно долго. Интересно, что этот возраст примерно совпадает с выходом на пенсию.

Необходимо заметить, что полученные предварительные результаты не дают представления о словарном запасе среднестатистического носителя русского языка, так как выборка респондентов, прошедших интернет-тест, не является репрезентативной ни для России, ни для всего множества русскоговорящих людей. Во-первых, уровень образования респондентов, прошедших тест, значительно выше общероссийского – 65 % респондентов имеют высшее образование, тогда как в России, по данным всероссийской переписи населения 2010 г., таковых только 23 %

[Росстат: электр. ресурс]. Во-вторых, интернет-тест нашли и прошли в основном активные пользователи интернета, что делает выборку специфичной (по крайней мере для пожилых людей). В-третьих, выборка респондентов была также значительно смещена самим желанием пройти тест на словарный запас. Логично предположить, что полученные по такой особенной выборке результаты словарного запаса должны быть несколько выше среднестатистических.

Для того чтобы представить полученные данные в перспективе, был рассчитан процент покрытия Национального Корпуса Русского Языка различным количеством слов, отсортированных по частотности (см. рис. 5). Из этих данных видно, например, что 450 самых частотных слов составляют половину всех словоупотреблений корпуса, в то время как на остальную половину приходится уже 135+ тысяч слов. Процент покрытия позволяет оценить, сколько слов нужно знать, чтобы понимать некий усредненный текст. Как было экспериментально установлено, для понимания текста необходимо знание от 95 до 98 % содержащихся в нем слов [Laufer, Ravenhorst-Kalovski 2010], что соответствует, по полученным данным, от 26 до 42 тысячам слов.

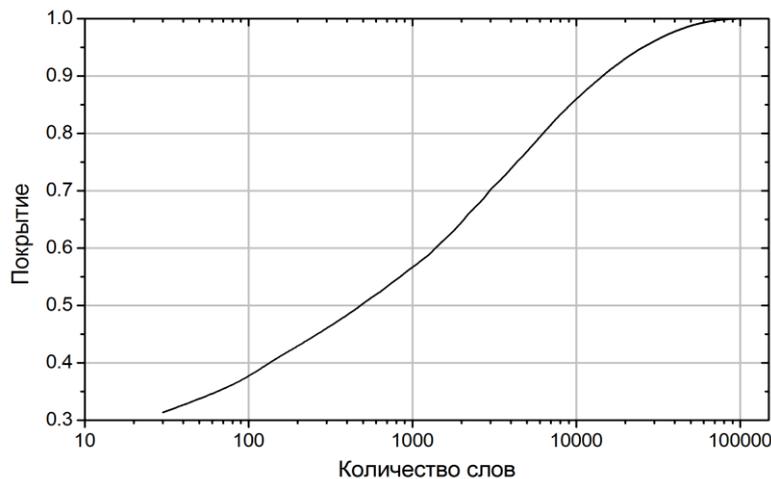


Рис. 5. Процент покрытия Национального корпуса русского языка различным количеством слов, отсортированных по частотности

5. Заключение

В заключение будут рассмотрены преимущества и недостатки теста, а также перспективы и методы его дальнейшего улучшения.

К преимуществам теста можно отнести следующее. Во-первых, тест очень быстр – прохождение занимает около 5 минут. Во-вторых, тест является адаптивным (за счет тестирования в два этапа) и подходит для проверки словарного запаса людей, очень сильно отличающихся друг от друга по уровню знаний. В-третьих, тест

полностью автоматизирован и не требует ручной обработки. Недостатки теста являются следствием его быстроты. Формат опроса «знаю/не знаю», позволяющий получить ответы на большое количество тестовых вопросов и делающий тест быстрым и точным, подразумевает некоторую вольность трактовки «знания» каждым респондентом. Даже если респондент пометил тестовое слово как знакомое, это не значит, что он владеет всеми аспектами знания (стилистическими, морфологическими, смысловыми и про-

чье) этого слова. Поэтому тест измеряет размер, но не качество словарного запаса. Тест также не может быть использован в тех случаях, когда респонденту выгодно получить высокий результат, так как контроль честности прохождения довольно прост и не позволит надежно выявить респондентов, целенаправленно пытающихся повысить свои результаты.

Дальнейшее улучшение теста возможно по двум направлениям. Во-первых, его надежность может быть повышена добавлением заданий с множественными вариантами ответа. Это позволит лучше контролировать, действительно ли респонденты понимают значения тестовых слов, которые они отмечают как знакомые. Во-вторых, математический аппарат теста может быть изменен в соответствии с современной теорией тестов (Item Response Theory – IRT) [Аванесов 2005: электр. ресурс]. Это позволит выявить наиболее «качественные» тестовые слова, использование которых приведет к сокращению длины теста при сохранении его точности.

Использование IRT в тестах на словарный запас имеет еще одно важное следствие. Результаты таких тестов сильно зависят от того, каким образом разработчики решили вопрос построения словарных семей (т. е. какие слова были посчитаны за самостоятельные, а какие – за производные). Это приводит к тому, что сравнивать между собой результаты разных тестов сложно – один и тот же респондент по разным тестам может получить значительно отличающиеся результаты. Если же тесты выполнены по методологии IRT, их результаты должны быть одинаковы, пусть и выраженные в нескольких абстрактных баллах (логитах), что может дать возможность сравнивать результаты разных тестов.

Список литературы

Аванесов В.С. Теория и методика педагогических измерений: материалы публикаций в открытых источниках и Интернет. 2005. [Электронный ресурс]. URL: <http://viperson.ru/data/200812/jbjejbjxjklmjje.pdf> (дата обращения: 26.11.2015).

Головин Г. Тест словарного запаса. [Электронный ресурс]. URL: www.myvocab.info (дата обращения: 26.11.2015).

Ефремова Т.В. Новый толково-словообразовательный словарь русского языка. М.: Русский язык, 2000. [Электронный ресурс]. URL: <http://www.speakrus.ru/dict> (дата обращения: 16.11.2015).

Засорина Л.Н. Частотный словарь русского языка. М.: Русский язык, 1977. 936 с.

Ким В.С. Тестирование учебных достижений. Уссурийск: Изд-во Уссурийск. гос. пед. ин-т, 2007. 214 с.

Ляшевская О.Н., Шаров С.А. Частотный словарь современного русского языка (на материалах национального корпуса русского языка). М.: Азбуковник, 2009. 1112 с.

Национальный корпус русского языка. [Электронный ресурс]. URL: <http://www.ruscorgpora.ru> (дата обращения: 26.11.2015).

Росстат [Электронный ресурс]. Данные Всероссийской переписи населения 2010, Москва. URL: moscow.gks.ru/wps/wcm/connect/rosstat_ts/moscow/ru/census_and_researching/census/national_census_2010 (дата обращения: 26.11.2015).

Anderson R.C., Freebody P. Reading Comprehension and the Assessment and Acquisition of Word Knowledge // *Advances in Reading/Language Research*. 1983. Vol. 2. P. 231–256.

Bauer L., Nation P. Word Families // *International Journal of Lexicography*. 1993. Vol. 6(4). P. 253–379.

Culligan B. A Comparison of Three Test Formats to Assess Word Difficulty // *Language Testing*. 2015. Vol. 32(4). P. 503–520.

Guy C., Browne C., Culligan B. V-Check and WordEngine Academic FAQ (Ver. 1.8). [Электронный ресурс]. URL: www.wordengine.jp/research/pdf/V-CheckAndWordEngineAcademicFAQ_rev1.8.pdf (дата обращения: 29.11.2015).

Laufer B., Ravenhorst-Kalovski G.C. Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension // *Reading in a Foreign Language*. 2010. Vol. 22(1). P. 15–30.

Meara P., Buxton B. An Alternative to Multiple Choice Vocabulary Tests // *Language Testing*. 1987. Vol. 4(2). P. 142–154.

Nation P. *Teaching and Learning Vocabulary*. Boston: Heinle Cengage Learning. 2013. 274 p.

Pellicer-Sanchez A., Schmitt N. Scoring Yes-No Vocabulary Tests: Reaction Time vs. Nonword Approaches // *Language Testing*. 2012. Vol. 29(4). P. 489–509.

Read J. The Development of a New Measure of L2 Vocabulary Knowledge // *Language Testing*. 1993. Vol. 10(3). P. 355–371.

Read J. *Assessing Vocabulary*. Cambridge: Cambridge University Press. 2000. 294 p.

Schmitt N., Schmitt D., Clapham C. Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Levels Test // *Language Testing*. 2001. Vol. 18(1). P. 55–88.

Webb S.A., Sasao Y. New Directions In Vocabulary Testing // *RELC Journal*. 2013. Vol. 44(3). P. 263–77.

**RECEPTIVE VOCABULARY SIZE MEASUREMENT
FOR RUSSIAN LANGUAGE**

Grigory V. Golovin

Senior research associate

University of Nebraska-Lincoln, USA

The paper presents the first adaptive receptive vocabulary test developed for Russian language. We discuss all aspects of the test design: building a frequency dictionary, forming word families, test words and test form selection, and use of fake words to assess credibility of test results. We introduce a novel algorithm to estimate vocabulary size, based on measuring and integration of knowledge function. We perform the test validation and precision analysis. The test reliability coefficient (0,95) indicates that it can be used for studies of groups as well as individuals. The test standard error is less than 7 % for adult (>18 years old) test takers. We show preliminary results on receptive vocabulary size of Russian native speakers as a function of their age, obtained from a sample of 123k test takers.

Key words: vocabulary; testing; Russian language.