

Качество образования

О.М. Карпенко, кандидат экономических наук

А.В. Абрамова, научный сотрудник

Т.Ю. Семенова, кандидат социологических наук

М.Е. Широкова, кандидат социологических наук

ЛИНГВОСТАТИСТИЧЕСКИЙ АНАЛИЗ СОДЕРЖАНИЯ ТЕКСТОВ ОТЗЫВОВ О ВУЗАХ МОСКВЫ В СЕТИ ИНТЕРНЕТ

Методами лингвостатистики проведен анализ отзывов студентов, размещенных в Интернете. Изучены количественные закономерности «живого» русского языка, проявляющиеся в современных текстах на примере текстов отзывов о вузах, размещенных в Интернете. Исследование позволило выдвинуть и подтвердить гипотезу о том, что авторы положительных отзывов имеют более высокий уровень образованности и позитивные жизненные установки.

Ключевые слова: образование, анализ текстов, частота употребления слова, образовательный портал, социолингвистика, лингвостатистика, лексико-семантическая группа

Перед будущим абитуриентом часто встает вопрос: как выбрать вуз? В числе рекомендаций, которые дают различные образовательные порталы (например, «Вузы России» <http://www.1000vuzov.ru/>, «ЕГЭметр» <http://www.egemetr.ru/>, «Учеба» <http://www.ucheba.ru/>, «Поступление» <http://postyplenie.ru/> и др.), предлагаются отзывы о вузе. Авторы подобных рекомендаций считают, что отзывы о вузах в Интернете позволяют потребителям образовательных услуг (студентам, их родителям, общественным организациям и работодателям) иметь объективную оценку вузов. На наш взгляд, отзывы о вузе в Интернете могут дать динамичную оценку о вузе на определенный момент времени, так как баланс положительных и отрица-

тельных отзывов постоянно колеблется в ту или иную сторону. Более того, вузы, отзывы о которых чаще всего просматриваются, лидируют как среди лучших, так и среди худших.

Специалистами СГА было проанализировано несколько сайтов, на которых размещены отзывы о вузах: «Всеобуч» www.edu-all.ru, «Указка» <http://www.ykazka.ru>, «Учеба-отзыв» <http://ucheba-otziv.ru>, форум сайта «Антимух» <http://forum.antimuh.ru>, сайт <http://www.galya.ru>. С целью соблюдения объективности исследования, для лингвостатистического анализа были выбраны отзывы на самые обсуждаемые на указанных сайтах вузы (государственные и негосударственные), которые в рейтинге позитивных и не-

гативных занимали практически равное положение:

- Российский государственный социальный университет (РГСУ);
- Институт туризма и гостеприимства (ИТИГ);
- Московская академия образования Натальи Нестеровой;
- Московский институт национальных и региональных отношений (МИНРО);
- Всероссийский заочный финансово-экономический институт (ВЗФЭИ);
- Современная гуманитарная академия (СГА).

Близость положительных и отрицательных отзывов об одном и том же вузе можно проиллюстрировать на следующих примерах: в рейтинге положительных отзывов РГСУ занимал третье место, а в рейтинге отрицательных – первое; Академия Натальи Нестеровой – седьмая среди положительных отзывов и вторая среди отрицательных; Институт туризма и гостеприимства занимал практически одинаковое место как в положительном, так и в отрицательном рейтинге – 14 и 15 соответственно. На сайте «Учеба-отзыв» рейтингование вузов по отзывам не ведется, а имеется только возможность увидеть, какое количество положительных («зеленых»), нейтральных («желтых») и отрицательных («красных») отзывов разместили пользователи о том или ином вузе.

Чтение различных сообщений о вузах даже при поверхностном анализе их текстов позволяет нам предположить, что **авторы положительных отзывов**, как правило, обладают более высоким уровнем культуры, в том числе более грамотной письменной речью, богатым сло-

варным запасом, употребляют в своих сообщениях меньше жаргонных слов, т. е. они **более образованные люди и имеют позитивные жизненные установки.**

Для проверки этого предположения была использована методика лингвостатистического анализа текстов с использованием современных автоматизированных средств, предложенная в статье «Статистические методы анализа текста в социолингвистике для оценки образованности личности» [1, с. 51–65]. Правомерность применения такого анализа научно подтверждена специалистами в области социолингвистики¹, которые установили, что тексты являются показателями определенных качеств людей [2]. При автоматизированной обработке текстов применялось разработанное в СГА специализированное программное обеспечение – интеллектуальный робот «Живой русский язык» (ИР ЖИЯЗ) (Свидетельство о государственной регистрации программы для ЭВМ № 2014614769 от 07.05.2014 «Электронный образовательный ресурс. Интеллектуальный робот “Живой русский язык”»). ИР ЖИЯЗ в автоматизированном режиме проводит лемматизацию² загруженных в него текстов, преобразуя употребляемые в тексте слова в различных словоформах к нормальной³

¹ Социолингвистика – это раздел языкознания, изучающий связь между языком и социальными условиями его бытования.

² Лемматизация – это преобразование словоформ к базовой (нормальной) форме слова.

³ Нормальная форма слова – это слово в именительном падеже единственном числе (для существительных), для прилагательного добавляется признак мужского рода, для глагола – неопределенная форма.

форме слова, а также проводит подсчет относительных частот использования слов в анализируемых текстах. Используя подсчет относительных частот употребления различных слов в текстах отзывов, специалисты СГА оценили уровень образованности и жизненные установки тех, кто размещает свое мнение о вузах.

С целью проведения автоматизированного лингвостатистического анализа текстов все размещенные отзывы скопированы с указанных выше сайтов и распределены на две группы: положительные и отрицательные. Специалистами в области лингвостатистики в процессе проведения научных исследований установлено [3], что для проведения объективного сравнительного анализа текстов, объемы их корпусов должны быть сопоставимы между собой. В результате сформировано два корпуса текстов: корпус отрицательных отзывов объемом 2,3 МБ, содержащий 59238 словоупотреблений и корпус положительных отзывов объемом 2,8 МБ, содержащий 61676 словоупотреблений. Таким образом, сформированные корпуса текстов сопоставимы по объемам, что позволяет провести объективный анализ частотных характеристик слов, употребляемых в данных текстах.

В результате обработки текстов ИР ЖИЯЗ, из отзывов выделено 5969 неповторяющихся слов – значимых частей речи (без учета предлогов, союзов, частиц и числительных), причем из текстов позитивных отзывов – 4161 слово, из текстов негативных отзывов – 4128 слов. По каждому из выделенных слов исследованы следующие показатели,

характеризующие частоту употребления слова в текстах отзывов: частота на миллион словоупотреблений¹ и ранг слова.

Частота на миллион словоупотреблений [4, с. 116–127] (относительная частота употребления слова) (ЧМ) рассчитывается как отношение количества употреблений слова в тексте (X) к общему количеству словоупотреблений в тексте (Y):

$$\text{ЧМ} = (X/Y) \cdot 10^6,$$

где X – количество употреблений слова в корпусе текстов (абсолютная частота); Y – общее количество словоупотреблений в тексте.

Частота на миллион словоупотреблений показывает долю употребления слова в тексте. Ранг слова – частотный порядковый номер слова. Слово с самой большой частотой будет иметь ранг 1.

Интересным, на наш взгляд, представляется анализ слов, встречающихся в текстах двух групп, но существенно отличающихся по частоте употребления. Такие слова распределены нами на несколько лексико-семантических групп², наиболее ярко характеризующих авторов отзывов: «Образование», «Карьера», «Деньги», «Жаргонизмы». Рассмотрим далее, как часто употребляют авторы позитивных и негативных отзывов различные слова в выделенных семантических группах.

¹ Словоупотребление – это любая единица текста, кроме знаков препинания и чисел в виде цифровых обозначений.

² Лексико-семантическая группа – совокупность слов, относящихся к той же части речи, объединенных внутриязыковыми связями на основе взаимообусловленных и взаимосвязанных элементов значения.

Самой многочисленной группой является «Образование». В употреблении авторами отзывов слов из этой группы наблюдается следующая тенденция: часть слов чаще употребляется авторами негативных отзывов, а другая часть – авторами позитивных. С учетом этого все слова группы разделены на две подгруппы:

- в первую подгруппу вошли слова, которые чаще употребляются авторами положительных отзывов;

- во вторую группу вошли слова, чаще употребляемые авторами отрицательных отзывов.

Анализ частотных характеристик слов семантической группы «Образование» в текстах позитивных и негативных отзывов показан в табл. 1.

Анализ слов, относящихся к разным подгруппам, позволяет сделать, интересные, на наш взгляд, наблюдения. В первую подгруппу вошли слова, связанные

Таблица 1

Сравнительный анализ частотных характеристик слов семантической группы «Образование», употребляемых в текстах отзывов о вузах

№ п/п	Слово	Частота на миллион словоупотреблений, ЧМ	
		положительные отзывы	отрицательные отзывы
1	2	3	4
Подгруппа I			
1.1	Учиться	13178	8368
1.2	Преподаватель	5285	3291
1.3	Знание	4290	2601
1.4	Образование	2056	1630
1.5	Учеба	1526	1066
1.6	Группа	1282	909
1.7	Закончить	1238	752
1.8	Читать	730	627
1.9	Понимать	884	407
1.10	Заниматься	663	376
1.11	Преподавать	464	376
1.12	Оценка	553	282
1.13	Заканчивать	619	188
1.14	Наука	354	94

Таблица 1. Продолжение

1	2	3	4
1.15	Научиться	199	63
1.16	Книга	133	63
1.17	Выучить	155	31
1.18	Изучение	111	31
1.19	ИТОГО по подгруппе I	33720	21155
Подгруппа II			
2.1	Диплом	2587	3448
2.2	Обучение	1680	2539
2.3	Специальность	907	1222
2.4	Учебный	862	1222
2.5	Занятие	730	1348
2.6	Зачет	708	972
2.7	Методист	708	972
2.8	Сайт	663	1128
2.9	Расписание	619	909
2.10	Практика	509	784
2.11	Аккредитация	464	1034
2.12	Заведение	464	1034
2.13	Декан	464	690
2.14	Абитуриент	442	784
2.15	Пара	420	533
2.16	Ректор	398	752
2.17	Комиссия	332	564
2.18	Поступление	310	784
2.19	Перевестись	287	784

Таблица 1. Окончание

1	2	3	4
2.20	Кафедра	221	439
2.21	Переводиться	177	251
2.22	Рейтинг	177	251
2.23	Договор	133	439
2.24	Справка	133	439
2.25	Перейти	133	282
2.26	Руководство	133	251
2.27	Отчислять	133	219
2.28	Отчислить	111	564
2.29	Семинар	111	282
2.30	Контракт	88	313
2.31	Дипломный	88	157
2.32	Тест	66	219
2.33	Аккредитованный	66	157
2.34	Приказ	44	345
2.35	Перевести	44	251
2.36	Отчисление	44	219
2.37	Пересдать	44	94
2.38	Пятерка	44	63
2.39	Перевод	22	282
2.40	Отчисленный	22	188
2.41	Заваливать	22	125
2.42	Репетитор	22	125
2.43	Завалить	22	125
2.44	Содрать	22	94
2.45	Переписывать	22	63
2.46	Незнание	22	63
2.47	Двоечник	22	63
2.48	Вылететь	22	63
2.49	Реферат	22	63
2.50	ИТОГО по подгруппе II	15786	27992

не просто с образованием, а именно с активным получением знаний, самообразованием (*учиться, знание, читать, образование, понимать, заниматься, научиться, выучить и пр.*). Суммарная частота употребления слов данной подгруппы в тек-

стах положительных отзывов в 1,6 раза выше, чем в текстах отрицательных отзывов (33720 ЧМ против 21155 ЧМ).

Иная картина складывается для слов второй подгруппы, куда вошли слова образовательной тематики, которые характеризуют образование не с позиции получения знаний, а с позиций оценки, проверки, сдачи различного рода процедур проверки знаний. Именно поэтому во вторую подгруппу вошли слова: *занятие, зачет, пара, тест, пятерка, двоечник, завалить, заваливать, пересдать, переписывать*. Большую частоту имеют у авторов негативных отзывов и такие слова, как *перевестись, перевести, перевод, переводиться, отчислить, отчисление, отчисленный, вылететь*. Так, слова *отчислить, отчисление, перевести* встречаются в текстах негативных отзывов в 5 раз чаще, чем в позитивных; слово *отчисленный* встречается в 8 раз чаще, а слово *перевод* – почти в 13 раз (частота употребления в позитивных отзывах – 22 ЧМ, в негативных – 282 ЧМ), т. е. проблема перевода и отчисления для авторов таких отзывов является актуальной. Как видно из табл. 1, суммарная относительная частота употребления слов этой подгруппы в 1,8 раза выше в отрицательных отзывах (27992 ЧМ), чем в положительных (15786 ЧМ). Это подтверждает наш вывод о том, что для тех, кто пишет негативные отзывы приоритетным является не получение образования, расширение знаний, умений, а возможность «отучиться», «сдаться», «списать», т. е. получить диплом посредством обмана, а не благодаря повышению уровня своей образованности.

Анализ частоты употребления слов, входящих в лексико-семантическую группу «Карьера» (табл. 2) показал, что будущая работа, дело, возможность совершенствования своего профессионализма и повышения квалификации гораздо больше интересуют авторов положительных отзывов. Суммарная частота употребления слов этой семантической группы в положительных отзывах в 1,3 раза выше, чем в отрицательных отзывах (5550 ЧМ против 4231 ЧМ).

Таблица 2

Сравнительный анализ частотных характеристик слов семантической группы «Карьера», употребляемых в текстах отзывов о вузах

№ п/п	Слово	Частота на миллион словоупотреблений, ЧМ	
		положительные отзывы	отрицательные отзывы
1	Работа	2300	1724
2	Работать	1968	1567
3	Специалист	575	376
4	Положение	88	157
5	Должность	66	125
6	Карьера	44	94
7	Труд	177	63
8	Квалифицированный	155	63
9	Высококвалифицированный	111	31
10	Профессионализм	66	31
11	ИТОГО	5550	4231

Как видно из табл. 2, авторов положительных отзывов больше волнует ра-

бота, квалификация, профессионализм (частота этих слов в 1,5 раза выше в этих отзывах). Авторы негативных отзывов больше волнует положение, должность, которую они получают после окончания вуза, как сложится их карьера. К такому выводу специалисты СГА пришли на основании того, что авторы негативных отзывов употребляют такие слова в 2 раза чаще авторов позитивных отзывов.

В табл. 3 показан сравнительный анализ частотных характеристик слов группы «Деньги», употребляемых авторами позитивных и негативных отзывов.

Таблица 3

Сравнительный анализ частотных характеристик слов семантической группы «Деньги», употребляемых в текстах отзывов о вузах

№ п/п	Слово	Частота на миллион словоупотреблений, ЧМ	
		положительные отзывы	отрицательные отзывы
1	Деньги	1614	3197
2	Платить	1150	1536
3	Оплата	243	1160
4	Стоимость	221	407
5	Купить	177	313
6	Плата	111	251
7	Заплатить	88	313
8	Доллар	44	157
9	Кредит	44	125
10	Евро	22	125
11	ИТОГО	3714	7584

Как видно из табл. 3, авторы отрицательных отзывов в 2 раза чаще употребляют слова, относящиеся к группе «Деньги», чем авторы положительных

отзывов. В число наиболее часто употребляемых слов этой группы вошли *деньги, платить, оплата, стоимость*. Вывод о меркантильности авторов негативных отзывов о вузах напрашивается сам собой.

Особенностью общения на форумах, в различных социальных сетях, в блогах и в процессе виртуального общения является употребление так называемых жаргонизмов, молодежного сленга. Характерно это и для авторов отзывов, причем как отрицательных, так и положительных. Однако в положительных отзывах таких слов употребляется гораздо меньше, и частота их употребления в 4 раза ниже, чем в отрицательных (табл. 4).

Таким образом, наша гипотеза о том, что позитивные отзывы пишут более образованные люди, которые стремятся учиться, постигать новые знания, добиваться карьерных успехов, научно подтверждена проведенным исследованием. По результатам проведенного исследования можно сделать еще несколько выводов о личностных качествах авторов анализируемых отзывов. Сравнительный анализ частоты употребления авторами положительных отзывов слов различной семантической окраски позволил нам выявить следующие личностные особенности этих людей (в сравнении с авторами негативных отзывов):

- целеустремленность (частота употребления ими слов *цель и стремиться* в 4 раза выше, слова *захотеть* – в 5 раз выше, чем в негативных отзывах, *добиться, добиваться* – в 2 раза выше, чем в негативных отзывах);

Таблица 4

Сравнительный анализ частотных характеристик слов семантической группы «Жаргонизмы», употребляемых в текстах отзывов о вузах

№ п/п	Слово	Частота на миллион словоупотреблений, ЧМ	
		положительные отзывы	отрицательные отзывы
1	Бардак	155	752
2	Ужасный	111	752
3	Ужас	88	501
4	Отстой	177	470
5	Пошлый	88	407
6	Кошмар	22	376
7	Тупой	155	345
8	Наплевать	199	313
9	Отстойник	22	282
10	Свалить	44	251
11	Отстойный	88	188
12	Хамский	22	188
13	Хамство	66	125
14	Шарашка	22	94
15	Наглый	22	94
16	Хамить	22	63
17	Срать	22	63
18	ИТОГО	1325	5264

- оптимизм (слова *желать, хотеть, хочется, успешно, уверенный* употребляются ими в среднем в 2 раза чаще, чем авторами негативных отзывов, слово *надежда* – в 4 раза чаще, *успех* в 5 раз чаще, а слово *сильный* – в 8 раз чаще);

- открытость и дружелюбие (слова *привет, друг, добрый, общение*, употребляются ими в среднем в 2 раза чаще, чем авторами негативных отзывов, а

слова *уважать, общаться, коллектив* – в 5 раз чаще);

- ответственность и исполнительность, так как слова *надо, делать, сделать* употребляются ими в среднем в 1,5 раза чаще, чем авторами негативных отзывов.

Негативные отзывы с грубыми словами, с грамматическими ошибками практически в каждом слове, вероятнее всего, пишут те, кому трудно учиться, кто плохо учился в школе, у кого нет интереса к получению знаний. Недаром у этой категории слова *придумывать* и *устраиваться* встречаются в 5 раза чаще, чем у авторов положительных отзывов. Кроме того, авторы негативных отзывов больше склонны надеяться на чужую помощь, чем на собственные силы. Подтверждением этого служат частое употребление ими таких слов, как *нельзя* (533 ЧМ против 199 ЧМ у авторов положительных отзывов) и *советовать* (1003 ЧМ против 420 ЧМ у авторов положительных отзывов). Также, вероятно, авторам негативных отзывов в большей степени присущ инфантилизм, так как действовать они предпочитают не по собственной инициативе, а по принуждению извне (частота употребления ими слова *пришлось* в 1,5 раза выше, чем у авторов положительных отзывов; а слова *самостоятельно* и *инициатива* они употребляют в 1,5 раза реже).

Сравнительные частотные характеристики слов различной семантической окраски в позитивных и негативных отзывах, по которым выявлены личностные особенности авторов отзывов, приведены в табл. 5.

Сравнительные частотные характеристики слов различной семантической окраски, употребляемых авторами позитивных и негативных отзывов

№ п/п	Слово	Частота на миллион словоупотреблений, ЧМ	
		положительные отзывы	отрицательные отзывы
1	2	3	4
1. Семантическая группа «Целеполагание»			
1.1	Возможность	553	376
1.2	Цель	243	63
1.3	Захотеть	155	31
1.4	Стремиться	133	31
1.5	Добиться	155	94
1.6	Добиваться	66	31
2. Семантическая группа «Оптимизм»			
2.1	Хотеть	4710	3103
2.2	Хочется	929	658
2.3	Желать	619	345
2.4	Сильный	531	63
2.5	Успех	310	63
2.6	Успешно	199	94
2.7	Уверенный	177	94
2.8	Замечательно	177	94
2.9	Надежда	133	31
2.10	Прекрасно	111	31
3. Семантическая группа «Открытость и дружелюбие»			
3.1	Привет	2277	1097
3.2	Друг	796	376
3.3	Добрый	376	219
3.4	Общаться	177	31
3.5	Уважать	177	31
3.6	Коллектив	155	31
3.7	Общение	66	31

Таблица 5. Окончание

1	2	3	4
4. Семантическая группа «Обязанность»			
4.1	Надо	3184	2633
4.2	Делать	1216	784
4.3	Сделать	597	345
5. Семантическая группа «Самостоятельность»			
5.1	Сам	2764	2006
5.2	Лично	575	345
5.3	Советовать	420	1003
5.4	Пришлось	376	564
5.5	Смочь	376	219
5.6	Решать	354	219
5.7	Самостоятельно	155	63
5.8	Инициатива	111	31

К сожалению, проведенное исследование показало, что существует большое количество людей, которые видят цель образования не в получении новых знаний и компетенций, а в получении формального документа – диплома. В процессе исследования обнаружилось и большое количество явно заказных материалов, часто одни и те же тексты размещены в разных отзывах и/или на разных сайтах. Нередки случаи откровенного «троллинга» – злонамеренного вмешательства в сетевую коммуникацию, выражающегося в нагнетании участником общения («троллем») гнева, конфликта путем скрытого или явного задиранья, принижения, оскорбления другого участника или участников, зачастую с нарушением этики сетевого взаимодействия [5]. В противовес «троллям» многие студенты стремятся защитить свой вуз, дать реальную оценку его деятельности, обменяться информацией об учебном процессе,

возможностях последующего трудоустройства.

Результаты исследования наглядно демонстрируют, что авторы положительных отзывов обладают не только более высоким уровнем общей культуры и образованности по сравнению с авторами отрицательных отзывов, но и такими качествами характера, как целеустремленность, оптимизм, дружелюбие, ответственность и самостоятельность. Именно таким абитуриентам и их родителям нужно смотреть не только на рейтинги вузов, но и не пренебрегать содержательной частью отзывов для выбора вуза или образовательной программы, что, на наш взгляд, поможет потребителям образовательных услуг ориентироваться в многообразии позитивных и негативных отзывов, размещенных в Интернете.

Литература

1. Фокина В.Н., Лукьянова А.В. Статистические методы анализа текста в социолингвистике для оценки образованности личности // Социология образования. 2012. № 4.
2. Тичер С. и др. Методы анализа текста и дискурса / Пер. с англ. Харьков: Гуманитарный Центр, 2009.
3. Ляшевская О.Н., Шаров С.А. Частотный словарь современного русского языка на материалах Национального корпуса русского языка. М.: Азбуковник, 2009.
4. Широкова М.Е., Потоцкая Е.Д. Формирование словаря «живого» русского языка с использованием средств информатизации // Труды СГА. 2011. № 12.

5. Троллинг. Википедия [Электронный ресурс] // Режим доступа: <https://ru.wikipedia.org/wiki/Троллинг>.

THE QUALITY OF EDUCATION

Karpenko O.M., *PhD in Economics*

Abramova A.V.

Semenova T.Yu., *PhD in Sociology*

Shirokova M.E., *PhD in Sociology*

Linguistic Statistical Text Analysis of Internet Comments on the Universities of Moscow

Using linguistic statistical methods there have been carried analysis of students' Internet comments on the universities. The article shows quantitative regularities of "living" Russian language, manifested in modern texts of the Internet comments made about the universities. The research allowed to develop and confirm the hypothesis that the authors of positive comments have a higher level of education and more positive attitude towards life.

Key words: *education, text analysis, word frequency, educational portal, sociolinguistics, linguistic statistics, lexical-semantic group.*