

КОРПУС ТЕКСТОВ УЧЕБНИКОВ РКИ КАК ИНСТРУМЕНТ АНАЛИЗА УЧЕБНЫХ МАТЕРИАЛОВ



А.Н. ЛАПОШИНА

ANLaposhina@pushkin.institute

специалист

по компьютерной лингвистике
лаборатории когнитивных
и лингвистических исследований
Государственного института
русского языка
им. А.С. Пушкина
Москва, Россия

Ключевые слова:

*учебник РКИ, корпус
учебников, педагогический
корпус, корпусная
лингводидактика,
учебниковедение*

DOI: 10.37632/PI.2020.283.6.003

В статье описан опыт сбора, разметки и использования корпуса текстов из учебников РКИ, RuFoLa. Такая коллекция текстов является удобным инструментом получения объективных данных о содержании современных учебников РКИ как в исследовательских, так и в практических целях. В качестве примеров задач, решаемых на материале корпуса RuFoLa, приведены два исследования. Во-первых, произведен расчет языковых признаков постепенного усложнения учебных материалов по РКИ по шкале уровней CEFR от A1 до C1. Во-вторых, приведены примеры частотного анализа лексики учебников и ее сравнение с содержанием лексических минимумов по РКИ – такие данные могут оказаться ценны при проверке соответствия учебника заявленному уровню или при обновлении минимума.

Центральным источником знаний студентов-иностранцев об изучаемом языке чаще всего является учебник, особенно на начальных этапах обучения. При этом учебник «отражает состояние теории и практики обучения в определенный исторический период» [9], формирует представление о культуре и традициях страны изучаемого языка, транслирует определенные социокультурные стереотипы. Обобщенные данные репрезентативной коллекции текстов учебников по РКИ могут служить своеобразным срезом современного состояния методики: если Государственный стандарт призван задокументировать «как должно быть», то подобная коллекция иллюстрирует состояние «как есть сейчас». Поэтому развитие методики и инструментов для объективного анализа учебников представляется важной задачей и имеет два направления применения: как оценка конкретного пособия (соответствие заявленному уровню, адекватность представления грамматического и лексического материала), так и составление общей картины состояния современной методики исходя из обобщенного анализа содержания пособий.

Анализ литературы

Анализ учебников для обучения языку как иностранному имеет богатую научную базу как в зарубежной [22, 27], так и в отечественной методике [7, 9, 13, 18]. Среди всего многообразия работ по анализу содержимого учебника иностранного языка стоит сфокусироваться на исследованиях, посвященных количественному анализу текста учебника, т.е. тех, в которых тот или иной методический вывод (представленность грамматической конструкции, лексический состав, наличие лингвокультурологической информации) делается на основании анализа репрезентативной выборки из нескольких пособий. Венингер [26] отмечает условное деление работ по количественному анализу учебника на две линии: первая представляет собой методический анализ – насколько материалы подходят по уровню сложности, соответствуют реальному языку; вторая линия – анализ социокультурной информации, транслируемой посредством учебника.

Примером исследования первой, методической, линии может служить работа Габриэлятос [19], посвященная количественному анализу

грамматических конструкций, предложенных в учебниках английского как иностранного. Автор приходит к выводу о том, что примеры традиционных типов условных конструкций, представленные в учебниках (нулевая условная конструкция, первый, второй и третий типы условной конструкции и смешанный тип), покрывают лишь 44% от 781 условной конструкции, случайно выбранной из Британского национального корпуса. Исследование также показывает, что конструкции смешанного типа составляют около 1,5% от этих примеров, что ставит под вопрос оправданность такого большого внимания к ним в учебном процессе.

В качестве примера анализа коллекции пособий по РКИ в методическом ключе можно привести работы о порядке слов в учебниках РКИ для монгольских учащихся [8]; видах глагола в учебных пособиях [10]; представленности в пособиях темы глагольно-предложного управления [5].

Социокультурные исследования на базе анализа учебных материалов чаще всего затрагивают вопросы культурной информации, транслируемой посредством материалов учебника, а также гендерных и расовых стереотипов. Так, например, Ким демонстрирует разницу методических школ Южной и Северной Кореи на материале учебников английского для детей этих стран [20]. Мукундан и коллеги исследуют гендерные стереотипы, транслируемые в учебниках английского как иностранного для детей в Малайзии [23].

Примерами исследований социокультурной информации, отраженной в учебниках РКИ, могут служить монография С.К. Милославской «Русский язык в истории становления европейского образа России» [17]; многочисленные работы, посвященные образу России и русских в учебниках РКИ [14–16]; выражение гендерных отношений в учебных материалах по РКИ [12].

Корпус учебников: определение понятия

Под нужды исследований, указанных выше, часто собираются коллекции учебных материалов различного объема и содержания. Если данная коллекция отвечает требованиям репрезентативности и сбалансированности по выбранным критериям, а также снабжена разметкой, она может называться *корпусом*.

Отправной точкой в разработке терминологии и методики создания таких корпусов стоит считать статью «Новые типы корпуса для новых учебных задач: сбор, разметка и использование корпуса учебников» [22]. Авторы вводят понятие *педагогического корпуса* как репрезентативной выборки письменной и устной речи, которая предполагается быть предъявленной студенту, изучающему данный язык в классе

или самостоятельно. Чаще всего учебные материалы являются текстами учебников, аудиозаписями диалогов и упражнениями. *Корпус учебников* при этом является одним из видов педагогического корпуса.

Среди примеров создания таких специализированных корпусов можно назвать корпус учебников английского для немецкоязычной аудитории (GEFL TC) [24], корпус учебников английского языка как иностранного [22], корпус учебников шведского языка как иностранного [25]. В следующей секции будет описан наш собственный опыт создания корпуса текстов учебников по РКИ.

Корпус текстов учебников русского РКИ RuFoLa

Проект **RuFoLa** (Russain as a Foreign Language) направлен на сбор и разметку репрезентативной выборки текстов из учебников и учебных пособий по РКИ. Для первой версии корпуса было принято решение отобрать тексты учебных пособий по общему курсу русского языка, изданных после 1990 г. Среди них есть как учебные комплексы, уже ставшие классикой («Дорога в Россию», «Жили-были», «Поехали!»), так и сравнительно новые пособия («Русский сувенир», «Точка Ру»). Кроме того, добавлены тексты из учебной текстотеки ЦМО¹ МГУ и вариантов теста ТРКИ. Пособия, направленные на изучение языка специальности (русский для медиков, русский язык в сфере туризма и т.п.), не включались в первую редакцию корпуса, так как очевидно, что они содержат специфическую лексику и это повлияет на результаты подсчетов.

На данный момент корпус представляет собой более 800 текстов, содержащих дополнительную информацию о названии и авторах пособия, годе выпуска издания и уровне сложности по шкале CEFR (от A1 до C1). Корпус сбалансирован по числу текстов на каждый уровень (таблица 1).

Таблица 1

Объем корпуса RuFoLa

	A	A2	B1	B2	C1
Количество текстов	220	137	144	158	143
Количество слов	29 422	42 529	53 619	89 076	51 444
Объем словаря (количество уникальных слов)	1690	4352	6685	11381	8715

¹ texts.cie.ru

Разметка корпуса по уровням сложности по шкале CEFR, вначале представляющаяся нам простейшей задачей поиска необходимой информации в методической справке пособия, оказалась не столь тривиальной по нескольким причинам:

- необязательность и туманность сообщения об уровне: для некоторых пособий уровень не указан вообще или выражен нечетко («для продвинутых студентов», «для начинающих», «пособие рассчитано на второй семестр первого года обучения»);
- размытость границ для текстов высоких уровней: многие пособия ориентированы на B1–B2, B2–C1, что также затрудняет отнесение текста к одному определенному классу.

Ниже мы приведем два примера методических задач, решаемых с помощью материалов корпуса RuFoLa. Первая задача посвящена расчету языковых признаков постепенного усложнения учебных материалов по РКИ по шкале уровней CEFR от A1 до C1. Вторая задача представляет собой выборочное сравнение лексики учебников и лексических минимумов по РКИ – такие данные могут оказаться ценны при проверке соответствия учебника заявленному уровню или при обновлении минимума.

Поиск формальных лингвистических показателей уровня сложности текста на материале корпуса учебников RuFoLa

Располагая коллекцией текстов, размеченной по уровням сложности CEFR, мы можем проследить путь постепенного усложнения учебных материалов и осуществить поиск релевантных формальных признаков постепенного увеличения сложности текстов от уровня к уровню.

Таблица 2 содержит данные некоторых структурных и лексических показателей, способных влиять на сложность текста. Значения средней длины слова и предложения являются базовыми показателями сложности текста как в многочисленных формулах читабельности, так и в государственных стандартах по РКИ. Количество знаков препинания на предложение также косвенно указывает на синтаксическую сложность текста.

Методический интерес представляют и грамматические показатели усложнения текстов: так, например, процент глаголов в 1 л. ед.ч. снижается по мере увеличения уровня сложности текста. Вероятно, это связано с тем, что на начальных этапах овладения «языком выживания» важно отработать умения

удовлетворения первичных нужд (я хочу/люблю, я не понимаю, я болен и т.д.), а также научиться навыкам самопрезентации на русском языке (лексико-грамматические темы «О себе», «Мой день», «Моя семья» и т.д.). На более высоких уровнях эти базовые потребности удовлетворены, включаются новостные, культурологические тексты, и глаголов в 1 л. ед.ч. становится меньше. Еще одним грамматическим показателем увеличения сложности является процент слов в Р.п. на предложение: скорее всего, это связано с усложнением синтаксических конструкций и появлением целых цепочек из родительного падежа Р.п.:

1) История *раскрытия секрета изготовления кефира* похожа на детектив, но каждое слово здесь – правда. [1]

Последние три ряда таблицы демонстрируют лексические показатели повышения сложности текста. Значение TTR (type/token ratio) иллюстрирует лексическое разнообразие текста и рассчитывается как отношение количества уникальных слов текста к общему количеству слов в тексте. Иными словами, чем значение TTR меньше, тем чаще в тексте повторяются одни и те же слова. И наоборот, увеличение этих значений говорит о большем лексическом разнообразии и, следовательно, усложнении текста.

Степень соответствия текстов лексическим минимумам также является важнейшим и очевидным сигналом повышения сложности текста. Из данных таблицы 2 видно, что процент незнакомых слов в целом выше, чем это предполагается Государственным стандартом (например, для чтения на уровне A1 допустимо до 1–2% незнакомой лексики [11], а мы наблюдаем среднее значение 16%). Это вызвано как ограничениями в подсчетах (система пока не может учитывать словообразование, поэтому, например, слова *фотограф*, *утренний* помечаются как незнакомые, однако их однокоренные слова *фотографировать* и *утро* есть в лексическом минимуме); появлением в текстах лексики на опережение из лексического минимума следующего уровня (лексема *особенно* появляется в текстах A1 14 раз, *раньше* – 22 раза; обе они появляются в лексических минимумах начиная с уровня A2); появлением лексики, совсем не входящей в лексические минимумы (*диета*, *смартфон*) или появляющейся значительно позже (*парень* встречается 29 раз в текстах всех уровней начиная от A2; *примерно* – 30 раз также начиная с уровня A2, эти лексеммы появляются в лексическом минимуме уровня C1).

Таблица 2

Лингвистические показатели уровней сложности текста по РКИ

	A1	A2	B1	B2	C1
Средняя длина слова в знаках	4,8	5	5,3	5,3	5,5
Средняя длина предложения в словах	8,8	10,7	12,7	12,7	13,1
Среднее количество знаков препинания на предложение	1,2	1,2	1,8	1,7	1,9
Процент глаголов в форме 1 л.	47%	25%	18%	22%	14%
Процент слов в Р.п. в предложении	15%	10%	40%	42%	48%
TTR (type/token ratio)	0,09	0,10	0,11	0,12	0,16
Процент текста, который не покрывается лексическим минимумом A1	16%	24%	33%	40%	44%

Эти и многие другие вычисляемые параметры текстов были использованы для обучения математической модели по автоматическому определению уровня сложности текста по РКИ [21]. Данная математическая модель легла в основу программы «Текстометр» Textometer, которая позволяет пользователю ввести любой текст и получить информацию о его уровне сложности в терминологии РКИ, от A1 до C1, а также некоторые статические данные о тексте².

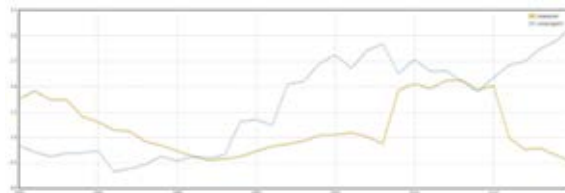
Сравнительный анализ лексики учебников и лексических минимумов

Государственные образовательные стандарты по РКИ и лексические минимумы, являясь официальными регулирующими документами, по понятным причинам консервативны и порой медленнее реагируют на изменения в языке, чем учебные пособия, которые должны постоянно ориентироваться на меняющиеся коммуникативные нужды учащихся. Поэтому статистическая информация, собранная на коллекции учебных текстов, может служить аргументом за или против включения/исключения слова в лексический минимум. Например, в государственном лексическом минимуме по РКИ 2017 г. [3] все еще можно увидеть слово *дискета*, давно отошедшее в разряд историзмов, тогда как во всей коллекции текстов учебников это слово не встречается ни разу. Лексема *универмаг* присутствует в лексических минимумах A1, A2, B1 [1–3], а *супермаркет* – нет. Интуиция носителя языка подсказывает, что в коммуникативной действительности последних двух десятилетий ситуация обратная: гораздо

чаще мы употребляем слово *супермаркет*, чем *универмаг*. Однако доверять интуиции в процессе отбора лексики – опасная практика: интуиция может подвести, может быть тесно связана с городом и страной жительства, профессией и возрастом человека. Поэтому обратимся к цифрам. В корпусе **RuFoLa** *универмаг* появляется всего 2 раза, а *супермаркет* – 10. Но при таких небольших значениях рискованно делать поспешные выводы, цифры в таком случае могут быть непоказательны, что связано с относительно небольшим объемом корпуса RuFoLa. Развеем сомнения с помощью сервиса построения графика частотности слова по годам на материале текстов НКРЯ³.

График 1

Сравнение частот слов *универмаг* и *супермаркет* с 1980 по 2015 г.



На графике 1 представлена встречаемость в корпусе слов *универмаг* (желтая линия) и *супермаркет* (голубая линия) с 1980 по 2015 г. Можно сделать вывод, что *супермаркет* является более употребляемой лексемой примерно с 1993 г., а к 2015 г. разрыв в частотности

² Программа доступна по адресу pushkin-lab.ru

³ ruscorpora.ru/new/ngram

становится еще больше. Таким образом, данные НКРЯ подтвердили гипотезу о постепенном замещении слова *универмаг* словом *супермаркет* с конца прошлого столетия, что можно принять во внимание при очередном обновлении лексических минимумов.

Стоит оговорить и риски такого подхода к отбору лексики: автор пособия может опираться на нужды определенной узкой категории учащихся (например, экспатов, приехавших работать в Россию, для которых крайне актуальны темы, связанные с работой в компании и налаживанием коммуникации с русскими коллегами), неактуальные для других групп студентов; или транслировать свои личные лексические привычки и

симпатии. Способом фильтрации таких случаев может стать проверка частоты слова по нескольким пособиям: вероятность ошибки или индивидуальных особенностей стиля автора при таком подходе становится гораздо ниже. Например, слово *офис* отсутствует в лексических минимумах А1 и А2, однако встречается 37 раз в корпусе учеников RuFoLa, из них 18 вхождений – в текстах уровня А1. Как видно из таблицы 3, есть учебник, в котором это слово оказалось особенно востребованным («Точка Ру»), однако оно встречается и в других комплексах, включая коллекцию текстов ЦМО МГУ. На основании таких данных можно включить лексему *офис* в качестве кандидата в обновленный лексический минимум.

Таблица 3

Распределение вхождений слова *офис* в подкорпусе уровня А1 корпуса RuFoLa

Название пособия	Количество вхождений
«Точка Ру» 1	12
«Поехали» часть 1.1	2
«5 элементов» часть 1	2
Текстотека МГУ (элементарный уровень)	2

Необходимо также заметить, что авторы пособий при составлении текстов в свою очередь опираются на существующие лексические минимумы, поэтому в некоторых случаях такое исследование рискует быть хождением по кругу. Однако, несмотря на перечисленные недостатки, учет встречаемости и востребованности лексики из лексических минимумов в реальных современных учебных пособиях представляется полезной практикой в русле доказательной педагогики.

Создание корпуса учебников РКИ и его развитие представляется новой страницей в теории учебника иностранного языка: корпус

становится ценным материалом как для исследователей в области цифровой лингводидактики, методистов, так и для самодиагностики авторов будущих учебных пособий по РКИ. Наиболее перспективными направлениями дальнейшего развития нам видятся пополнение корпуса учебниками более ранних годов издания для изучения диахронических изменений в учебниках; пополнение коллекции учебниками, изданными в других странах для сравнительного анализа методических подходов; пополнение текстами упражнений для анализа представленности различных грамматических явлений в учебниках. ■

ЛИТЕРАТУРА

1. Андрияшина Н.П., Козлова Т.В. Лексический минимум по русскому языку как иностранному. Элементарный уровень. Общее владение. СПб., 2012.
2. Андрияшина Н.П., Козлова Т.В. Лексический минимум по русскому языку как иностранному. Базовый уровень. Общее владение. СПб., 2011.
3. Андрияшина Н.П. и др. Лексический минимум по русскому языку как иностранному. Первый сертификационный уровень. Общее владение. СПб., 2017.
4. Андрияшина Н.П., Макова М.Н., Працук Н.И. Тренировочные тесты по русскому языку как иностранному. I сертификационный уровень. М., 2004.
5. Арбузова И.И. Методика обучения иностранных студентов глагольно-предложному управлению: на материале русских пространственных предлогов: Автореф. дис. ... канд. пед. наук. СПб., 2013.
6. Ардатова Е.В. Образ России в современных учебниках по русскому языку как иностранному // APRIORI. Серия: Гуманитарные науки. 2015. № 3.

7. Арутюнов А.Р., Трушина Л.Б., Чеботарев П.Г. Многофакторный анализ учебников иностранных языков // Содержание и структура учебника русского языка как иностранного. М., 1981.
8. Батсуурь Б. Вопрос о порядке слов в учебниках и учебных программах по русскому языку для монгольских студентов // Русистика. 2012. № 1.
9. Бим И.Л. Методика обучения иностранным языкам как наука и проблемы школьного учебника. М., 1977.
10. Ван С. Обучение видам русского глагола в китайской аудитории (уровень А2–В1): Автореф. дис. ... канд. пед. наук. М., 2018.
11. Владимирова Т.Е. Государственный стандарт по русскому языку как иностранному. Элементарный уровень. СПб., 2001.
12. Вострякова Н.А. Русские гендерные отношения в учебных материалах по РКИ как предмет осмысления в мусульманской аудитории // Вестник ВолГУ. Серия 2: Языкознание. 2019. № 3.
13. Вятютнев М.Н. Теория учебника русского языка как иностранного (методические основы). М., 1984.
14. Дзюба Е.В. Образ России в болгарских учебниках по русскому языку как иностранному. Ч. 1: На материале учебников для школьников среднего звена // Педагогическое образование в России. 2019. № 6.
15. Зайцева О.А., Паустьян Е.В. Этнические стереотипы о русских в китайских учебных изданиях по русскому языку как иностранному // Коммуникативные исследования. 2019. № 3.
16. Ивлева Д.А. Учебник русского языка как иностранного – источник формирования и средство поддержания стереотипов // Вестник ЧелГУ. 2013. № 2 (293).
17. Милославская С.К. Русский язык как иностранный в истории становления европейского образа России. М., 2012.
18. Шукин А.Н. Русский язык как иностранный: основы учебниковедения. М., 2018.
19. Gabrielatos C. Corpus-based Evaluation of Pedagogical Materials If-conditionals in ELT Coursebooks and the BNC. Paper presented at the 7th Teaching and Language Corpora (TALC) Conference, 1–4 July 2006. Paris, 2006.
20. Kim J.Y., Lee J.Y., Kim J.-r. A synchronic lexical analysis of the South and North Korean English textbooks. Linguistic Science. 2017. 24 (2).
21. Laposhina A.N., Veselovskaya T.S., Lebedeva M.U., Kupreshchenko O.F. Automated Text Readability Assessment For Russian Second Language Learners // Komp'yuternaja Lingvistika i Intellektual'nye Tehnologii / Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialogue 2018». 2018. 17 (24).
22. Meuener F., Gouverneur, C. New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material // Corpora and Language Teaching. 2009.
23. Mukundan J., Nimehchisalem V. Gender representation in Malaysian Secondary School English Language Textbooks. Indonesian Journal of English Language Teaching. 2008.
24. Römer U. Textbooks: a Corpus-driven Approach to Modal Auxiliaries and their Didactics. In Sinclair, J. (ed.) How to Use Corpora in Language Teaching. Amsterdam / Philadelphia: Benjamins, 2004.
25. Volodina E. et al. You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a Second Language // Proceedings of the third workshop on NLP for computer-assisted language learning. 2014.
26. Weninger C. Textbook analysis. In Chapelle, C.A. (Ed.), The Encyclopedia of Applied Linguistics. Hoboken, NJ, 2018.
27. Williams D., Developing criteria for textbook evaluation, ELT Journal. Vol. 37, Issue 3, 1983.

References

1. Andryushina N.P., Kozlova T.V. Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Elementarnyy uroven'. Obshchee vladenie. SPb., 2012.
2. Andryushina N.P., Kozlova T.V. Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Bazovyy uroven'. Obshchee vladenie. SPb., 2011.
3. Andryushina N.P. i dr. Leksicheskiy minimum po russkomu yazyku kak inostrannomu. Pervyy sertifikatsionnyy uroven'. Obshchee vladenie. SPb., 2017.
4. Andryushina N.P., Makova M.N., Prashchuk N.I. Trenirovochnye testy po russkomu yazyku kak inostrannomu. I sertifikatsionnyy uroven'. M., 2004.
5. Arbuzova I.I. Metodika obucheniya inostrannykh studentov glagol'no-predlozhnomu upravleniyu: na materiale russkikh prostranstvennykh predlogov: Avtoref. dis. ... kand. ped. nauk. SPb., 2013.
6. Ardatova E.V. Образ России в современных учебниках по русскому языку как иностранному // APRIORI. Seriya: Gumanitarnye nauki. 2015. № 3.

7. Arutyunov A.R., Trushina L.B., Chebotarev P.G. *Mnogofaktornyj analiz uchebnikov inostrannyh yazykov / Soderzhanie i struktura uchebnika russkogo yazyka kak inostrannogo*. M., 1981.
8. Batsuur' B. *Vopros o poryadke slov v uchebnikah i uchebnyh programmah po russkomu yazyku dlya mongol'skih studentov* // *Rusistika*. 2012. № 1.
9. Bim I.L. *Metodika obucheniya inostrannym yazykam kak nauka i problemy shkol'nogo uchebnika*. M., 1977.
10. Van S. *Obuchenie vidam russkogo glagola v kitajskoj auditorii (uroven' A2–V1): Avtoref. dis. ... kand. ped. nauk*. M., 2018.
11. Vladimirova T.E. *Gosudarstvennyj standart po russkomu yazyku kak inostrannomu. Elementarnyj uroven'*. SPb., 2001.
12. Vostryakova N.A. *Russkie gendernye otnosheniya v uchebnyh materialah po RKI kak predmet osmysleniya v musul'manskoj auditorii* // *Vestnik VolGU. Seriya 2: Yazykoznanie*. 2019. № 3.
13. Vyatyutnev M.N. *Teoriya uchebnika russkogo yazyka kak inostrannogo (metodicheskie osnovy)*. M., 1984.
14. Dzyuba E.V. *Obraz Rossii v bolgarskih uchebnikah po russkomu yazyku kak inostrannomu. Chast' 1: Na materiale uchebnikov dlya shkol'nikov srednego zvena* // *Pedagogicheskoe obrazovanie v Rossii*. 2019. № 6.
15. Zajceva O.A., Paust'yan E.V. *Etnicheskie stereotipy o russkikh v kitajskih uchebnyh izdaniyah po russkomu yazyku kak inostrannomu* // *Kommunikativnye issledovaniya*. 2019. № 3.
16. Ivleva D.A. *Uchebnik russkogo yazyka kak inostrannogo istochnik formirovaniya i sredstvo podderzhanija stereotipov* // *Vestnik ChelGU*. 2013. № 2 (293).
17. Miloslavskaya S.K. *Russkij yazyk kak inostrannyj v istorii stanovleniya evropejskogo obraza Rossii*. M., 2012.
18. Shchukin A.N. *Russkij yazyk kak inostrannyj: osnovy uchebnikovedeniya*. M., 2018.
19. Gabrielatos C. *Corpus-based Evaluation of Pedagogical Materials If-conditionals in ELT Coursebooks and the BNC*. Paper presented at the 7th Teaching and Language Corpora (TALC) Conference, 1–4 July 2006, Paris. 2006.
20. Kim J.Y., Lee J.Y., Kim J.-r. *A synchronic lexical analysis of the South and North Korean English textbooks*. *Linguistic Science*, 2017. 24 (2).
21. Laposhina A.N., Veselovskaya T.S., Lebedeva M.U., Kupreshchenko O.F. *Automated Text Readability Assessment For Russian Second Language Learners* // *Komp'yuternaja Lingvistika i Intellektual'nye Tehnologii / Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialogue 2018»*. 2018. 17 (24).
22. Meuener F., Gouverneur C. *New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material* // *Corpora and Language Teaching*. 2009.
23. Mukundan J., Nimehchisalem V. *Gender representation in Malaysian Secondary School English Language Textbooks*. *Indonesian Journal of English Language Teaching*. 2008.
24. Römer U. *Textbooks: a Corpus-driven Approach to Modal Auxiliaries and their Didactics*. In Sinclair, J. (ed.) *How to Use Corpora in Language Teaching*. Amsterdam; Philadelphia: Benjamins, 2004.
25. Volodina E. et al. *You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a Second Language* // *Proceedings of the third workshop on NLP for computer-assisted language learning*. 2014.
26. Weninger C. *Textbook analysis*. In Chappelle, C.A. (Ed.), *The Encyclopedia of Applied Linguistics*. Hoboken, NJ, 2018.
27. Williams D., *Developing criteria for textbook evaluation*, *ELT Journal*, Vol. 37. Issue 3, 1983.

A.N. Laposhina

Pushkin State Russian language Institute
Moscow, Russia

A CORPUS OF RUSSIAN TEXTBOOK MATERIALS FOR FOREIGN STUDENTS AS AN INSTRUMENT OF AN EDUCATIONAL CONTENT ANALYSIS

RFL textbook, corpus of textbooks, pedagogical corpus, corpus-driven language learning, textbook analysis.

The paper describes the experience of collecting, marking up and utilising the corpus of texts from textbooks of Russian as a foreign language, RuFoLa. Such a collection of texts is an effective tool for getting objective information about the content of modern RFL textbooks for both research and practical purposes. Two studies are given as examples of tasks solved on the basis of the RuFoLa corpus. First, we calculated the data on linguistic features of the complexity level of educational materials on the CEFR scale from A1 to C1. Secondly, we showed examples of frequency analysis of textbook vocabulary and its comparison with the RFL vocabulary lists – it provides valuable information for evaluating textbook content and for vocabulary lists updating.