

*Антонина Лапошина, Мария Лебедева*

*Москва, Россия*

## **КОРПУСНЫЙ ПОДХОД К РЕШЕНИЮ ПРОБЛЕМЫ ОТБОРА ЛЕКСИКИ В ОБУЧЕНИИ РКИ**

### **1. Введение**

«Сколько слов и какие слова мне нужно выучить в первую очередь, чтобы начать говорить по-русски и понимать русскую речь?» – закономерный вопрос студента, начинающего изучать РКИ. Предполагается, что ответ на этот вопрос дают лексические минимумы – списки слов, которые должны быть усвоены на определенном этапе обучения для решения соответствующих этому этапу коммуникативных задач.

Однако на деле практикующий преподаватель РКИ сталкивается с недостатком словников, актуальных с точки зрения современного языка и оптимальных с точки зрения лингводидактики. Так, лексические минимумы, разработанные в рамках государственной системы тестирования по русскому языку как иностранному (ТРКИ), ориентированы прежде всего на иноязычных абитуриентов, студентов подготовительных отделений российских вузов, а потому обнаруживают заметный крен в сторону академической и учебно-бытовой лексики (Андрюшина и др. 2011, 2012, 2017а, 2017б, 2018). Словники ТРКИ, а также большинство других списков слов (например, размещенных в учебниках РКИ) создавались в «докорпусную эпоху», что ставит под сомнение их способность отражать реальную коммуникативную практику носителей русского языка. Кроме того, на современном этапе не решена задача разграничения списков слов для активных и пассивных видов речевой деятельности на объективных основаниях («хочу говорить» vs. «хочу понимать» в вопросе студента).

Таким образом, современная лингводидактика РКИ находится в поиске оптимальных лексических минимумов. Актуальность задачи осознается практикующими преподавателями и авторами учебно-методических пособий по РКИ, российскими и зарубежными русистами – об этом свидетельствует, в частности, появление нескольких проектов по созданию частотных словников для изучающих русский язык. Некоторые принципы и подходы к составлению лексических минимумов и аналогичных им списков слов<sup>1</sup>, используемые в русистике, рассматриваются ниже.

Не вызывает сомнений, что на современном этапе развития лингвистики задача создания лексических минимумов должна решаться с опорой на корпусные данные. В статье мы анализируем некоторые случаи применения корпуса для принятия решения о лингводидактической ценности того или иного слова с позиций РКИ и обсуждаем возможные подходы к созданию лексических минимумов, соответствующих следующим критериям:

- ориентация на реальное употребление (узус),
- учет как частотности слова, так и его методической ценности,
- учет коммуникативных задач и ситуаций, определенных уровнем владения русским языком,
- дифференциация лексики для активных и пассивных видов речевой деятельности.

---

<sup>1</sup> Мы разграничиваем понятия «лексический минимум» и «список слов для изучения», понимая под первым минимальный набор лексических единиц, предназначенный для усвоения на определенном этапе изучения языка и оптимальный для решения коммуникативных задач этого этапа («максимальный с точки зрения возможностей учащегося и минимальный с точки зрения системы языка» (Балыхина 2006: 40). Однако в статье будут рассмотрены как лексические минимумы, так и другие учебные списки слов, в том числе словники учебников и частотные словари, ориентированные на изучающих РКИ.

## 2. Подходы к созданию лексических минимумов в лингводидактике РКИ

Одна из первых градуальных систем словников для изучающих РКИ, созданная с опорой на статистическую ценность лексики, была предложена в 1985 году коллективом под руководством В. В. Морковкина. Словарь «Лексические минимумы современного русского языка» (Морковкин 1985) содержал в себе 8 частотных списков, включая «нулевой» (перечень служебных, полужнаменательных и др. слов русского языка, формирующих синтаксическую и модальную структуру текста), а также тематический словник. 8 градуальных минимумов были созданы на основе частотных словарей русского языка, которые в свою очередь, опирались на письменные тексты. Это позволило авторам закономерно предположить, что такие минимумы могут обеспечить прежде всего рецептивное владение языком. Отметим, что объем каждого минимума определялся довольно механистически: так, в первый список включались 500 наиболее употребительных единиц русского языка, во второй – 1 000 (включая слова первого списка) и т.д., таким образом, седьмой список представлял собой перечень 3 500 наиболее употребительных слов<sup>1</sup>. В отличие от градуальной серии минимумов, тематический словник, по замыслу авторов, был ориентирован на продуктивные виды речевой деятельности: слова для этого списка отбирались не на основе их частотности в текстах, а на основе их значимости в различных ситуациях общения. Идеографическая организация словника давала возможность находить единицы, объединенные тематически, но не позволяла оценивать их с точки зрения соответствия тому или иному этапу обучения. Обратим, однако, внимание на принципиальное стремление авторов разграничить лексические минимумы для пассивных и активных видов речевой деятельности, причем решить эту задачу с помощью ресурсов, доступных на докорпусном этапе развития лингвистики.

В 90-е годы под эгидой Совета Европы реализовывался проект по созданию лингводидактического описания уровней владения русским языком (под руководством Е. М. Степановой). Одним из компонентов такого описания стал лексический минимум. Так, в 1996 г. вышел том «Русский язык. Повседневное общение. Пороговый уровень» (Волкова и др. 1996), содержащий словник общего владения для уровня В1. Отбор единиц, вошедших в словник, проводился на основе гипотетической модели речевого поведения иностранца в среде носителей русского языка.

Следующим системным проектом в российской лингводидактике РКИ стало создание лексических минимумов как компонента Российской системы тестирования по русскому языку как иностранному<sup>2</sup>. Принципиальной особенностью лексических минимумов является их строгое соответствие интенциям и ситуациям общения, определенным для каждого уровня владения русским языком. При этом значительное место в требованиях к уровням ТРКИ занимает обучение коммуникации в учебной сфере; как отмечают авторы, на содержание словников значительно повлияла лексика, отобранная в процессе обучения на подготовительных факультетах российских вузов (Андрюшина 2011: 649). В связи с этим лексические минимумы ТРКИ не в полной мере решают задачу определения объема лексики для других целей изучения РКИ (личных, туристических, деловых). Отдельно в общем списке отмечаются слова, рекомендованные для активного усвоения, однако критерии для дифференциации лексики активного и пассивного запаса не эксплицированы. Объем каждого лексического минимума методически обоснован.

Авторы выделяют следующие критерии отбора лексических единиц, которые применялись при составлении лексических минимумов ТРКИ: 1) семантическая ценность, т.е.

---

<sup>1</sup> Обновленный вариант лексических минимумов под редакцией В. В. Морковкина (2003) был расширен до 5 000 слов.

<sup>2</sup> Первая редакция списков для 3 уровней (А1-В1) была предложена в 1997 году, а в дальнейшем были разработаны минимумы и для других уровней.

способность слова обозначать часто встречающиеся предметы и явления; 2) высокая словообразовательная способность слова; 3) способность слова входить в различные словосочетания; 4) на первых трёх уровнях владения – стилистическая немаркированность (при этом в отдельных случаях допускается включение в словник стилистически маркированных слов, таких как *картошка*, *тройка* и т.д.); на высоких уровнях владения РКИ – стилистическая принадлежность слова; 5) частотность (по данным частотных словарей, а также частота использования слова в учебниках по РКИ и «тематическая» частотность слова) (см. Андрюшина и др. 2011, 2012, 2017а, 2017б, 2018).

Отметим, что частотность является лишь одним из 5 критериев для включения единицы в лексический минимум, причем, как следует из предисловий, авторы не опирались на информацию о частотности по корпусным данным. Нет никаких сомнений, что употребимость слова не может быть единственным основанием для включения слова в содержание обучения языку, однако очевидна и критическая важность этого критерия для оптимизации списка изучаемых единиц (см. Leech 1997: 16; Aston 2000: 8; Römer 2008: 115). В ряде случаев – некоторые мы рассмотрим в следующем разделе – более частотное слово оказывается и семантически, и стилистически, и коммуникативно более ценным, чем, например, его менее частотный синоним.

Среди работ последних лет, посвященных разработке списков лексики для учебных целей, стоит отметить Kelly – масштабный проект по созданию параллельных лексических списков для 9 языков на основе корпусных данных (Kilgarriff 2010; Johansson, Volodina 2011). По замыслу участников проекта, эти списки должны удовлетворять следующим условиям: отражать современное состояние языка, содержать наиболее частотные лексические единицы, отбор лексики должен производиться по объективным критериям (Johansson, Volodina 2011). Для этих целей были созданы или выбраны из существующих сбалансированные сравнимые корпуса текстов, на основе которых были построены частотные списки из 9 000 слов. Далее каждый такой частотный список переводился несколькими переводчиками на другие языки проекта и обратно – это позволяло выявить наиболее универсальные слова, чаще всего используемые при переводах. Именно такие слова, которые встречались в переводе не менее одного раза, и оказывались в финальных параллельных списках. Кроме этого критерия, учитывалась также значимость в обучении языку (см. Johansson, Volodina 2011: 135).

Исследования для русского языка были выполнены на материале Русского Интернет-корпуса I-Ru и опубликованы в открытом доступе (Frequency Lists For Language Learners, <http://corpus.leeds.ac.uk/serge/kelly>). Результатом стали списки слов, размеченные в соответствии с уровнями владения русским языком. Объемы каждого словника распределены неравномерно: словник для уровня А1 включает 939 слов, А2 – 1 999, В1 – 3 561, В2 – 6 984, С1 – 8 377, С2 – 8 959. Заметно наличие вверху списков слов, называющих актуальные для повседневной жизни реалии: *банкомат*, *супермаркет*, *скорая помощь*, *водительские права* и др. При внимательном рассмотрении, однако, возникают вопросы к тому, насколько тщательно проводилась методическая проработка списков. Так, лексемы *ресторан*, *вокзал*, *кафе* оказались размещены в списке для уровня А2, хотя они, очевидно, обеспечивают коммуникативные задачи элементарного уровня: позволяют ориентироваться в городе (на рецептивном уровне – читать вывески, на продуктивном – спросить дорогу). Словник уровня А1 включает, например, слова *лифчик*, *разведенный*, методическая ценность которых для элементарного уровня вызывает сомнение. Наличие этих слов (и их отсутствие в лексических минимумах и учебниках РКИ) показывает в том числе и разность методических традиций: вероятнее всего, эти слова оказались в словнике русского языка под влиянием списков для других языков.

На материале финального перечня слов для русского языка Kelly был создан также частотный словарь для изучающих РКИ (Sharoff et al. 2013). Он включает в себя 5 000 слов,

сопровождаемых индексом частотности, и 26 тематических списков наиболее употребительных слов. В предисловии к словарю С. Шаров отмечает, что при использовании «объективной» частотности на основании корпуса часто возникает «проблема бананов и апельсинов», при которой в частотные списки не попадают *mots disponibles* – слова повседневной речи, лексемы, обозначающие бытовые явления (Sharoff et al. 2013: 1). Подобная проблема возникает, в частности, из-за особенностей корпусов, на основе которых создаются такие частотные списки: скажем, интернет-корпус I-Ru не способен представить особенности устной повседневной речи, но, очевидно, содержит значительный объем текстов общественно-политической сферы. Такие единицы, как *право, деятельность, процесс, развитие, власть, народ* входят в список 300 наиболее частотных слов, но вряд ли могут быть рекомендованы к изучению на начальном этапе овладения языком.

Таким образом, анализ существующих подходов к отбору лексики для лексических минимумов показывает, что для создания жизнеспособных словариков необходимо учитывать две группы критериев – статистическую ценность слова (желательно, рассчитанную на большом объеме современных корпусных данных) и лингводидактическую ценность слова. При этом первая иллюстрирует дескриптивный подход к отбору языковых единиц, а вторая – прескриптивный. Представляется, что проблема отбора лексики для методических целей заключается, прежде всего, в определении оптимального баланса между двумя подходами. Крен в сторону корпусной частотности ведет к «проблеме бананов и апельсинов», а предпочтение лингводидактической ценности в ущерб учета объективной частотности, напротив, уводит от современного живого языка. На наш взгляд, оптимальные лексические минимумы по РКИ должны создаваться **не на основе** корпусных данных, а **с учетом** корпусных данных (ср. *corpus-based materials vs. corpus-informed materials* в терминологии McCarthy (2008: 566)), то есть при тщательном методическом анализе корпусной информации.

### 3. Примеры учета корпусных данных при отборе лексики в системе РКИ

Закономерно, что, будучи прикладной областью языкознания, методика преподавания языка адаптирует достижения лингвистики к своим нуждам. Так, с развитием корпусной лингвистики разрабатывались идеи использования корпусов в лингводидактических целях, которые в общем можно разделить на две группы: прямой способ (*data-driven learning*), при котором учащийся сам исследует язык с помощью корпуса, и непрямой, при котором на основе корпуса создаются различные средства обучения, от словарей до учебников и учебных пособий (Leech 1997). В лингводидактике РКИ педагогами активно осваивается прямой способ (см., например, Добрушина, Левинзон 2006; Добрушина 2007; Культепина 2018 и пр.). При этом другое направление – разработка содержания обучения РКИ (в том числе лексических минимумов) с учетом корпусных данных – находится в самом начале становления.

Ниже мы рассмотрим несколько практических примеров применения корпусных подсчетов для оптимизации лексических минимумов по РКИ и покажем, как в частных случаях могут приниматься решения о методической ценности слова с учетом информации о его корпусной частотности. Информация о частотности может быть аргументом в пользу включения слова в лексический минимум или, напротив, его исключения из него. Так, в Таблице 1 представлены наиболее яркие примеры сопоставления информации о частотности слова из лексического минимума уровня В1 на материале нескольких корпусов: НКРЯ, корпуса текстов интернет-страниц Russian Web Corpus и корпуса учебных текстов по РКИ (подробнее состав и особенности этих корпусов обсуждаются в 3-м разделе статьи).

Таблица 1. Данные о наличии слова в лексическом минимуме уровня В1 и его частотности по НКРЯ, Russian Web Corpus и RuFoLa

Слово	Присутствие в лексическом минимуме ТРКИ (уровень В1)	Частотность по НКРЯ (1900 - 2017 г.), ipm <sup>1</sup>	Частотность по Russian Web Corpus, ipm	Частотность по RuFoLa <sup>2</sup> , число вхождений
мужество	есть	10.6	12	5
навещать	есть	10	10.1	1
гражданка	есть	7.6	4.1	2
бандероль	есть	2.1	1	0
решение	нет	457.7	557.8	71
компания	нет	447.9	970.6	80
уровень	нет	385.9	537.2	74
туалет	нет	39.4	25.6	20

Даже подобные простейшие вычисления ставят под вопрос целесообразность присутствия отдельных слов в минимуме и «выдвигают кандидатов», более ценных с точки зрения употребимости.

Корпусные данные могут быть также полезны при отборе коллокаций как для лексических минимумов, так и в качестве иллюстрации грамматического материала. Таблица 2 содержит результаты сравнения запроса «ХОДИТЬ<sup>3</sup> В + объект в винительном падеже» с помощью инструмента SketchEngine на материале четырех популярных учебных комплексов для обучения русскому языку (уровни А1, А2 и В1), НКРЯ и Russian Web Corpus. В анализе УМК участвовали все варианты, которые встретились более 1 раза, в Таблице 2 представлены 10 самых частотных (в порядке убывания).

Таблица 2. Сравнение данных запроса «ходить в + объект в вин. п.» на материале учебников и корпусных данных

Дорога в Россию	Первые шаги	5 элементов	Поехали	Корпус Russian Web 2011	Устный подкорпус НКРЯ 1900-2017 г.
театр бассейн гости поликлин ика музей кино банк цирк магазин клуб	театр ресторан школа магазин библиотека гости поликлиника кино университет бассейн	магазин сад театр школа гости бассейн баня клуб университет горы	баня театр университет музей магазин ресторан кафе горы комната кино	<b>туалет</b> садик спортзал кино <b>поход</b> <b>баня</b> <b>церковь</b> сауна школа солярий	школа <b>церковь</b> театр <b>поход</b> бассейн <b>баня</b> магазин клуб ресторан храм

<sup>1</sup> Item per million, среднее число вхождений на миллион.

<sup>2</sup> RuFoLa — корпус учебных текстов по РКИ из учебников и учебно-методических пособий по чтению. Из-за его сравнительно небольшого размера мы указываем абсолютное число вхождений слова в корпус как более информативное.

<sup>3</sup> В поиске участвовали все личные и временные формы глагола *ходить*.

Результаты этого запроса, безусловно, не стоит воспринимать как готовую рекомендацию по включению того или иного слова в содержание обучения. Во многом они отражают особенности корпуса: по всей видимости, женские форумы занимают большой сегмент интернет-сайтов, попавших в Russian Web Corpus, этим может объясняться высокая частотность сочетаний *ходить в туалет / садик / солярий*. Тем не менее интересно отметить некоторые особенности. Первое и довольно комичное несовпадение учебной лексики и информации из корпуса касается лексемы *туалет*. Несмотря на его частотность (см. Таблица 1) и очевидную необходимость в коммуникативных ситуациях элементарного уровня (например, разделы «Ориентация в городе», «Рестораны/кафе», «Здоровье»), *туалет* появляется в лексическом минимуме лишь с уровня В2, словосочетание *ходить в туалет* в приведенных выше УМК отсутствует совсем.

Также в корпусе оказались частотными сочетания, связанные с особенностями русской культуры и уклада жизни: *ходить в баню/церковь/поход*. Из них в учебниках встретился только вариант *ходить в баню* в комплексах «Поехали» и «5 элементов». Лексема *церковь* появляется в лексических минимумах начиная с уровня В1, *баня* и *поход* – с В2. Эта ситуация ярко иллюстрирует несовпадение данных о частотности и мнения авторов о методической ценности данных слов/выражений. В подобных случаях может оказаться полезным разграничение списков активной и пассивной лексики: слова *баня* и *поход* предположительно более ценны для рецептивных видов речевой деятельности. В целом же Таблица 2 позволяет нам сделать вывод, что сочетание *ходить в...* представлено ближе всего к корпусным данным в комплексе «5 элементов», дальше всего от корпуса в описании этого сочетания отстоит комплекс «Дорога в Россию».

Языковые корпуса и другие инструменты также могут помочь в выборе слова из синонимического ряда. Эта задача представляется особенно актуальной при составлении списков для начального этапа обучения, где важно минимизировать объем языкового материала без ущерба для коммуникативной компетенции. В Таблице 3 представлены 3 пары синонимов/когипонимов: *больница/поликлиника, отель/гостиница, бизнесмен/предприниматель*. Для этих пар указано их присутствие в лексических минимумах и частотность по нескольким коллекциям текстов.

**Таблица 3. Встречаемость синонимов и когипонимов лексических минимумах и корпусах**

Слово	Первое появление в лексических минимумах ТРКИ	Частотность по НКРЯ (1900 -2017 г.), ipm	Частотность по Russian Web Corpus, ipm	Частотность по RuFoLa, число вхождений
больница	A2	94.5	67	49
поликлиника	A1	13.5	20	15
отель	нет	35.1	109	9
гостиница	A1	67.1	67	47
бизнесмен	A1	32.9	26	20
предприниматель	B2	53.5	71	16

Из первой пары когипонимов лексические минимумы ТРКИ предлагают сначала, на уровне А1, вводить слово *поликлиника*, а слово *больница* изучать на следующем этапе. Корпусный анализ при этом показывает, что *поликлиника* встречается в несколько раз реже по всем корпусам, включая и корпус учебных текстов РКИ. Кроме того, *больница* «выгоднее» и в методическом плане: это слово короче, имеет словообразовательный потенциал (*болеть — больной — больно*), иллюстрирует частотную словообразовательную модель (ср. *гостиница, теплица*), и наконец, как нельзя более точно отвечает

потребностям «уровня выживания»<sup>1</sup>. На основании комплекса этих фактов можно рекомендовать включать единицу *больница* в словник уровня А1, а единицу *поликлиника* вводить позже.

Ситуация во второй паре синонимов (*отель/гостиница*) более противоречива: слово *отель* вообще не включено в систему лексических минимумов, тогда как *гостиница* появляется в словнике А1. Данные о сравнительной частотности синонимов, как видно из таблицы, разнятся; причина этого явления может крыться в наполнении корпусов. В таком случае можно обратиться к дополнительному инструменту, чтобы сравнить и визуализировать современное состояние употребления слов, – Google Trends. Сервис анализирует популярность запросов в разных сегментах интернета. Информация из таких источников может быть очень ценна, т.к. она объективно отражает интересы и потребности современного человека, использование языка в реальных жизненных обстоятельствах (в данном случае это ситуация бронирования номера в путешествии, которая является одной из коммуникативных задач элементарного уровня). На Рис. 1 представлено сравнение запросов «отель» и «гостиница» в строке информационного поиска в период с 2004 по 2018, где с 2008 года уверенно лидирует запрос «отель». Интернациональность и фонетическая простота этого слова, вместе с информацией о его частотности в современном языке, кажутся весомыми аргументами в пользу включения слова в лексические минимумы начального этапа обучения.

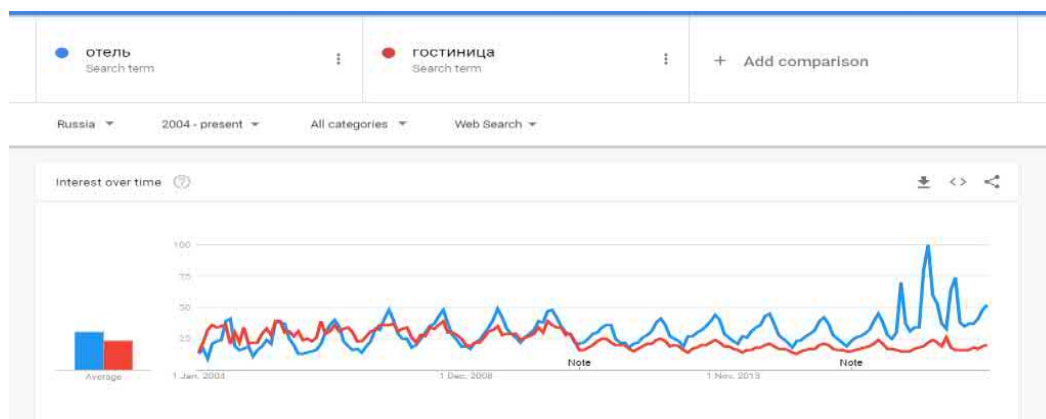


Рис. 1. Сравнение частотности запросов «отель» и «гостиница» в Google Trends.

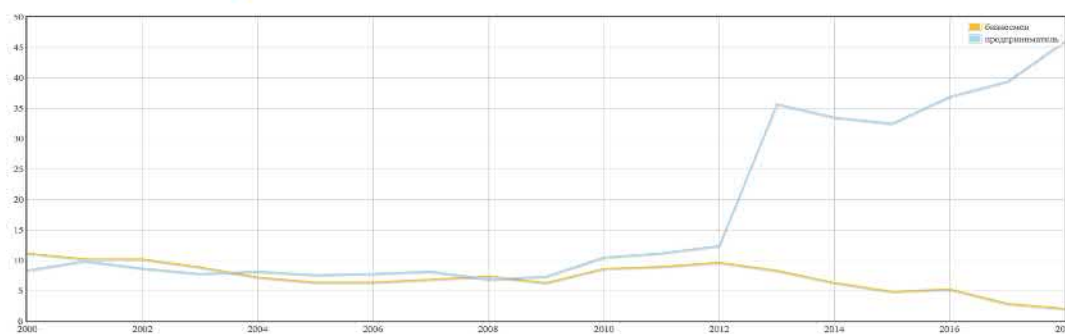


Рис. 2. Сравнение частотности слов *бизнесмен/предприниматель* в период с 2000 по 2018 гг. по данным НКРЯ

Наконец, в третьей паре синонимов – *бизнесмен/предприниматель* – частотнее оказывается более сложное слово (рис. 2). Вероятно, по причине сложности и учебники по РКИ, и лексические минимумы явно отдают предпочтение международному *бизнесмену*,

<sup>1</sup> В то же время понятны и аргументы для предпочтения слова *поликлиника* на начальном этапе: во-первых, его интернациональность, во-вторых, его важность для иностранных студентов российских вузов (*прикрепиться к поликлинике* должны все, а вот с *больницей* столкнутся, как мы надеемся, немногие).

несмотря на его некоторую стилистическую и семантическую ограниченность в современном языке. В таком случае частотность не является решающим аргументом в вопросе о том, какой из синонимов вводить раньше и на каком именно уровне.

#### **4. Корпус для решения лингводидактических задач: в поисках идеала**

Преимущества использования корпусных данных для решения методических задач неоспоримы при условии, что его содержание соответствует целям конкретного исследования. В данном случае это объективная репрезентация языкового наполнения основных коммуникативных ситуаций, с которыми иностранец может столкнуться в общении с носителями языка.

Самым известным и подробно аннотированным корпусом русского языка является НКРЯ. Его основной подкорпус представлен преимущественно произведениями художественной литературы, публицистическими, научно-популярными и учебными текстами, и при использовании этих данных нужно помнить о преобладании определенных стилей и жанров. Подкорпус устной речи включает в себя расшифровки магнитофонных записей публичной и частной устной речи, а также транскрипты кинофильмов, что, безусловно, является ценнейшим материалом для методистов, однако не может использоваться как единственный источник статистической информации о слове.

Russian Web Corpus (ruTenTen) и I-RU (Sharoff 2006) состоят из текстов интернет-страниц. С одной стороны, такие источники могут предоставить более объективную информацию о современном русском языке, с другой стороны, присутствие в источниках тематических форумов может также привести к крену в сторону языковых реалий тех или иных групп пользователей интернета, что проиллюстрировано в Таблице 2.

RuFoLa – корпус учебных текстов по РКИ из учебников и учебно-методических пособий по чтению. Он интересен тем, что содержит актуальные учебные материалы и поэтому отражает реальное положение дел в практике преподавания РКИ. С другой стороны, авторы российских учебников в той или иной степени ориентируются на лексические минимумы ТРКИ, из-за чего возникает своего рода замкнутый круг: слово может быть частотно в учебнике не вследствие высокой методической ценности, а благодаря вхождению в утвержденные словники.

Одним из решений проблемы несовершенства корпуса может стать использование данных нескольких источников. Так, например, В. В. Морковкин (2003) использует усредненную оценку слова сразу по 8 частотным словарям, однако в этом случае успех зависит от источников, на основе которых создавались эти частотные словари. Вероятно, при сборе текстовых коллекций частотных словарей произошел перевес в сторону технических и газетных публикаций, из-за чего, например, слова *бригада* и *аппарат* оказались намного ценнее в статистических рангах, чем *автобус* и *банк*, также в этой работе видна проблема «апельсинов и бананов» в буквальном смысле: *апельсин* попал в ранг 36–60 (самый низший), а *банан* не попал в финальный список вообще (что, впрочем, может быть связано со временем создания частотных словарей и описываемых минимумов).

Другим решением может стать создание специального корпуса текстов для задач лингводидактики РКИ. Опыт создания такого корпуса описан для шведского языка (Johansson, Volodina 2011). Нам представляется перспективным направлением работы создание корпуса аутентичных текстов на основе требований Общеευропейских компетенций, который может быть использован в исследованиях по лингводидактике РКИ как самостоятельный источник, так и в сочетании с информацией из других корпусов. Особенно актуальным является создание такого источника информации для начального уровня обучения, где в сжатые сроки необходимо освоить наиболее полезные и частотные слова и выражения.

Основной особенностью такого корпуса должен стать отбор текстов в соответствии с коммуникативными ситуациями, описанными в требованиях Common European Framework of Reference (CEFR) и требованиях к владению РКИ. Например, для создания



подкорпуса раздела «Ресторан. Кафе. Бар» потребуется коллекция текстов меню и различных рекламных предложений, отзывы о кафе и ресторанах. Для репрезентации диалогов с официантом возможно использование расшифровок субтитров соответствующих фрагментов фильмов/телепередач и устного подкорпуса НКРЯ. Также перспективным в формировании корпуса для нужд РКИ нам представляется использование корпуса «Один речевой день» – коллекции расшифрованных записей речи людей разного возраста и профессий в течение целого дня (Богданова-Бегларян и др. 2017).

Работа над созданием такого корпуса ведется в настоящее время. Предполагается, что в результате будет собрана коллекция текстов, покрывающих жанры, темы и ситуации с опорой на требования Общеввропейских компетенций владения иностранным языком для каждого уровня. Такая коллекция может служить как материалом для лингводидактического описания РКИ, так и выступать «золотым стандартом» репрезентации той или иной темы, что может быть полезным при составлении пособий и учебников по РКИ. В дальнейшем она может быть автоматически расширена за счет семантически близких документов.

## 5. Выводы

Обновление лексических минимумов для русского языка как иностранного с учетом новых возможностей современной лингвистики представляется актуальной задачей. Решение этой задачи заключается в определении оптимального баланса между дескриптивным (статистическим) и прескриптивным (лингводидактическим) подходами к созданию учебных словарей. На наш взгляд, перспективным шагом в этом направлении станет специальный корпус аутентичных текстов, отобранных в соответствии с коммуникативными ситуациями и интенциями, востребованными на каждом из уровней владения иностранным языком.

## Литература

- Андрюшина, Н.П. (2011) Лексические минимумы по русскому языку как иностранному: проблема отбора лексических и фразеологических единиц. *Проблемы истории, филологии, культуры*, 3(33), 648–652.
- Андрюшина, Н.П. (ред.) (2018) *Лексический минимум по русскому языку как иностранному. Третий сертификационный уровень. Общее владение*. Санкт-Петербург: Златоуст.
- Андрюшина, Н.П. (ред.) (2017) *Лексический минимум по русскому языку как иностранному. Первый сертификационный уровень. Общее владение*. Санкт-Петербург: Златоуст.
- Андрюшина, Н.П.; Афанасьева, И.Н.; Битехтина, Г.А.; Клобукова, Л.П.; Яценко, И.И. (2017) *Лексический минимум по русскому языку как иностранному. Второй сертификационный уровень. Общее владение*. Санкт-Петербург: Златоуст.
- Андрюшина, Н.П.; Козлова, Т.В. (2011) *Лексический минимум по русскому языку как иностранному. Базовый уровень. Общее владение*. Санкт-Петербург: Златоуст.
- Андрюшина, Н.П.; Козлова, Т.В. (2012) *Лексический минимум по русскому языку как иностранному. Элементарный уровень. Общее владение*. Санкт-Петербург: Златоуст.
- Балыхина, Т.М. (2006) *Словарь терминов и понятий тестологии*. Москва: Русский язык. Курсы.
- Богданова-Бегларян, Н.В.; Блинова, О.В.; Мартыненко, Г.Я.; Шерстинова, Т.Ю. (2017) Корпус "Один речевой день" в исследованиях социолингвистической вариативности русской разговорной речи. *Сборник трудов семинара «Анализ разговорной русской речи (арЗ - 2017)»*. Санкт-Петербург: СПбГУ.
- Волкова, Т.Г.; Корчагина, Е.Л.; Кузнецов, А.Л.; Орлова, Е.М.; Самуйлова, Н.И.; Степанова, Е.М.; Трушина, Л.Б.; Чеботарев, П. (1996) *Пороговый уровень. Русский язык. Том I. Повседневное общение*. Страсбург: Совет Европы Пресс Г.
- Добрушина, Н.Р.; Левинзон, А.И. (2006) Информационные технологии в гуманитарном образовании: Национальный корпус русского языка. *Вопросы образования*, 4, 163–180.
- Добрушина, Н.Р. (ред.) (2007) *Национальный корпус русского языка и проблемы гуманитарного образования*. Москва: Издательский дом ГУ-ВШЭ.
- Культепина, О.А. (2018) Возможности корпусного инструмента RuSkELL в преподавании РКИ. В: Никунласси, А., Протасова, Е. (ред.) *Многоязычие и семья*. Берлин: Retorika, 170–183
- Морковкин, В.В. (ред.) (2003) *Система лексических минимумов современного русского языка: 10 лексических списков: от 500 до 5000 самых важных русских слов*. Москва: АСТ.
- Морковкин, В.В. (ред.) (1985) *Лексические минимумы современного русского языка*. Москва: Русский язык.

- Aston, G. (2000). Corpora and language teaching. In: Burnard, L.; McEnery, T. (eds.) *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang, 7–17
- Johansson Kokkinakis, S.; Volodina, E. (2011) Corpus-based approaches for the creation of a frequency-based vocabulary list in the EU project KELLY – issues on reliability, validity and coverage. In: *Proceedings of eLex 2011*, Slovenia: Trojina, Institute for Applied Slovene Studies, 129–139.
- Kilgarriff, A. (2010). Comparable corpora within and across languages, word frequency lists and the Kelly project. *Proceedings of the 3<sup>rd</sup> Workshop on Building and Using Comparable Corpora at LREC 2010*, Malta.
- Leech, G. (1997). Teaching and language corpora: A convergence. In: Wichmann, A.; Fligelstone, S.; McEnery, A.M.; Knowles, G. (eds.) *Teaching and Language Corpora*, pp. 1–23. London: Longman.
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4), 563–574.
- Römer, U. (2008). Corpora and language teaching. In: Lüdeling, A.; Kytö, M. (eds.) *Corpus Linguistics. An International Handbook*. Berlin: De Gruyter, 1, 112–131.
- Sharoff, S. (2006). Open-source corpora: Using the net to fish for linguistic data. *International Journal of Corpus Linguistics*, 11(4), 435–462.
- Sharoff, S.; Umanskaya E.; Wilson J. (2013) *A Frequency Dictionary of Russian: Core vocabulary for learners*. New York: Routledge.