

УДК 159.922.72

ББК 81, 88.4

DOI 10.30982/2077-5911-2023-57-3-92-110

Научная статья

ОЦЕНКА ДОСТОВЕРНОСТИ ДАННЫХ О ВОЗРАСТЕ УСВОЕНИЯ СЛОВ, ПОЛУЧАЕМЫХ ПУТЕМ ОПРОСА ВЗРОСЛЫХ РЕСПОНДЕНТОВ¹

Сугоняев Константин Владимирович

Институт психологии Российской академии наук, Москва, Россия

Григорьев Николай Андреевич

Независимый исследователь, Москва, Россия

Аннотация

В связи с трудоемкостью и ограниченными возможностями объективного оценивания возраста усвоения слов, проведено исследование, направленное на изучение возможности использования *субъективных* оценок в качестве нормативного материала для этой характеристики. Субъективные оценки собирались для 7500 слов русского языка путем предъявления восьми сотням участников 25 списков слов с инструкцией указать предположительный возраст их усвоения. Объективные оценки были получены путем тестирования школьников 2, 4, 6 и 8 классов с помощью словарных тестов, включающих суммарно 720 заданий, специально разработанных и адаптированных к каждому классу с учетом результатов субъективного оценивания. Надежность тестовых оценок в исследованных выборках варьировала в диапазоне 0,85–0,95. Сопоставление субъективных и объективных оценок осуществлялось несколькими способами, основной из которых состоял в коррелировании субъективного возраста усвоения с точностью идентификации значения целевого слова. После надлежащей коррекции на ограничение диапазона варьирования переменной получены корреляции в диапазоне 0,6–0,9 (в среднем 0,79), что позволяет сделать вывод о достаточно высокой валидности метода нормирования возраста усвоения слов с помощью субъективных оценок. Намечены пути дальнейших исследований.

Ключевые слова: возраст усвоения слов, субъективные и объективные оценки, надежность, валидность, коррекция на ограничение варьирования переменной

Статья Джона Кэрролла и Маргарет Уайт, опубликованная в 1973 году [Carroll, White 1973], задала новое направление в исследованиях факторов протекания процессов восприятия и продуцирования речи. С момента ее выхода возраст, в котором дети усваивают слова, стал рассматриваться как важный, если не важнейший, детерминант выполнения экспериментальных задач, вовлекающих эти процессы. Был проведен ряд исследований, в которых возраст усвоения (ВУ) слов выступал как экспериментальная переменная. Соответственно, было выполнено большое число нормативных исследований, в которых собирались оценки ВУ для разных языков.

¹ Исследование выполнено при финансовой поддержке гранта Российского научного фонда (РНФ) № 22-18-00269 «Сбор норм по возрасту усвоения для большого корпуса слов русского языка».

Эти оценки распадаются на два класса, которые иногда называют «субъективными» и «объективными» оценками ВУ. Первые собирают, спрашивая взрослых людей, в каком возрасте они (или их дети) усвоили те или иные слова. Вторые получают, предъявляя детям разного возраста изображения и прося назвать то, что изображено, предъявляя детям разного возраста словарный тест и др. Объективные оценки вызывают больше доверия, однако их сбор намного затратнее, так что норм субъективных оценок ВУ слов куда больше, чем норм оценок объективных. И, конечно же, к обоснованию права на существование первых было приложено немало усилий.

Чтобы обосновать право на существование того или иного показателя, нужно продемонстрировать его надежность и валидность. Полученные к настоящему времени данные свидетельствуют о достаточно высокой надежности субъективных оценок ВУ слов. Так, в одной работе [Gilhooly, Nay 1977: 13] корреляция между средними оценками респондентов-мужчин и респондентов-женщин оказалась равной 0,96. В другом исследовании [Gilhooly, Logie 1980: 396–397] корреляция между подгруппами респондентов, сбалансированными, насколько было возможно, по полу и количеству, составила 0,98. Моррисон с соавт. [Morrison, Chappell, Ellis 1997: 540] рассчитали корреляцию полученных предыдущими авторами оценок с собственными; она оказалась равной 0,85. В работе Дианы Цапариной с соавт. [Tsaparina, Bonin, Méot 2011: 1090] корреляция между средними оценками ВУ, сделанными респондентами с четными и нечетными номерами, составила 0,95. В масштабной нормативной работе Виктора Купермана с соавт. [Kuperman, Stadthagen-Gonzalez, Brysbaert 2012: 981] значение аналогичного показателя оказалось равным 0,843.

Была продемонстрирована и валидность субъективных оценок ВУ слов. В работе Кэрролла и Уайт [Carroll, White 1973] эти оценки были скоррелированы с данными других авторов о том, какие слова известны учащимся разных классов. Корреляция оказалась равной 0,847. Лайонс с соавт. [Lyons, Teer, Rubenstein 1978] проверяли знание первоклассниками 30 слов, оцененных как рано усваиваемые, и 30 слов, оцененных как поздно усваиваемые. Ни одно из слов второй группы не было известно первоклассникам, в то время как слова первой группы были известны, в среднем, 74% детей. В работе Моррисон с соавт. [Morrison, Chappell, Ellis 1997: Table 3] корреляции субъективных оценок ВУ с объективными составили 0,759 и 0,747 (авторы использовали два способа выведения объективных оценок). Корреляция субъективных оценок ВУ, полученных Цапариной с соавт. [Tsaparina, Bonin, Méot 2011] с объективными оценками для соответствующих слов оказалась равной 0,673 [Grigoriev, Oshhepkov 2013: 1213]. Корреляция субъективных оценок Купермана с соавт. с объективными оценками, полученными ранее другими авторами путем предъявления вербального теста, составила 0,757 [Brysbaert, Biemiller 2017: Table 1].

В то же время, было показано, что субъективные оценки ВУ слов находятся под искажающим влиянием других психолингвистических переменных. Моррисон с соавт. [Morrison, Chappell, Ellis 1997] провели три регрессионных анализа, в которых зависимыми переменными выступали субъективные оценки ВУ и выведенные разными способами два набора объективных оценок. Предикторами во всех трех анализах были число фонем в слове, образность и частотность слов, знакомость обозначаемых словами предметов, показатель степени согласованности названия изображений этих предметов и визуальная сложность этих изображений. Рассматривались также корреляции зависимых переменных с предикторами. Авторы заключили, в частности,

что респонденты, выставляя субъективные оценки ВУ слов, чрезмерно ориентируются на их длину (число фонем) и частотность, а также на их знакомость. Таким образом, любые собираемые нормы для ВУ слов должны пройти проверку на надежность и валидность собранных оценок; целесообразным является также исследование воздействия на эти оценки других психолингвистических переменных.

В настоящей работе предпринята попытка определить надежность и валидность субъективных оценок ВУ 7500 слов, полученных на первом этапе выполнения проекта, направленного на оценку ВУ большого корпуса слов русского языка (около 30000 слов); столь масштабное исследование ВУ слов проводится в России впервые.

Для проверки валидности субъективных оценок ВУ слов используются объективные данные о возрасте их усвоения, как полученные в предыдущих работах, так и специально собранные в ходе выполнения настоящего проекта.

Метод

Получение субъективных оценок возраста усвоения слов

На начальном этапе проекта было отобрано 30000 слов русского языка. Для обеспечения представительности выборки слов было, во-первых, проконтролировано, чтобы количество слов, принадлежащих к различным частям речи, соответствовало их частотности в частотном словаре О.Н. Ляшевской и С.А. Шарова [Ляшевская, Шаров 2009]. Это определило следующее распределение слов по частям речи: 13400 существительных, 7600 глаголов, 7000 прилагательных, 1700 наречий и 300 слов, принадлежащих к другим частям речи. Во-вторых, основываясь на англоязычных данных [Van Overschelde, Rawson, Dunlosky 2004; Brysbaert, Biemiller 2017], подбирались слова, усваиваемые в разном возрасте. В стимульный материал не включались некоторые классы слов, присутствующие в частотном словаре: узкоспециальная лексика, просторечия, жаргонизмы, обценная лексика, слова, обозначающие национальную принадлежность и некоторые другие.

Из этого множества было отобрано 7500 слов, которые были разделены на 25 списков по 300 тестовых слов. В каждом списке было 134 существительных, 76 глаголов, 70 прилагательных, 17 наречий и три слова из других частей речи. При формировании списков, помимо равенства долей разных частей речи и их равномерного распределения по списку, были предприняты усилия, направленные на то, чтобы избежать включения в один список однокоренных слов и обеспечить сопоставимый разброс возраста усвоения, при этом слова из одной лексико-семантической категории распределялись, по возможности, по разным спискам.

Кроме того, каждый список был дополнен сорока словами с ранее рассчитанными субъективными [Tsaparina, Bonin, Méot 2011; Пашнева 2014] и объективными [Grigoriev, Oshhepkov 2013] оценками ВУ. Десять из этих слов предьявлялись в начале каждого списка в качестве ознакомительных (представляли широкий диапазон ВУ), а 30 других были в случайном порядке распределены среди тестовых слов и использовались в качестве контрольных. Таким образом, в финальной форме каждый список состоял из 340 слов, из которых 300 были оригинальными, а 40 – общими для всех списков².

Каждый список планировалось предьявить для оценивания не менее чем 25 респондентам (с учетом возможной отбраковки части протоколов). Согласно инструкции, респонденту предлагается указать предполагаемый возраст усвоения

² Отбор слов и подготовка списков выполнены А.А. Григорьевым, О.В. Балясниковой и С.А. Пашневой.

каждого слова, вписав число полных лет, или выбрать ответ «не знаю», введя напротив оцениваемого слова букву «Н» или прочерк. В инструкции разъяснено, что усвоенным считается не только то слово, которое испытуемый использует в своей речи, но и то, смысл которого он понимает, когда слышит его от окружающих или читает.

Сбор данных о субъективном ВУ слов осуществлялся в нескольких популяциях, которые, предположительно, должны были представлять широкий спектр лексического развития граждан России, а именно, студенты (и некоторое число преподавателей) одного из вузов г. Курска, студенты московского региона, курсанты 1-го курса одного из военных вузов, военнослужащие по призыву. Примерно в 40% случаев списки предъявлялись в форме таблиц Excel, в остальных случаях – в печатной форме³. Каждый респондент работал лишь с одним списком.

Получение объективных оценок возраста усвоения

Основной метод получения объективных оценок ВУ слов состоял в следующем. Из 7500 слов, ВУ которых оценивался на первом этапе проекта субъективным методом, было выбрано 770 «тестопригодных» слов, охватывающих диапазон ВУ от 7 до 17 лет, для которых: а) в доступных электронных ресурсах (ru.m.wiktionary.org, dic.academic.ru и др.) имеются синонимы, гипонимы или гиперонимы проверяемых слов или относительно краткие и доступные пониманию школьников словосочетания, наиболее точно передающие смысл оцениваемого слова и пригодные на роль правильного ответа; б) может быть сгенерировано не менее четырех дистракторов, отличающихся по смыслу от проверяемого слова или отражающих лишь частный аспект его значения. В число тестовых не включались однокоренные слова (выбиралось одно слово, как правило, с *большим* ВУ); слова, для которых были зафиксированы слишком большие (более 2,5 лет) гендерные различия ВУ, а также слова, вероятность узнавания которых оказалась менее 0,5.

При формировании тестовых заданий образцами служили словарные (суб)тесты из таких известных интеллектуальных тестовых батарей, как ШТУР, Р. Амтхауэра, Дж. Фланагана и т.п. Синтез пяти вариантов ответа для каждого задания (опознаваемого слова) представляется разумным компромиссом между стремлением снизить влияние вероятности случайного угадывания и трудоемкостью конструирования тестовых заданий⁴. Задача респондента состояла в выборе варианта ответа, наиболее точно передающего смысл тестового слова.

Отобранные слова распределялись по спискам, предположительно адаптированным к уровням лексического развития учащихся 2, 4, 6, 8 и 10 классов. При распределении тестовых слов по классам, учитывались данные о субъективном ВУ слов, полученные в ходе выполнения 1 этапа проекта. При этом отбираемые слова для 2–6 классов центрировались, соответственно, в районе ВУ 8, 10 и 12 лет ($\pm 1,2$ года). Вследствие дефицита слов с ВУ выше 13 лет (по данным опроса), набор слов для восьмиклассников охватил более широкий диапазон ВУ (11,8–15,4 лет) со средним значением около 13 лет. В еще большей степени эта проблема отразилась на формировании тестового набора для 10-классников, однако ввиду недостаточности объективных данных для этой категории учащихся результаты их обследования в настоящей статье не рассматриваются.

³ Сбор субъективных оценок ВУ выполнен С.А. Пашневой, Е.В. Бородиной, О.Е. Солохой и С.К. Сугоняевой, за что авторы выражают им глубокую признательность.

⁴ В подготовке тестовых заданий, помимо авторов данной статьи, активное участие приняла С.К. Сугоняева.

Для 6 и 8 классов тестовые наборы включали по 300 слов; для 4 класса – 240 слов; для 2 класса – 160 слов. В каждом наборе часть слов перекрывалась с наборами для смежных классов. Так, между наборами 2 и 4 классов, а также 4 и 6 классов имелось по 50 общих слов; наборы 6 и 8 классов имели 180 общих слов. Общее число оригинальных заданий (слов), включенных в субтесты для 2–8 классов, составило 720. Средний субъективный ВУ всего массива тестовых слов составил $11,06 \pm 1,96$; аналогичные показатели для слов, включенных в тестовые наборы для отдельных классов, составили: $12,9 \pm 0,87$ для 8 класса; $11,9 \pm 0,61$ для 6 класса; $10,1 \pm 0,64$ для 4 класса и $8,26 \pm 0,79$ для 2 класса. Таким образом, вариативность ВУ слов в классных тестовых наборах была в 2,25–3,21 раза ниже, чем в общем массиве.

Примеры заданий, репрезентирующих тестовые наборы для каждого из классов, представлены в табл. 1:

Таблица №1

Примеры тестовых заданий

Класс	Тестовое слово	Варианты ответа				
		А	Б	В	Г	Д
2	коверкать	искажать	шить пальто	переворачиваться через голову	мастерить	раскапывать
4	танго	дикая собака	сольный танец	парный танец	тропический фрукт	злаковая культура
6	пафосный	большой	праздничный	многолюдный	красивый	напыщенный
8	категорически	изначально	Уверенно	шутливо	безусловно	спорно
10	сепаратный	заметный	обособленный	мирный	зависимый	молочный

Примечание: жирным шрифтом выделены ключевые варианты ответа

Учитывая большое число заданий и высокую когнитивную нагрузку при выполнении словарного теста, тестовые наборы для каждого класса делились на два субтеста, уравненных по числу заданий и трудности (субъективному ВУ), что позволяло администраторам тестирования гибко подстраиваться к школьному графику и возможностям учащихся.

Все необходимые для проведения обследования материалы (контент, ключи, инструкции) были введены в программу компьютерного психодиагностического комплекса (КПДК), что обеспечило возможность тиражирования тестовых материалов (буклетов и бланков) в стандартном дизайне. Учащиеся 6 и 8 классов при выполнении словарных субтестов фиксировали свои ответы в формируемых программой регистрационных бланках; учащиеся 2 и 4 классов обводили выбранный вариант ответа в тестовых буклетах (с которых впоследствии информация переносилась в бланк). Заполненные регистрационные бланки обрабатывались путем сканирования; результаты распознавания ответов оцифровывались, индивидуальные протоколы паспортизировались и сохранялись в базе данных КПДК. Полученные при этом данные позволяли, во-первых, оценить надежность объективных оценок ВУ, во-вторых, сопоставить вероятность правильного опознания слова среди респондентов

определенного возраста с субъективными оценками ВУ; в третьих, сравнить успешность опознания слов респондентами различного возраста в перекрывающихся подмножествах слов.

В качестве дополнительного источника объективных оценок ВУ использовались оценки из работы Григорьева и Ощепкова [Grigoriev, Oshhepkov 2013]. В этой работе оценивался возраст усвоения наименований 286 предметов путем предъявления детям разного возраста изображений этих предметов с просьбой их назвать. 63 названия оказались среди 7500 слов, использованных в первый год выполнения проекта, что позволило сопоставить субъективные и объективные оценки ВУ слов.

Для обработки количественных данных использовались вычислительные возможности табличного процессора Microsoft Excel, программы КПКД, а также статистического программного пакета STATISTICA v.8.

Результаты и обсуждение

1) Характеристики исследовательских выборок

В рамках сбора *субъективных* оценок ВУ слов было получено свыше 800 протоколов, из которых пригодными к анализу (содержащих не более 10% непомеченных слов) было признано 802.

Поскольку процедура получения 340 субъективных оценок сопряжена с определенной когнитивной нагрузкой на респондентов и, исходя из опыта сбора аналогичных данных за рубежом, предполагалось, что не все респонденты готовы приложить достаточные усилия для добросовестного выполнения задания и не все обладают достаточной рефлексией для точного выполнения инструкции [Desimone, Harms, Desimone 2015], были предложены и использованы следующие критерии качества протокола:

1. Рассчитывалась автокорреляция в ответах респондентов. Данный показатель оценивает степень независимости каждой частной оценки от предшествующего выбора [Gottfried, Ježek, Králová, Řiháček 2022]. В качестве максимально допустимого было принято значение 0,33, хотя в отдельных случаях допускалось некоторое превышение этой величины.

2. Корреляция с контрольными словами. Как правило, браковались ответы респондентов при значениях этой корреляции ниже 0,7, но в отдельных случаях признавались удовлетворительными ответы при корреляции несколько ниже этого значения.

3. Средние значения оценок респондентов. При значительном (свыше двух стандартных отклонений от среднего по списку) повышении или понижении среднего уровня оценок результаты респондента могли быть забракованы. Уровневые смещения дополнительно контролировались коэффициентом внутриклассовой корреляции с контрольными словами [Furt 2010], чувствительным не только к нарушениям сходства соотношений оценок, что характерно для корреляции Пирсона, но и к уровневым различиям этих оценок (в качестве нижней границы был принят коэффициент ICC=0,4).

4. Минимальный ВУ слов в списке не должен превышать 4⁵.

5. Количество ответов «не знаю» не должно превышать 60 (20% от общего числа оцениваемых слов).

⁵ Среднее значение минимального ВУ по выборке составило $1,98 \pm 0,90$, таким образом, минимальный ВУ ≥ 5 выходит за пределы трех сигм.

6. Корреляция всех 340 оценок респондента со средними оценками по данному списку, рассчитанными по ответам всех остальных респондентов (не менее 0,7).

Помимо перечисленных формальных критериев применялся также визуальный просмотр всей истории ответов; в отдельных случаях, несмотря на приемлемые значения формальных индексов соответствия, выявлялась заметная доля экзотических ответов (нереалистично заниженный ВУ явно сложных слов или завышенный для простых), что также могло стать основанием для отбраковки протокола.

В результате число протоколов, включенных в итоговые расчеты, сократилось до 597 (из них 265 женщин), т.е. было отбраковано по различным основаниям немногим более четверти протоколов. Средний возраст участников, представлявших 57 регионов России⁶, составил $20,50 \pm 3,84$ года (17–53 года), при этом 573 респондента (96%) имели возраст от 17 до 25 лет.

Более детально распределение респондентов по группам, базовые сведения демографического характера и результаты отбраковки протоколов сомнительной достоверности представлены в табл. 2.

Таблица №2

Характеристики выборки респондентов, включенных в исследование субъективного ВУ 7500 слов

Контингент	До чистки		После чистки	
	Кол-во (в т.ч. женщин)	Возраст (диапазон)	Кол-во (в т.ч. женщин)	Возраст (диапазон)
Студенты и преподаватели гг. Курск, Москва	355 (289)	24,4±5,7 (17,3–53,3)	325 (265)	21,2±4,9 (17,3–53,3)
Курсанты военного института	229 (-)	19,3±1,7 (16,9–27,1)	163 (-)	19,1±1,4 (16,9–24,3)
Военнослужащие по призыву	218 (-)	20,6±1,4 (18,0–25,4)	109 (-)	20,7±1,4 (18,1–24,9)

В рамках сбора *объективных* оценок ВУ слов из шести школ от 335 респондентов было получено 483 протокола, содержащих ответы на задания субтестов для 2, 4, 6 и 8 классов⁷. В 6 и 8 классах большинство участников выполнило оба субтеста; во 2-х классах большая часть учащихся выполнила один из субтестов; в 4-х классах каждый респондент выполнил по одному субтесту. Во всех случаях, когда учащиеся выполняли оба субтеста, порядок их предъявления балансировался; кроме того, с целью снижения когнитивной нагрузки парные субтесты, как правило, выполнялись в разные дни.

Как и в случае сбора субъективных оценок ВУ, полученные данные были подвергнуты отбраковке, поскольку не все учащиеся отнеслись к процедуре обследования достаточно добросовестно; кроме того, в ряде случаев отмечалось снижение качества выполнения задания от первого по порядку выполнения субтеста ко второму. Основные причины отбраковки протоколов – невыполнение субтеста

⁶ По месту получения диплома о среднем образовании.

⁷ Обследование учащихся проводилось в четырех школах г. Курска и двух школах Московской области. Сбор первичных данных выполнен С.А. Пашневой, В.Г. Володичевой и О.А. Белошицкой, за что авторы выражают им глубокую признательность.

до конца или наличие признаков хаотического реагирования (вероятность правильного ответа существенно не отличается от вероятности случайного угадывания: $0,20 \pm 0,05$). В результате чистки количество протоколов, признанных пригодными для включения в анализ, сократилось до 428 (отсев составил 11,4%), а число респондентов – до 313. Распределение респондентов по классам, субтестам, местонахождению и базовые сведения демографического характера, отражающие конечный массив данных, отражены в табл. 3.

Таблица №3

**Характеристики выборки респондентов, включенных в исследование
объективного ВУ 720 слов**

Класс	Субтест	Кол-во протоколов по источникам данных				Характеристики респондентов	
		Власиха	Воскресенск	Курск	Всего	Кол-во (из них девочек)	Возраст (M±СтО)
8	Словарь-3	47	23	-	70	72 (42)	14,8±0,34
	Словарь-4	40	21	-	61		
6	Словарь-5	18	28	-	46	54 (30)	12,9±0,37
	Словарь-6	20	27	-	47		
4	Словарь-7	-	-	48	48	94 (47)	10,9±0,34
	Словарь-8	-	-	46	46		
2	Словарь-9	-	-	54	54	93 (47)	8,96±0,34
	Словарь-10	-	-	56	56		

Примечание: М – средняя; СтО – стандартное отклонение; в скобках указано число девушек

2) Результаты сбора субъективных оценок ВУ слов

Из более чем 240 тыс. субъективных оценок ВУ, полученных в ходе опроса, в анализ было включено 179100 оценок, извлеченных из 597 протоколов, при этом 12168 оценок (6,8%) пришлось на вариант «не знаю». После чистки исходных данных на отдельные списки пришлось от 20 до 30 респондентов (в среднем 23,9); отбраковка по отдельным спискам варьировала от трех до 14 протоколов.

Основные описательные статистики итогового массива субъективных оценок ВУ, а также контрольных индексов, используемых для оценки достоверности данных, представлены в табл. 4.

Таблица №4

**Описательные статистики массива субъективных оценок ВУ 7500 слов
и контрольных индексов, рассчитанных по выборке 597 респондентов**

Индекс	ВУ	СтО ио ВУ	Мин ВУ	Макс ВУ	Н	АК	$r_{\text{ксл}}$	ИСС	$r_{\text{инт пр}}$
Средняя	9,09	3,58	1,98	17,84	20,38	0,18	0,87	0,74	0,82
СтО	1,41	0,57	0,90	2,33	13,68	0,09	0,06	0,11	0,04

Примечание: СтО ио ВУ – стандартное отклонение индивидуальных оценок ВУ; Мин, Макс – минимальное и максимальное значения ВУ; Н – сумма ответов «не знаю»; АК – автокорреляция; $r_{\text{ксл}}$ – корреляция с контрольными словами; ИСС – внутриклассовая корреляция; $r_{\text{инт пр}}$ – корреляция с интегральным профилем по списку.

Следует отметить, что все расчетные показатели очищенной выборки отличались от исходной (n=802) незначительно, например, среднее значение автокорреляции в результате чистки снизилось с 0,21 до 0,18; средняя корреляция с контрольными словами возросла с 0,84 до 0,87 и т.п.

Более полное представление о характере распределении оценок субъективного ВУ дает рисунок 1. Из-за высокой погрешности в гистограмму не включены оценки ВУ слов с низкой вероятностью узнавания (0,5 и ниже); таких оказалось 327⁸. На этом же рисунке отражено распределение оценок ВУ слов, включенных в состав словарных тестов для школьников.

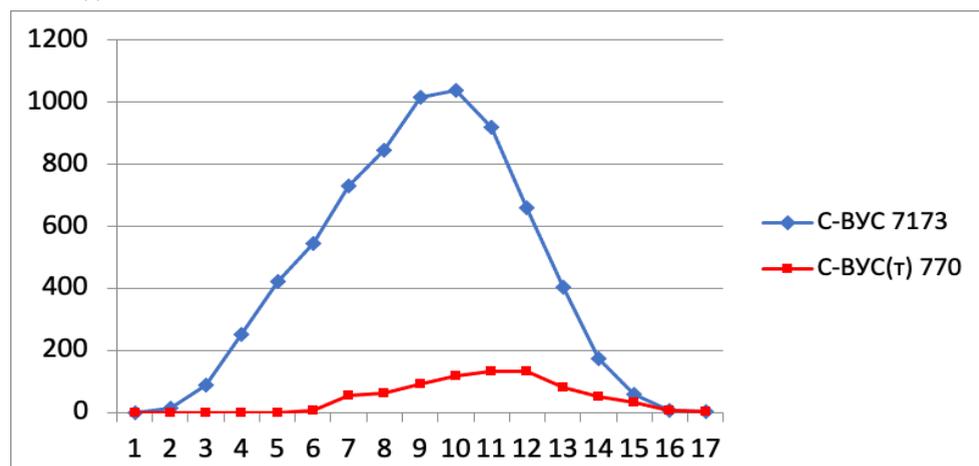


Рис. 1. Распределение субъективных оценок ВУ.

Примечание: С-ВУС 7173 – субъективные оценки ВУ слов с вероятностью узнавания более 0,5; С-ВУС(т) 770 – субъективные оценки ВУ слов, включенных в состав тестовых наборов.

Надежность субъективных оценок ВУ определялась путем расчета корреляции между средними оценками респондентов с четными и нечетными номерами. Из расчета исключались слова, у которых суммарная доля ответов «не знаю» и пропусков превышала 50%. Полученный коэффициент корреляции составил 0,925, что свидетельствует о высокой надежности субъективных оценок ВУ слов.

3) Результаты сбора объективных оценок ВУ слов

Учитывая примерно равную сложность субтестов и в целях упрощения дальнейшего изложения, анализ полученных результатов будет привязан преимущественно к отдельным субтестам, а не к индивидам.

Всем включенным в анализ участникам тестирования было предложено выполнить в общей сложности 53680 заданий; получено 53229 ответов и в 451 случае (0,84%) были зафиксированы пропуски.

Более детально (с дифференциацией по классам и субтестам) массив полученных в ходе тестирования данных, включая характеристики реагирования, точность ответов и надежность полученных оценок, представлен в табл. 5.

⁸ Включая два слова с нулевой узнаваемостью: *тенета* и *рулада*.

Таблица №5

Характеристики реагирования и точности выполнения заданий на опознание слов по классам и субтестам

Класс	Субтест	Кол-во заданий	Доля пропусков, в %	Правильные ответы			Надежность (KR-20)
				Размах	В среднем		
					кол-во	доля, %	
8	Словарь-3	150	0,8	41–133	83,6±23,8	0,557	0,954
	Словарь-4	150	0,7	47–136	84,4±21,0	0,563	0,938
6	Словарь-5	150	1,3	39–115	77,1±23,9	0,514	0,952
	Словарь-6	150	0,8	33–118	72,8±20,1	0,485	0,930
4	Словарь-7	120	1,4	28–108	68,1±20,4	0,568	0,951
	Словарь-8	120	0,6	40–99	71,1±15,5	0,593	0,918
2	Словарь-9	80	0,6	30–75	53,2±9,75	0,665	0,867
	Словарь-10	80	0,5	29–71	55,7±8,83	0,696	0,854

Примечание: KR-20 – коэффициент Кьюдера–Ричардсона №20.

Представленные в табл. 5 данные позволяют сделать следующие выводы. Включенные в анализ данные являются достаточно качественными, что подтверждается высокой надежностью тестовых оценок. Вместе с тем, если бы субъективные оценки ВУ точно отражали истинный возраст усвоения слов, следовало бы ожидать доли правильных ответов в районе 0,65–0,75, причем, исходя из условий подбора слов для тестов, наибольшей она должна была быть в 8-м классе. Однако результат оказался иным: во 2-м классе она оказалась наибольшей (0,68), тогда как в 8-м составила лишь 0,56.

Означают ли возможные смещения уровня субъективных оценок в отдельных диапазонах ВУ их неспособности к относительной дифференциации слов по возрасту усвоения? Для получения ответа на этот вопрос мы рассчитали корреляцию между субъективными оценками ВУ слов, включенных в субтест, и долей правильных опознаний этих слов респондентами. Все рассчитанные таким образом коэффициенты оказались, как и ожидалось, отрицательными, но умеренными по силе связи (диапазон от –0,319 до –0,547; средняя – 0,43). Однако эти величины могут ввести в заблуждение, поскольку они не учитывают того факта, что в субтесты включались слова, диапазон вариации субъективного ВУ которых был существенно ограничен.

В 1949 г. Р. Торндайком – на основе более ранних работ К. Пирсона – была предложена формула для коррекции *прямого* ограничения диапазона варьирования признака (range restriction) [Thorndike 1949, Case II formula], известная также как уравнение №18 Х. Галликсена [Gulliksen 1950: Eq. 18, p. 137]. С помощью этой формулы можно рассчитать «истинный» коэффициент корреляции между двумя переменными в условиях, когда диапазон варьирования одной из них ограничен:

$$R_{\text{корр}} = Ur / \sqrt{(U^2r^2 - r^2 + 1)}, \text{ где}$$

$R_{\text{корр}}$ – истинная корреляция между двумя переменными;

U – отношение вариативности переменной в неселективной (полной) выборке к таковой в селективной;

r – наблюдаемая корреляция между двумя переменными, вариативность одной из которых искусственно ограничена.

Данная формула полностью уместна в рассматриваемом случае, поскольку для тестирования учащихся каждого класса была использована лишь относительно небольшая часть общего тестового набора с ограниченным (как уже отмечалось выше – в 2,25–3,21 раза) диапазоном варьирования субъективных оценок ВУ слов [Wiberg, Sundström 2009].

Результат коррекции коэффициентов корреляции Пирсона для отдельных субтестов по формуле Торндайка представлен в табл. 6.

Таблица №6

Расчет скорректированных коэффициентов корреляции между успешностью выполнения тестовых заданий на опознание слов и субъективными оценками ВУ

Класс	Субтест	N	СтО с/т	U	r	R корр.
8	Словарь-3	70	0,86	2,279	-0,537	-0,823
	Словарь-4	61	0,89	2,202	-0,325	-0,603
6	Словарь-5	46	0,62	3,161	-0,358	-0,771
	Словарь-6	47	0,61	3,213	-0,481	-0,870
4	Словарь-7	48	0,64	3,063	-0,547	-0,895
	Словарь-8	46	0,64	3,063	-0,386	-0,788
2	Словарь-9	54	0,79	2,481	-0,348	-0,677
	Словарь-10	56	0,79	2,481	-0,419	-0,753

Примечание: СтО кл – стандартное отклонение субъективных оценок ВУ слов, включенных в состав субтеста; значения индексов U, r , Rкорр. объяснены в тексте.

Как можно видеть из данных, представленных в табл. 6, после коррекции значения коэффициентов корреляции оказались в диапазоне от -0,603 до -0,895. Средняя корреляция по всей выборке, рассчитанная с помощью z-преобразования частных коэффициентов, составила -0,79 для коэффициентов r Пирсона и -0,80 для ρ Спирмена, что соответствует зарубежным данным.

В то же время исследование выявило в составе каждого списка ряд слов, декларируемая знакомость которых (оцениваемая по субъективному ВУ) не подтверждается объективно, т.е. не соотносится с правильным пониманием их смысла. В частности, в отдельных субтестах можно было найти от двух до 14 слов, доля правильных ответов на которые составляла 20% и менее (т.е. не превышала вероятности случайного угадывания). Наиболее яркие примеры таких слов:

- *оригинал:* субъективный ВУ – 9,05 лет; однако в 4 классе доминирующим среди предложенных вариантов ответа оказался вариант «лидер» (63%), тогда как вариант «чудак» был выбран лишь однажды;

- *проходимец:* субъективный ВУ – 10,4 лет; доминирующим выбором в 4 классе оказался «незнакомый человек» (52%), тогда как вариант «мошенник» был выбран лишь однажды;

- *общепринятый*: субъективный ВУ – 10,9 лет; наиболее частыми выборами в 4 классе являются «многолюдный» (35%) и «достойный» (23%), тогда как вариант «обычный» оказался самым редким (4%);

- *безвозмездный*: субъективный ВУ – 10,0 лет; наиболее частым выбором в 4 классе является «бессмысленный» (41%), тогда как вариант «бесплатный» оказался самым редким (9%);

- *статный*: субъективный ВУ – 12,2 лет; наиболее частым выбором в 6 классе является «обладающий высоким статусом» (42%), тогда как вариант «стройный» оказался самым редким (8%);

- *злободневный*: субъективный ВУ – 13,3 лет; наиболее частыми выборами в 8 классе являются «агрессивный» (43%) и «угрожающий» (33%), тогда как вариант «актуальный» был выбран лишь в 7% случаев;

- *компрометировать*: субъективный ВУ – 13,75 лет; доминирующим выбором в 8 классе оказался «договариваться» (59%), тогда как вариант «порочить» был выбран лишь в 7% случаев.

Таким образом, хотя высокий уровень корреляций между субъективными и объективными оценками свидетельствует о валидности субъективного метода оценки ВУ по отношению к большим массивам слов, в отдельных случаях он может приводить к недооценке *фактического* возраста, когда слово стало интерпретироваться адекватно его реальному значению.

Второй способ валидации субъективных оценок ВУ предполагал сопоставление успешности выполнения одинаковых заданий учащимися смежных классов (за счет частичного совпадения слов, предъявляемых для опознания). Анализ показал, что во всех случаях имеет место существенный прирост успешности опознания в старшем классе. Более того, случаи более успешного опознания какого-либо слова в младшем классе по сравнению со старшим (инверсии) оказались крайне редки. Результаты сравнения успешности выполнения сходных заданий в парах ближайших классов обобщены в табл. 7.

Таблица №7

Характеристики прироста успешности опознания одинаковых слов в старшем классе по сравнению с младшим

Сравниваемые классы	Средняя успешность		Прирост успешности, в %	Число инверсий	Критерий Стьюдента	
	в младшем классе	в старшем классе			T	p
6 → 8	43,2	62,8	45,5	5 из 180	9,49	<0,001
4 → 6	45,6	61,3	35,9	3 из 50	3,54	<0,001
2 → 4	54,8	73,4	34,5	2 из 50	4,13	<0,001

Данный результат также свидетельствует в пользу валидности субъективных оценок ВУ слов в качестве средства изучения возрастной последовательности овладения словами русского языка.

Наконец, третий способ проверки валидности субъективных оценок ВУ слов состоял в сопоставлении их с объективными оценками ВУ слов детьми от 2,5 до 10 лет. С этой целью определялась корреляция 63 субъективных оценок, полученных

на первом этапе настоящего проекта, с объективными оценками ВУ этих слов детьми, полученными ранее в работе Григорьева и Ощепкова [Grigoriev, Oshhepkov 2013]. Как уже отмечалось выше, объективные оценки ВУ в этом случае были рассчитаны по результатам предъявления детям разного возраста изображений предметов с просьбой их назвать; в качестве ВУ принимался возраст, в котором вероятность правильного названия слова составляла не менее 75%. Корреляция двух рядов оценок оказалась равной 0,742; scatterграмма на рис. 2 иллюстрирует соотношение между объективными и субъективными оценками на уровне отдельных слов.

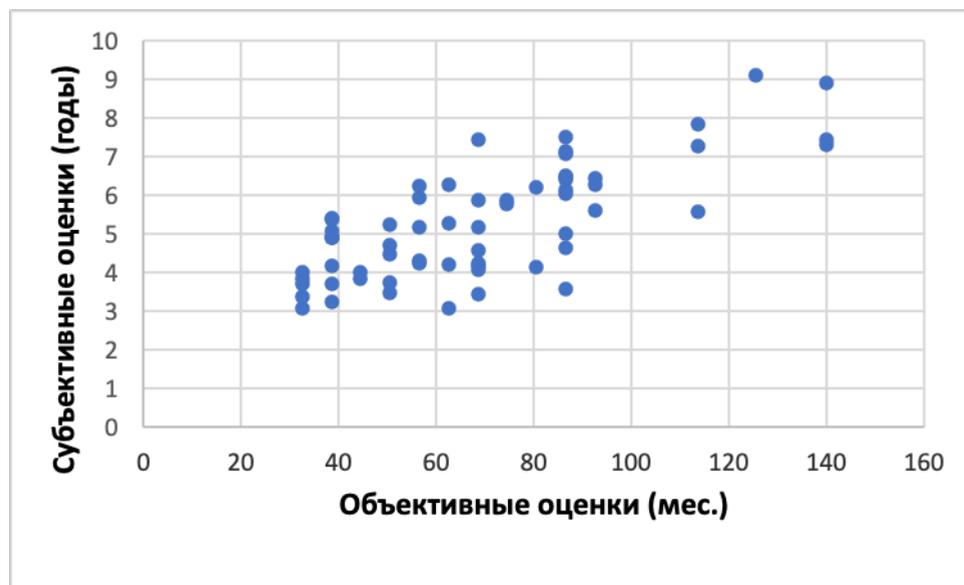


Рис. 2. Соотношение субъективных оценок возраста усвоения с объективными оценками (из [Grigoriev, Oshhepkov 2013])

При рассмотрении рисунка 2 обращает на себя внимание следующее. ВУ рано усваиваемых слов обычно переоценивается, в то время как ВУ поздно усваиваемых слов, напротив, недооценивается. Другими словами, средняя субъективная оценка рано усваиваемых слов бывает, как правило, выше действительной, в то время как субъективная оценка поздно усваиваемых слов, наоборот, часто оказывается ниже действительной. Таблица 1 в Приложении показывает это отчетливо. Из 26 слов, усваиваемых до 5 лет, только четыре (*лев, лиса, гора* и *свисток*) получили субъективные оценки ниже действительных, тогда как из 37 слов, усваиваемых позже 5 лет, субъективные оценки меньше действительных получили 32 слова (за исключением слов *колодец, зебра, сигарета, кактус, джинн*). Данный результат согласуется с продемонстрированным выше (табл. 5) отклонением в меньшую сторону (относительно прогнозируемой) доли правильных ответов при интерпретации смысла тестовых слов в старших классах школы. О тенденции немного преуменьшать возраст усвоения поздно усваиваемых слов говорится также в работе Моррисон с соавт. [Morrison, Chappell, Ellis 1997: 545].

Заклучение

Настоящее исследование не свободно от некоторых ограничений. В частности, нами не рассматривался вопрос о возможном влиянии на тесноту связей с объективными оценками повышающей коррекции субъективных оценок ВУ для тех слов, для которых часть ответов была представлена вариантом «не знаю». При выборе слов для формирования тестовых форм учитывалась такая слабоформализуемая характеристика, как «тестопригодность», что неизбежно могло приводить к нарушению соотношения между частями речи в разных субтестах. Некоторые задания могут быть не идеальны с точки зрения подбора дистракторов. Хотя все субтесты продемонстрировали приемлемую надежность (согласованность), при подготовке к следующему этапу исследований предполагается осуществить детальный анализ всех заданий, которые оказались недостаточно дискриминативными, с целью их возможной коррекции или замены. Более пристального внимания требуют выявленные факты смещения оценок в некоторых возрастных поддиапазонах, а также ошибочного позиционирования отдельных слов на временной шкале как следствие их иллюзорного «усвоения».

Тем не менее, полученные результаты свидетельствуют о том, что субъективные оценки дают неплохое представление о ВУ слов; об этом говорят довольно высокие корреляции субъективных оценок с объективными. Результаты исследования могут быть полезны для развития психолингвистики и формирования возрастных норм овладения словами, что позволит скорректировать практику написания учебно-методических, рекламных, тестовых, инструктивных и прочих текстовых материалов с учетом уровня языкового развития целевой аудитории.

В какой мере учет упомянутых во Введении (а возможно, и каких-то других) дополнительных факторов способен привести к сближению субъективных и объективных оценок ВУ? Решение этого вопроса может стать предметом дальнейших исследований.

Приложение

Таблица №1

Соотношение субъективных оценок возраста усвоения с объективными оценками (из [Grigoriev, Oshhepkov, 2013])

Слово	Объективная оценка ВУ в мес.	Субъективная оценка ВУ в годах	Субъективная оценка ВУ в мес.	«меньше» = субъективная оценка возраста меньше объективной
собака	32,5	3,08	36,96	
качели	32,5	3,38	40,50	
медведь	32,5	3,69	44,31	
мышь	32,5	3,83	46,00	
гриб	32,5	4,00	48,00	
улитка	38,5	4,92	59,04	
корова	38,5	3,70	44,35	
лошадь	38,5	3,23	38,73	

Теоретические и экспериментальные исследования

очки	38,5	4,91	58,96	
часы	38,5	5,09	61,04	
слон	38,5	4,18	50,14	
жираф	38,5	4,92	59,00	
молоток	38,5	5,39	64,79	
телефон	38,5	5,36	64,32	
тигр	44,5	4,00	48,00	
белка	44,5	3,83	45,91	
игла	50,5	4,70	56,35	
лев	50,5	3,48	41,74	меньше
светофор	50,5	5,23	62,77	
клоун	50,5	4,48	53,76	
лиса	50,5	3,73	44,73	меньше
гора	56,5	4,23	50,73	меньше
забор	56,5	5,93	71,14	
свисток	56,5	4,30	51,65	меньше
пила	56,5	6,25	75,00	
вешалка	56,5	5,16	61,92	
мальчик	62,5	3,09	37,04	меньше
колодец	62,5	6,27	75,23	
сова	62,5	4,22	50,61	меньше
зебра	62,5	5,26	63,13	
сигарета	68,5	7,44	89,28	
ваза	68,5	4,57	54,78	меньше
олень	68,5	4,23	50,80	меньше
свинья	68,5	3,43	41,22	меньше
кактус	68,5	5,87	70,43	
осел	68,5	4,07	48,86	меньше
кролик	68,5	4,17	50,00	меньше
верблюд	68,5	5,17	62,09	меньше
линейка	74,5	5,87	70,43	меньше
кошелек	74,5	5,77	69,27	меньше
овца	80,5	4,13	49,57	меньше
гвоздь	80,5	6,20	74,40	меньше
пушка	86,5	5,00	60,00	меньше
рубашка	86,5	6,04	72,52	меньше

пальто	86,5	6,50	78,00	меньше
коза	86,5	3,57	42,86	меньше
труба	86,5	6,14	73,64	меньше
грузовик	86,5	4,64	55,64	меньше
горилла	86,5	7,09	85,04	меньше
вулкан	86,5	6,39	76,70	меньше
джинн	86,5	7,52	90,24	
леопард	86,5	6,48	77,76	меньше
палатка	86,5	7,13	85,57	меньше
павлин	92,5	6,26	75,13	меньше
кувшин	92,5	6,43	77,14	меньше
кольцо	92,5	5,60	67,20	меньше
гайка	113,5	7,29	87,43	меньше
свитер	113,5	5,57	66,78	меньше
термос	113,5	7,85	94,20	меньше
микроскоп	125,5	9,09	109,09	меньше
рояль	140	7,44	89,28	меньше
конверт	140	7,32	87,82	меньше
телескоп	140	8,91	106,91	меньше

© Сугоняев К.В., Григорьев Н.А., 2023

Литература

Brysbaert M., Biemiller A. Test-based age-of-acquisition norms for 44 thousand English word meanings // *Behavior Research Methods*. 2017. Vol. 49. N. 4. P. 1520–1523. DOI: 10.3758/s13428-016-0811-4

Carroll J.B., White M.N. Word frequency and age of acquisition as determinants of picture-naming latency // *Quarterly Journal of Experimental Psychology*. 1973. Vol. 25. P. 85–95.

Desimone J.A., Harms P.D., Desimone A.J. Best practice recommendations for data screening // *Journal of Organizational Behavior*. 2015. Vol. 36. N. 2. P. 171–181. DOI: 10.1002/job.1962

Furr R.M. The double-entry intraclass correlation as an index of profile similarity: meaning, limitations, and alternatives // *Journal of Personality Assessment*. 2010. Vol. 92(1). P. 1–15. DOI: 10.1080/00223890903379134

Gilhooly K.J., Hay D. Imagery, concreteness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single-solution anagram // *Behavior Research Methods & Instrumentation*. 1977. Vol. 9. N. 1. P. 12–17.

Gilhooly K.J., Logie R.H. Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words // *Behavior Research Methods & Instrumentation*, 1980. Vol. 12. N. 4. P. 395–427.

Gottfried J., Ježek S., Králová M., Řiháček T. Autocorrelation screening: A potentially efficient method for detecting repetitive response patterns in questionnaire data // *Practical Assessment, Research, and Evaluation*. 2022. Vol. 27. Article 2. DOI: 10.7275/vyxb-gt24

Grigoriev A., Oshhepkov I. Objective age of acquisition norms for a set of 286 words in Russian: Relationships with other psycholinguistic variables // *Behavior Research Methods*. 2013. Vol. 45. N. 4. P. 1208–1217. DOI: 10.3758/s13428-013-0319-0

Gulliksen H. Theory of mental tests. New York: Wiley, 1950. DOI:10.1037/13240-000

Kuperman V., Stadthagen-Gonzalez H., Brysbaert M. Age-of-acquisition ratings for 30,000 English words // *Behavior Research Methods*. 2012. Vol. 44. N. 4. P. 978–990. DOI: 10.3758/s13428-012-0210-4.

Lyons A.W., Teer P., Rubenstein H. Age-at-acquisition and word recognition // *Journal of Psycholinguistic Research*. 1978. Vol. 7. N. 3. P. 179–187.

Morrison C.M., Chappell T.D., Ellis A.W. Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables // *The Quarterly Journal of Experimental Psychology*. 1997. Vol. 50A. N. 3. P. 528–559.

Thorndike R.L. Personnel selection: Test and measurement techniques. New York: Wiley, 1949.

Tsaparina D., Bonin P., Méot A. Russian norms for name agreement, image agreement for the colorized version of Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names // *Behavior Research Methods*. 2011. Vol. 43. N. 4. P. 1085–1099. DOI: 10.3758/s13428-011-0121-9

Wiberg M., Sundström A. A comparison of two approaches to correction of restriction of range in correlation analysis // *Practical Assessment, Research, and Evaluation*. 2009. Vol. 14. N. 5. URL: <http://pareonline.net/getvn.asp?v=14&n=5>. Дата обращения: 07.07.2023.

Словари

Ляшевская О.Н., Шаров С.А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009. 1090 с.

Сведения об авторах:

Сугоняев Константин Владимирович – кандидат технических наук, доцент, ассоциированный сотрудник Института психологии Российской академии наук (ФГБУН «ИП РАН»)

Контактная информация:

129366 Москва, ул. Ярославская, д.13

ORCID: 0000-0002-6207-7228

e-mail: skv-354@yandex.ru

Григорьев Николай Андреевич – независимый исследователь

Контактная информация:

127576, г. Москва, ул. Абрамцевская, д. 3А, кв. 11

ORCID: 0000-0001-9885-8088

e-mail: nikolay_grigorev_92@mail.ru

Для цитирования:

Сугоняев К.В., Григорьев Н.А. Оценка достоверности данных о возрасте усвоения слов, получаемых путем опроса взрослых респондентов // *Вопросы психолингвистики* №3(57) 2023, С. 92–110, doi: 10.30982/2077-5911-2023-57-3-92-110

UDC 159.922.72

LBC 81, 88.4

DOI 10.30982/2077-5911-2023-57-3-92-110

Research article

VALIDATION OF ADULT AGE-OF-ACQUISITION RATINGS FOR LARGE CORPUS OF RUSSIAN WORDS

Konstantin V. Sugonyaev

Institute of Psychology, Russian Academy of Sciences, Moscow, Russia

Nikolay A. Grigoriev

Independent scientist, Moscow, Russia

Abstract

Given the difficulty to establish an objective measure of age of acquisition (AoA), this study focuses on the subjective measures of this parameter, i.e., adult ratings, or judgment of word learning age. The question under study is the validity of subjective adult estimates of AoA as a standard norming tool for this characteristic in Russia. AoA estimates were gathered for 7,500 words by presentation to eight hundred participants. The latter were given 25 lists of words with the instruction to specify supposed age of their acquisition. A total of 179,100 estimates collected from 597 participants were included in the analysis (from 20 to 30 protocols on each list). The middle age of the participants was 20.50 ± 3.84 years, with 96% of them being at the age of 17–25 years old. Objective estimates were received through the examination of school children from the 2nd, 4th, 6th and 8th grade by means of the dictionary-type tests. The words for each test were selected so that their subjective AoAs were nearest to the actual age of students. Total number of generated tasks amounted to 720. The subset of test items for each grade was divided into two subtests equated on complexity. The subtest length varied from 80 (for the 2nd grade) to 150 (the 6th and 8th grades) tasks. Reliability of subtest scores varied in a range of 0.85–0.95. The comparison of subjective and objective estimates was carried out in several ways, first of all, by calculating the correlation coefficients between subjective AoA of test words and accuracy of their identification. After appropriate correction on direct range restriction of subjective AoA variation, corrected correlations in a range of 0.6–0.9 (0.79 on average) were received. This figure (along with the additional data) allows to draw a conclusion on a reasonable validity of subjective estimates of AoA as a norming tool for this characteristic. The article outlines the prospects for further research of factors influencing accuracy of adult estimates of word learning age in Russia.

Keywords: age of acquisition, subjective (ratings) and objective measures, reliability, validity, correction for range restriction.

Bionotes:

Konstantin V. Sugonyaev – Candidate of Technical Sciences, Associate Professor, Associate Fellow at the Institute of Psychology of the Russian Academy of Sciences.

Contact information:

Yaroslavskaya Str., 13, Moscow, Russia, 129366

ORCID: 0000-0002-6207-7228

e-mail: skv-354@yandex.ru

Nikolay A. Grigoriev – Independent scientist, Moscow, Russia

Contact information:

Abramtsevsкая St., 3A, apt. 11, Moscow, Russia, 127576

ORCID: 0000-0001-9885-8088

e-mail: nikolay_grigorev_92@mail.ru

For citation:

Konstantin V. Sugonyaev, Nikolay A. Grigoriev. Validation of adult age-of-acquisition ratings for large corpus of Russian words // Journal of Psycholinguistics. 3(57), 2023. P. 92–110. Available from: doi: 10.30982/2077-5911-2023-57-3-92-110 (In Russian)