

## ***Аспекты образованности***

---

**Фокина В.Н.**, кандидат социологических наук

**Лукьянова А.В.**, кандидат технических наук

**Широкова М.Е.**, кандидат социологических наук

## **ЛЕКСИЧЕСКОЕ БОГАТСТВО ЧЕЛОВЕКА И СТЕПЕНЬ ЕГО ОБРАЗОВАННОСТИ**

В статье представлены результаты исследования влияния возраста и уровня образованности человека на объем и качество его словарного запаса. Установлено, что зависимость объема словарного запаса человека от продолжительности образования носит практически линейный характер.

**Ключевые слова:** частотный словарь, словарный запас, образованность.

Образование и образованность сегодня становятся основными составляющими профессионального успеха и материального благополучия человека.

Понятие «образованность» используется в различных значениях. Так, в Толковом словаре Ушакова Д.Н. [1] образованность означает степень наличия образования, культурность. В энциклопедическом словаре педагога [2, с. 530] под образованностью понимается качество личности, характеризующее способность человека использовать в своей жизнедеятельности освоенные им продукты материального и духовного труда людей, т.е. их прошлого опыта и включать их в свой собственный опыт, руководствуясь действующими социальными нор-

мами и духовными ценностями в обществе. Существует и ряд других определений понятия «образованность». В своем исследовании авторы статьи остановились на определении, которое, на наш взгляд, отражает сущность понятия: образованность – уровень образования как совокупности знаний, полученных в процессе обучения [3].

Наиболее ярко образованного человека характеризует богатый словарный запас. Различают два вида словарного запаса: активный и пассивный. *Активный* словарный запас включает слова, которые используются в устной речи и письме. *Пассивный* словарный запас включает в себя слова, которые человек узнает при чтении и на слух, но не использует их сам в устной речи и письме.

Пассивный словарный запас больше активного в несколько раз.

Так, О. Генри в рассказе «Русские соболя» говорит о членах банды, которые, пользуясь лексиконом в двести слов, могут «вести между собой непринужденную беседу». У Г. Белля в его романе «И не сказал ни единого слова» герой, от имени которого ведется повествование, убежден, что словарный запас офицера вермахта составляют 120–140 слов, а в телефонных переговорах, которые ведутся через канцелярию епископа, он тоже не превышает у собеседников ста пятидесяти слов. Л. Фейхтвангер в романе «Успех» более точен в статистике: землевладелец располагал словарем в 612 слов, лексикон генерал-майора состоял из 412 слов, агент по перевозке пива обладал словарем из 724 слов, директор департамента пользовался словарем из 1453 немецких слов, 103 латинских, 22 французских, 12 английских и 1 русского слова [4, с. 6].

Богатый словарный запас свидетельствует об уровне интеллектуального развития человека. Общество воспринимает человека с богатым словарным запасом как умную и творческую личность.

Люди, получившие высшее образование имеют большой объем активного словаря, такую речь характеризуют разнообразие морфологических форм и синтаксических

конструкций. Образованный человек умеет наилучшим образом отбирать наиболее подходящие в каждом конкретном случае слова и профессиональные термины.

С объемом словарного запаса непосредственно связано лексическое богатство индивидуума как одно из слагаемых совокупного богатства речи – умение пользоваться разнообразными лексическими единицами [5, с. 3]. Лексическое богатство проявляется в том, что слова, не несущие специального коммуникативного намерения, применяются как можно реже. Этого можно достигнуть лишь при наличии у говорящего/пишущего большого словарного запаса.

«Лексическое богатство человека не остается постоянным и изменяется во времени» [6, с. 190]. Кроме того, культурный уровень, а следовательно, и словарный запас человека растет по мере освоения им новых знаний в процессе получения образования. Мы согласны с мнением авторов монографии «Качество образования: приглашение к размышлению»: «Можно говорить о качестве образования специалиста по тому, какое лексическое богатство он использует» [6, с. 194]. Так, на основе наблюдений над большим количеством респондентов установлено, что словарь среднего ребенка составляет около 3600 слов, а в 14

лет – уже 9000. Средний взрослый употребляет 11700 слов, а человек «повышенной интеллигентности» до 13500 [7, с. 7].

Однако, на наш взгляд, интереснее было бы проанализировать не количество слов в языке, а влияние возраста и уровня образованности человека на формирование и увеличение объема и качества его словарного запаса. Поэтому цель данного исследования – показать наличие зависимости лексического богатства личности от уровня его образованности, а также определить вид этой зависимости.

Первым шагом исследования стало создание частотного словаря общекультурных слов, сформированного в результате автоматического анализа текстов видных ученых и политиков – Д.С. Лихачева, Л.И. Абалкина, В.В. Путина, Д.А. Медведева и др., а также текстов газет общеполитической направленности: «Аргументы и факты», «Итоги», «Коммерсантъ», «Новая газета», «Правда», «Культура», «Российские вести» за 2008–2012 гг.

Теоретические рекомендации, выработанные в 1970-е годы Пиотровским и др., доказывали, что для достоверного описания 1600–1700 наиболее частотных слов достаточно использовать корпус размером 400 тыс. словоупотреблений [8, с. 3]. Общий объем текстов, подобран-

ных нами для формирования словарей, составляет около 20 миллионов словоупотреблений, т.е. можно говорить о достоверности полученных значений частот использования данных понятий.

Обработка текстов проводилась с помощью программного обеспечения, разработанного в Современной гуманитарной академии (СГА) интеллектуального робота «Живой русский язык» (далее – ИР ЖИЯЗ), которое позволяет в автоматизированном режиме проводить лемматизацию<sup>1</sup> загруженных в него текстов, преобразуя употребляемые в тексте слова в различных словоформах к нормальной<sup>2</sup> форме слова, а также проводить подсчет относительных частот использования слов в анализируемых текстах.

Для автоматического расчета частоты использования слова (ЧИС) в ПО «ИР ЖИЯЗ» применяется следующая формула:

$$\text{Частота использования (ЧИС)} = \frac{X}{Y},$$

где  $X$  – количество вхождений слова в корпус текстов;  $Y$  – общее количество

<sup>1</sup> Лемматизация – это преобразование словоформ к базовой (нормальной) форме слова.

<sup>2</sup> Нормальная форма слова – это слово в именительном падеже единственного числа (для существительных), для прилагательного добавляется признак мужского рода, для глагола – неопределенная форма.

тво словоупотреблений в корпусе текстов.

За единицу измерения ЧИС принято ЧМ – частота употребления на 1 миллион словоупотреблений.

Программное обеспечение ИР ЖИЯЗ позволило также исключить из дальнейшего рассмотрения повседневные слова, которые употребляются в обыденной речи.

В результате был сформирован частотный словарь общекультурных слов объемом 4059 слов, значения ЧИС которых лежат в интервале от 0,17 до 75,71 ЧМ. В словарь вошли понятия, которые, на наш взгляд, должен знать любой современный человек, считающий себя достаточно образованным и культурным.

Следующим шагом исследования стало выделение из общего объема полученного словаря наиболее трудных слов. На наш взгляд, использование таких слов является одним из ключевых показателей лексического богатства индивидуума.

Простейший экспертный анализ словаря показывает, что некоторые слова, в основном имеющие высокое значение ЧИС и стоящие в начале частотного списка, как правило, используются студентами и аспирантами СГА в нужном контексте. К таким словам, например, относятся: «сопровождение», «зачитывание», «интегрирование», «допустимость», «генерация», «декларирование»,

«ущемление», «сдержанность», «игнорирование», «трактоваться», «вовлеченность», «провозглашение» и др.

Однако, с уменьшением значения ЧИС, слова в частотном словаре становятся все более трудными для запоминания и редко используемыми. Поэтому интересно проанализировать как люди, имеющие различный уровень образованности, понимают и используют в своем непосредственном общении именно трудные слова, не являющиеся общеупотребимыми, значения которых не всегда понятно и, как правило, требует специального заучивания.

Для дальнейшего исследования словаря разобьем его на пять кластеров. Принципиально важным является определение границ кластеров.

Для определения цены деления шкалы интервал, в котором лежат указанные значения ЧИС, был разделен на пять равных отрезков (по предполагаемому количеству кластеров).

В результате получили цену деления  $(7,571 \times 10^{-5} - 1,695 \times 10^{-7})/5 = 1,5 \times 10^{-5}$ .

При такой цене деления в первый кластер из 2141 понятия словаря попало всего одно слово – «интервью» (табл. 1).

Во второй кластер не попало ни одного слова, так как в интервале ЧИС от  $4,5 \times 10^{-5}$  до  $6,0 \times 10^{-5}$  нет ни одного понятия.

Таблица 1

**Средняя частота употребления  
понятий в частотном словаре  
общекультурных слов (фрагмент  
частотного словаря)**

№ п/п	Понятия из корпусов текстов	Средняя частота использования понятий в корпусах текстов 1990–2010 годов
1	интервью	7,571*10 <sup>-5</sup>
2	аккредитация	4,304*10 <sup>-5</sup>
3	вторгаться	4,236*10 <sup>-5</sup>
4	контекст	3,728*10 <sup>-5</sup>
5	равномерность	3,504*10 <sup>-5</sup>
6	валентность	3,437*10 <sup>-5</sup>
7	сопроводить	1,746*10 <sup>-5</sup>
8	зачитывание	1,593*10 <sup>-5</sup>
9	интегрирование	1,593*10 <sup>-5</sup>
10	допустимость	1,542*10 <sup>-5</sup> E
11	генерация	1,424*10 <sup>-5</sup>
...	.....	.....
4059	явность	1,695*10 <sup>-7</sup>

В третьем кластере с величиной ЧИС от  $3,0 \times 10^{-5}$  до  $4,5 \times 10^{-5}$  лежат пять понятий (строки 2–6 табл. 1).

В четвертом кластере с величиной ЧИС от  $1,5 \times 10^{-5}$  до  $3,05 \times 10^{-5}$  лежат четыре понятия (строки 7–10 табл. 1).

Все остальные понятия словаря (4059) вошли в пятый кластер.

Таким образом, данные таблицы 1 показывают, что рассчитанная цена деления шкалы не удачна для последующего разбиения словаря на кластеры.

Поэтому было принято решение для расчета цены деления отбросить максимальные значения ЧИС и взять диапазон ЧИС в интервале

$$1,695 \times 10^{-7} \leq \text{ЧИС} \leq 1,542 \times 10^{-5}$$

В результате получаем цену деления, равную

$$(1,542 \times 10^{-5} - 1,695 \times 10^{-7}) / 5 \approx 3 \times 10^{-6}$$

Используя полученную равномерную шкалу с ценой деления  $3 \times 10^{-6}$ , словарь достаточного уровня общей культуры разбили на пять кластеров.

В кластер №1 вошло 40 понятий с ЧИС в интервале  $1,271 \times 10^{-5} \leq \text{ЧИС} \leq 7,571 \times 10^{-5}$ .

В кластер №2 вошло 68 понятия с ЧИС в интервале  $9,153 \times 10^{-6} \leq \text{ЧИС} \leq 1,135 \times 10^{-5}$ .

В кластер №3 вошло 282 понятие с ЧИС в интервале  $6,097 \times 10^{-6} \leq \text{ЧИС} \leq 8,986 \times 10^{-6}$ .

В кластер №4 вошло 1050 понятий с ЧИС в интервале  $5,931 \times 10^{-6} \leq \text{ЧИС} \leq 3,05 \times 10^{-6}$ .

В кластер №5 вошло 2619 понятие с ЧИС в интервале  $1,695 \times 10^{-7} \leq \text{ЧИС} \leq 3,042 \times 10^{-6}$ .

Таким образом, по результатам анализа кластеров можно сказать, что с увеличением номера кластера в него попадают более редкие слова. Чем меньше значение ЧИС, тем реже слово употребляется в повседневной речи культурного человека.

Для определения степени «узнаваемости» понятий словарь,

разбитый на кластеры, представили 10 экспертам, из которых 7 чел. имеют ученые степени и звания, 3 чел. – высшее образование, и предложили им отнести каждое понятие к одной из пяти групп по уровню «узнаваемости».

Были предложены следующие уровни «узнаваемости»:

1. Знаю точное определение.
2. Известно примерное определение.
3. Определяю значение в конкретном тексте.
4. Понятие встречалось, но смысл не совсем ясен.

5. Впервые встречаю понятие.

Свое мнение об уровне «узнаваемости» понятий словаря эксперты заносили в таблицу (табл. 2), где цифрой 1 обозначено, к какому уровню «узнаваемости» относится, по мнению эксперта, то или иное понятие.

Эксперты установили, что знают точное определение всех слов, входящих в кластер № 1. Для кластера № 5, в который вошли слова с наименьшей величиной ЧИС, степень точного «узнавания» экспертами наименьшая – 91,3% от объема кластера.

Таблица 2

**Пример отчета эксперта об уровне «узнаваемости» понятий из словаря общекультурных слов (фрагмент)**

№ п/п	Понятия из корпусов текстов	Знаю точное определение	Известно примерное определение	Определяю значение в конкретном тексте	Понятие встречалось, но смысл не совсем ясен	Впервые встречаю понятие
1	Интервью	1				
2	Аккредитация	1				
...	....	....	...	....	....	....
1203	Афронт			1		
1204	Безжизненный	1				
...	....	....	...	....	....	....
1258	Корреляция		1			
1259	Неизменность	1				
...	....	....	...	....	....	....
2139	Элиминация					1
2140	Энтропия	1				
...	...	...	...	...	...	...

Впервые экспертам встречаются слова только из четвертого (всего одно понятие – «венчурный») и из пятого кластеров (в среднем – 22 понятия – «адгезия», «бифуркация», «атараксия», «андеррайтинг», «аффинаж», «адиабатический», «элиминация», «интроспекция», «интенция», «инсайт», «инкорпорировать», «имманентный», «вигоросный», «вальвация», «анизотропия», «интерстициальный», «охлократия», «инволюция», «штудия», «гедонизм», «блистер» и «агломерация»).

В результате обработки данных о степени «узнаваемости» экспертами понятий общекультурного словаря из него выделено 436 слов, вызывающих наибольшее затруднение у экспертов.

Следующая задача исследования – установить степень «узнаваемости» трудных понятий из словаря «достаточного уровня культуры» различными группами респондентов, отличающихся по совокупной продолжительности обучения.

Для исследования выбрано пять групп респондентов:

- выпускники 9-го класса (продолжительность обучения – 9 лет),
- выпускники 11-го класса (продолжительность обучения – 11 лет),
- выпускники вуза (продолжительность обучения, начиная со школьного возраста, – 15 лет),
- аспиранты, завершившие обучение (общая продолжительность обучения – 18 лет),

- профессорско-преподавательская элита (общая продолжительность обучения – от 23 лет).

В опросе в качестве респондентов первых двух групп приняли участие школьники- выпускники 9-го и 11-го классов школы. Продолжительность обучения респондентов этих групп составляет 9 и 11 лет соответственно.

Респондентами третьей группы выступали выпускники вузов, завершившие обучение в 2011–2013 гг. Совокупная продолжительность обучения выпускников вуза (с учетом обучения в школе) – 15 лет.

Респонденты четвертой группы – аспиранты, завершившие обучение в аспирантуре, и «молодые» кандидаты наук, защитившие диссертации в 2011–2013 гг. Общая продолжительность обучения респондентов этой группы с учетом окончания школы, вуза и аспирантуры – 18 лет.

В качестве респондентов пятой группы была привлечена профессорско-преподавательская «элита», т. е. доктора наук, непрерывно обучающиеся в процессе своей профессиональной деятельности. За общую продолжительность обучения данной группы респондентов принято минимальное значение – 23 года, учитывающее, что защита докторской диссертации возможна не ранее чем через 5 лет после защиты кандидатской диссертации.

Всего было опрошено 365 респондентов, в среднем по 73 респондента в каждой группе.

Каждый респондент оценивал понятия из словаря по следующим критериям:

- «знаю»,
- «не уверен»,
- «не знаю».

Первоначально опрос респондентов вели по двум критериям: «знаю», «не знаю». Однако большинство респондентов затруднялись отнести некоторые слова к той или иной категории, поэтому было принято решение – ввести дополнительный критерий – «не уверен».

Свое мнение об уровне «узнаваемости» понятий словаря респондент заносил в соответствующий столбец таблицы (табл. 3), где цифрой 1 или любым другим символом обозначил, к какому уровню «узнаваемости», по его мнению, относится то или иное понятие.

Результаты опроса респондентов каждой группы проанализированы, крайние значения (минимальные и максимальные) в каждой группе отброшены. Средние значения о количестве «трудных» понятий и процентном отношении к общему количеству анализируемых слов по уровню их «узнаваемости» приведены в табл. 4.

Дальнейшую обработку полученных результатов вели для двух критериев «знаю», «не знаю», учитывая, что ответы респондентов по критерию «не уверен» нельзя назвать достаточно достоверными.

Анализ данных табл. 4 показал, что с увеличением продолжительности обучения количество трудных слов, смысл которых понимает респондент, возрастает. График зависимости количества трудных слов, которое знают респонденты ( $Y$ ), от продолжительности обучения ( $X$ )

Таблица 3

### Пример отчета респондента об уровне «узнаваемости» понятий из словаря общекультурных слов (фрагмент)

№ п/п	Понятие	Степень «узнаваемости» понятия		
		знаю	не уверен	не знаю
1	Интервью	1		
2	Аккредитация		1	
3	Вторгаться	1		
...	....	....	...	....
436	Энтропия			1

### Результаты опроса групп респондентов об уровне узнаваемости ими понятий из словаря общекультурных слов

№ п/п	Группа респондентов	Продолжительность обучения, лет	Количество понятий, распределенных респондентами по уровням «узнаваемости»					
			знаю		не уверен		не знаю	
			количество понятий	% от общего объема трудных понятий	количество понятий	% от общего объема трудных понятий	количество понятий	% от общего объема трудных понятий
1	Выпускники 9-го класса	9	158	36	77	18	201	46
2	Выпускники 11-го класса	11	212	49	76	17	148	34
3	Выпускники вуза	15	265	61	90	21	81	19
4	Выпускники аспирантуры	18	340	78	33	8	63	14
5	Профессорско-преподавательская «элита»	от 23	416	95	10	2	11	3

носит практически линейный характер (рис. 1).

Используя возможности электронной таблицы Excel, построим линию регрессии. С достоверностью  $R^2 = 0,9908$  зависимость  $Y$  от  $X$  описывается линейной функцией

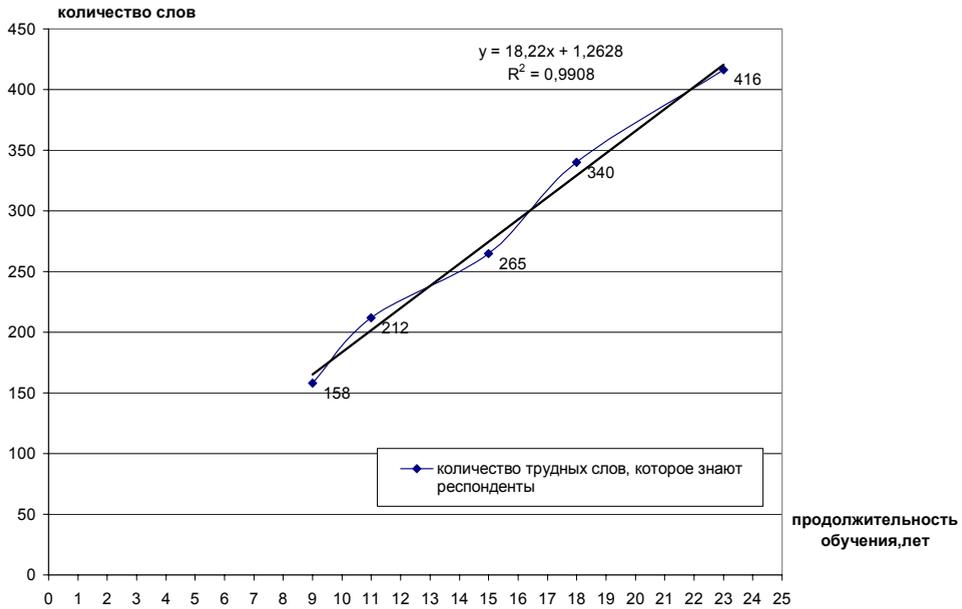
$$Y = 18,22X + 1,2628, \quad (1)$$

где  $Y$  – количество трудных слов, «узнаваемых» респондентом;  $X$  – продолжительность обучения.

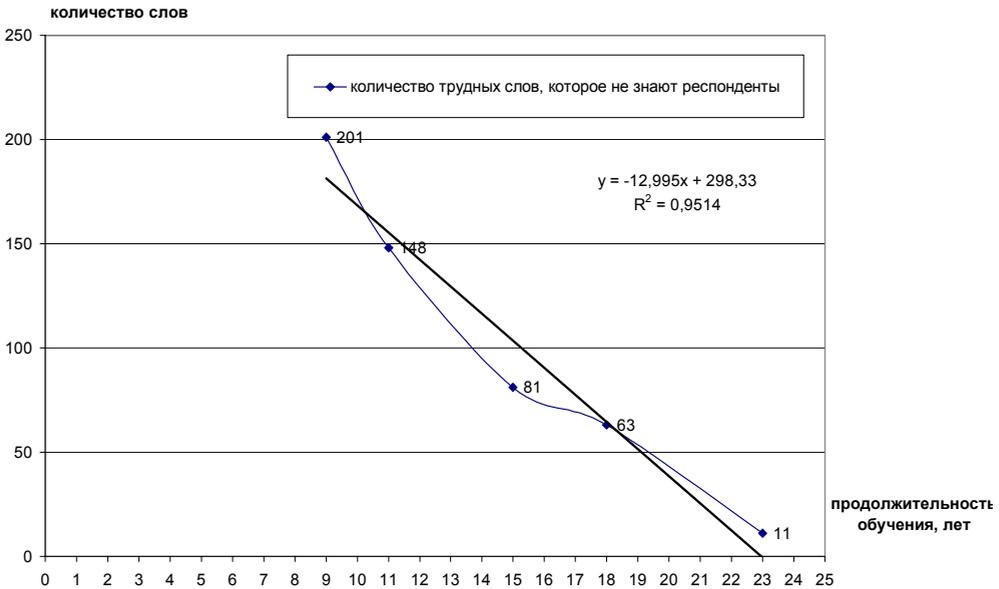
Таким образом, зная совокупную продолжительность обучения, и подставив это значение в формулу (1), можно легко рассчитать степень образованности респондента.

На рис. 2 показана зависимость количества слов, которое не знают респонденты ( $Y$ ), от продолжительности обучения ( $X$ ).

Из рис. 2 видно, что с увеличением общей продолжительности обучения, количество слов, значе-



**Рис. 1.** Зависимость количества трудных слов, которые знают респонденты, от совокупной продолжительности их обучения



**Рис. 2.** Зависимость количества трудных слов, которое не знают респонденты, от совокупной продолжительности их обучения

ние которое респондент не знает, а следовательно, не может применять при вербальном общении, написании сочинений, рефератов, курсовых, дипломных работ, статей и диссертаций уменьшается.

С достоверностью  $R^2 = 0,9514$  график зависимости аппроксимируется уравнением регрессии вида:

$$Y = -12,995X + 298,33. \quad (2)$$

Приведенная зависимость и график позволяют нам с высокой достоверностью сказать, что чем дольше человек учится, чем разнообразнее круг его интересов, тем богаче его словарный запас, так как в процессе получения новых знаний его «живой» язык обогащается все новыми и новыми понятиями.

По результатам проведенного опроса были сформированы перечни

слов, вызывающие затруднение для каждой группы респондентов.

Вероятно, у респондентов одной группы затруднение в распознавании могут вызывать разные слова. Поэтому нами проанализированы слова, вызывающие затруднение у респондентов группы и по каждой из них составлены два перечня:

- слов, которые не знает хотя бы один респондент из группы,
- слов, которые не знают все респонденты группы.

Сводные данные по количеству таких слов для каждой группы показаны в табл. 5.

Из табл. 5 видно, что перечень трудных слов, которые не знают респонденты первой группы (школьники 9-го класса) содержит 314 неповторяющихся слов, тогда как среднее

Таблица 5

### Сведения о количестве слов, которое не знают респонденты различных групп

№ п/п	Группа респондентов	Количество слов, которое не знают респонденты		
		среднее по группе	не знают все	не знает хотя бы 1 из группы
1	Выпускники 9-го класса	201	104	314
2	Выпускники 11-го класса	148	77	224
3	Выпускники вуза	81	20	159
4	Выпускники аспирантуры	63	11	168
5	Профессорско-преподавательская «элита»	11	4	41

количество слов, которое не знает выпускник 9-го класса – 201. Перечень слов, которые не знают все респонденты этой группы, содержит 104 слова.

Перечень трудных слов, которые не знает хотя бы один респондент пятой группы (профессорско-преподавательская «элита») содержит 41 неповторяющиеся слово, при том, что среднее количество понятий, которое не знает каждый из этой группы равно 11. На наш взгляд, такое расхождение связано с различиями в профессиональной направленности респондентов этой группы (в опросе принимали участие доктора педагогических, технических, экономических, философских, медицинских наук).

Интересно отметить, что каждый респондент группы «профессорско-преподавательская элита» не знает 4 слова: атараксия, аффинаж, интерстициальный, рестайлинг.

На основе проведенного исследования можно сделать вывод, что чем дольше человек учится, занимается научной деятельностью, тем большим лексическим богатством он обладает. У каждого индивидуума лексическое богатство свое, и это, прежде всего, либо заслуга, либо упущение образования, так как качество образования – это создание условий для максимального

развития личности, и прежде всего в части овладения всем богатством речевого общения.

## Литература

1. Толковый словарь русского языка: в четырех томах / Под ред. Д.Н. Ушакова. М.: ОГИЗ, 1935–1940.
2. Безрукова В.С. Основы духовной культуры (энциклопедический словарь педагога). Екатеринбург, 2000.
3. Ефремова Т.Ф. Новый словарь русского языка. Толково-словообразовательный. М.: Русский язык, 2000.
4. Алексеев П.М. Частотные словари. СПб.: Изд-во СПб. ун-та, 2001.
5. Жеребило Т.В. Словарь лингвистических терминов: 5-е изд., испр. и доп. Назрань: Пилигрим, 2010.
6. Пугач В.Н., Кирсанов К.А., Алимова Н.К. Качество образования: приглашение к размышлению: Монография. М.: Изд-во Торговая корпорация «Дашков и К», 2012.
7. Звегинцев В.А. Очерки по общему языкознанию. 2-е изд., испр. М.: Либроком, 2009.
8. Ляшевская О.Н., Шаров С.А. Частотный словарь современного русского языка на материалах Национального корпуса русского языка. М.: Азбуковник, 2009.

## ASPECTS OF EDUCATION

**Fokina V.N.**, *PhD in Sociology*

**Lukyanova A.V.**, *PhD in Technical Sciences*

**Shirokova M.E.**, *PhD in Sociology*

### **Lexical wealthy of man and level of education of the person**

The article presents the results of studies of the effect of age and level of education of the person on the volume and quality of his vocabulary. The dependence of the volume of the human vocabulary of the duration of education is almost linear.

**Key words:** *frequency dictionary, vocabulary, education.*