

## КОЛИЧЕСТВЕННЫЙ АНАЛИЗ КОМПЬЮТЕРНОЙ ТЕРМИНОЛОГИИ В РАЗЛИЧНЫХ ТИПАХ ДИСКУРСА

*Е.В. Дмитриева*

*Аспирант кафедры иностранных языков и профессиональной коммуникации  
e-mail: [ellen\\_dmitrieva@mail.ru](mailto:ellen_dmitrieva@mail.ru)*

*Курский государственный университет*

*Данное исследование нацелено на апробирование методики количественного анализа компьютерной терминологии (далее – КТ) в разных типах дискурса, т.е. выявления ее частотности с опорой на психолингвистическую парадигму.*

*Объектом нашего исследования стали разные по тематике дискурсы: экономический, образовательно-педагогический и развлекательный. Изучению подверглись материалы российских и зарубежных периодических интернет-изданий, а также развлекательные и образовательные веб-порталы и видеохостинги.*

*Проведенное исследование указало на возможность применения использованной нами методики для изучения КТ. Интерпретация результатов статистического анализа позволила установить, что компьютерная терминология существует в разных типах дискурса в различном процентном соотношении, что ранее для данного типа лексики не являлось типичным.*

***Ключевые слова:** компьютерная терминология, количественный анализ, экономический дискурс, образовательно-педагогический дискурс, развлекательный дискурс, психолингвистический подход, теория проксиматики.*

### **Введение**

Эпоха научно-технического прогресса (далее – НТП), стремительно развивающаяся информатизация экономики, социума и культуры спровоцировала появление большого количества терминологии и профессионализмов, которые обслуживают все сферы жизни и деятельности современного человека. Особое место в ряду новых терминологий занимает компьютерная терминология (далее – КТ), которая является реакцией на процессы внедрения и использования компьютера в социуме. Заметим, что номинация КТ в настоящее время достаточно условна, поэтому хотим пояснить, что лексемы КТ, компьютерная лексика, компьютерный термин используются нами как синонимичные.

Междисциплинарный подход к изучению КТ исключает, на наш взгляд, унификацию в ее трактовке, поскольку постоянное участие КТ в развитии языка, речи и культуры в целом является ее отличительной особенностью. Это связано с формированием компьютерной терминологии под воздействием современных источников знаний и информации, использованием компьютера в определении объектов технических достижений человечества. Современная КТ переживает новый этап своего

развития и становления, требующий, на наш взгляд, изучения ее в рамках психолингвистической парадигмы, а также поиска определенных методик ее выявления в разных типах дискурса, иными словами, количественного и качественного анализа. В нашем исследовании КТ понимается как ментальная сущность, возникающая в процессе речемыслительной деятельности человека и отражающая взаимодействие человека с окружающим миром.

*Задача предлагаемого исследования* – апробирование методики количественного анализа компьютерной терминологии в разных типах дискурса, т.е. выявления ее частотности с опорой на психолингвистическую парадигму. Мы полагаем, что КТ может быть «правильно» идентифицирована и интерпретирована лишь в дискурсивном окружении, от которого она непосредственно зависит и которое она тематически модифицирует. При этом не требуется ее соотнесения с прямым инвариантом.

Следует отметить, что, несмотря на длительную историю, в традиционной лингвистике проблема терминологии, терминосистем и термина до сих пор вызывает острые дискуссии (см. например, [Бабалова 2009, Орлова 2008, Комлева 2006, Лобанова 2009]). Объективные условия и потребности современной цивилизации определяют актуальность изучения специфики КТ в современном мире. *Объектом* нашего исследования стали разные по тематике дискурсы: материалы периодических интернет-изданий различной тематики как российские («КОММЕРСАНТЪ», «Эксперт»), так и зарубежные («Forbes», «The Economist», «British Council Learn English teens», «Mental Floss»), а также развлекательные и образовательные веб-порталы (интернет-блог «LinguaLeo») и видеохостинги («YouTube»). Очевидно, что приведённые выше материалы представляют одну сферу профессиональной коммуникации, поскольку весь выбранный материал объединён одной общей темой «Технологии» (“Technologies”). С нашей точки зрения, это позволит соблюсти валидность и тождественность при проведении исследования. Отметим, что категория «Технологии», представленная в интерфейсе веб-сайтов отдельным блоком, показывает важность влияния научно-технических инноваций на современный мир. Наличие в текстах специальной лексики предполагает владение участниками коммуникации понятийно-категориальным аппаратом сферы информатики, их готовность к «переходу на профессиональный уровень сознания» [Культура русской речи и эффективность общения 1996: 238].

Считаем необходимым акцентировать внимание на использовании в качестве теоретической основы исследования психолингвистической парадигмы, а именно теории проксиматики С.В. Лебедевой (см. подробно [Лебедева 2002; 2011; 2012; 2013]). Как известно, специфика осмысления

феноменов любого дискурса отражается через обращение к обыденным представлениям человека о мире. Согласно высказыванию С.В. Лебедевой, «использование конвенциональных социальных, профессиональных и «наивных» значений приводит к тому, что информация, как правило, не верифицируется, поскольку значима опора на сложившийся консенсус, заданный конкретной предметной областью» [Лебедева 2013]. В нашем случае предметной областью является сфера компьютерных технологий. Также мы разделяем точку зрения М.В. Орловой, исследования которой позволило автору утверждать, что КТ – это условное название обширной терминологической области, описывающей различные виды деятельности индивида, в которой встречаются слова или словесные сочетания, описывающие отдельные фрагменты жизни человека и социума [Орлова 2008].

Опора на теорию проксиматики позволяет нам выделить в исследуемых текстах проксонимические проекции, а именно, «разноуровневую систему, обладающую особой структурированностью и диффузными границами» [Лебедева 2014]. Считаем уместным привести подробную цитату С.В. Лебедевой для прояснения нашей теоретической позиции: «Эффективность действия данной проекции в ситуации «для меня, здесь и сейчас» определяется как качеством и количеством формирующих ее элементов, так и их взаимосвязью. Важно заметить, что оператор сходства соотносится с фрагментом действительности (знания о котором конвенционально аккумулирует) и может формироваться как результат социального, чувственного, культурного, в том числе и личностного освоения индивидом окружающей действительности. Основу проксонимической проекции составляет субъективный семантический функциональный центр, позволяющий индивиду предъявлять различные виды социального и индивидуального знания через когнитивно-лингвистическую схему связи между понятийными сферами, подвергающимися проксимации» [Лебедева 2014]. Экстраполируя сказанное выше на цели и результаты нашего исследования, полагаем, что в качестве основы субъективного семантического центра можно рассматривать компьютер как когнитивный феномен, который «задает вектор «раскодирования» основного смысла, выраженного вербальными и невербальными единицами» [там же].

Считаем обязательным акцентировать внимание на том, что изучаемые нами слова и словесные сочетания обладают свойством «компьютерной терминологичности» (в рамках определенного контекста их употребления) и семантического единства. Мы предполагаем, что слова, имеющие вышеупомянутые признаки, входят в проксонимическую проекцию и «притягиваются» когнитивной доминантой (термин С.В. Лебедевой), т.е. концентрируются вокруг феномена

«компьютер» и используются в качестве компьютерных терминов, функционирующих в текстах различной тематики.

### **Методология исследования**

Методом исследования послужил статистический анализ. Он предполагает определение частотности употребления языковой единицы в определенном тексте или языке в целом, типичность конкретных языковых явлений для определённого периода развития языка или для определённого языкового стиля. С этой целью часто используются различные способы индексации, определяющие количество употреблений анализируемой формы на каждые 100 слов текста.

Инструментом для проведения статистического анализа текста выступает сервис Miratext [Miratext [http](http://miratext.com)]. Он позволяет проанализировать текст по следующим параметрам: общая статистика текста, позволяющая оценить суммарное количество символов и слов, качество текста по закону Ципфа, а также морфологические повторы и общее количество повторений слов и словосочетаний в тексте. Помимо инструмента Miratext, мы также использовали сервис Advego [Advego [http](http://advego.com)]. Одним из преимуществ данной платформы является проведение семантического анализа без включения в итоговые результаты служебных слов («стоп-слов»).

Мы считаем оправданным демонстрировать результаты анализа наглядно, используя таблицы и диаграммы. Также мы хотим подчеркнуть, что ввиду ограничений по объёму публикации, мы не сможем привести все имеющиеся результаты полученных количественных показателей, поэтому мы продемонстрируем только наиболее показательные примеры: «фотогалерея», «технология», «машинный», «социальная сеть», «искусственный интеллект», “post”, “camera”, “disposable cameras”, “apple watch”, “airpod” и общую статистику отобранного для исследования материала. Основная задача при анализе – избавиться от побочных бытовых элементов значения, которые придают особую когнитивную окраску объектам и явлениям реальности и способствуют вариативности значения, что препятствует проявлению компьютерной составляющей. Полагаем, что КТ выступает как наиболее продуктивное средство пополнения языка индивида на основании уже имеющихся в языке значений слов, морфем, сочетаний слов и даже синтаксических конструкций и является частью языковой компетенции.

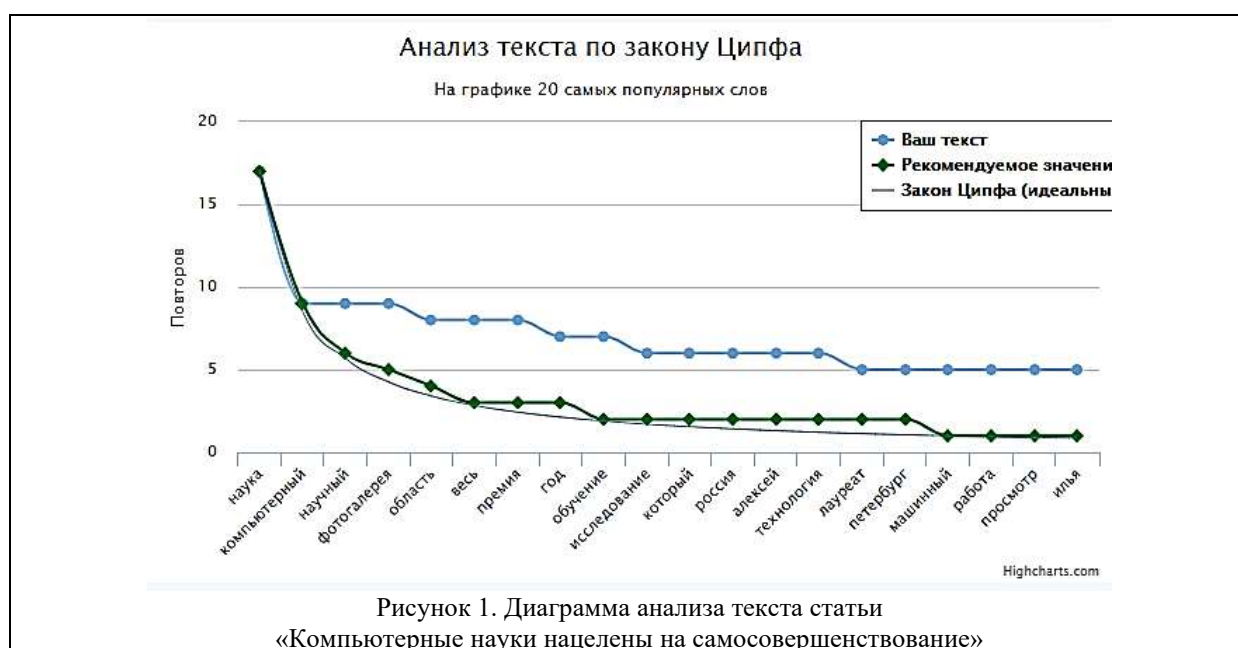
Заметим, что субъективный семантический функциональный центр позволяет дескриптивно представить разнообразные тематические группы, функционально-семантические и ассоциативно-семантические поля, предметные парадигмы и их контекстуальные границы

## Ход исследования

В первую очередь мы намерены проанализировать экономический дискурс. Выбор данного типа дискурса был осуществлен по причине того, что научно-технический прогресс оказал непосредственное влияние на мировую экономику, и созданные научно-технические инновации повсеместно используются в разнообразных экономических процессах.

Проведем анализ статьи под названием «Компьютерные науки нацелены на самосовершенствование», российского периодического издания «КОММЕРСАНТЪ» [Компьютерные науки нацелены...http].

На рисунке 1 представлена диаграмма анализа текста по закону Ципфа, на которой отражено 20 наиболее популярных слов, используемых в тексте публикации.



Согласно таблице 1, представленная выше диаграмма состоит из следующих лексем и процента частотности их употребления.

Таблица 1. Количество повторов в тексте статьи «Компьютерные науки нацелены на самосовершенствование» и их соответствие анализу по закону Ципфа

Слово и его морфологические повторы	Кол-во повторов	По Ципфу	Соответствие
<b>наука</b> (наука, науки, наук, науками)	17	17	100%
<b>компьютерный</b> (компьютерные, компьютерных, компьютерного, компьютерное)	9	9	100%
<b>научный</b> (научные, научным, научное, научный, научными, научных)	9	6	67%
<b>фотогалерея</b>	9	5	56%
<b>область</b> (области, область)	8	4	50%
<b>весь</b> (все, всеми)	8	3	38%
<b>премия</b> (премии, премию, премия)	8	3	38%

Продолжение таблицы 1

<b>год (года, году, год, лет)</b>	7	3	43%
<b>обучение (обучения, обучение)</b>	7	2	29%
<b>исследование (исследования, исследованиями, исследование, исследований)</b>	6	2	33%
<b>который (которые, который, которая)</b>	6	2	33%
<b>Россия (России, Россия)</b>	6	2	33%
<b>Алексей (Алексей, Алексею, Алексея)</b>	6	2	33%
<b>технология (технологий, технологии)</b>	6	2	33%
<b>лауреат (лауреаты, лауреатами, лауреат)</b>	5	2	40%
<b>Петербург (Петербург, Петербурге)</b>	5	2	40%
<b>машинный (машинного, машинное)</b>	5	1	20%

При анализе полученного материала обратим внимание на высокий процент употребления компьютерных терминов в тексте статьи экономической тематики («компьютерный», «фотогалерея», «технология», «машинный»), включая их морфологические повторы («технологии», «компьютерных» и другие). Мы хотим сделать акцент на том, что лексемы «фотогалерея», «технология», «машинный» мы относим к сфере КТ, так как в тексте статьи идет речь о программном и аппаратном обеспечении персонального компьютера на основании контекста употребления, т.е. осуществляется «переход значения», связанного с одним конкретным понятием, в сферу обозначения другого понятия, следовательно, видоизменяется «исходная» семантика [Лебедева 2011].

Следует отметить, что для того чтобы понять что-то новое, человек обращается к хорошо известному. Так, например, лексема «фотогалерея» была заимствована из словаря, в котором зафиксировано следующее значение: «подборка фотографий, представленная на обозрение» и имеющая сходство с печатным фотоальбомом [Викисловарь [http](http://)].

В Федеральном законе от 27.07.2006 № 49-ФЗ (ред. от 09.03.2021) «Об информации, информационных технологиях и о защите информации» лексема «технология» трактуется как «совокупность процессов, методов поиска, сбора, хранения, обработки, предоставления, распространения информации и способы осуществления таких процессов и методов» [Федеральный закон [http](http://)].

Лексема «машинный» имеет несколько значений: 1) связанный с машиной; 2) свойственный машине, характерный для нее; напоминающий машину автоматизмом; 3) предназначенный для машины [Толковый словарь русского языка [http](http://)].

Мы придерживаемся мнения, что в данной ситуации присутствует феномен проксимии, т.е. наличие неких «вербальных единиц, сходство/различие которых обусловлено социальными, культурными, житейскими и личностными конвенциями в ситуации “здесь и сейчас”» [Кондратенко 2014]. По всей видимости, у данных слов имеются

некие сходства с зафиксированными в словарях феноменами. Следовательно, данное сходство позволило автору использовать лексемы «фотогалерея», «технология» в качестве компьютерных терминов.

Помимо вышеупомянутых лексем отметим, что в тексте статьи присутствуют и термины других тематических областей: «наука», «научный», «область», «премия», «год», «обучение», «исследование», «лауреат», а также онимы «Россия», «Алексей», «Петербург», которые, по нашему мнению, не относятся к сфере компьютерных технологий.

Далее приступим к анализу статьи под заголовком «Интернет-гигантов приземлят в России», опубликованной на сайте российского экономического издания «Эксперт» [Интернет-гигантов приземлят ... http].

На рисунке 2 представлена диаграмма анализа текста статьи.

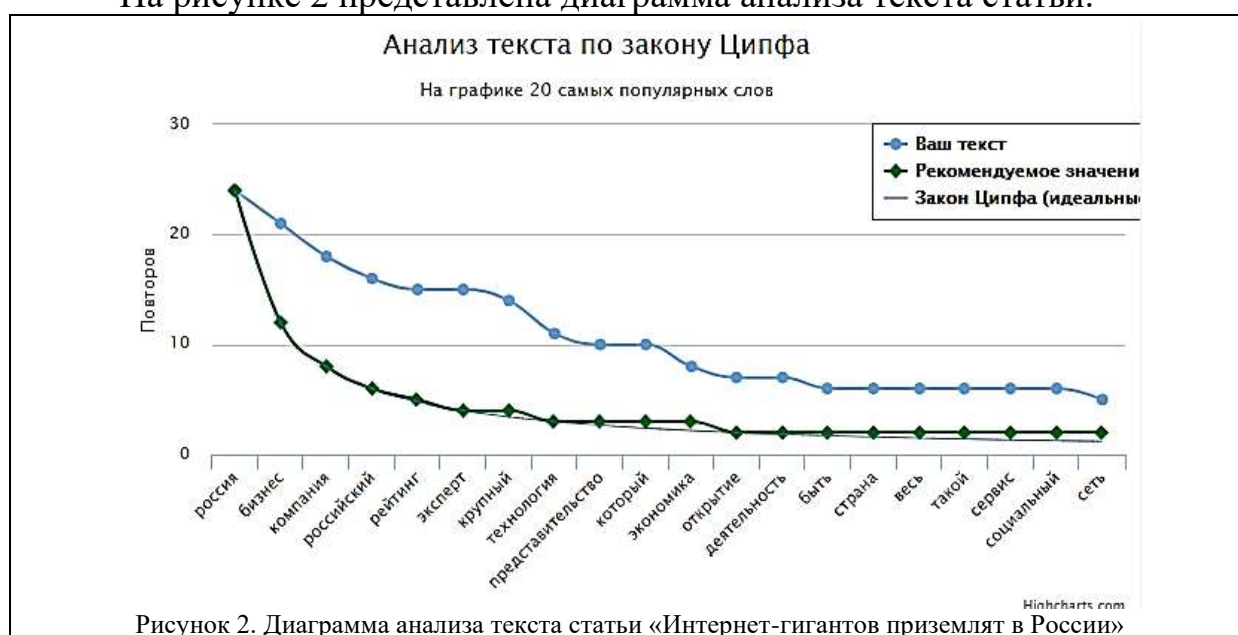


Рисунок 2. Диаграмма анализа текста статьи «Интернет-гигантов приземлят в России»

Согласно таблице 2, представленная выше диаграмма состоит из следующих лексем и процента частотности их употребления.

Таблица 2. Количество повторов в тексте статьи «Интернет-гигантов приземлят в России» и их соответствие анализу по закону Ципфа

Слово и его морфологические повторы	Кол-во повторов	По Ципфу	Соответствие
<b>Россия</b>	24	24	100%
<b>бизнес</b> (бизнес, бизнеса, бизнесе)	21	12	57%
<b>компания</b> (компаний, компании, компания, компаниями)	18	8	44%
<b>российский</b> (российский, российские, российских, российской, российского, российское, российским, российском)	16	6	38%
<b>рейтинг</b> (рейтинг, рейтинги)	15	5	33%
<b>эксперт</b> (эксперт, эксперте, эксперту)	15	4	27%
<b>крупный</b> (крупных, крупного, крупнейших, крупная, крупной, крупные)	14	4	29%

Продолжение таблицы 2

<b>технология</b> (технологии, технологий, технологиями)	11	3	27%
<b>представительство</b> (представительств, представительства, представительство)	10	3	30%
<b>который</b> (которых, которые, который, которой)	10	3	30%
<b>экономика</b> (экономика, экономике)	8	3	38%
<b>открытие</b> (открытия, открытием, открытия)	7	2	29%
<b>деятельность</b> (деятельности, деятельность)	7	2	29%
<b>быть</b> (будут, быть)	6	2	33%
<b>сервис</b> (сервисы, сервисов)	6	2	33%
<b>социальный</b> (социальная, социальных, социальной, социальные, социальную)	6	2	33%
<b>сеть</b> (сетями, сетях, сети, сетью)	6	2	33%

Примечательно, что название статьи непосредственно указывает на узкоспециализированную компьютерную тематику, но компьютерных терминов использовано не так много, как мы предполагали. Судя по таблице и диаграмме, компьютерными терминами являются понятия «технология» и «сеть», а также их морфологические повторы «технологиями», «сетью» и другие.

Лексема «сеть» в «Толковом словаре русского языка» С.И. Ожегова трактуется как «приспособление, изделие из закреплённых на равных промежутках, перекрещивающихся нитей, верёвок, проволоки» [Толковый словарь Ожегова [http](http://)]. Лексема «сеть» в контексте анализируемой статьи рассматривается нами в совокупности с лексемой «социальный», ввиду того, что лексема «сеть» приобретает определенную характеристику на основании знаний человека об этом феномене, а также имеет довольно большое количество повторений в тексте статьи, также как и термин «социальный» («сеть» – 6, «социальный» – 6). Совокупность двух лексем позволила сгенерировать новое лексическое словосочетание – «социальная сеть».

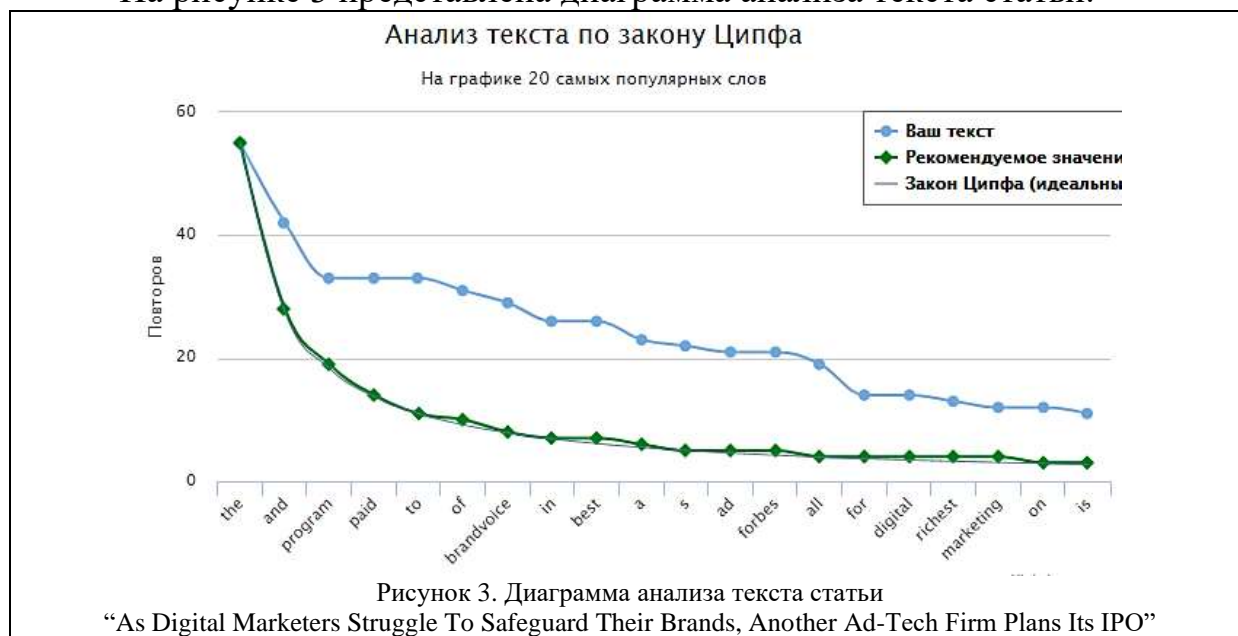
Данное словесное сочетание, согласно словарю SEO терминологии, трактуется как «Интернет-площадка, сайт, который позволяет зарегистрированным на нем пользователям размещать информацию о себе и коммуницировать между собой, устанавливая социальные связи» [Словарь SEO терминов [http](http://)]. В данном случае также прослеживается сходство с «сетью» в ее первоначальном значении.

Помимо вышеупомянутых лексем, которые могут обладать компьютерной терминологичностью и являться терминами области компьютерных технологий, также отметим, что в тексте статьи присутствуют и термины других тематических областей: «бизнес», «компания», «деятельность», «сервис», «экономика», «открытие», «представительство», «эксперт», «рейтинг», не являющиеся компьютерными терминами.



Следующей статьей, которую мы намерены проанализировать в рамках нашего исследования, является публикация Марти Свонта под названием “As Digital Marketers Struggle To Safeguard Their Brands, Another Ad-Tech Firm Plans Its IPO”, размещенная на странице журнала «Forbes» [As Digital Marketers...http].

На рисунке 3 представлена диаграмма анализа текста статьи.



Согласно таблице 3, представленная выше диаграмма состоит из следующих лексем и процента частотности их употребления.

Таблица 3. Количество повторов в тексте статьи “As Digital Marketers ...” и их соответствие анализу по закону Ципфа

Слово	Кол-во повторов	По Ципфу	Соответствие
<b>the</b>	55	55	100%
<b>and</b>	42	28	67%
<b>program</b>	33	19	58%
<b>paid</b>	33	14	42%
<b>to</b>	33	11	33%
<b>of</b>	31	10	32%
<b>brandvoice</b>	29	8	28%
<b>in</b>	26	7	27%
<b>best</b>	26	7	27%
<b>a</b>	23	6	26%
<b>forbes</b>	21	5	24%
<b>all</b>	19	4	21%
<b>for</b>	14	4	29%
<b>digital</b>	14	4	29%
<b>richest</b>	13	4	31%
<b>marketing</b>	12	4	33%
<b>on</b>	12	3	25%
<b>is</b>	11	3	27%

Считаем необходимым акцентировать внимание на том, что анализ зарубежных источников не всегда может быть качественно выполнен инструментом Miratext, что подтверждается результатами таблицы 3.

Согласно результатам таблицы 3 и рисунка 3, методика статистического анализа позволила выявить также большое количество служебных частей речи: артиклей (“*the*” – 55 вхождений, “*a*” – 23 вхождения), предлогов (“*to*” – 33; “*of*” – 31 “*in*” – 26; “*for*” – 14; “*on*” – 12), союзов (“*and*” – 42) и глагольных форм (“*is*” – 11), что делает наши результаты недостаточно достоверными. Поэтому мы обратимся ко второму инструменту – Advego, который позволит нам провести статистический анализ, не включая в итоговые результаты слова, не несущие в себе семантического потенциала [Advego [http](http://)].

Результаты анализа при помощи инструмента Advego представлены в таблице 4.

Таблица 4. Количество повторов в тексте статьи “*As Digital Marketers Struggle To Safeguard Their Brands, Another Ad-Tech Firm Plans Its IPO*” без учета служебных слов

Слово	Кол-во повторов	Частота употребления
<b>company</b>	16	1.86
<b>digital</b>	14	1.63
<b>million</b>	10	1.16
<b>market</b>	6	0.70
<b>more</b>	6	0.70
<b>platform</b>	6	0.70
<b>advertising</b>	5	0.58
<b>billion</b>	5	0.58
<b>brand</b>	5	0.58
<b>IPO</b>	5	0.58
<b>like</b>	5	0.58
<b>public</b>	5	0.58
<b>said</b>	5	0.58
<b>there</b>	5	0.58
<b>advertiser</b>	4	0.46
<b>digital advertising</b>	4	0.46 / 0.93

При анализе полученного материала можно сделать вывод о том, что в экономической статье присутствует компьютерная терминология, которая имеет довольно высокое процентное соотношение по сравнению с остальными понятиями. Например, “*digital*”, “*platform*”, “*digital advertising*”, “*net*”. Также хотим отметить, что в указанных примерах есть взаимосвязь двух дискурсов: компьютерного (“*digital*”) и рекламного (“*advertising*”), которая способствовала генерации нового лексического сочетания.

По нашему мнению, остальные лексемы нельзя отнести к сфере компьютерных технологий: “*company*”, “*million*”, “*market*”, “*more*”, “*billion*”, “*brand*”, “*like*”, “*public*”, “*said*”, “*there*”, “*advertiser*”.

Далее приступим к анализу статьи “Computer-generated realities are becoming ubiquitous”, опубликованной на сайте американского зарубежного издательства «The Economist» [Computer-generated ... http].

В таблице 5 представлены результаты анализа статьи.

Таблица 5. Количество повторов в тексте статьи “Computer-generated realities are becoming ubiquitous” без учета служебных слов

Слово	Кол-во повторов	Частота употребления
<b>game</b>	32	1.61
<b>world</b>	28	1.41
<b>that</b>	27	1.36
<b>virtual</b>	27	1.36
<b>engine</b>	15	0.76
<b>this</b>	14	0.70
<b>can</b>	12	0.60
<b>more</b>	12	0.60
<b>other</b>	10	0.50
<b>reality</b>	10	0.50
<b>build</b>	9	0.45
<b>headset</b>	9	0.45
<b>virtual worlds</b>	9	0.45 / 0.91
<b>game engines</b>	8	0.40 / 0.81
<b>many</b>	7	0.35
<b>new</b>	7	0.35
<b>will</b>	7	0.35
<b>software</b>	6	0.30

Отметим, что в тексте статьи задействовано довольно большое количество повторений компьютерных терминов по отношению к суммарному количеству слов в тексте публикации. Например, «*virtual*» (27), «*headset*» (9), «*virtual worlds*» (9), «*game engines*» (8), «*software*» (6), что является одним из доказательств миграции компьютерной терминологии в экономический дискурс.

Но помимо вышеупомянутых лексем, обладающих свойством компьютерной терминологичности, также отметим, что в тексте статьи присутствуют такие слова, как: “*game*”, “*world*”, “*that*”, “*engine*”, “*this*”, “*can*”, “*more*”, “*other*”, “*reality*”, “*build*”, “*many*”, “*new*”, не являющиеся компьютерными терминами.

Следующий тип дискурса, проанализированный нами, – образовательно-педагогический дискурс. И в качестве материала для исследования была выбрана статья “Disposable cameras are the future” автора MeganBlogger, опубликованная на сайте образовательного портала British Council (Learn English teens) [Disposable cameras...http].

В таблице 6 представлены результаты анализа статьи.

Таблица 6. Количество повторов в тексте статьи “Disposable cameras are the future” без учета служебных слов

Слово	Кол-во повторов	Частота употребления
<b>picture</b>	9	2.09
<b>photograph</b>	8	1.86
<b>this</b>	8	1.86
<b>photo</b>	7	1.62
<b>can</b>	6	1.39
<b>post</b>	5	1.16
<b>that</b>	5	1.16
<b>camera</b>	4	0.93
<b>social</b>	4	0.93
<b>disposable</b>	3	0.70
<b>disposable cameras</b>	3	0.70 / 1.39
<b>look</b>	3	0.70
<b>meaningful</b>	3	0.70
<b>our</b>	3	0.70
<b>really</b>	3	0.70
<b>share</b>	3	0.70

В тексте статьи присутствуют такие термины, как “*post*”, “*camera*”, “*disposable cameras*”. Мы считаем, что данные лексемы можно отнести к сфере компьютерных технологий ввиду того, что они связаны с компьютерными и мобильными устройствами, которые возникли благодаря появлению новых современных технологий. Также отметим наличие иных слов, не принадлежащих области КТ: “*picture*”, “*photograph*”, “*this*”, “*photo*”, “*can*”, “*that*”, “*social*”, “*look*”, “*meaningful*”, “*our*”, “*really*”, “*share*”.

Следующей статьей, отобранной в качестве материала для исследования, выступает публикация под названием “Amazon Is Having a Sale on AirPods Pro and Apple Watches”, размещенная на сайте журнала «Mental Floss» [Amazon Is Having ... [http](#)].

В таблице 7 представлены результаты анализа статьи.

Таблица 7. Количество повторов в тексте статьи «Amazon Is Having a Sale on AirPods Pro and Apple Watches» без учета служебных слов

Слово	Кол-во повторов	Частота употребления
<b>apple</b>	6	2.17
<b>for</b>	6	2.17
<b>airpod</b>	4	1.45
<b>can</b>	4	1.45
<b>pair</b>	4	1.45
<b>apple watch</b>	3	1.09 / 2.17
<b>for percent</b>	3	1.09 / 2.17
<b>for percent off</b>	3	1.09 / 3.26
<b>more</b>	3	1.09
<b>off</b>	3	1.09
<b>percent</b>	3	1.09
<b>percent off</b>	3	1.09 / 2.17

Продолжение таблицы 7

price	3	1.09
this	3	1.09
watch	3	1.09
amazon	2	0.72
apple watch series	2	0.72 / 2.17
call	2	0.72
case	2	0.72
continue work	2	0.72 / 1.45
down	2	0.72
feature	2	0.72
headphone	2	0.72
here	2	0.72
like	2	0.72
look	2	0.72
more about airpods here	2	0.72 / 2.90
music	2	0.72
music podcasts	2	0.72 / 1.45
need	2	0.72
noise	2	0.72
noise can	2	0.72 / 1.45
podcast	2	0.72
price down	2	0.72 / 1.45
sale	2	0.72
series	2	0.72
standard	2	0.72
that	2	0.72
through	2	0.72
watch series	2	0.72 / 1.45
will	2	0.72
wireless	2	0.72

Согласно таблице 7, содержащей итоговые результаты анализа, в тексте публикации встречаются такие сочетания как *“apple”*, *“apple watch”*, *“apple watch series”*. Данные слова и словосочетания могут по-разному интерпретироваться. Например, *“apple”* имеет значение «яблонный, яблонный плод, плод дерева яблонь», согласно «Толковому словарю живого великорусского языка» В.И. Даля [Словарь В. Даля [http](http://)]. Но так как эта статья опубликована в блоке «Технологии», то мы подозреваем, что лексема *“apple”* будет иметь иное значение. Для того чтобы корректно интерпретировать значение слова *“apple”* в тексте статьи попробуем проанализировать словосочетание *“apple watch”*. Оно состоит из двух лексем: *“apple”* (прим. «яблоко») и *“watch”* (прим. «часы», «смотреть»). Лексема *“apple watch”* имеет данную номинацию ввиду того, что часы имеют сходство с плодом яблока по цвету и, в некоторых случаях, по форме. Также, согласно тексту статьи, это инновационная разработка компании Apple Inc., изображением бренда которой является яблоневый плод. Анализ лексемы *“airpod”* позволил

установить, что данный объект также является инновационной разработкой Apple Inc. и был назван таким образом из-за сходства функций и значений двух лексем “air” (в значении «воздушный, легкий (прим. по весу), беспроводной») и “pods” (в значении «стандарт для хранения данных»). Помимо этого, в тексте статьи также присутствуют такие компьютерные термины, как “headphone”, “music podcasts”, “podcast” и др.

Заметим, что в данной статье преобладают слова иной тематики, суммарное количество повторений которых превышает суммарное количество вхождений в текст вышеупомянутых лексем. К данным словам мы относим: “pair”, “price”, “call”, “case”, “down”, “feature” и др.

Далее проведем статистический анализ статьи “Computer System Validation. Chapter 1”, опубликованной в образовательном Интернет-блоге «LinguaLeo» [Computer System...http].

В таблице 8 представлены результаты анализа статьи.

Таблица 8. Количество повторов в тексте статьи «Computer System Validation. Chapter 1» без учета служебных слов

Слово	Кол-во повторов	Частота употребления
validation	27	5.66
system	23	4.82
policy	14	2.94
computer systems validation	13	2.73 / 8.18
business	12	2.52
computer	12	2.52
computer systems	12	2.52 / 5.03
unit	12	2.52
business unit	10	2.10 / 4.19
site	10	2.10
system validation	10	2.10 / 4.19
business unit policies	6	1.26 / 3.77
departmental	6	1.26
protocol	6	1.26
responsibility	6	1.26
site departmental	6	1.26 / 2.52
that	6	1.26
this	6	1.26
department	5	1.05
inventory	5	1.05
requirement	5	1.05
unit policies	5	1.05 / 2.10
comply	4	0.84
corporate	4	0.84
department site	4	0.84 / 1.68
management	4	0.84
must	4	0.84
protocols system specific validation	4	0.84 / 3.35
shall	4	0.84
SOP	4	0.84

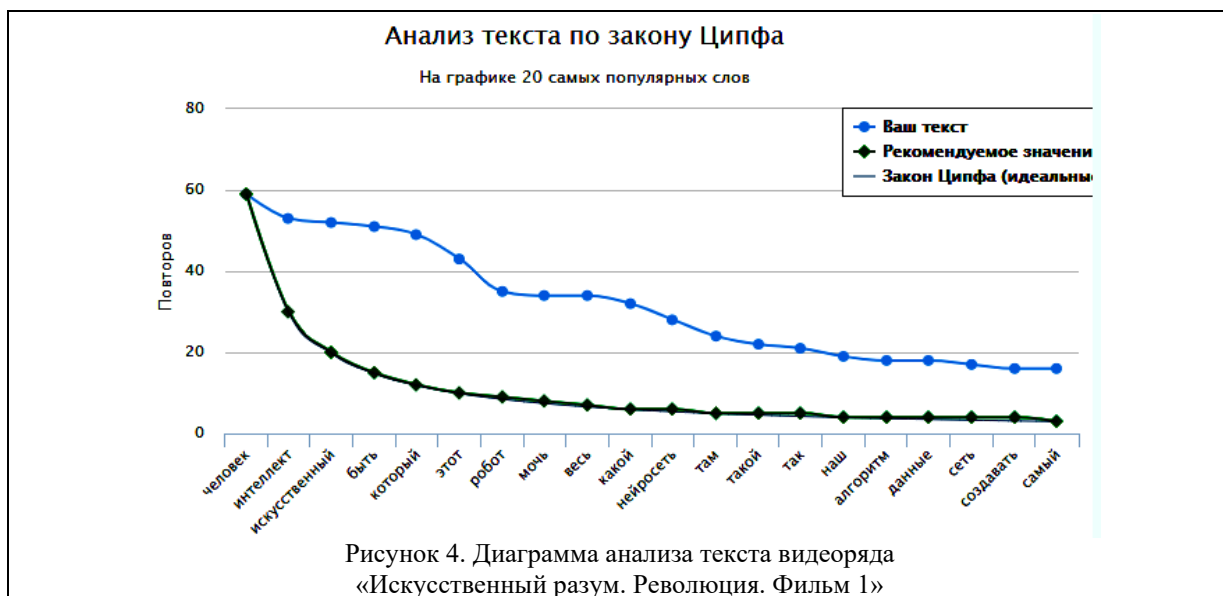
Примечательно, что в тексте данной статьи, имеющей узкоспециализированную компьютерную тематику, но опубликованной в образовательном блоге в рамках образовательно-педагогического дискурса, используется большое количество компьютерных терминов, имеющих высокий процент частотности и суммарное количество повторений в тексте. Например, “*computer systems validation*” (13), “*computer*” (12), “*computer systems*” (12), “*site*” (10), “*department site*” (4), “*protocols system specific validation*” (4). Также встречаются слова, не относящиеся к области КТ: “*policy*”, “*business*”, “*unit*”, “*departmental*” “*responsibility*”, “*protocol*”, “*inventory*”, “*requirement*” и другие.

Подводя итог всему вышесказанному, можно утверждать, что компьютерная терминология также употребляется в образовательно-педагогическом дискурсе в различном процентном соотношении и с разной частотностью.

В рамках нашего исследования мы также проанализировали функционирование КТ в развлекательном дискурсе. Особый интерес изучения функционирования компьютерной терминологии в этом типе дискурса был вызван его популярностью и повсеместным использованием. К развлекательному дискурсу мы относим видеохостинги, социальные сети, интернет-блоги и другое.

В качестве материала для исследования развлекательного дискурса был выбран видеоряд, носящий название «Искусственный разум. Революция. Фильм 1», из аккаунта Наука 2.0, созданного на базе видеохостинга YouTube [Искусственный разум...http]. Отметим, что в отличие от периодических изданий, данный видеохостинг не имеет таких тематических категорий как «Наука», «Спорт», «Технологии», поэтому выбор материала для исследования осуществлялся при помощи поисковой строки, основным критерием которого был запрос на самый популярный аккаунт научной тематики на видеохостинге YouTube. Так как мы анализировали встроенные в видеоряд субтитры, то сочли необходимым использовать инструмент извлечения субтитров DOWNSUB, который позволил нам проанализировать текстовый материал вышеупомянутого видеоряда [DOWNSUB http].

На рисунке 4 представлена диаграмма анализа текста статьи.



Согласно таблице 9, представленная выше диаграмма состоит из следующих лексем и процента частотности их употребления.

Таблица 9. Количество повторов в субтитрах видеоряда «Искусственный разум. Революция. Фильм 1» и их соответствие анализу по закону Ципфа

Слово	Кол-во повторов	По Ципфу	Соответствие
<b>человек</b> (человек, люди, людям, людей, человека, человеком, людях, человеку, людьми)	59	59	100%
<b>интеллект</b> (интеллектом, интеллект, интеллекта, интеллекту, интеллектов, интеллекты, интеллекте)	52	30	57%
<b>искусственный</b> (искусственным, искусственный, искусственные, искусственного, искусственная, искусственных, искусственному)	52	20	38%
<b>быть</b> (есть, было, быть, будут, была, был, будем)	51	15	29%
<b>который</b> (которой, которые, которым, которую, который, которая, которых, которого, которое)	49	12	24%
<b>этот</b> (эти, этим, этой, этих, этот, эта, этими)	43	10	23%
<b>робот</b> (робота, роботы, роботов, робот, роботу, роботом)	35	9	26%
<b>мочь</b> (могут, может, можем, можешь, можете)	34	8	24%
<b>весь</b> (все, весь, всему, всю, всем, всех, всей)	34	7	21%
<b>какой</b> (какой, каким, каких, какие, какое, каком, какую, какими)	32	6	19%
<b>нейросеть</b> (нейросети, нейросетей, нейросеть, нейросетями)	28	6	21%
<b>там</b>	24	5	21%
<b>такой</b> (такой, такие, таким, такую, такая, такого, таких, такое)	22	5	23%
<b>так</b>	21	5	24%



Продолжение таблицы 9

<b>наш</b> (нашей, наше, наши, наш, нашего, наших, нашу)	19	4	21%
<b>алгоритм</b> (алгоритм, алгоритмы, алгоритмом, алгоритма, алгоритмов, алгоритмам)	18	4	22%
<b>данные</b> (данных, данные, данными)	18	4	22%
<b>сеть</b> (сети, сеть, сетей, сетям)	17	4	24%
<b>создавать</b> (создать, созданные, создает, создают, создали, создавать, создаем, созданы, создано)	16	4	25%
<b>самый</b> (самом, самый, самым, самых, самая, самой, самого, самое)	16	3	19%

Необходимо сделать акцент на том, что в данном видеоряде присутствуют следующие компьютерные термины: «робот», «нейросеть», «алгоритм», «данные», «сеть». Также при анализе видеоряда были замечены многочисленные повторения таких лексем как «интеллект» (52) и «искусственный» (52), имеющие одинаковое количество повторов. Мы считаем обоснованным не разделять эти понятия, а рассматривать их в комплексе, так как по отдельности данные дефиниции не имеют ничего общего с компьютерной терминологией, но при рассмотрении их как единого целого, они напрямую связаны с термином, принадлежащим области КТ. Так, в «Большом энциклопедическом словаре» термин «искусственный интеллект» трактуется как «раздел информатики, включающий разработку методов моделирования и воспроизведения с помощью ЭВМ отдельных функций творческой деятельности человека, решение проблемы представления знаний в ЭВМ, построение баз знаний, создание экспертных систем, разработку так называемых интеллектуальных роботов» [Большой энциклопедический словарь <http>]. Помимо компьютерных терминов в данном видеоряде встречаются такие слова, как «человек», «создавать», «самый», «мочь», «наш», «такой», «который», «быть», которые нельзя отнести к области КТ.

### Результаты исследования

Проведенное исследование указало на возможность применения использованной нами методики для изучения компьютерной терминологии. Интерпретация результатов статистического анализа позволила установить, что КТ существует в разных типах дискурса в различном процентном соотношении, что ранее для данного типа лексики не являлось типичным. Также мы хотим сделать акцент на том, что в отдельных случаях термины, относящиеся к сфере КТ, были образованы при помощи слов, которые близки человеку по каким-либо параметрам и критериям, что свидетельствует о наличии феномена проксимии в процессе формирования современной компьютерной лексики.

## Библиографический список

*Бабалова Г.Г.* Системно-аспектуальное функционирование компьютерной терминологии: дис. ... д-ра филол. наук: 10.02.19. Москва, 2009. 380 с.

*Большой энциклопедический словарь* [Электронный ресурс]. URL: <https://dic.academic.ru/dic.nsf/enc3p/141814> (дата обращения: 13.06.2021).

*Викисловарь* [Электронный ресурс]. URL: <https://ru.wiktionary.org/wiki/фотогалерея> (дата обращения: 15.06.2021).

*Интернет-гигантов приземлят в России* [Электронный ресурс] // Эксперт. URL: <https://expert.ru/2021/06/1/internet-gigantov-prizemlyat-v-rossii/> (дата обращения: 10.06.2021).

*Искусственный разум. Революция. Фильм 1* [Электронный ресурс]. URL: <https://www.youtube.com/watch?v=WBBWOMkixTs&t=332s> (дата обращения: 10.06.2021).

*Комлева И.Л.* Принципы формирования русской компьютерной терминологии: дис. ... канд. филол. наук: 10.02.01. Москва, 2006. 221 с.

*Компьютерные науки нацелены на самосовершенствование* [Электронный ресурс] // КОММЕРСАНТЪ. URL: <https://www.kommersant.ru/doc/4837210> (дата обращения: 10.06.2021).

*Кондратенко Е.Н.* Идеологические проксимы в индивидуальном лексиконе (экспериментальное исследование): дис. ... канд. филол. наук: 10.02.19. Курск, 2014. 175 с.

*Культура русской речи и эффективность общения* / отв. ред. Л.К. Граудина, Е.Н. Ширяев. Рос. акад. наук, Ин-т рус. яз. им. В.В. Виноградова. М.: Наука, 1996. 439 с.

*Лебедева С.В.* Близость значения слов в индивидуальном сознании: дис. ... д-ра филол. наук: 10.02.19. Тверь, 2002. 311 с.

*Лебедева С.В.* Возможность моделирования процесса проксимации [Электронный ресурс] // Теория языка и межкультурная коммуникация. 2013. № 1 (13). С. 40–44. URL: <http://tl-ic.kursksu.ru/pdf/013-008.pdf> (дата обращения: 05.06.2021).

*Лебедева С.В.* Проксимы в языке профессии [Электронный ресурс] // Вестник Челябинского государственного университета, 2011. № 24 (239). С. 26–28. URL: [https://elibrary.ru/download/elibrary\\_17685725\\_20207345.PDF](https://elibrary.ru/download/elibrary_17685725_20207345.PDF) (дата обращения: 21.06.2021).

*Лебедева С.В.* Психолингвистические характеристики проксимии [Электронный ресурс] // Ученые записки. Электронный научный журнал Курского государственного университета, 2012. № 3-2 (23). С.41–44. URL: [https://elibrary.ru/download/elibrary\\_18334498\\_71565682.pdf](https://elibrary.ru/download/elibrary_18334498_71565682.pdf) (дата обращения: 19.06.2021).

*Лебедева С.В.* Субъективный семантический функциональный центр: вопросы теории [Электронный ресурс] // Теория языка и межкультурная

коммуникация. 2014. № 2 (16). С. 34–38. URL: <http://tl-ic.kursksu.ru/pdf/016-006.pdf> (дата обращения: 05.06.2021).

*Лобанова М.А.* Структурно-семантические особенности современной компьютерной терминологии: на материале испанского языка : дис. ... канд. филол. наук: 10.02.19. Челябинск, 2009. 242 с.

*Орлова М.В.* Специфика идентификации компьютерной терминологии: экспериментальное исследование: дис. ... канд. филол. наук: 10.02.19. Курск, 2008. 168 с.

*Словарь SEO терминов* [Электронный ресурс]. URL: <https://seo.slovonline.com/> (дата обращения: 22.06.2021).

*Толковый словарь В. Даля* – Толковый словарь живого великорусского языка В.И. Даля [Электронный ресурс]. URL: <https://gufo.me/dict/dal> (дата обращения: 15.06.2021).

*Толковый словарь Ожегова* – Ожегов С.И. Толковый словарь русского языка [Электронный ресурс]. URL: <https://gufo.me/dict/ozhegov/> (дата обращения: 17.06.2021).

*Толковый словарь русского языка* [Электронный ресурс]. URL: <https://www.vedu.ru/expdic/> (дата обращения: 22.06.2021).

*Федеральный закон от 27.07.2006 N 149-ФЗ* (ред. от 09.03.2021) «Об информации, информационных технологиях и о защите информации» [Электронный ресурс]. URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_61798/c5051782233acca771e9adb35b47d3fb82c9ff1c/](http://www.consultant.ru/document/cons_doc_LAW_61798/c5051782233acca771e9adb35b47d3fb82c9ff1c/) (дата обращения: 22.06.2021).

*ADVEGO* – Инструмент для SEO-анализа [Электронный ресурс]. URL: <https://advego.com/text/seo/> (дата обращения: 06.06.2021).

*Amazon Is Having a Sale on AirPods Pro and Apple Watches* [Электронный ресурс] // Mental Floss. URL: <https://www.mentalfloss.com/article/629722/amazon-sale-airpods-pro-and-apple-watch-series-3> (дата обращения: 10.06.2021).

*As Digital Marketers Struggle To Safeguard Their Brands, Another Ad-Tech Firm Plans Its IPO* [Электронный ресурс] // Forbes. URL: <https://www.forbes.com/sites/martyswant/2021/06/04/as-digital-marketers-struggle-to-safeguard-their-brands-another-ad-tech-firm-plans-its-ipo> (дата обращения: 10.06.2021).

*Computer System Validation. Chapter 1* [Электронный ресурс] // LinguaLeo. URL: <https://lingualeo.com/nl/jungle/computer-system-validation-chapter-1-validation-resposibilities-101468> (дата обращения: 10.06.2021).

*Computer-generated realities are becoming ubiquitous* [Электронный ресурс] // The Economist. URL: <https://www.economist.com/technology-quarterly/2020/10/01/computer-generated-realities-are-becoming-ubiquitous> (дата обращения: 10.06.2021).

*Disposable cameras are the future* [Электронный ресурс] // British Council. Learn English teens. URL: <https://learnenglishteens.>

[britishcouncil.org/magazine/science-technology/disposable-cameras-are-future](https://britishcouncil.org/magazine/science-technology/disposable-cameras-are-future)  
(дата обращения: 10.06.2021).

*DOWNSUB* – Инструмент для извлечения субтитров [Электронный ресурс]. URL: <https://downsub.com/> (дата обращения: 10.06.2021).

*Miratext* – Инструмент для SEO-анализа [Электронный ресурс]. URL: [https://miratext.ru/seo\\_analiz\\_text](https://miratext.ru/seo_analiz_text) (дата обращения: 05.06.2021).