

УДК 81'23

## ВЫДЕЛЕНИЕ КЛЮЧЕВЫХ СЛОВ В ВЕРБАЛЬНЫХ И НЕВЕРБАЛЬНЫХ ПАТТЕРНАХ<sup>1</sup>

**Татьяна Евгеньевна Петрова**

к. филол. н., доцент кафедры русского языка

как иностранного и методики его преподавания

Санкт-Петербургский государственный университет

199034, г. Санкт-Петербург, Университетская наб., 11. t.e.petrova@spbu.ru

**Елена Игоревна Риехакайнен**

к. филол. н., доцент кафедры общего языкознания

Санкт-Петербургский государственный университет

199034, г. Санкт-Петербург, Университетская наб., 11. e.riehakajnen@spbu.ru

**Анастасия Сергеевна Кузнецова**

участник программы «Большие вызовы»

Образовательный центр «Сириус»

354349, г. Сочи, Олимпийский просп., 40

**Артём Валерьевич Мараев**

участник программы «Большие вызовы»

Образовательный центр «Сириус»

354349, г. Сочи, Олимпийский просп., 40

**Михаил Андреевич Шаталов**

участник программы «Большие вызовы»

Образовательный центр «Сириус»

354349, г. Сочи, Олимпийский просп., 40

В статье представлены результаты экспериментального психолингвистического исследования, направленного на сравнение восприятия информации, представленной в только вербальной форме (в обычном текстовом формате) и в формате инфографики. В ходе исследования 77 носителям русского языка было дано задание ознакомиться с двумя вербальными текстами и с двумя текстами в формате инфографики и выделить ключевые слова в каждом из них. Согласно полученным результатам, при определении ключевых слов, т. е. той информации, которая связана с цельностью текста, носители русского языка опираются прежде всего на вербальный компонент даже в текстах в формате инфографики, вследствие чего наборы ключевых слов практически не различаются для разных форматов текста.

**Ключевые слова:** психолингвистика; восприятие письменной речи; текст; инфографика; методика ключевых слов; русский язык.

Тамара Ивановна Ерофеева является основоположником Пермской школы социолингвистики, которая возникла на основе развития идей Петербургской лингвистической школы и опирается на теоретические положения Л.В. Щербы, Б.А. Ларина, Л.В. Вербицкой, Л.В. Бондарко, А.С. Штерн, Л.В. Сахарного и др. С именами двух последних тесно связана теория текстопрIMITИВОВ и психолингвистические методики анализа текста. Одним из направлений деятельности Пермской школы социо- и психолингвистики также является изучение текста, его свойств, внутреннего и внешнего устройства, процессов порождения и восприятия [Мурзин, Штерн, 1991; Ерофеева Е.В., Ерофеева Т.И. 2010].

Было показано, что цельность, которая является одним из основных свойств текста и соотносится с его смысловой, содержательной стороной, может быть отражена в тексте с большей или меньшей степенью подробности [Сахарный 1991]. При этом описание цельности коррелирует с иерархичностью структуры восприятия текста, и само восприятие может быть более или менее конкретным, полным, детальным – в зависимости от целей, условий коммуникации, знакомства с предметной областью, индивидуальных особенностей читающего. С точки зрения психолингвистики, цельность, как парадигматический феномен, позволяет считать сопоставимые по смыслу тексты-синонимы членами одного ряда и говорить о «парадигме текстов» – от текстопрIMITИВОВ до развернутых текстов больших размеров. Разница между членами такой парадигмы заключается в характере расчленения цельности, степени детализации такого расчленения. Тексты-прIMITИВЫ, несмотря на «экзотичность» их структуры с точки зрения традиционной лингвистики, представляют собой вполне естественные результаты работы механизма порождения речи. В них реализуются все основные характеристики текста, которые и позволяют с его помощью осуществить акт коммуникации [там же]. Выделяют первичные тексты-прIMITИВЫ (тексты детской речи, речи при афазии, реплики на иностранном языке, реплики в диалоге в разговорной речи, вывески и т. п.), которые являются результатом «первичного»,

непосредственного кодирования или вербализации цельности, и вторичные тексты-прIMITИВЫ – как компрессированный вариант развернутого текста, как результат «вторичного», «свернутого» кодирования, т. е. перифразы к исходному, «первичному» развернутому тексту (тезисы-конспекты готового текста, предметные рубрики, наборы ключевых слов). По мнению Л.В. Сахарного, вторичные тексты-прIMITИВЫ являются эффективным материалом для экспериментальных исследований, благодаря возможности сопоставления с исходным развернутым текстом [Сахарный 1984].

**Цель** настоящего исследования заключалась в изучении специфики восприятия графических (инфографики) и традиционных текстов и определении того, можно ли их считать членами одной парадигмы текстов.

Выбор текстов в формате инфографики (т. е. сопровождающихся краткими подписями рисунков, графиков, диаграмм и т. п.) в качестве материала исследования не случаен. В настоящее время в связи с тенденцией к использованию визуальной информации во многих сферах жизни особую актуальность получает исследование поликодовых текстов, интегрирующих в себе вербальные и невербальные средства, позволяющих с помощью инфографики максимально четко представить сложную информацию [Баранова 2013: электр. ресурс]. Учебная литература, согласуясь с современными образовательными технологиями, должна быть направлена на то, чтобы в нее включались все виды восприятия обучающегося – и зрительное восприятие в том числе [Соколова 2016; Соловьева 2010]. Тем не менее, несмотря на многие очевидные преимущества текстов в графическом формате, их доля в школьных учебниках и пособиях по русскому языку для иностранцев все еще остается небольшой, а механизмы процесса восприятия и понимания инфографики не изучены.

В качестве **материала** исследования были использованы тексты в формате инфографики из научно-популярного журнала «Кот Шрёдингера» (<http://kot.sh>) и их «переводы» в обычный текстовый формат<sup>2</sup>. Примеры подобных текстов приведены на рис. 1.

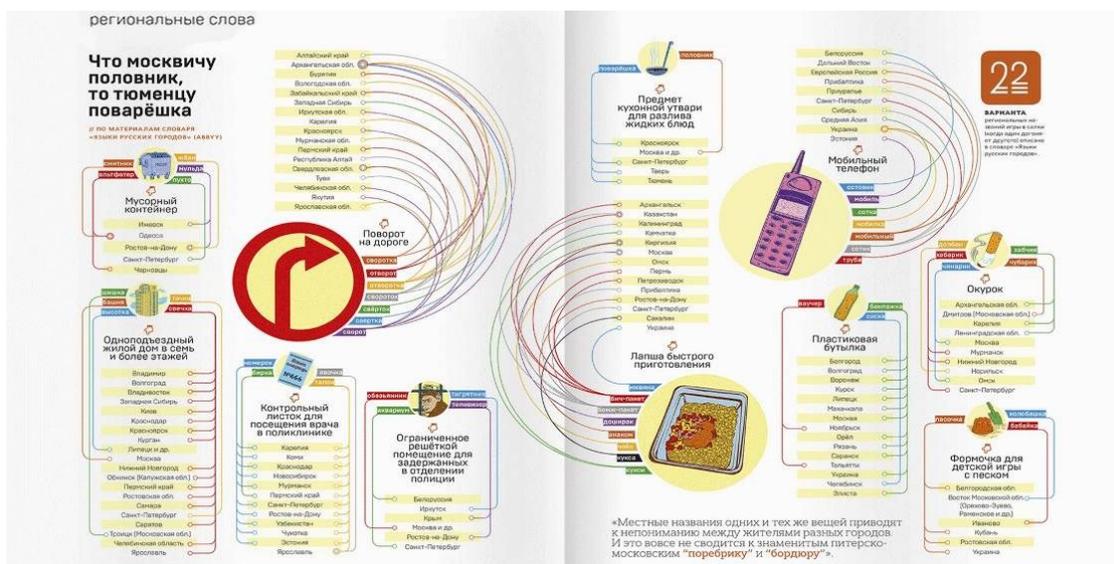
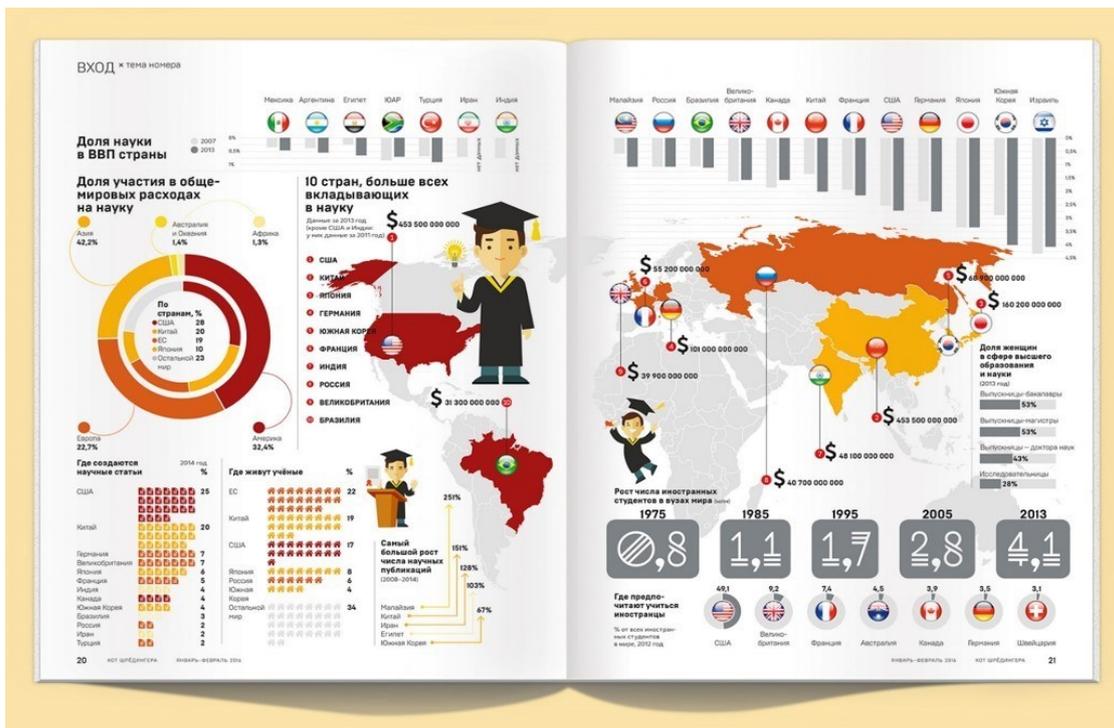


Рис. 1. Примеры текстов в формате инфографики

Предварительный этап исследования заключался в отборе из имевшихся в нашем распоряжении 11 пар текстов (инфографика – текстовый формат) тех, которые, по мнению экспертов, были наиболее релевантны для целей нашего исследования. Для субъективной оценки текстов пятью экспертами применялась методика семантического дифференциала. Семантический дифференциал служит для качественного и количественного определения значений с помощью двухполюсных шкал, задаваемых парами антонимичных прилагательных [Snider, Osgood 1969]. В нашем случае использовалась шкала с 5 градациями. Было выбрано четыре шкалы для оценивания текстов: инфорграфичный/неинфорграфичный, удачный/неудачный, интересный/не-

интересный и понятный/непонятный. В результате было отобрано шесть пар текстов, получивших наиболее высокие оценки экспертов по вышеперечисленным шкалам. Тематика этих текстов была близка либо гуманитарной, либо естественно-научной. Мы дали им условные названия: «ВВП<sup>3</sup>», «Перевод», «Работа», «Половник», «Будущее», «МГУ». Для их чтения и понимания не требовалось наличие особых знаний в какой-либо из научных областей. Их объем в текстовом формате был приблизительно одинаков.

**Гипотеза** состояла в следующем: вторичные тексты-перифразы, а именно наборы ключевых слов к тексту в формате инфографики и аналогичному вербальному тексту, будут совпадать.

**Методика исследования**

Нами был использован алгоритм определения истинного набора ключевых слов (НКС) текста, предложенный в работах А.С. Штерн и С.А. Сиротко-Сибирского [Мурзин, Штерн 1991; Сиротко-Сибирский 1988], в которых показано, что для выделения НКС достаточно 25–30 чел. Испытуемые выделяют свои наборы, а затем высчитывается частота индексирования (т. е. появления в качестве ключевого) каждого слова. Данная методика широко используется в психолингвистических исследованиях по сей день и может быть положена в основу экспертизы успешности/неуспешности понимания различных типов текста [Психофизиологические и нейролингвистические аспекты процесса распознавания вербальных и невербальных паттернов коммуникации 2016; Петрова 2006]. Экспериментальных исследований восприятия инфографики на материале русского языка, насколько нам известно, до сих пор не проводилось.

В основном эксперименте приняли участие 77 человек в возрасте от 14 до 27 лет (участники образовательной программы «Большие вызовы» Образовательного центра «Сириус» и кураторы, работающие в ОЦ «Сириус»). Каждый участник эксперимента читал в случайном порядке четыре текста на разные темы: два текста в обычном формате и два текста в формате инфографики. Задание состояло в следующем: прочитать тексты и выписать 10 ключевых слов для каждого из этих них. Каждый текст читали не менее 25 чел. (т. е. было составлено три протокола, в каждом по четыре разных текста – два в формате инфографики, два – в обычном текстовом формате). В итоге было получено 3 014 ключевых слов для 12 текстов.

**Обработка и анализ результатов**

Обработка данных проводилась следующим образом: слова сводились к словарной форме (*работают – работать*), объединялись однокоренные слова (*научный – наука*), словосочетания

делились на слова, входящие в них (*перевод книг – перевод и книги*), словосочетания «существительное + числительное» сводились к существительному (*700 школьников – школьники*), аббревиатуры считались за одно слово (*МГУ*), если слово встречалось в словосочетании и вне его, то это словосочетание делилось на два слова и более (*лауреат, нобелевский лауреат – нобелевский и лауреат*). После обработки мы получили 1 123 разных ключевых слова.

По результатам эксперимента для каждого из 12 текстов был выявлен истинный НКС, в который вошли слова, имеющие наибольшую частоту индексирования, а именно те, которые встретились в ответах не менее 33% испытуемых.

Анализ результатов показал, что ключевые слова, вошедшие в эти наборы, изначально присутствуют в исходном стимульном тексте или текстовой части инфографики. Это наблюдение соотносится с представленными в [Rayner et al. 2001] данными о том, что если изображение не отличается по смыслу от текста, то основное внимание уделяется текстовой части, а правильность ответов на вопросы по содержанию текста не связана с типом изображений. Кроме того, полученные нами данные совпадают с результатами исследования движения глаз при разглядывании геометрических фигур и подписей к ним [Lee, Wu 2017], в котором было показано, что картинка привлекает внимание значительно меньше, чем текст (лишь 40% времени читающий задерживает свое внимание на изображении, остальное время он тратит на текст). Мы можем сделать вывод о том, что при обработке текста любого формата, а именно при построении вторичного текста-перефразы, ключевая информация берется из текстовой части.

Затем наборы ключевых слов для каждой из 6 пар текстов (в формате инфографики и в обычном формате) были сопоставлены между собой. Результаты этого попарного сравнения представлены в таблице.

**Результаты анализа наборов ключевых слов к текстам разных форматов**

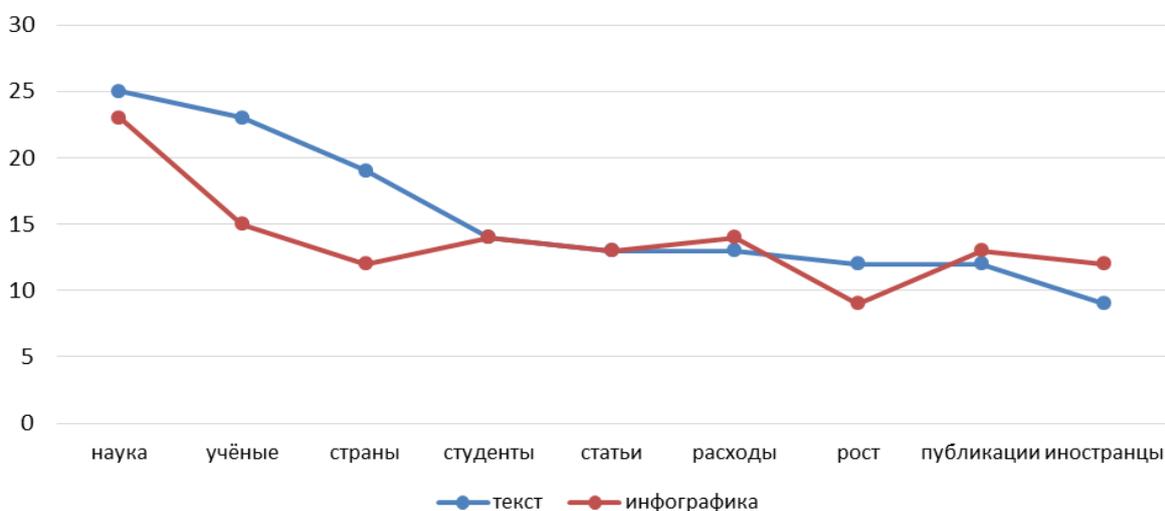
Параметры	ТЕКСТЫ											
	«ВВП»		«Перевод»		«Работа»		«Будущее»		«Половник»		«МГУ»	
	текст	инф.	текст	инф.	текст	инф.	текст	инф.	текст	инф.	текст	инф.
Кол-во разных ответов, абс.	73	130	58	68	140	71	135	76	151	178	142	168
Кол-во КС, абс.	49	70	41	69	109	74	110	67	101	162	89	131
Частотность самого индексированного КС, абс.	24	42	16	50	25	10	25	12	22	16	18	19
Совпадение парных текстов по 10 КС, %	70,0		71,4		58,3		36,4		60,0		60,0	

Как видно из таблицы, большой разброс выделенных КС получен к текстам в формате инфографики. Согласие оценок в выборе того или иного КС у участников эксперимента выше, если они имеют дело с обычным текстовым форматом. Мы можем говорить об иерархии инвариантных и вариантных моментов в передаче цельности в каждом конкретном тексте. Ожидаемо, что словосочетаний и предложений, выделенных в качестве КС к тексту, встретилось больше после прочтения привычного печатного формата.

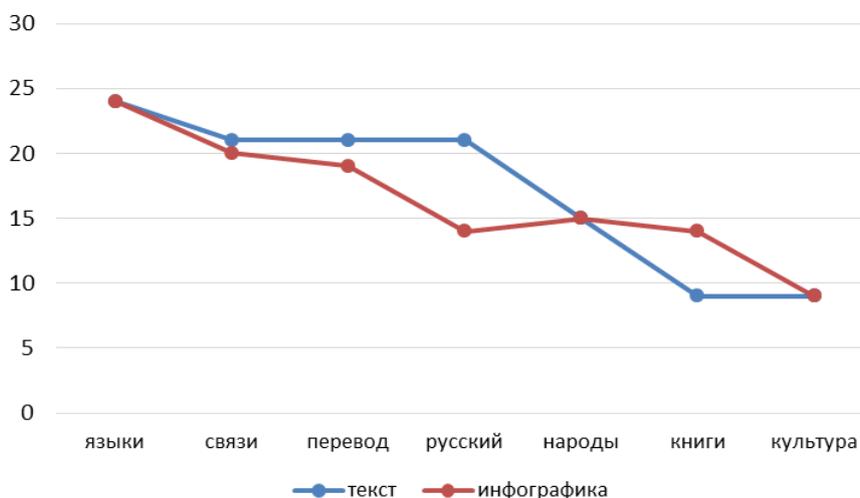
Процент пересечений наборов ключевых слов к тексту в формате инфографики и аналогичному вербальному тексту довольно высок, но не для всех текстов. Анализ материала позволяет отобрать наиболее тесно пересекающиеся

по составу НКС пары «текст – инфографика», внутри которых расхождения по содержанию минимальны (это тексты «ВВП» и «Перевод»). Что касается пары текстов «Будущее», то говорить о том, что инфографика передает то же самое содержание, что и текстовый формат, в данном конкретном случае не представляется возможным.

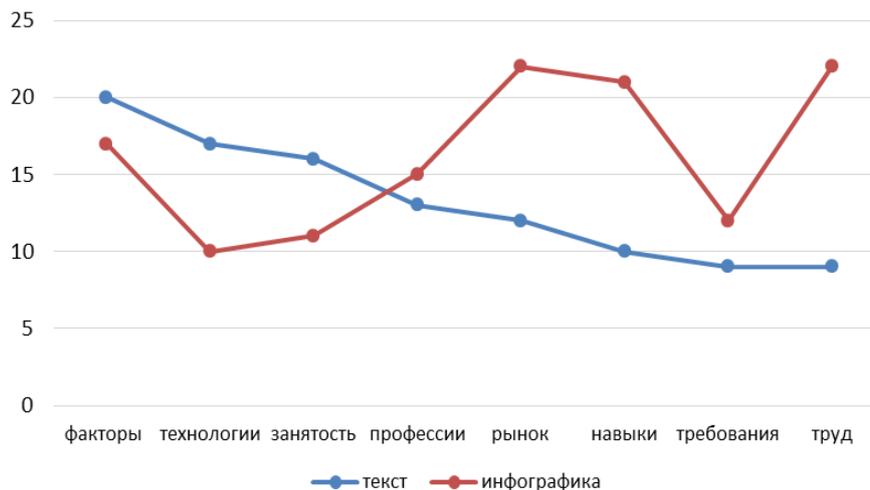
Далее для каждого текста был определен порядок следования КС в частотном списке и построен график частоты встречаемости КС в ответах испытуемых. Графики частоты встречаемости КС в ответах испытуемых представлены на рис. 2–5. Наиболее удачные (в смысле совпадения содержания) пары текстов представлены на рис. 2 и 3, наименее совпадающие – на рис. 4 и 5.



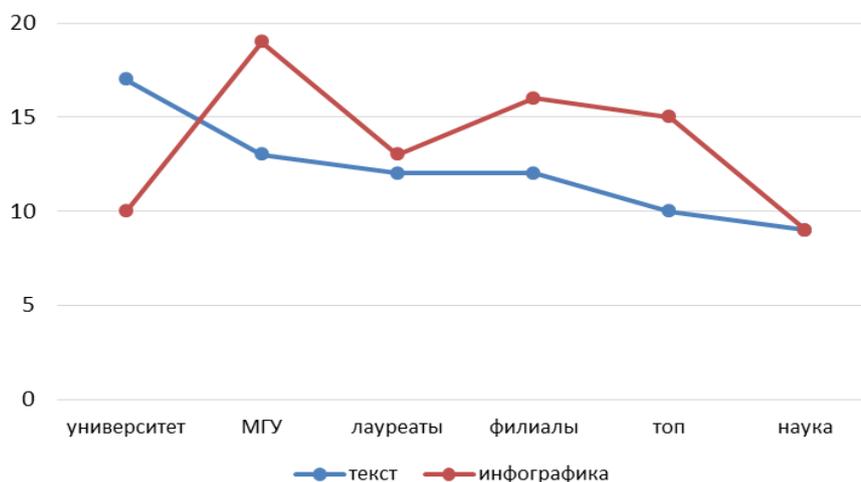
**Рис. 2. График НКС текста «ВВП» в обычном формате и в формате инфографики:** по оси абсцисс – КС в порядке их следования в частотном списке, по оси ординат – количество испытуемых, выделивших то или иное КС



**Рис. 3. График НКС текста «Перевод» в обычном формате и в формате инфографики:** по оси абсцисс – КС в порядке их следования в частотном списке, по оси ординат – количество испытуемых, выделивших то или иное КС



**Рис. 4. График НКС текста «Работа будущего» в обычном формате и в формате инфографики:**  
 по оси абсцисс – КС в порядке их следования в частотном списке,  
 по оси ординат – количество испытуемых, выделивших то или иное КС



**Рис. 5. График НКС текста «МГУ» в обычном формате и в формате инфографики:**  
 по оси абсцисс – КС в порядке их следования в частотном списке,  
 по оси ординат – количество испытуемых, выделивших то или иное КС

По результатам проведенного исследования каждый из участников эксперимента получил обратную связь, т.е. информацию о том, насколько (в процентах) его индивидуальный НКС к каждому из прочитанных им текстов совпадал с истинным набором, полученным на всей группе участников, что позволило ему/ей, во-первых, узнать, насколько хорошо он/она владеет техникой выделения ключевых слов при обработке информации, представленной в письменном виде, во-вторых, узнать, какой тип текста для него/нее предпочтителен (обычный текстовый формат или формат инфографики), в-третьих, получить рекомендации по улучшению навыков восприятия и понимания текстов (были предложены некоторые типы тренировочных упражнений).

Анализ полученных результатов показал, что лишь 10 из 77 участников эксперимента (менее 10%) одинаково успешно справляются с выделением КС в текстах любого формата: процент совпадений их истинных НКС к исходному развернутому тексту с истинным НКС очень высок по всем четырем предъявленным им стимульным текстам. Десять участников успешнее справились с заданием при работе с текстовым форматом, 13 – с инфографикой. Треть участников (29 чел.) плохо владеют техникой выделения ключевых слов, смысловых вех текста как в формате инфографики, так и в обычном текстовом формате. Для почти 20% участников при выполнении задания на выделение ключевых слов значимым оказался не фактор формата, а специфика самого текста (тематики, особенностей оформления и т. п.).

### Заключение

Процесс обработки основного содержания текста в формате инфографики и аналогичном вербальном тексте происходит сходным образом, при этом при анализе инфографики человек опирается преимущественно на информацию, выраженную вербально.

В когнитивном плане вырисовывается следующая картина: когда текст задан как объект восприятия (не важно, в каком формате) – он интерпретируется (перефразируется) с учетом универсальных стратегий. Если бы этих стратегий не было, то иллюзии восприятия не совпадали бы и не были бы однотипными у разных людей; не наблюдалось бы семантической адекватности, которая имеет место в том случае, когда очерчивается «коридор смыслов» (А.А. Брудный), совпадающий у автора и реципиента.

Именно поэтому тексты с совпадающими наборами ключевых слов можно рассматривать как входящие в одну парадигму текстов. И именно этот принцип следует учитывать при создании инфографики, понятной школьникам, в учебниках нового образца.

Методика НКС, использованная в исследовании, может быть положена в основу проверки того, насколько предпочтителен тот или иной формат текста для конкретного человека.

Дальнейшее изучение восприятия и обработки текстов в формате инфографики представляет не только практический, но и теоретический интерес для теории текста и психолингвистики, которая от узко лингвистического толкования текста как письменного произведения больших размеров [Гальперин 1981] переходит к более широкому – семиотическому – пониманию этого феномена коммуникации [Мурзин, Штерн 1991].

### Примечания

<sup>1</sup> Исследование выполнено при поддержке гранта РФФИ № 14-18-02135 «Психофизиологические и нейролингвистические аспекты процесса распознавания вербальных и невербальных паттернов».

<sup>2</sup> Авторы исследования выражают благодарность д. п. н., проф. Елене Ивановне Казаковой за предоставленные материалы.

<sup>3</sup> ВВП – валовой внутренний продукт.

### Список литературы

Баранова Е.А. Все, что вы должны знать, если хотите развивать инфографику на газетном сайте // Электронный научный журнал МГУ имени М.В. Ломоносова «Медиаскоп». 2013. № 4. [Электронный ресурс]. URL: <http://www.mediascope.ru/node/1435> (дата обращения: 11.08.2017)

Брудный А.А. Психологическая герменевтика. М.: Лабиринт, 1998. 201 с.

Гальперин И.Р. Текст как объект лингвистического исследования. М.: Наука, 1981. 140 с.

Ерофеева Е.В., Ерофеева Т.И. Человек и текст: антропоцентрический подход к исследованию // Вестник Пермского университета. Российская и зарубежная филология. 2010. № 4(10). С. 28–33.

Мурзин Л.Н., Штерн А.С. Текст и его восприятие. Свердловск: Изд-во Урал. гос. ун-та, 1991. 172 с.

Петрова Т.Е. Контекстная предсказуемость ключевых слов текста // ...Слово отзовется: памяти А.С. Штерн и Л.В. Сахарного / Перм. гос. ун-т. Пермь, 2006. 312 с.

Сахарный Л.В., Сибирский С.А., Штерн А.С. Набор ключевых слов как текст // Психолого-педагогические и лингвистические проблемы исследования текста. Пермь, 1984. С. 81–83.

Сахарный Л.В. Тексты-примитивы и закономерности их порождения // Человеческий фактор в языке: язык и порождение речи. М.: Наука, 1991. С. 221–237.

Сиротко-Сибирский С.А. Смысловое содержание текста и его отражение в ключевых словах (на материале русских текстов публицистического стиля): автореф. дис. ... канд. филол. наук. Л., 1988. 242 с.

Соколова Ю.В. Инфографика как продукт графического дизайна: проблема определения понятия // Культурологические чтения – 2016: материалы междунар. науч.-практич. конф. / Урал. федер. ун-т. Екатеринбург, 2016. С. 254–261.

Соловьева Т.В. Инфографика в медийном и учебном текстах // Вестник Новгородского государственного университета им. Ярослава Мудрого. 2010. № 57. С. 76–79.

Психофизиологические и нейролингвистические аспекты процесса распознавания вербальных и невербальных паттернов коммуникации / под науч. ред. Т.В. Черниговской, Ю.Е. Шелепина, О.В. Заширинской. СПб.: Изд-во ВВМ, 2016. 203 с.

Lee W.-K., Wu Ch.-J. Eye Movements in Integrating Geometric Text and Figure: Scanpaths and Given-New Effects // International Journal of Science and Mathematics Education. 2017. Pp. 1–16.

Rayner K. et al. Integrating text and pictorial information: eye movements when looking at print advertisements / Rayner K., Rotello C. M., Stewart A. J., Keir J., Duffy S. A. // Journal of Experimental Psychology: Applied. 2001. Vol. 7(3). Pp. 219–226.

Snider J. G., Osgood Ch. E. Semantic Differential Technique: A sourcebook. Aldine: Aldine Publishing Company, 1969. 681 p.

## KEYWORDS IN VERBAL AND NON-VERBAL PATTERNS

**Tatiana Ye. Petrova**

Assistant Professor of Russian as a Foreign Language Department  
Saint-Petersburg State University

**Elena I. Riekhakaynen**

Assistant Professor of General Linguistics Department  
Saint-Petersburg State University

**Anastasia S. Kuznetsova**

Participant of the Programme “Big Challenges”  
“Sirius” Educational Centre

**Artem V. Maraev**

Participant of the Programme “Big Challenges”  
“Sirius” Educational Centre

**Mikhail A. Shatalov**

Participant of the Programme “Big Challenges”  
“Sirius” Educational Centre

The article presents the results of an experimental psycholinguistic study aimed to compare the processing of information presented in text format and in the format of infographics. In the study, 77 native speakers of Russian were given the task to read two verbal texts and two infographics and to write out the key words for each of them. According to the results, when defining key words, i.e. the information that is associated with the coherence of the text, Russian speakers rely primarily on the text component even in infographics, whereby the sets of keywords are practically the same for different formats of the text.

**Keywords:** psycholinguistics; written text recognition; text; infographics; keywords method; Russian.