

ЧТО ЗНАЧИТ «НЕ ВХОДИТ В ЛЕКСИЧЕСКИЙ МИНИМУМ»? ПОДСЧЕТ ПРОЦЕНТА НЕЗНАКОМОЙ ЛЕКСИКИ В ТЕКСТЕ ПО РКИ С УЧЕТОМ ДОСТУПНЫХ СЛОВООБРАЗОВАТЕЛЬНЫХ МОДЕЛЕЙ

А.Н. Лапошина

Аннотация. Статья посвящена поиску оптимальной методики подсчета процента незнакомой лексики при автоматизированном анализе текста для изучающих русский язык как иностранный (РКИ). В частности, обсуждаются вопросы неоднозначности методики учета дериватов от слов, присутствующих в лексическом минимуме, а также учета географических названий и имен собственных. Прямым следствием подобной ситуации является несоответствие значений процента незнакомой лексики текста, являющегося важнейшим параметром отбора учебных и контрольных материалов, и представлений практикующих преподавателей о знакомости слов студентам. В качестве решения данной проблемы предлагается создание расширенных лексических списков на базе лексического минимума теста по русскому языку как иностранному (ТРКИ) с помощью словообразовательных моделей, доступных на данном уровне владения русским языком. Данная задача в работе решается в рамках уровня А2, однако в дальнейшем она может быть применена и к другим уровням. Объем списка при этом был увеличен на 243 лексические единицы и составил суммарно 1762 слова. В статье также приведены примеры анализа текста на материале расширенных списков и проиллюстрирована разница в проценте незнакомой лексики в зависимости от учета географических названий и имен собственных. Проблемы, описанные в данной статье, иллюстрируют необходимость пересмотра и максимальной формализации накопленных знаний о принципах обучения лексике РКИ для их корректного использования в автоматизированных системах.

Ключевые слова: лексический минимум, уровневая система РКИ, процент незнакомой лексики, анализ текста по РКИ, сложность текста, методика преподавания РКИ.

Для цитирования: Лапошина А.Н. Что значит «не входит в лексический минимум»? Подсчет процента незнакомой лексики в тексте по РКИ с учетом доступных словообразовательных моделей // Преподаватель XXI век. 2021. № 4. Часть 2. С. 473–483. DOI: 10.31862/2073-9613-2021-4-473-483

473

© Лапошина А.Н., 2021



Контент доступен по лицензии Creative Commons Attribution 4.0 International License
The content is licensed under a Creative Commons Attribution 4.0 International License

“IS OUT OF THE VOCABULARY LIST”, WHAT DOES IT MEAN?
CALCULATING THE PERCENTAGE OF UNKNOWN WORDS IN A TEXT
FOR FOREIGN STUDENTS CONSIDERING THEIR DERIVATIVES

A.N. Laposhina

Abstract. *The article focuses on the issues of finding the optimal methodology for calculating the percentage of unfamiliar vocabulary in the automated text analysis for learners of Russian as a foreign language. The questions of the ambiguity of the methodology of accounting derivatives of words present in the lexical minimum, as well as of geographical names and proper names are considered in particular. A direct consequence of this situation is the discrepancy between the results of automated estimation of the percentage of unfamiliar vocabulary in the text, which is the most important parameter in selecting teaching and testing materials, and practicing teachers' ideas about the students' familiarity with the words. As a solution to this problem the authors of the article propose to create extended vocabulary lists based on the vocabulary minimum of Russian as a foreign language using the word-formation models available at this level of proficiency in Russian. This problem is solved within the framework of level A2, but in the future it can be applied to other levels as well. In this case, the list was increased by 243 lexical units to a total of 1762 words. The article also presents examples of text analysis on the basis of extended lists and illustrates the difference in the percentage of unfamiliar vocabulary depending on the consideration of place names and proper names. The problems described in this article demonstrate the need to revise and maximally formalize the accumulated knowledge of the principles of vocabulary training in the Russian as a foreign language for their correct implementation in automated systems.*

Keywords: *lexical minima, CEFR levels, percentage of unknown words, text analysis according to Russian as a foreign language method, text difficulty, methods of teaching Russian as a foreign language.*

Cite as: Laposhina A.N. “Is Out of the Vocabulary List”, What Does It Mean? Calculating the Percentage of Unknown Words in a Text for Foreign Students Considering Their Derivatives. *Prepodavatel XXI vek. Russian Journal of Education*, 2021, No. 4, part 2, pp. 473–483. DOI: 10.31862/2073-9613-2021-4-473-483

474

Одним из важнейших требований к тексту для работы с ним в иностранной аудитории является доступность лексики, составляющей этот текст. При этом основным параметром для оценки соответствия лексики текста уровню сложности является процент лексики, которая не входит в лексический минимум для данного уровня, т. е. предположительно незнакомой студентам. Так, система требований к уровню владения русским языком как иностранным (далее Требования) допускает на уровне А1 1–2% незнакомой лексики, А2 — 3–4%, В1 — 5–7% и на

всех последующих уровнях допускается до 10% незнакомой лексики.

Материалом для подобных расчетов могут быть разные варианты лексических списков, ранжированных по шкале уровней Общевропейских компетенций (далее — CEFR) [2–4], однако самой распространенной и влиятельной в методическом плане является линейка лексических минимумов ТРКИ (далее ЛМ ТРКИ), поскольку она является звеном официального комплекса материалов Российской государственной системы тестирования граждан зарубежных стран по русскому языку. Следовательно,

на эти словарные списки чаще всего ориентируются авторы и редакторы учебных программ [5], новых пособий, практикующие преподаватели РКИ.

С развитием цифровой лингводидактики признаки текста, основанного на проценте слов из лексических списков, стали также широко использоваться в задачах автоматического определения уровня сложности текста [6; 7], изучения возможностей автоматической адаптации текста [8] и др. В частности, разрабатываемый нами сервис Текстометр¹, предлагающий для любого введенного текста информацию о его ориентировочном уровне сложности по шкале уровней CEFR, также использует в качестве одной из мер процент лексики, отсутствующей в ЛМ [9].

Однако при автоматизации расчета этой меры было отмечено несколько неясных моментов в методике ее подсчета, которым и будет посвящена настоящая статья. В частности, самым неоднозначным оказывается вопрос, как следует считать дериваты от слов, присутствующих в ЛМ и образованных по доступным на данном уровне словообразовательным моделям: знакомыми студенту или нет? Например, слова *небольшой, фотограф, узнать, прилететь, красиво* отсутствуют в ЛМ уровня А2, однако почти все из них имеют однокоренные слова, включенные в ЛМ (*большой, фотографировать, знать, лететь, красивый*). Рассмотрим подробнее случай лексики *красивый*. Она появляется в ЛМ на элементарном уровне (А1), а ее родственное наречие *красиво* появляется в списках только на уровне В1. Это расходится с интуитивными

представлениями преподавателей и авторов пособий о знакомости этой лексики: в корпусе учебных пособий по РКИ RuFoLa [10] наречие *красиво* встречается 11 раз уже на уровне А1 в 6 различных пособиях, включая 2 вхождения текстов из раздела «Чтение теста» ТРКИ (см. примеры 1 и 2), и 6 раз в 4 различных пособиях на уровне А2. Образование наречий на *-о* доступно студентам, начиная с уровня А2 по данным Требований [1].

(1) *Я думала тогда, что это было красиво*².

(2) *Осенью листья в парке красные, желтые, зеленые. Это очень красиво*³.

О нерегулярной представленности, в частности, прилагательных и образованных от них наречий предупреждают и сами авторы линейки ЛМ: «<...> в силу предельной минимизации объема лексики на элементарном и базовом уровнях не прослеживалась представленность в них всех словообразовательных моделей, перечисленных в соответствующих Требованиях. Тем не менее в определенной мере эти модели присутствуют, обеспечивая понимание и продукцию потенциальной лексики» [11, с. 649].

С одной стороны, действительно, отбор посильного объема лексики представляется крайне актуальной проблемой, особенно на начальных этапах обучения. Однако, с другой стороны, складывается ситуация, при которой заявленный объем лексики ЛМ А2 (1300 единиц) рискует не соответствовать фактически ожидаемому от студента объему знаний.

Данная проблема тесно связана с выбранным авторами способом учета лексики

¹ Свидетельство о государственной регистрации программы для ЭВМ № 2021661785 от 15.07.2021. URL: <https://textometr.ru> (дата обращения: 07.03.2021).

² Дубских, А.В. Тексты для чтения по русскому языку как иностранному: элементарный уровень: методические указания. Казанский национальный исследовательский технологический университет, 2018. 24 с.

³ Антонова, В.Е., Нахабина, М.М., Сафронова, М.В., Толстых, А.А. Дорога в Россию: учебник русского языка (элементарный уровень). СПб.: Златоуст, 2013. 344 с.

в списке. Так, часть авторов [2; 12] оперирует понятием лексического гнезда или семьи, указывая при этом доступные студенту дериваты и видовые пары в описании к основной лексеме (например, завод [прилаг. заводской]). Однако при создании системы ЛМ ТРКИ, как и большинства других списков для русского языка [13–15], каждая лексема в списке представлена отдельно, что может приводить к случаям нерегулярной представленности дериватов.

Кроме того, в пользу учета всех доступных на данном уровне дериватов говорит и такой критерий отбора лексики ЛМ, как словообразовательная ценность/потенциал/способность слова [16–18]. Он подразумевает, что освоение той или иной лексемы дает студенту возможность понимать смысл ее производных.

Отсутствие общепринятой методики учета такой лексики порождает неоднозначность в подсчете незнакомой лексики текста: с одной стороны, слова, отсутствующие в ЛМ, формально считаются незнакомыми студенту; с другой же стороны, исследователи и практикующие преподаватели зачастую рассматривают дериваты, образованные по словообразовательным моделям, доступным на данном уровне согласно Требованиям, как знакомые студентам (например, в [19]).

Вторым неясным моментом методики подсчета незнакомой лексики текста является учет имен собственных и географических названий. Так, например, в списках проекта KELLY содержатся некоторые частотные географические названия (на уровне А2 есть *Америка* и *Россия*, но нет *Франции* и *Германии*) [4]; в систему лексических минимумов под ред. В.В. Морковкина имена собственные и географические названия не включены

[2]. В линейке ЛМ ТРКИ данные группы слов не встречаются в основном алфавитном списке, но наиболее частотные страны и имена приведены в виде отдельных списков, однако, например, российские города, включая *Москву*, *Санкт-Петербург* и др., отсутствуют. Остается неясным, стоит ли считать эти лексемы незнакомыми студенту или не учитывать данные группы лексики при подсчетах.

Описанные проблемы ярко иллюстрируют необходимость пересмотра и максимальной формализации накопленных знаний о принципах обучения лексике РКИ для их использования в автоматизированных системах, поскольку то, что может быть очевидно для практикующего преподавателя при ручном подсчете (если студент знает лексему *большой*, то на уровне А2 поймет и слово *небольшой*; подавляющему большинству студентов знакомо слово *Москва*), совсем не является таковым для алгоритма.

Таким образом, целью работы мы ставим поиск оптимальной методики подсчета незнакомой лексики при автоматизированном анализе текста для изучающих РКИ. Для достижения этой цели мы создадим расширенные лексические списки на базе ЛМ ТРКИ, используя словообразовательные модели, доступные на данном уровне владения русским языком, а также оценим влияние учета имен собственных и географических названий на итоговый процент незнакомой лексики текста. Данные задачи в работе решаются в рамках лексических списков уровня А2, однако в дальнейшем она может быть применена и к другим уровням.

В качестве материалов для работы была использована линейка лексических минимумов ТРКИ от издательства «Златоуст» [16; 20]⁴. Отправной точкой для

⁴ Автор сердечно благодарит издательство «Златоуст» и лично А.В. Голубеву за помощь и поддержку проекта.

создания списка возможных словообразовательных моделей стал раздел «Словообразование» Требований соответствующего уровня, содержащий перечень аффиксов и описание моделей образования слов, доступных на данном уровне владения русским языком [1]. Так, например, на уровне А2 речь идет о распознавании ограниченного числа словообразовательных моделей:

- существительных, обозначающих лиц по национальности с суффиксами *-ец, -анин*; лиц женского пола с суффиксами *-к(а), -иц(а), -ниц(а)*; лиц по профессии с суффиксами *-ист, -тель*; действий с нулевым суффиксом; отглагольных существительных с суффиксами *-ение*;

- прилагательных с суффиксами *-н, -ск, -ичек*;

- наречий с суффиксом *-о*, мотивированных качественных прилагательных с префиксом *по-* и суффиксом *-ски*;

- глаголов с суффиксами *-ыва/ива, -ва, -а, -и, -ова/ева, -ну*; глаголов движения с префиксами *в-, вы-, у-, по-, при-*.

Поскольку создаваемый список направлен на подготовку материалов по чтению, т. е. исключительно перцептивный тип речевой деятельности, мы позволили себе расширить данный список моделями, примеры которых широко представлены в пособиях по РКИ уровней А1 и А2 по данным корпуса RuFoLa:

- прилагательными и наречиями с отрицательным префиксом *не-*. Данный способ появляется в учебниках, начиная с самых ранних уровней, однако нам не удалось найти упоминание этого префикса в возможных словообразовательных моделях Требований для уровней А1 и А2;

- прилагательными с суффиксом *-ов* (*апельсиновый*);

- существительными с суффиксом *-ость* (*молодость, специальность*).

На следующем этапе работы списки слов из лексических минимумов были переведены в цифровой формат. При этом были объединены слова из алфавитного списка и из тематических групп лексики. Далее списки вручную просматривались аннотаторами⁵ для проверки на возможные ошибки при оцифровке и расширения списка по заданным словообразовательным моделям. Предложенные аннотаторами новые лексические единицы были проверены вручную по следующим критериям:

Схожесть лексического значения однокоренного слова с исходным. Так, например, цепочки *пример — примерный, звонить — звонкий* не прошли этот этап отбора.

Соблюдение принципа системности. Например, присутствие в списке лексем *овоц* и *овощной* являлось аргументом в пользу добавления прилагательного **фруктовый** к присутствующей в списке лексеме *фрукт*. Другим примером может стать восстановление логических цепочек лексем, связанных с определенной областью науки: *философ — философия — философский* присутствуют в минимуме, это обосновывает добавление прилагательного **архитектурный** в имеющуюся цепочку *архитектура — архитектор*.

Результаты проделанной работы по расширению списков для уровня А2 представлены в таблице 1, отсортированной по количеству полученных примеров на каждый тип действий по расширению списка.

Из таблицы 1 видно, что самым частым объектом добавления в список стали отыменные прилагательные. Среди добавленных лексических единиц, оказавшихся

⁵ Автор выражает благодарность магистрантам ГИРЯ им. А.С. Пушкина Елизавете Бельковой, Ангелине Кучеровой, Валерии Таброско и Ле Ван Ань за их чрезвычайно важный вклад в проект.

Типы и количество лексем, добавленных в список на основании словообразовательных моделей

Часть речи исходного слова из ЛМ	Часть речи добавляемого слова	Количество примеров	Пример
существительное	прилагательное	144	архитектура → архитектурный, апельсин → апельсиновый
прилагательное	прилагательное + не-	38	красивый → некрасивый, трудный → нетрудный
прилагательное	наречие	23	бесплатный → бесплатно, грязный → грязно
наречие	прилагательное	16	опасно → опасный, скучно → скучный
наречие	наречие + не-	9	красиво → некрасиво, дорого → недорого
существительное	существительное	7	турист → туризм, политика → политик
прилагательное	существительное	4	модный → мода, национальный → национальность
глагол	видовая пара к глаголу	2	знать → узнать, готовить → приготовить
Итого		243	

наиболее востребованными в коллекции текстов пособий РКИ RuFoLa, отметим прилагательные *открытый* (13 вхождений), *нормальный* (12), *литературный* (12), *опасный* (11), *мировой*⁶ (9), *скучный* (8); среди прилагательных, образованных от названий стран и континентов, лидируют *итальянский* (23), *японский* (10) и *американский* (9); среди наречий отметим *красиво* (10), *честно* (7) и *удобно* (7), существительных — *путешествие* (11), *фотограф* (10), *мода* (4).

Самое большое количество упоминаний в текстах пособий по РКИ среди добавленной в список лексики демонстрируют отрицательные прилагательные и наречия: *немного* (39), *небольшой* (20), *необычный* (14), *недорогой* (8). Дальнейший анализ таких слов наталкивает на интересные методические и лингвокультурные наблюдения: поскольку 80% качественных прилагательных из списка ЛМ,

выражающих какую-либо оценку, оказались положительно окрашенными, кандидатами на добавление в список стали, соответственно, разнообразные отрицательные свойства предметов или явлений, часть из которых оказалась востребована в текстах пособий: *неудобно* (5), *непонятно* (4), *неинтересный* (4), *некрасивый* (4). Внимание к этим словарным единицам для предоставления студенту языкового инструментария для выражения эмоций и оценок, в том числе отрицательных (недовольства и фрустрации), представляется важным пунктом в русле современной лингводидактики и отвечает принципу учета коммуникативных потребностей студента [21].

Алфавитный список ЛМ ТРКИ для уровня составляет 1 300 слов. Общее количество добавленных с помощью словообразовательных моделей слов составило 243 единицы. Так получился

⁶ Интересно, что в оригинальной версии ЛМ А2 предлагается оба значения существительного *мир* и одно прилагательное — *мирный*, однако оно появляется в коллекции значительно реже, 1 раз.

расширенный список 1 объемом 1 543 слова. Однако не все лексемы из тематических списков и списка синонимов и антонимов ЛМ присутствуют в алфавитной версии списка, (например, *биологический, парламент, власть, партия, батон, булочка* и др.). Добавление подобных слов из всех отдельных списков ЛМ, включая также списки географических названий и наиболее употребительных русских имен, дает нам расширенный список 2 объемом 1 762 слова.

Рассмотрим результаты работы оригинальной и расширенных версий списка на материале текста из пособия «Дорога в Россию» базового уровня (А2). Его начальный фрагмент приведен в примере:

(3) *Маша Новикова — студентка исторического факультета МГУ — прекрасно сдала летнюю сессию и стала думать, куда можно поехать отдохнуть. Девушке очень нравилось море. Бабушка Маши — Ольга Петровна — посоветовала своей внучке поехать в Крым, в Ялту, на Черное море. Маша не хотела ехать одна, поэтому она позвонила своей подруге Лене и предложила ей отдохнуть на юге вместе.* <...>⁷

Текст имеет длину 618 слов, в нем присутствует 6 имен собственных, упоминающихся в тексте всего 31 раз и 3 географических названия, упоминающихся 12 раз. Для демонстрации разницы результатов в зависимости от методики учета в расчетах имен собственных и

географических названий были произведены две серии подсчетов: с учетом всех слов текста (т. е. общее количество слов = 618) и без учета имен и географических названий (т. е. общее количество слов = 575). Для расчета незнакомой лексики текста все слова были приведены к их начальной словарной форме. Процент незнакомой лексики текста рассчитывался как отношение количества слов, отсутствующих в лексическом списке, к общему количеству слов текста. При этом учитывалось каждое употребление слова в тексте (например, лексема *загорать* встретилась в тексте 2 раза в различных формах и отсутствует в лексическом списке; значит, она добавляет $2/618 = 0,00324\%$ к общему проценту незнакомой лексики).

В таблице 2 приведено сравнение результатов подсчетов незнакомой лексики текста по оригинальному алфавитному списку ЛМ ТРКИ, расширенному с помощью словообразования (расширенный список 1) и общему расширенному списку (расширенный список 2).

Из таблицы 2 видно, что расширение списка с учетом словообразования закономерно уменьшает процент незнакомой лексики. В данном тексте словами, перешедшими в разряд знакомых с помощью доступных словообразовательных моделей, стали прилагательные *небольшой, недорогой, утренний, живописный, тихий, жаркий, немолодой*, а также наречие *непонятно* и глагол *узнать*.

Таблица 2
Различные варианты подсчета процента незнакомой лексики текста с помощью оригинальной и расширенной версий списка

	Оригинальный список	Расширенный список 1	Расширенный список 2
Учитываются все слова текста	16%	13%	11%
Имена собственные и географические названия не участвуют в расчете	11%	10%	9%

⁷ Антонова, В.Е., Нахабина, М.М., Толстых, А.А. Дорога в Россию: учебник русского языка (базовый уровень). 4-е изд. М.: ЦМО МГУ им. М.В. Ломоносова; СПб.: Златоуст, 2009. 256 с.

Становится очевидна также разница в проценте незнакомой лексики текста в зависимости от методики подсчета: учет/неучет имен собственных, географических названий и словообразования в данном тексте является причиной разброса меры в 7% (от 16 до 9), что является существенной цифрой, особенно для текстов начальных уровней. Эта разница рискует быть еще больше при оценке информационных текстов, избилующих упоминанием персон и локаций или текстов с иностранными именами героев. Поскольку ресурс Текстометр, для нужд которого инициирована данная работа, связан с оценкой текстов для чтения, т. е. относительно простым видом речевой деятельности, для дальнейшей работы сервиса нами выбрана самая «нестрогая» методика подсчета (географические названия и имена не участвуют в расчете процента незнакомой лексики).

Отметим, что даже эта нестрогая методика и новые расширенные списки дают

минимальный процент незнакомой лексики (9%), тогда как требования для уровня А2 содержат рекомендации о 3–4%. Это может говорить о необходимости дальнейшего изучения способностей современных студентов и возможного пересмотра норм в Требованиях.

В заключение повторим, что предложенное в статье решение все же является локальным и ориентированным, прежде всего, на пользователей сервиса Текстометр, которым важен расчет процента незнакомой лексики по классической линейке лексических минимумов ТРКИ. Основным же вектором развития данного исследования является создание собственной системы лексических списков на основе корпусного подхода, когда принятие решения о включении слова в список базируется на его частотности в релевантных коллекциях текстов, а также его методической востребованности, выраженной во встречаемости слова в корпусе учебников РКИ и корпусах речи студентов.

СПИСОК ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Государственный стандарт по русскому языку как иностранному. Базовый уровень / М.М. Нахабина и др. СПб.: Златоуст, 2001. 23 с.
2. Система лексических минимумов современного русского языка: 10 лексических списков от 500 до 5000 самых важных русских слов / Г.Ф. Богачева, Н.М. Луцкая, В.В. Морковкин, З.П. Попова; под ред. проф. В.В. Морковкина. М.: АСТ: Астрель, 2003. 768 с.
3. Волкова, Т.Г. Пороговый уровень. Русский язык. Т. I. Повседневное общение / Т.Г. Волкова, Е.Л. Корчагина, А.Л. Кузнецов, Е.М. Орлова, Н.И. Самуйлова, Е.М. Степанова, Л.Б. Трушина, П. Чеботарев. Страсбург, 1996. 285 с.
4. Kilgarriff, A. Comparable Corpora Within and Across Languages, Word Frequency Lists and the Kelly Project // Proceedings of the 3rd Workshop on Building and Using Comparable Corpora at LREC. Malta, 2010. P. 1–5.
5. Глазунова, О.И., Колесова, Д.В., Попова, Т.И. Программа по русскому языку как иностранному. Уровни А1–С2. Основной курс. М.: Русский язык. Курсы, 2017. 216 с.
6. Karpov, N., Baranova, J., Vitugin, F. Single-Sentence Readability Prediction in Russian // Proceedings of Analysis of Images, Social Networks, and Texts Conference. AIST, 2014. P. 91–100.
7. Reynolds, R. Insights from Russian Second Language Readability Classification: Complexity-Dependent Training Requirements, and Feature Evaluation of Multiple Categories // Proceedings

- of the 11th Workshop on the Innovative Use of NLP for Building Educational Applications. San Diego, 2016. P. 289–300.
8. Дмитриева, А.Д., Лапошина, А.Н., Лебедева, М.Ю. Квантитативное исследование стратегий упрощения на материале адаптированных текстов для изучающих РКИ // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог». М.: РГГУ, 2021. С. 191–204.
 9. Laposhina, A.N. Automated Text Readability Assessment for Russian Second Language Learners / A.N. Laposhina, T.S. Veselovskaya, M.Yu. Lebedeva, and O.F. Kupreshchenko // Dialogue 2018: Proceedings of the International Conference. Vol. 17. Issue 24. Moscow, 2018. P. 396–406.
 10. Лапошина, А.Н. Корпус текстов учебников РКИ как инструмент анализа учебных материалов // Русский язык за рубежом. 2020. № 6 (283). С. 22–28.
 11. Андриюшина, Н.П. Лексические минимумы по русскому языку как иностранному: проблема отбора лексических и фразеологических единиц // Проблемы истории, филологии, культуры. 2011. № 3 (33). С. 648–652.
 12. Сидорова, М.Ю., Шматко, А.С. От «Лексического минимума» к «Лексико-грамматической основе»: новый подход к представлению языка предметной области // Мир русского слова. 2019. № 3. С. 83–91.
 13. Власова, Е.А., Карпова, Е.Л., Ольшевская, М.Ю. Лексический минимум по языку специальности: сколько слов достаточно? Разработка принципов минимизации // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. № 4. С. 63–77.
 14. Ильина, О.А. Лексический минимум по языку специальности «Робототехника» как основа формирования лингвокоммуникативной компетенции иностранных магистрантов // Гуманитарный вестник. 2013. № 2 (4). С. 1–16.
 15. Маркина, Е.И. Лингводидактические основы разработки лексических минимумов по русскому языку как иностранному (для разных уровней и профилей обучения): дис. ... канд. пед. наук. М., 2011. 235 с.
 16. Андриюшина, Н.П., Козлова, Т.В. Лексический минимум по русскому языку как иностранному. Базовый уровень. Общее владение. СПб.: Златоуст, 2013. 116 с.
 17. Костомаров, В.Г. Принципы отбора лексического минимума // Русский язык в национальной школе. 1963. № 1. С. 29–35.
 18. Штейнфельдт, Э.А. Частотный словарь современного русского литературного языка: 2500 наиболее употребительных слов: Пособие для преподавателей рус. яз. Таллин, 1963. 316 с.
 19. Голубева, А.В. Лингвокультурологический комментарий к текстам: «Библиотека Златоуста» как пример текстов для градуированного чтения. Устный доклад на Неделе русского языка и культуры в Сербии. Белград, 2014.
 20. Андриюшина, Н.П., Козлова, Т.В. Лексический минимум по русскому языку как иностранному. Элементарный уровень. Общее владение. СПб.: Златоуст, 2012. 80 с.
 21. Lebedeva M. Instructional Design of Skill-Balanced LMOOC: a Case of the Russian Language MOOC for Beginners // Journal of Universal Computer Science. Vol. 27 (5). 2021. P. 485–497.

REFERENCES

1. Gosudarstvennyj standart po russkomu jazyku kak inostrannomu. Bazovyj uroven [State Standard for Russian as a Foreign Language. Basic level], M.M. Nahabina et al. St. Petersburg, 2001, 23 p. (in Russ.)

2. Bogacheva, G.F., Luckaya, N.M., Morkovkin, V.V., Popova, Z.P. *Sistema leksicheskikh minimumov sovremennoogo russkogo yazyka: 10 leksicheskikh spiskov ot 500 do 5000 samykh vazhnykh russkikh slov* [The System of Lexical Minima of the Modern Russian Language: 10 Lexical Lists from 500 to 5000 of the Most Important Russian Words]. Moscow: AST: Astrel, 2003, 768 p. (in Russ.)
3. Volkova, T.G., Korchagina, E.L., Kuznecov, A.L., Orlova, E.M., Samujlova, N.I., Stepanova, E.M., Trushina, L.B., Chebotarev, P. *Porogovyj uroven. Russkij jazyk. T. I. Povsednevnoe obshhenie* [Threshold Level. Russian Language. Vol. I. Common Language]. Strasburg, 1996. 285 p. (in Russ.)
4. Kilgarrieff, A. Comparable Corpora Within and Across Languages, Word Frequency Lists and the Kelly Project. In: *Proceedings of the 3rd Workshop on Building and Using Comparable Corpora at LREC*. Malta, 2010, pp. 1–5.
5. Glazunova, O.I., Kolesova, D.V., Popova, T.I. *Programma po russkomu yazyku kak inostrannomu. Urovni A1–C2. Osnovnoj kurs* [Program of Russian as a Foreign Language. Levels A1–C2. The Main Course]. Moscow: Russkij yazyk. Kursy, 2017, 216 p. (in Russ.)
6. Karpov, N., Baranova, J., Vitugin, F. Single-Sentence Readability Prediction in Russian. In: *Proceedings of Analysis of Images, Social Networks, and Texts conference*. AIST, 2014, pp. 91–100.
7. Reynolds, R. Insights from Russian Second Language Readability Classification: Complexity-Dependent Training Requirements, and Feature Evaluation of Multiple Categories. In: *Proceedings of the 11th Workshop on the Innovative Use of NLP for Building Educational Applications*. San Diego, 2016, pp. 289–300.
8. Dmitrieva, A.D., Laposhina, A.N., Lebedeva, M.Yu. Kvantitativnoe issledovanie strategij uproshcheniya na materiale adaptirovannykh tekstov dlya izuchayushchih RKI [Quantum research strategy of improvement based on the material of adapted texts for students of RCT]. In: *Kompyuternaya lingvistika i intellektualnye tekhnologii: po materialam mezhdunarodnoj konferencii “Dialog”* [Computational Linguistics and Intelligent Technologies: Based on the Materials of the International Conference “Dialogue”]. Moscow, Rossijskij gosudarstvennyj gumanitarnyj universitet, 2021, pp. 191–204. (in Russ.)
9. Laposhina, A.N., Veselovskaya, T.S., Lebedeva, M.Yu. and Kupreshchenko, O.F. Automated Text Readability Assessment for Russian Second Language Learners. In: *Proceedings of the International Conference “Dialogue 2018”*, vol. 17, iss. 24. Moscow, 2018, pp. 396–406.
10. Laposhina A.N. Korpus tekstov uchebnikov RKI kak instrument analiza uchebnykh materialov [Corpus of Texts of Russian Language Textbooks as a Tool for the Analysis of Educational Materials], *Russkij yazyk za rubezom* = Russian Language Abroad, 2020, No. 283 (6), pp. 22–28. (in Russ.)
11. Andrjushina, N.P. Leksicheskie minimumy po russkomu jazyku kak inostrannomu: problema otbora leksicheskikh i frazeologicheskikh edinic [Lexical Minima in Russian as a Foreign Language: The Problem of Selecting Lexical and Phraseological Units], *Problemy istorii, filologii, kultury* = Problems of History, Philology, Culture, 2011, No. 3 (33), pp. 648–652. (in Russ.)
12. Sidorova, M.Yu., Shmatko, A.S. Ot “Leksicheskogo minimum” k “Leksiko-grammaticheskoy osnove”: novyj podhod k predstavleniyu yazyka predmetnoj oblasti [From “Minimized Word List” to “Lexico-Grammatical Base”: A New Approach Towards Representation of the Language of a Scientific Discipline], *Mir russkogo slova* = The World of the Russian Word, 2019, No. 3, pp. 83–91. (in Russ.)
13. Vlasova, E.A., Karpova, E.L., Olshevskaya, M.Yu. Leksicheskij minimum po yazyku specialnosti: skolko slov dostatochno? Razrabotka principov minimizacii [Vocabulary: How Many Words are

- Enough? Principles of Minimizing Learners' Vocabulary], *Vestnik NGU. Seriya: Lingvistika i mezhkulturnaya kommunikaciya* = Bulletin of Novosibirsk State University. Series: Linguistics and Intercultural Communication, 2019, vol. 17, No. 4, pp. 63–77. (in Russ.)
14. Ilina, O.A. Leksicheskiy minimum po yazyku specialnosti "Robototekhnika" kak osnova formirovaniya lingvokommunikativnoj kompetencii inostrannykh magistrantov [Lexical Minimum in the Language of the Specialty "Robotics" as the Basis for the Formation of Linguocommunicative Competence of Foreignmaster Students], *Gumanitarnyj vestnik* = Humanitarian Bulletin, 2013, No. 2(4), pp. 1–16. (in Russ.)
 15. Markina, E.I. *Lingvodidakticheskie osnovy razrabotki leksicheskikh minimumov po russkomu yazyku kak inostrannomu (dlya raznykh urovnej i profilej obucheniya)* [Linguodidactic Basic for the Lexical Minima for Russian as a Foreign Language (For Different Levels and Profiles of Education)]: PhD Dissertation (Pedagogy). Moscow, 2011, 235 p. (in Russ.)
 16. Andryushina, N.P., Kozlova, T.V. *Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Bazovyy uroven. Obshchee vladenie* [Lexical Minimum of Russian as a Foreign Language. Level A2. Common Language]. St. Petersburg: Zlatoust, 2013, 116 p. (in Russ.)
 17. Kostomarov, V.G. Principy otbora leksicheskogo minimuma [The Principles of the Vocabulary Selection for the Lexical Minimum], *Russkiy yazyk v nacionalnoj shkole* = Russian Language in the National School, 1963, No. 1, pp. 29–35. (in Russ.)
 18. Shtejfeldt, E.A. *Chastotnyj slovar sovremennogo russkogo literaturnogo yazyka: 2500 naibolee upotrebitelnykh slov: Posobie dlya prepodavatelej russkogo yazyka* [Frequency Dictionary of Modern Russian Literary Language: 2500 Most Common Words]. Tallin, 1963, 316 p. (in Russ.)
 19. Golubeva, A.V. *Lingvokulturologicheskij kommentarij k tekstam: "Biblioteka Zlatousta" kak primer tekstov dlya graduirovannogo chteniya. Ustnyj doklad na Nedele russkogo yazyka yazyka i kultury v Serbii* [Linguistic and Cultural Commentary on the Texts: "Zlatoust's Library" as an Example of Texts for Graded Reading. Oral Presentation at the Week of the Russian Language and Culture in Serbia]. Belgrad, 2014. (in Russ.)
 20. Andryushina, N.P., Kozlova, T.V. *Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Elementarnyj uroven. Obshchee vladenie* [Lexical Minimum of Russian as a Foreign Language. Level A1. Common Language]. St. Petersburg: Zlatoust, 2012, 80 p. (in Russ.)
 21. Lebedeva, M. Instructional Design of Skill-Balanced LMOOC: a Case of the Russian Language MOOC for Beginners, *Journal of Universal Computer Science*, vol. 27 (5), 2021, pp. 485–497.

Лапошина Антонина Николаевна, ведущий эксперт, лаборатория когнитивных и лингвистических исследований, Государственный институт русского языка имени А.С. Пушкина, ANLaposhina@pushkin.institute

Antonina N. Laposhina, Leading Expert, Laboratory of Cognitive and Linguistic Research, Pushkin State Russian Language Institute, ANLaposhina@pushkin.institute

Статья поступила в редакцию 15.09.2021. Принята к публикации 27.09.2021

The paper was submitted 15.09.2021. Accepted for publication 27.09.2021