

УТВЕРЖДАЮ

Первый проректор федерального
государственного автономного
образовательного учреждения высшего
образования «Национальный
исследовательский университет
«Высшая школа экономики»

д.э.н. профессор Вадим Валерьевич Радаев



В.В. Радаев

« 17 » *апреля* 2023 г.

ОТЗЫВ

ведущей организации на диссертацию Лапошиной Антонины
Николаевны
«Лингводидактическое обоснование применения автоматической оценки
сложности учебного текста в преподавании РКИ»,
представленную на соискание учёной степени кандидата педагогических
наук по специальности 5.8.2 – Теория и методика обучения и воспитания
(русский язык как иностранный, уровень общего, профессионального,
дополнительного образования, профессионального обучения)

Актуальность темы диссертации

Диссертация Антонины Николаевны Лапошиной посвящена
актуальной проблеме применения автоматической оценки сложности
учебного текста в преподавании русского языка как иностранного (далее
РКИ). Отечественная лингводидактика формировалась еще в
докомпьютерную эпоху, поэтому методы автоматической оценки сложности
текста были просто не доступны методистам, в результате чего наиболее
востребованные учебные пособия, лексические минимумы и государственные
стандарты создавались с опорой на интуицию экспертов, без использования
автоматических методов обработки и анализа текста.

Новизна полученных результатов и выводов

Работа носит ярко выраженный междисциплинарный характер и
соединяет идеи современной теоретической лингвистики и науки об анализе
данных, с одной стороны, и достижения лингводидактики, с другой. Это

позволило соискателю получить совершенно **новые** нетривиальные результаты и выводы.

Новизна исследования заключается в том, что в нем был представлен комплексный анализ учебного текста с точки зрения формальных показателей его сложности для иностранных учащихся. Автор впервые обобщает и формализует систему лингвистических признаков сложности текста, релевантных при обучении русскому языку как иностранному. В работе также предложена и реализована авторская методика сбора и разметки сбалансированного корпуса образцов учебных текстов. Автором создана и апробирована математическая модель автоматической оценки сложности текста на русском языке для изучающих русский язык как иностранный, а также методика верификации применимости автоматической системы оценки сложности текстов в практике обучения РКИ. Наконец, автором впервые был предложен комплект методических материалов, позволяющих интерпретировать формальные лингвистические характеристики текста в зависимости от уровня владения русским языком и/или типа методической задачи.

Апробация работы и публикации

Работа была апробирована в группах иностранных студентов подготовительного факультета и факультета обучения русскому языку как иностранному Государственного института русского языка им. А.С. Пушкина, а также среди и преподавателей русского языка как иностранного (78 студентов и 7 преподавателей). Кроме того, в апробации приняли участие российские и зарубежные преподаватели РКИ, привлеченные посредством краудсорсинга (41 человек).

Основные положения диссертации были изложены автором на шести научно-практических конференциях и вебинарах:

1. Ежегодная международная конференция по компьютерной лингвистике и интеллектуальным технологиям «Диалог 2017» (РГГУ, 31 мая – 3 июня 2017 г.);
2. Международная научно-практическая интернет-конференция «Актуальные вопросы описания и преподавания русского языка как иностранного/неродного» (Москва, 27 ноября – 1 декабря 2017 г.);
3. Ежегодная международная конференция по компьютерной лингвистике и интеллектуальным технологиям «Диалог 2018» (РГГУ, 30 мая – 2 июня 2018 г.);

4. Международная научно-практическая конференции «Корпусные и компьютерные технологии и лингвистические проблемы» (Нижний Новгород, 12–14 октября 2018 г.);

5. XLVIII международная филологическая конференция (СПбГУ, 26 марта 2019);

6. Международный форум «РКИ-перезагрузка 2021: уроки пандемии» (Москва, 18 – 19 июня 2021); 7. VII конгресс РОПРЯЛ «Динамика языковых и культурных процессов в современной России» (УрФУ, 7 – 8 октября 2021 г.); 8. Вебинар «Текстометр: новый инструмент для подготовки текста к занятию по РКИ» на портале «Образование на русском» (27 марта 2019).

По теме диссертации подготовлено 8 публикаций, 3 из которых опубликованы в изданиях, включенных в перечень ВАК РФ, 1 – в издании, приравненном к перечню ВАК РФ. Кроме того, получено 1 свидетельство о государственной регистрации программы для ЭВМ в Федеральной службе по интеллектуальной собственности, приравниваемое к перечню ВАК при Минобрнауки России.

Автореферат и публикации автора полностью отражают содержание и структуру диссертации.

*Обоснованность научных положений и выводов,
сформулированных в диссертации*

Научные положения, выводы и рекомендации, сформулированные в диссертации, представляются **обоснованными**. Исследование проводится с применением современных методов компьютерной лингвистики, которыми автор владеет на высоком профессиональном уровне. Анализ полученных данных осуществляется с помощью статистических методов, которыми автор также блестяще владеет и уверенно пользуется. **Достоверность** научных положений и полученных результатов не вызывает сомнений.

*Соответствие содержания диссертации автореферату и указанной
специальности*

Диссертация и автореферат **полностью соответствуют паспорту научной специальности 5.8.2 – Теория и методика обучения и воспитания (русский язык как иностранный, уровень общего, профессионального, дополнительного образования, профессионального обучения)**: автором разработан новый методологический подход к построению предметного образования (обучению русскому языку как иностранному), представлена модель автоматизированной оценки уровня сложности учебного текста,

методической науки и успешно применяемая в лингводидактической практике.

Значимость результатов для науки и производства

Идея применить к лингводидактическим разработкам методы автоматической оценки сложности текста имеет огромную **практическую ценность** и выводит методику преподавания РКИ принципиально на новый уровень, облегчая труд преподавателей и делая его результаты более валидными. Так, применение разработанной А. Н. Лапошиной системы оценки сложности текста «Текстометр» позволяет за считанные секунды понять, какому уровню владения языком соответствует текст, предлагаемый для изучения, какие слова и грамматические единицы в этом тексте наиболее полезны для изучения, а какие редки и, возможно, нуждаются в адаптации. Кроме того, автоматизированная оценка сложности текста позволяет стандартизировать маркировку учебных пособий и контрольно-измерительных материалов по РКИ в соответствии с уровнями владения языком, минуя расхождения, неизбежные при использовавшейся ранее экспертной оценке.

Исследование А. Н. Лапошиной имеет не только безусловную практическую ценность, но и немалую **теоретическую значимость**, заключающуюся в том, что, здесь, опыт исследователей в области определения сложности текста был систематизирован и представлен в новом ракурсе – с позиций лингводидактики; кроме того, при написании диссертации был создан эталонный корпус текстов, представленных в учебных пособиях по РКИ, где уровень определен заранее; наконец, формальные признаки текста, применяемые при оценке его сложности лингвистами-теоретиками, были обобщены и сопоставлены с критериями оценки сложности текста, традиционно используемыми в лингводидактике, и те и другие – проверены на эффективность в преподавании РКИ. В результате обогатились такие области научного знания, как теоретическая лингвистика и методика преподавания русского языка как иностранного.

Замечания по диссертационной работе

1. Несколько странным представляется определение сложности текста как характеристики, не зависящей от качеств читателя, и при этом основывающейся среди прочего на количестве *незнакомых* слов (с. 15). Ведь слова не могут быть незнакомыми сами по себе – они всегда знакомы или не знакомы тому, кто воспринимает текст. А значит, пример противоречит

определению. Но это, по-видимому, вопрос не столько к А. Н. Лапошиной, сколько к автору определения, на которое она опирается, - Я. А. Микку.

2. На с. 65 автор упоминает о проблеме, с которой столкнулись, как кажется, все пользователи системы «Текстометр»: при оценке сложности текстов не учитываются дериваты, не вошедшие в лексические минимумы (даже если производящее слово в минимум вошло). К этой проблеме автор возвращается также в третьей главе – при описании работы сервиса «Текстометр» (с. 115, с. 130). А. Н. Лапошина пишет о том, что эта проблема была решена и отсылает читателя к статье 2021 г. Однако гораздо удобнее было бы прочитать об основных принципах решения проблемы (о том, какие дериваты учитывались, а какие нет, были ли это закрытые списки слов или списки аффиксов) непосредственно в диссертации. Интересно также узнать, насколько принятые решения реализованы в системе: как кажется, дериваты до сих пор учитываются далеко не все – хочется понять, насколько можно предсказать, что будет учтено системой, а что нет.

3. Также встает вопрос об учете когнатов (заимствованных из европейских языков слов), не вошедших в лексические минимумы, но без труда узнаваемых студентами, знакомыми с английским языком, а также, возможно, другими европейскими языками. Учет когнатов (хотя бы верхней части частотного списка), как кажется, повысил бы эффективность модели, т. к. позволил бы предсказывать догадку студентов (о том, что система пока не умеет этого, знает и сам автор – с. 98). А. Н. Лапошина на с. 115 говорит о том, что заимствования из европейских языков могут не распознаваться студентами, не знакомыми с этими языками, но таких, по-видимому, с каждым годом будет становиться все меньше: английский язык изучается в качестве первого иностранного практически повсеместно.

4. На с. 66 описывается метод работы с абстрактными словами. Автор говорит о двух применявшихся методах, один из которых предполагал работу со списками аффиксов. К сожалению, полный список использованных в работе аффиксов не приведен – перечисление заканчивается стандартным «и др.». Представляется, что полный список морфем, использованных при анализе, украсил бы работу и был бы очень полезен читателям. (Особенно это важно учесть, если работу впоследствии опубликовать). То же касается списка абстрактной лексики, разработанного командой АВВУУ: если это не закрытая информация, то стоило бы ее привести хотя бы в приложении.

5. Автор доказывает (на с. 70 и выше), что такие стандартные параметры, используемые при оценке сложности текста, как коэффициенты лексического разнообразия и плотности, для учебных текстов РКИ оказываются

малорелевантными (p -value больше 0,05). Интересно, есть ли у автора какие-то гипотезы о том, почему так вышло.

6. По данным автора, одним из значимых синтаксических признаков текста является процент местоименных прилагательных, в то время как процент местоименных существительных и наречий не оказывает существенного влияния на сложность текста (с. 76). Похоже, что причина такого результата в своего рода размножении сущностей: местоименные прилагательные *который* и *чей*, а также относительное *какой* по определению влияют на глубину дерева (вводят относительное предложение или сентенциальный акт). С другой стороны, легко представить себе текст уровня A0-A1 на тему, скажем, «Моя семья», изобилующий притяжательными местоименными прилагательными (*мой, твой, наш...*). Удалось ли посмотреть, какие именно местоименные прилагательные попали в выборку и насколько разнообразным оказалось это множество в эталонных текстах?

7. Также хочется, пользуясь случаем, высказать предложение по улучшению работы «Текстометра»: пользователю, занимающемуся адаптацией текста, было бы удобнее, если бы система выделяла (например, цветом) слова, не вошедшие в минимум того или иного уровня, **прямо в тексте** (а не после него, как это делается сейчас). Без такого выделения слова, найденные системой, приходится потом искать в тексте самостоятельно, что существенно замедляет работу.

8. Текст написан хорошим научным языком – читать его действительно интересно и совсем не трудно. Тем не менее некоторые **недостатки чисто формального свойства** хочется исправить:

1. Композиционно диссертация выглядит несколько разнородно. Так, вторую главу открывает синопсис, который украсил бы и другие главы, но, к сожалению, там отсутствует.

2. Насколько нам известно, ЦМО МГУ (упоминаемый на с. 57) сейчас называется иначе и имеет другую аббревиатуру.

3. В тексте встретилось несколько фрагментов, нуждающихся в перефразировании:

- В качестве уровня сложности текстов принималась информация, отображённая в аннотации пособия (с. 56) – вряд ли информацию можно принимать *в качестве уровня сложности*.
- Отношение количества глаголов на количество существительных в предложении (с. 63) – *отношение* управляет группой К ЧЕМУ, а не *НА ЧТО.

- Таблица 11 содержит данные корреляционного анализа группы грамматических признаков на уровень сложности текста (с. 71) – кажется, НА УРОВЕНЬ в данном предложении сказать нельзя: нет элемента, который управлял бы такой предложной группой.
- Самым прямым способом получить такую оценку является прямая просьба отнести текст к одному из 6 возможных уровней нескольких экспертов (с. 92) – неудачный порядок слов: НЕСКОЛЬКИХ ЭКСПЕРТОВ стоит переместить ближе к вершине (ПРОСЬБА) и, видимо, поменять падеж.
- Далее происходила сама «партия»: эксперту демонстрировались на экране два текста и предлагалось выбрать, какой из текстов является сложнее (с. 93) – ЯВЛЯЕТСЯ лишнее.

4. В работе встречаются повторы фраз и отдельных лексем, от которых, как кажется, легко можно было бы избавиться:

- Оговорка о том, что изучаться будут не любые учебные тексты, а лишь прозаические тексты монологического характера, приводится сначала в разделе 1.1 (с. 14), а потом в разделе 1.2 (с. 15), далее мы видим ее во второй главе – на с. 56. Кажется, достаточно было одного упоминания.
- На с. 116 в первом абзаце, состоящем из шести строк, слово МАТЕРИАЛ(Ы) использовано пять раз. На той же странице слово ТАК, вводящее пример, используется в начале двух соседних предложений.

5. Текст не везде хорошо вычитан. Если готовить диссертацию к публикации, то стоит обратить внимание на следующие моменты:

- Размытость границ уровней (например, ориентирован на В1-С1) **также затрудняет что затрудняет** (здесь и далее выделение жирным шрифтом наше – А. Л.) отнесение текста к одному определенному классу сложности (с. 58).
- Например, существительное СТОЛЕ **получал** такой набор тэгов: существительное, единственное число, мужской род, предложный падеж, начальная форма – *стол* (с. 60).
- Первую группу составляют лексические признаки текста, которые **свою очередь** можно разделить на несколько подгрупп... (с. 63)
- Вслед за большинством исследователей, мы также **используем рассчитываем** коэффициент лексического разнообразия (англ. lexical diversity) с помощью мер TTR и MLTD, и **коэффициент** лексической плотности текста (lexical density) (с. 64).
- В качестве источника данных о частотности слов для расчета лексических признаков, основанных на частотности слов, мы

использовали данные Нового частотного **словарь** современного русского языка [Ляшевская, Шаров 2009] (с. 64),

- Для расчета этой меры была 65 использована линейка лексических минимумов ТРКИ (далее ЛМ ТРКИ) от А1 до С1 [Лексический минимум 2013; 2015; 2017а; 2017б; 2018] (с. 65) – отсутствует закрывающая квадратная скобка.

Наконец, в работе встречается незначительное количество пунктуационных ошибок, легко устранимых при корректорской вычитке. На с. 104 вместо тире поставлен дефис (списки ключевых слов, и слов-наилучших кандидатов в словарик)

Высказанные замечания или незначительны, или носят дискуссионный характер и ни в коем случае не снижают безусловно положительного впечатления от работы.

Вывод

Диссертационная работа Лапошиной Антонины Николаевны «Лингводидактическое обоснование применения автоматической оценки сложности учебного текста в преподавании РКИ», соответствует требованиям пунктов 9-14 Положения о присуждении ученых степеней, утвержденного постановлением Правительства РФ от 24 сентября 2013 года № 842, предъявляемым к кандидатским диссертациям, а её автор заслуживает присуждения учёной степени кандидата педагогических наук по специальности 5.8.2 – Теория и методика обучения и воспитания (русский язык как иностранный, уровень общего, профессионального, дополнительного образования, профессионального обучения).

Отзыв подготовлен кандидатом филологических наук, доцентом Школы лингвистики факультета гуманитарных наук федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики» Леонтьевой Анной Леонидовной.

Отзыв рассмотрен и одобрен на заседании Школы лингвистики факультета гуманитарных наук федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики», протокол № 2.4.12-01/0404-01 от «4» апреля 2023 года.

Сведения о ведущей организации: Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ)

Адрес: 101000 г. Москва, ул. Мясницкая, 20.
Тел.: (495) 771-32-32
Электронная почта: hse@hse.ru
Сайт: <http://www.hse.ru>

Зам. руководителя Школы лингвистики
факультета гуманитарных наук,
кандидат филологических наук
по специальности 10.02.01 – «Русский язык»,
доцент Школы лингвистики
факультета гуманитарных наук
НИУ ВШЭ

Ахапкина Яна Эмильевна

Доцент
Школы лингвистики факультета
гуманитарных наук
кандидат филологических наук
по специальности 10.02.19 – Теория языка

Леонтьева Анна Леонидовна



Подпись заверяю



Подпись заверяю

10.06.2023

Сведения о ведущей организации
по диссертации Лапошиной Антонины Николаевны
«Лингводидактическое обоснование применения автоматической оценки
сложности учебного текста в преподавании РКИ»
по научной специальности 5.8.2 – Теория и методика обучения и воспитания
(русский язык как иностранный, уровень общего, профессионального,
дополнительного образования, профессионального обучения)
на соискание ученой степени кандидата педагогических наук

Полное наименование организации в соответствии с уставом	Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет «Высшая школа экономики»
Сокращенное наименование организации в соответствии с уставом	Национальный исследовательский университет «Высшая школа экономики», Высшая школа экономики, НИУ ВШЭ, ВШЭ
Ведомственная принадлежность	Правительство Российской Федерации
Почтовый индекс, адрес организации	101000, г. Москва, ул. Мясницкая, д. 20
Веб-сайт	https://www.hse.ru
Телефон	(495) 771-32-32
Адрес электронной почты	hse@hse.ru
Список основных публикаций работников структурного подразделения, составляющего отзыв, за последние пять лет по теме диссертации (не более 15 публикаций).	<ol style="list-style-type: none"> 1. Lyashevskaya O., Irina Panteleeva, Olga Vinogradova. Automated assessment of learner text complexity // <i>Assessing Writing</i> – 2021. – No. 49. – Article 100529. doi: http://doi.org/10.1016/j.asw.2021.100529 https://publications.hse.ru/view/313424453 2. Toldova S., Slioussar N., Bonch-Osmolovskaya A. A. Discourse complexity in the light of eye-tracking: a pilot Russian language study // <i>Russian Journal of Linguistics</i>. – 2022. – Vol. 26. – No. 2. – P. 449-470. doi: http://doi.org/10.22363/2687-0088-30140 https://publications.hse.ru/view/732256649 3. Khotimsky M., Leontyeva, A. Lessons of the Pandemic: A critical look at digital tools for online and face-to-face instruction of Russian as a foreign language, in: <i>Student-Centered Approaches to Russian Language Teaching: Insights, Strategies, and Adaptations</i>. Routledge, 2023. Ch. 9. P. 135-155. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003289333-13/lessons-pandemic-maria-khotimsky-anna-leontyeva 4. Lyashevskaya O., Vinogradova O. I.,

- Scherbakova A. Accuracy, syntactic complexity and task type at play in examination writing: A corpus-based study, in: Complexity, Accuracy and Fluency in Learner Corpus Research. Amsterdam : John Benjamins Publishing Company. – 2022. –Ch. 10. – P. 241-272. <https://publications.hse.ru/view/779488665>
5. Lyashevskaya O., Pyzhak J., Vinogradova O. I. Word-formation complexity: a learner corpus-based study // Russian Journal of Linguistics. – 2022. – Vol. 26. – No. 2. – P. 471-492. <https://publications.hse.ru/view/779489900>
 6. Виноградова О. И., Смирнова Е. А., Виклова А. В., Пантелеева И. М. Синтаксическая сложность академического текста: корпусное исследование работ русскоязычных студентов, изучающих английский язык, в сравнении с текстами профессиональных авторов. // Вестник РГГУ. Серия «Литературоведение. Языкознание. Культурология». – 2020. – № 7. – С. 107-129. <https://publications.hse.ru/view/332785968>
 7. Власова Е. А., Бец Ю. В., Каллистратидис Е. Создание устного учебного корпуса русского как иностранного: первые результаты и перспективы // Актуальные проблемы филологии и педагогической лингвистики. – 2022. – № 2. – С. 118-130. <https://publications.hse.ru/view/800515114>
 8. Власова Е. А., Еремина О. С., Культепина О. А. Об опыте создания массового онлайн-курса по русскому языку как иностранному для студентов элементарного уровня // Мир русского слова. – 2018. – № 2. – С. 93-98. <https://publications.hse.ru/view/222492232>
 9. Ольшевская М. Ю., Карпова Е. Л., Власова Е. А. Лексический минимум по языку специальности: сколько слов достаточно? Разработка принципов минимизации // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2019. – Т. 17. – № 4. – С. 63-77. <https://linggu.elpub.ru/jour/article/view/104>
 10. Литвинцева К. В., Леонтьева А. Л. Стало быть, по-русски так говорят: вводная конструкция стало быть в русском языке // Труды института русского языка им. В.В.

Первый проректор
федерального государственного автономного
образовательного учреждения высшего образования
«Национальный исследовательский
университет «Высшая школа экономики»


В.В.Радаев

« 13 » 03. 2023 г.

