

Некорректные заимствования в российских докторских диссертациях: сколько, где и у кого?^{1,2}

А.А. МАКЕЕВА*, А.О. ЦИВИНСКАЯ**, М.М. СОКОЛОВ***,
Н.А. СОКОЛОВА****, К.С. ГУБА*****

***Александра Алексеевна Макеева** – лаборант, Центр институционального анализа науки и образования, Европейский университет в Санкт-Петербурге. Адрес: 191187, Санкт-Петербург, ул. Гагаринская, д. 6/1а. E-mail: amakeeva@eu.spb.ru

****Анжелика Олеговна Цивинская** – младший научный сотрудник, Центр институционального анализа науки и образования, Европейский университет в Санкт-Петербурге. Адрес: 191187, Санкт-Петербург, ул. Гагаринская, д. 6/1а. E-mail: atsivinskaya@eu.spb.ru

*****Михаил Михайлович Соколов** – кандидат социологических наук, научный сотрудник, Центр институционального анализа науки и образования, Европейский университет в Санкт-Петербурге. Адрес: 191187, Санкт-Петербург, ул. Гагаринская, д. 6/1а. E-mail: msokolov@eu.spb.ru

******Надежда Александровна Соколова** – младший научный сотрудник, Центр институционального анализа науки и образования, Европейский университет в Санкт-Петербурге. Адрес: 191187, Санкт-Петербург, ул. Гагаринская, д. 6/1а. E-mail: nsokolova@eu.spb.ru

*******Катерина Сергеевна Губа** – кандидат социологических наук, директор, Центр институционального анализа науки и образования, Европейский университет в Санкт-Петербурге. Адрес: 191187, Санкт-Петербург, ул. Гагаринская, д. 6/1а. E-mail: kguba@eu.spb.ru

Цитирование: Макеева А.А., Цивинская А.О., Соколов М.М., Соколова Н.А., Губа К.С. (2020) Некорректные заимствования в российских докторских диссертациях: сколько, где и у кого? // Мир России. Т. 29. № 2. С. 27–48. DOI: 10.17323/1811-038X-2020-29-2-27-48

Тема плагиата в диссертациях политиков и иных публичных фигур неоднократно попадала в фокус внимания общественности в последние годы. Активистские расследо-

¹ Исследование финансировалось в рамках НИР «Где искать академическую нечестность?», номер госрегистрации – АААА-А18-118042590111-9. Авторы благодарят Юрия Чеховича («Антиплагиат»), Виктора Глухова (eLibrary) и Михаила Копотева (Университет Хельсинки) за помощь и плодотворные обсуждения.

² Статья опубликована в рамках проекта НИУ ВШЭ по поддержке публикаций авторов российских образовательных и научных организаций «Университетское партнерство».

вания, направленные на изобличение конкретных нарушителей, не позволяют, однако, составить общую картину распространенности плагиата в России. Мы предприняли попытку восполнить этот пробел, проанализировав случайную выборку из 2468 докторских диссертаций, защищенных в 2006–2015 гг. (9,3% от генеральной совокупности) с помощью скриптов программы «Антиплагиат». Наши данные позволяют не только оценить общее распространение плагиата, но и проверить несколько гипотез о том, какие факторы стоят за практикой заимствования. Мы рассматриваем три возможных объяснения, основанные на теоретических моделях альтернативных конвенций (группы ученых разделяют разные представления о допустимых заимствованиях), аномии (за плагиатом стоят сдвиги в академической социализации) и рационального выбора (за заимствованием стоит рациональное взвешивание издержек и выгод нарушения нормы). Мы проанализировали влияние на вероятность появления плагиата ряда независимых переменных, соотносимых с этими гипотезами – дисциплины, региона и типа организации, где прошла защита, года защиты и наукометрических показателей диссертанта. Средняя доля обнаруженных заимствований составила 19,1%. Медиана – 13,9%. Только четверть диссертаций имеют долю заимствований менее 7% текста, при этом, однако, 6,1% авторов заимствовали более 50%. В целом гипотеза альтернативных конвенций в свете полученных результатов выглядит наиболее убедительно: переменные, которые являются вероятными коррелятами аномии (экспансия дисциплины, высокие показатели имитативности научной деятельности), незначительно влияют на вероятность появления плагиата значимого воздействия, а переменные, существенные с точки зрения модели рационального выбора (индивидуальная сила ученого, сила институции, распространение «Антиплагиата»), оказывают лишь слабое влияние.

Ключевые слова: социология науки, наука в России, диссертации, плагиат, академическая нечестность, научная недобросовестность, академические нормы, автоматический анализ текстов

Случаи обнаружения некорректных заимствований в диссертациях публичных фигур часто становятся причинами громких скандалов. Только за последние годы подобные скандалы, в которые были вовлечены известные политики, разразились в Германии, США и Финляндии³. Разоблачения группы «Диссернет» коснулись и российских политиков, чиновников, многочисленных преподавателей и администраторов от науки и образования (ректоры вузов, директора школ). Вместе с тем отсутствуют данные, которые позволили бы оценить ситуацию с плагиатом в российской науке в целом, проследить ее динамику и ответить на вопрос о том, какие факторы стоят за появлением некорректных заимствований в диссертациях. По результатам своих исследований «Диссернет» публиковал распределения выявленных текстов с признаками обширного плагиата по дисциплинам. Однако при отсутствии систематической процедуры отбора диссертаций для проверок и четких критериев того, что заслуживает публичного осуждения как плагиат, а что – нет,

³ В 2011 г. министр обороны ФРГ Карл-Теодор цу Гуттенберг вынужден был оставить свой пост после того, как в его диссертации было обнаружено порядка 75% плагиата. В США в 2017 г. в центре скандала оказалась Моника Коули, которую президент Трамп хотел видеть своим представителем в Национальном совете безопасности (назначение не состоялось). В 2018 г. в Финляндии в плагиате в магистерской диссертации была обвинена Лаура Хухтасаари, кандидат в президенты от правой партии «Истинные финны» (степень не была отозвана, поскольку скопированная была лишь часть литературного обзора, но скандал мог поспособствовать поражению Хухтасаари на выборах).

возможность делать выводы на основании этих данных весьма ограничена [Структура, объем и ключевые игроки 2018]. Нисколько не умаляя общественно-политического значения отдельных разоблачений научной недобросовестности, мы считаем, что для получения общей картины необходим более систематический подход.

В этой статье мы делаем попытку реализовать такой подход, представляя результаты проверки случайной выборки докторских диссертаций, защищенных в 2006–2015 гг., используя компьютерные алгоритмы системы «Антиплагиат». Это позволяет оценить общие координаты распространения некорректных заимствований, узнать, как изменяется их объем, и выявить, какие факторы приводят индивидов в «группу риска». Помимо этого, подобный анализ позволяет обнаружить, насколько ситуация с некорректными заимствованиями в российских диссертациях вписывается в модели объяснения девиантного поведения, существующие в мировой литературе (рационального выбора, аномии, альтернативных конвенций).

Теоретические подходы к объяснению некорректных заимствований

Проблема определения некорректного заимствования

Распространение современного представления о научном авторстве совпадает с возникновением экспериментальной науки, и это совпадение кажется слишком точным, чтобы быть случайным. Связь собственнического отношения к открытиям с духом современной науки интересовала, среди прочих, Роберта Мертона [Merton 1973], утверждавшего, что интеллектуальная собственность автора на свои открытия гарантирует индивиду вознаграждение за внесение вклада в копилку общего знания. Собственности в данном случае соответствует не возможность предотвращать распространение информации (как в случае с более древними производственными секретами), а право требовать, чтобы каждый ее получатель в будущем использовал ее исключительно с упоминанием оригинального производителя. Только это, предполагал Р. Мертон, может служить условием, на котором ученые будут соглашаться производить информацию и помещать ее в открытый доступ. По этой причине рождение современной науки сопровождается спорами о приоритете и авторстве (иногда довольно уродливые). Та же идея встречается в институциональной экономике науки, возводящей нашу концепцию авторства к потребности в интернализации экстерналий производства коллективных благ общедоступной информации [Stephans 1996].

По Р. Мернтону, плагиат представляет собой симптом распада научного этоса, на котором зиждутся институты науки. Плагиатор, с этой точки зрения, совершает сразу два преступления – кражу и мошенничество, присутствующие в разных пропорциях в конкретном акте некорректного заимствования: во-первых, он крадет у подлинного автора, отнимая у того право на заслуженное признание; во-вторых, он мошеннически вводит в заблуждение общество, принимающее за «настоящего» доктора наук человека, который получил степень, не пройдя всех необходимых испытаний. В данном случае жертвами становятся будущие работодатели, клиенты

или студенты фальшивого ученого, а также коллеги (другие обладатели той же степени), которые могут пострадать в результате того, что репутация дисциплины ухудшится. Поэтому плагиат остается проступком, даже если автор присвоенного текста не возражает и сам предлагает использовать его таким образом. Элемент кражи в этом случае отсутствует, однако оба оказываются замешаны в мошенничестве.

Единодушная поддержка нормы, запрещающей плагиат, не означает, однако, единства толкования того, что такое плагиат. Действительно, что составляет авторство и что может пониматься как корректное заимствование, а что – нет, является предметом разных трактовок. В случае с научным текстом обвинения в некорректных заимствованиях могут принять три формы. «А» может обвинить «В» в том, что последний заимствовал его идеи, данные или слова (текст как таковой). Разумеется, это не взаимоисключающие вещи. Если «В» украл текст целиком, заменив только первую страницу с именем, то украдено и первое, и второе, и третье. Однако до определенных пределов возможна и кража чего-то одного: предположим, «А» украл идею революционного эксперимента, но поставил его сам и написал текст самостоятельно; или взял большие куски текста с обзором литературы и описанием процедур сбора данных, но сами данные и их анализ были полностью оригинальными. Под большинство определений академического плагиата, которыми пользуются соответствующие инстанции (включая российский ВАК [Шахрай, Аристер, Тедеев 2015]), подпадает и первое, и второе, и третье, однако хорошо известно, что им приписывается разная степень непростительности. Заимствование и использование чужой идеи, особенно тем способом, когда ею уже не сможет воспользоваться законный автор, интуитивно кажутся более важными и тяжелыми прегрешениями, а заимствование технических частей текста (скажем, «А» скопировал из статьи «В» описания классического эксперимента, поставленного «В») – относительно мелким. К несчастью, кражу идей почти невозможно доказать в судебном порядке. Напротив, благодаря программному обеспечению распознавание текстуальных заимствований является разрешимой задачей, хотя само по себе оно и наименее важно. Тем не менее только оно может быть использовано активистами в качестве судебного доказательства или включено в статистический анализ как зависимая переменная.

В статье мы обратимся к изучению случаев следования или нарушений именно этого аспекта запрета плагиата, который можно назвать *нормой текстуальной аутентичности*. Эта норма требует оригинального сочинения любого академического текста, написанного с использованием естественного языка. Иными словами, в случае с диссертантами она требует собственноручного написания диссертации от первого до последнего слова. Именно случаи нарушения этой нормы (благодаря развитию машинного анализа текстов) могут быть выявлены со значительной точностью на больших массивах данных, и мы можем проанализировать, кто следует или не следует ей.

Теоретические объяснения академических девиаций и исследовательские гипотезы

С чем связано следование академической норме или отказ от нее? В социальных науках есть несколько готовых объяснений, которые обычно применяются для случаев поведения, определяемого как нарушение нормы. Во-первых, нарушители могут

не разделять эту норму в принципе: норма действенна в одной группе и не принимается в другой; если эти группы соприкасаются и если одна из них обладает большей властью, то слабая группа помечается ярлыком девианта (напр., [Becker 2000]). Применительно к плагиату мы можем предполагать, что разные группы ученых по-разному понимают, что такое «оригинальность» и «корректное заимствование».

Во-вторых, даже в группе, в которой есть консенсус по поводу нормы, находятся индивиды, нарушающие ее в корыстных целях. Объяснения девиации распадаются на два больших класса: социализационные, с одной стороны, и рационально-структурные, с другой. В первом случае девиантным индивидам известно о норме, но у них нет моральных ограничений, которые должны сдерживать пренебрежение ею [Thomas 1923]. В отношении плагиата такое объяснение подразумевает, что его появление является результатом сбоя в академической социализации или в уровне морального контроля над индивидом со стороны сообщества: по каким-то причинам некоторые не воспринимают этос, который должен уберечь их от подобных проступков. Основное эмпирическое ожидание, следующее из этой теории, будет состоять в том, что нечестные индивиды более склонны и к другим способам «срезать углы» – фейковым публикациям, публикациям в «мусорных» журналах, созданию «цитатных» картелей и применению иных способов накачивания наукометрических показателей. Из этого следует *Гипотеза 1: чем интенсивнее нарушения нормы текстовой аутентичности, тем выше вероятность встретить другие формы академической нечестности.*

Многие из тех, кто писал о распространении диссертационного плагиата в России, исходили из того, что он типичен для дисциплин, действие академических норм в которых было слабым в силу их (дисциплин) недавней экспансии. Скажем, под воздействием взрывного студенческого спроса программы по экономике и менеджменту множились, создавая параллельный спрос университетов на преподавателей с соответствующими степенями. Этот спрос множил ряды диссертантов, которые оказались в академическом мире случайно и не интернализировали научный этос; соответственно, они легко нарушали все мыслимые нормы научной жизни. Эти рассуждения подводят нас к *Гипотезе 2: чем сильнее была экспансия дисциплины после 1991 г., тем выше доля некорректных заимствований в ней.*

Рационально-структурные теории (напр., [Becker 1968]) утверждают, что все индивиды одинаково оппортунистичны и похожи друг на друга в отношении испытываемых ими моральных ограничений, но сильно отличаются в области относительных затрат и выгод, ожидаемых от нарушения нормы. Если социализационная теория предполагает, что индивиды придерживаются нормы, поскольку она представляется им правильной и легитимной, то рационально-структурная теория предусматривает, что норме с большей вероятностью следуют те, для кого затраты на следование ей ниже, а риски обнаружения и потери от наказания, наоборот, – выше. Применительно к плагиату рационально-структурная теория допускает, что к некорректным заимствованиям будут склонны те, кому (а) сложнее всего написать оригинальный текст (потому что выгода от копирования чужой работы для них больше, чем для тех, для кого это не проблема) и (б) кто по каким-то причинам меньше других боится разоблачения. Первое предполагает, что распространение плагиата будет характеризовать «слабую» науку, и тогда следует ожидать, что наличие некорректно цитируемого текста будет сопровождаться низкими значениями показателей цитирования и прочих характеризующих интеллектуальное влияние метрик. Кроме

того, мы предполагаем, что заимствования будут присутствовать в работах диссертантов, представляющих скорее слабые в исследовательском плане институции, чем сильные, а также в работах диссертантов из академически более слабых локаций. Это позволяет нам выдвинуть *Гипотезу 3: чем выше доля заимствования, тем ниже показатели, характеризующие научное влияние и признание индивида.*

Отметим отличие этой гипотезы от *Гипотезы 1*, согласно которой, в публикационной биографии плагиаторов найдутся факты, свидетельствующие о попытке «накачивать» показатели, но это не исключает, что эти показатели будут высокими. Напротив, *Гипотеза 3* утверждает, что мы найдем среди плагиаторов индивидов с низким влиянием, хотя и необязательно использующих «мусорные» журналы или иные способы обмана наукометрии.

Важно отметить: наше исследование не было направлено на то, чтобы определить, является ли строгое соблюдение нормы текстуальной аутентичности примером «хорошего института» в науке, который способствует ее развитию. Тем не менее наши данные позволяют проверить валидность по крайней мере одного из рациональных обоснований этой нормы. В духе теории рыночных сигналов [Spence 1973; Stiglitz 2000] норма текстуальной аутентичности может быть обоснована тем, что для более сильных ученых выигрыши, связанные с заимствованием чужого текста, ниже, чем для более слабых авторов, т.е. хороший ученый потратит на то, чтобы корректно описать другими словами те же теории, методы или данные, смысл которых ему понятен, сравнительно небольшое время, а плохой ученый потратит на это месяцы и, возможно, вовсе откажется от замысла дописывать диссертацию. Если это так, то насаждение нормы текстуальной аутентичности приведет к повышению сортирующей способности диссертационного сита. Результаты проверки нашей *Гипотезы 3* позволят определить, менее ли сильные ученые склонны к плагиату и, соответственно, валиден ли этот аргумент.

Далее мы можем предположить, что в течение изучаемого периода риски разоблачения увеличивались: все более широкое распространение получил «Антиплагиат», «Диссернет» публиковал результаты своих расследований, а ВАК стремился усилить контроль. И поскольку риски разоблачения непрерывно возрастали, наши рациональные оппортунисты должны отказываться от копи-пэйста в пользу оригинального написания текста. Отсюда следует *Гипотеза 4: чем позже период, к которому относится анализируемая диссертация, тем ниже в ней доля заимствований.*

Последняя из наших трех теорий, теория альтернативных конвенций, предполагает, что ответственные за то, что «Антиплагиат» определяет как некорректные заимствования, на самом деле просто следуют альтернативным нормам. Сами они считают, что ведут себя сообразно научному этосу, и не ожидают каких-либо отрицательных санкций. Они и их окружение не рассчитывают, что копирование фрагментов чужих текстов обязательно подпадает под категорию нарушения прав интеллектуальной собственности или демонстрации интеллектуальной слабости. Возможно, что в их понимании плагиат – допустимая демонстрация пренебрежения к процессу защиты (или к ее составляющим, таким как написание раздела «актуальность работы»⁴), они считают ее пустой формальностью. Возможно также,

⁴ Интересно, что это понимание разделяет сам «Диссернет», который не проверяет вводные части работ на оригинальность, – видимо, предполагая, что ни один человек в здравом уме не будет писать раздел про актуальность и практическую значимость своей работы самостоятельно.

что они разделяют идеологию, отвергающую идею индивидуального авторства в принципе (в этом случае особенно вероятно заимствование у своих учеников или учителей). При этом такие представления могут разделяться не только соискателями, но и всеми возможными участниками процесса защиты, включая экспертов ВАК по определенной специальности.

На основании этой теории сложно выдвинуть позитивные гипотезы о том, кто будет замечен в нарушении норм текстуальной аутентичности. Скорее, она может объяснить, почему предположения, следующие из других теорий, не находят подтверждения. Мы можем лишь связать распространение более строгих конвенций, характерных для англо-американской науки, с большей глобализованностью академической среды. Так, в случае с плагиатом доля заимствований будет ниже в более интернационализированных институциях, скажем, в университетах Проекта 5-100 по сравнению с обычными вузами. Это предположение, впрочем, не противоречит ожиданиям, основанным на двух других теориях. Мы можем ожидать, что в лучших университетах будут сконцентрированы также индивиды, впитавшие ценности этоса и не имеющие проблем с самостоятельным написанием статей. Скорее, мы можем интерпретировать как аргумент в пользу модели альтернативных конвенций то, что влияние глобализованности институций на объем заимствований будет единственным наблюдаемым эффектом в условиях, когда все остальные окажутся незначимыми. Продолжая эту линию рассуждений, мы также можем предполагать, что вузы в «мировых городах» – прежде всего Москве и Петербурге – первыми воспримут импортированные нормы. Из этого вытекает *Гипотеза 5: чем выше глобализация академической среды, тем ниже уровень некорректных заимствований.*

Данные и методы

Большинство исследований академической недобросовестности (как и других форм девиации) опираются на выявленные случаи такого поведения⁵. При этом возникает существенный риск перепутать факторы, ведущие к совершению проступка, с факторами, делающими вероятным разоблачение⁶.

Единственным надежным способом оценить масштабы и детерминанты нечестного поведения (например, некорректного заимствования) представляется проверка случайных выборок ученых. В настоящее время существуют технические средства, позволяющие проверять тексты работ на наличие некорректных текстовых заимствований. Использование такого программного обеспечения сопряжено с определенными техническими и порой юридическими сложностями – начиная с необходимости просматривать отчеты программы вручную и заканчивая вопросом авторского права и открытого доступа к публикациям. Несмотря на это,

⁵ За исключением опросных исследований, при которых индивидах опрашивают о совершенных ими проступках. По понятным причинам, однако, этот источник страдает от значительных искажений.

⁶ К сожалению, такая же проблема существует с результатами проверок «Диссернетом», которые направлялись подозрениями участников группы в отношении того, где может иметь место плагиат, и соображениями политической важности изобличения той или иной крупной фигуры.

в последние годы появляются статьи, использующие такое ПО (преимущественно *Turnitin*) для оценки некорректных заимствований, в том числе самоплагиата [Honig, Bedi 2012; Horbach, Halffman 2017], однако такие работы в основном посвящены отдельным группам ученых (например, участникам конференции Академии менеджмента [Honig, Bedi 2012] и голландским ученым из четырех областей знаний [Horbach, Halffman 2017]). Насколько нам известно, до сих пор не предпринимались попытки оценить масштабы заимствований во всей популяции ученых той или иной страны, и в настоящей работе мы стремимся заполнить этот пробел на примере российских докторов наук.

Исследование строится на проверке случайной выборки диссертаций программой «Антиплагиат»⁷. Обращение именно к российским докторским диссертациям было во многом обусловлено техническими соображениями: авторефераты докторских диссертаций вывешиваются на сайте ВАК с 2006 г. (в отличие от авторефератов кандидатских, доступных с 2011), что позволяло собрать данные о параметрах генеральной совокупности и на этом основании сформировать случайную выборку за более длинный исторический период. Кроме того, обычно ученый за свою жизнь пишет только одну докторскую диссертацию (что исключает повторения), диссертация имеет единственного автора и ее статус примерно одинаков во всех дисциплинах. Если бы для проверки была взята, например, случайная выборка статей в журналах, то мы столкнулись бы с выборочным смещением в пользу плодовитых авторов, проблемами с соавторствами и тем, что журнальные статьи не являются основными формами публикации для некоторых специальностей. В пользу нашего выбора можно привести, однако, и содержательные соображения. Докторская диссертация пишется в зрелом академическом возрасте; если аспирант, сажающийся за кандидатскую, может быть еще недостаточно знаком с нормами своего сообщества, то относительно доктора наук мы можем предполагать, что он знает правила игры и соблюдает или нарушает их вполне осознанно. Защита докторской, кроме того, открывает путь к руководству аспирантами и работе в диссертационных советах, и защитившие ее индивиды сами становятся хранителями и ретрансляторами норм сообщества.

Была собрана случайная выборка из 2600 докторских диссертаций за период 2006–2015 гг. по всем специальностям, сформированная на основе списка защищенных диссертаций, зарегистрированных на сайте ВАК. Поскольку авторефераты начали регулярно размещаться только в конце 2006 г., то он оказался недостаточно представлен в выборке; тем не менее мы не стали исключать соответствующие случаи. Мы также решили поставить верхнюю границу

⁷ «Антиплагиат» – система для поиска текстовых заимствований. Сначала текст извлекается из загруженного документа, затем ищутся совпадения в границах подключенных модулей. Итогом становится отчет о проверке документа, в котором указывается процент заимствования, процент оригинальности и процент процитированных источников. В сыром виде отчет не идентифицирует направление заимствования: если мы проверяем диссертацию 2000 г., а в источнике за 2001 г. встречается фрагмент из нее, то «Антиплагиат» покажет совпадение. Другая важная проблема – это заимствование текста из собственной работы. Если автор вставил в диссертацию фрагменты текста из собственных ранее опубликованных статей, то программа также покажет совпадающие фрагменты. И в первом и во втором случае мы не можем сказать, что автор ответственен за некорректные заимствования, так как в одном случае текст взяты у него, а в другом случае он вставил свой собственный текст. Мы решали эти проблемы следующим образом: в программу «Антиплагиат» были включены специально разработанные алгоритмы, которые считали заимствованием только те совпадения, которые были сделаны у автора с другой фамилией и если источники были опубликованы ранее года защиты диссертации.

на конце 2015 г., предполагая, что к тому времени доля заимствований резко снизится вследствие распространения программы «Антиплагиат». Как читатель увидит дальше, это предположение могло быть ошибочным. Всего за этот период было защищено 27735 докторских диссертаций (исключая, вероятно, небольшое число диссертаций по закрытой тематике, которые миновали публикацию в Интернете); наша выборка, таким образом, составила 9,3% от генеральной совокупности. Затем диссертации были проверены в системе «Антиплагиат». С помощью алгоритма отсекались тексты, выпущенные позже того года, в котором была написана диссертация, а также тексты, среди авторов которых встречался данный диссертант. Некоторые диссертации (132, т.е. приблизительно 5,5%) системе загрузить корректно не удалось. Правильно были распознаны 2468, которые и использовались в основных подсчетах. Мы сравнивали тексты диссертаций по двум модулям «Антиплагиата» – Российской научной библиотеки и модулю eLibrary.ru, поскольку они содержат большую часть релевантных академических текстов, и, помимо этого, тексты, размещенные в них, имеют однозначную датировку, что позволяет установить направление заимствования. Первый модуль дает возможность учесть заимствования в основном из диссертаций, второй – в основном из научных статей.

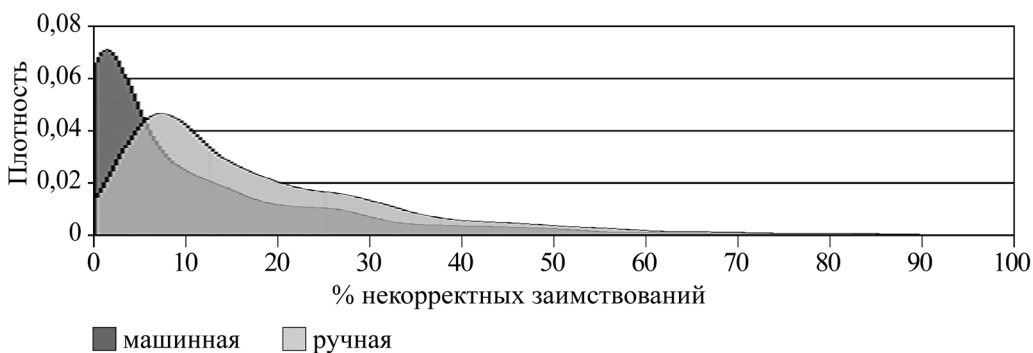


Рисунок 1. Распределение процента некорректных заимствований в зависимости от типа проверки

Для оценки надежности «Антиплагиата» мы сравнили результаты автоматической проверки с ручной проверкой для 860 диссертаций. Ручная проверка заключалась в просмотре всех источников, опубликованных позднее года защиты диссертации, и всех источников, авторы которых носили ту же фамилию. Кроме того, просматривался сам источник с тем, чтобы исключить мелкие совпадения в виде титульного листа или списка литературы. Тем самым в целом мы повторяли действия алгоритма «Антиплагиата». При ручной проверке средний процент некорректных заимствований оказался на 7% ниже, чем при автоматической проверке (медианный – на 7,5%). При этом только четверть работ имели разность в результатах проверок более 8 процентных пунктов. Распределение машинной и ручной проверки смотрите на *рисунке 1*. Однако следует отметить, что ручная проверка была заведомо консервативной (мы не отслеживали мелкие заимствования размером

менее 0,1% текста (примерно пять-семь предложений), которые иногда составляли приличную часть заимствований)⁸. В дальнейшем анализе в качестве независимой переменной использовались данные машинной проверки. Расчеты на диссертациях, проверенных вручную, дают сходные коэффициенты (доступно по требованию).

Результаты

Общее распределение

Рисунок 1 дает общее представление о том, как распределены доли некорректных заимствований. Про значительное большинство диссертаций мы можем сказать, что в них присутствует нечто, что может квалифицироваться как нарушение нормы текстуальной оригинальности. Таким образом, мы можем утверждать, что заимствования частей чужого текста является широко распространенной, едва ли не универсальной практикой. С другой стороны, ее радикальные нарушения встречаются также сравнительно редко. Заимствования свыше 40% охватывают примерно 10% диссертаций, а заимствования свыше 50% мы обнаружили в 157 случаях (3,64%). В этом смысле один из выводов нашего исследования может состоять в том, что норма текстуальной аутентичности, видимо, нарушается повсеместно. Однако это не значит, что нарушаются все нормы, регулирующие оригинальность вклада автора; фактически тотальный плагиат сравнительно редко проходит через жернова диссертационных советов. В принципе, эта картина сама по себе лучше вписывается в теорию альтернативной конвенции, когда нормы действуют, но не совсем те, которые предписывают ВАК и иные контролирующие органы. Если бы консенсус по поводу того, что текстуальные заимствования неприемлемы, действительно существовал, мы могли бы ожидать встретить бимодальное распределение с одним пиком около нуля (честные ученые) и вторым ближе к 100% заимствования (плагиаторы). Действительно, кража 10% текста, когда можно украсть 100%, сродни ограблению, в ходе которого грабитель забирает у жертвы 1/10 денег из кошелька: учитывая, что квалификация преступления от этого не изменится и на наказание это повлияет слабо, рациональнее брать все, если вообще пошел на преступление.

Структурные факторы

На первой фазе нашего анализа мы попытались связать количество некорректных заимствований с такими характеристиками защиты и диссертанта, как год

⁸ Более детальный анализ, проведенный на подборке диссертаций с заметным расхождением в проценте найденных заимствований, показал, что автоматическая идентификация заимствования допускала ошибки в части установления тождественности автора и неверного распознавания годов. Также в текстах могут не распознаваться легитимные цитаты (частая проблема для юридических наук). Совпадения также могут включать библиографические описания или приложения. Однако мы нашли всего два случая, в которых машинная проверка нашла значительные заимствования (свыше 60%), а ручная нашла их в пределах 10%. Эти случаи были удалены из дальнейших подсчетов.

защиты⁹, федеральный округ и организация (институт РАН, ведущий университет¹⁰, другой университет, прочие организации), где проходила защита, специальность и пол диссертанта. Тип организации может считаться переменной, характеризующей академическую силу среды, из которой происходит претендент на степень (*Гипотеза 3*), а также глобализованность этой среды (*Гипотеза 5*). Дисциплины в соответствии с широко распространенным предположением различаются уровнем интернализации норм научного этоса их представителями (*Гипотеза 2*). Пол также характеризует вероятность девиации: литература по академической недобросовестности (напр. [Honig, Bedi 2012]) предполагает, что мужчины в целом более склонны к нарушению социальных норм в достижительных целях (*Гипотеза 1*). Федеральный округ может характеризовать влияние локальной академической культуры и глобализованность норм (Центральный и Северо-Западный округа, в которых свыше 90% защит приходится на Москву и Петербург, предположительно более глобализованные, чем остальные – *Гипотеза 5*). Наконец, год характеризует эффекты контроля: распространение «Антиплагиата» и активность «Диссернета», по идее, должны были оказать сдерживающий эффект на активность рациональных злоумышленников (*Гипотеза 4*).

Количество наблюдений в каждой группе, средние и медианные значения, а также характеристики распределения числа некорректных заимствований представлены в *таблице 1*.

Из *таблицы 1* видно, что распределения количества некорректных заимствований не подчиняются нормальному закону распределения, медианы и средние не совпадают, асимметрия зачастую превышает пороговые значения, характерные для данных из нормального распределения. В связи с неприменимостью для данного случая параметрических методов мы использовали тест Краскела-Уоллиса с поправкой Бонферрони, являющейся консервативной оценкой. Так как одним из требований для критерия Краскела-Уоллиса является однородность распределений, для всех категориальных переменных был проведен тест Флингера-Килена, во всех случаях показавший наличие неоднородности. При интерпретации результатов теста Краскела-Уоллиса стоит учитывать нарушение этого требования, поэтому мы предлагаем ориентироваться на наиболее выдающиеся различия между группами.

Нам удалось найти существенные различия в количестве некорректных заимствований для организаций, федеральных округов и специальностей; также выявлена частично важная разница в заимствованиях в зависимости от года; между мужчинами и женщинами статистически значимых различий по количеству некорректных заимствований обнаружено не было. Так, рядовые университеты и другие организации имеют медианное значение количества некорректных заимствований 16,2 и 14,8% соответственно и статистически весомо между собой не различаются. В то же время медианное значение для институтов РАН и ведущих университетов совпадает и составляет 10,1%. Эти институции отличаются от рядовых университетов и других организаций на уровне значимости $<0,001$, при этом не различаясь между собой.

⁹ Для года защиты характеристики распределения не приводятся для 2006 г., который был исключен из последующего анализа по годам в связи с малым числом наблюдений (7).

¹⁰ В эту группу вошли все университеты Проекта 5-100, МГУ и СПбГУ.

Таблица 1. Описательные статистики категориальных переменных анализа

| | Среднее | Медиана | Минимум | Максимум | Стандартное отклонение | Асимметрия | N |
|-------------------------------------|---------|---------|---------|----------|------------------------|------------|------|
| Специальность | | | | | | | |
| Экономические | 22,9 | 18,3 | 2,3 | 85,8 | 16,7 | 1,3 | 433 |
| Технические | 19,7 | 14,3 | 0,7 | 74,2 | 17,2 | 1,1 | 363 |
| Медицинские | 19,1 | 15,1 | 2 | 88,8 | 13,9 | 1,6 | 330 |
| Педагогические | 20,3 | 15,6 | 2,5 | 89,2 | 16 | 1,5 | 202 |
| Физико-математические | 12,6 | 6,2 | 0,5 | 94,4 | 15,3 | 2,3 | 166 |
| Исторические | 14,9 | 10,3 | 1,4 | 91,8 | 14,9 | 3,1 | 129 |
| Биологические | 19,1 | 11,9 | 1,9 | 80,8 | 16,5 | 1,3 | 121 |
| Филологические | 9,2 | 6,9 | 0,8 | 61 | 8,5 | 3,5 | 117 |
| Юридические | 28,1 | 25,6 | 4,5 | 67,6 | 13,3 | 0,6 | 106 |
| Сельскохозяйственные и ветеринарные | 31,3 | 29,8 | 3 | 84,5 | 18,7 | 0,7 | 99 |
| Философские | 12,1 | 8,4 | 1,9 | 47,9 | 10,1 | 1,7 | 91 |
| Химические | 26,2 | 20,8 | 2,5 | 81,1 | 19 | 0,9 | 66 |
| Геологические | 11,5 | 8,7 | 0,9 | 57,9 | 11,1 | 2,6 | 57 |
| Психологические | 17 | 12,5 | 3,9 | 62,9 | 14,1 | 1,6 | 53 |
| Социологические | 14,5 | 12,3 | 2,2 | 41,5 | 9,7 | 1,2 | 45 |
| Политические | 16,1 | 11,9 | 3,1 | 45,5 | 11 | 1,1 | 40 |
| Искусствоведческие | 4,6 | 4,1 | 1,4 | 9,9 | 2,3 | 0,8 | 20 |
| Культурологические | 9,6 | 7 | 2,9 | 32,9 | 7,7 | 1,7 | 16 |
| Тип организации | | | | | | | |
| Университет | 20,6 | 16,2 | 0,7 | 94,4 | 16,4 | 1,3 | 1466 |
| Ведущий университет | 14,6 | 10,1 | 0,6 | 66 | 13 | 1,6 | 342 |
| Другие организации | 20,3 | 14,8 | 0,8 | 89,2 | 17,1 | 1,5 | 334 |
| Институты РАН | 15,5 | 10,1 | 0,5 | 89,1 | 15 | 2 | 311 |
| Федеральный округ | | | | | | | |
| Центральный | 18,6 | 13,3 | 0,5 | 91,8 | 15,9 | 1,5 | 1170 |
| Приволжский | 22,6 | 17,3 | 1,3 | 94,4 | 18,8 | 1,2 | 322 |
| Северо-Западный | 16,3 | 12,2 | 1,8 | 73,6 | 13,6 | 1,6 | 313 |
| Сибирский | 18,7 | 13,2 | 0,9 | 89,1 | 15,5 | 1,4 | 266 |
| Южный | 22 | 18,9 | 2,1 | 80,8 | 16,5 | 1,2 | 177 |
| Уральский | 16,8 | 11,9 | 1,9 | 65,3 | 14,2 | 1,4 | 112 |

| | | | | | | | |
|------------------------|------|------|-----|------|------|-----|------|
| Северо-Кавказский | 23,7 | 18,4 | 0,8 | 73,7 | 17,6 | 0,9 | 59 |
| Дальневосточный | 15,5 | 12 | 2 | 54,3 | 12,2 | 1,3 | 34 |
| Год защиты | | | | | | | |
| 2007 | 19,6 | 16,9 | 1,4 | 89,1 | 15,9 | 1,8 | 68 |
| 2008 | 19,4 | 13,9 | 2 | 74 | 15,8 | 1,2 | 329 |
| 2009 | 20 | 13,9 | 0,5 | 88,8 | 17,2 | 1,4 | 534 |
| 2010 | 19,7 | 14,5 | 0,6 | 81,1 | 16,6 | 1,2 | 314 |
| 2011 | 19,7 | 13,9 | 0,8 | 84,5 | 17 | 1,3 | 371 |
| 2012 | 19,7 | 15,7 | 1,3 | 89,2 | 15,4 | 1,6 | 254 |
| 2013 | 19,4 | 15,6 | 2 | 94,4 | 15,5 | 1,5 | 184 |
| 2014 | 17,3 | 13,3 | 1,6 | 80,7 | 13,8 | 1,9 | 252 |
| 2015 | 13,7 | 9,6 | 0,9 | 91,8 | 12,5 | 3 | 124 |
| Пол диссертанта | | | | | | | |
| Мужчины | 19,2 | 14,1 | 0,5 | 91,8 | 16,8 | 1,4 | 1358 |
| Женщины | 18,9 | 13,8 | 0,7 | 94,4 | 15,1 | 1,5 | 1095 |

Если говорить о федеральных округах, то здесь картина сложнее. По большей части статистически существенные различия между ними отсутствуют, исключение составляет разница между Центральным округом, с одной стороны, и Приволжским и Южным, с другой (в последних медианное количество некорректных заимствований больше на 4 и 5,6% соответственно, значимо на уровне $p < 0,05$). Мы также нашли разницу для Северо-Западного против Северо-Кавказского (в последнем количество некорректных заимствований выше на 6,2%, значимо на уровне $p < 0,05$), Приволжского и Южного (в обоих количестве некорректных заимствований выше на 5,1 и 6,7% соответственно, значимо на уровне $p < 0,01$). При том что мы не можем выделить явных лидеров по количеству некорректных заимствований, стоит отметить, что Южный, Северо-Кавказский и Приволжский округа демонстрируют высокие медианные значения по сравнению с остальными. По годам некорректные заимствования практически не различаются, и хотя может показаться, что произошло существенное снижение к 2015 г., статистически ощутимые различия обнаруживаются только между, с одной стороны, 2015 г. и, с другой – 2008, 2009, 2012 и 2013 гг., причем разница между 2015 и 2008–2012 гг. меньше, чем между 2013 и 2015 гг. Мы не видим никакой однозначной тенденции к снижению доли заимствований, которую можно было бы объяснить появлением «Антиплагиата» (получившего широкое распространение уже к 2010 г.) или деятельностью «Диссернета», начавшейся в 2013 г.

Нам удалось обнаружить определенное количество специальностей, статистически существенно отличающихся от других по количеству некорректных заимствований. Наименьшее их число мы выявили в искусствоведческих, физико-

математических, филологических и культурологических науках (но надо иметь в виду, что количество диссертаций по культурологии и искусствоведению в нашей выборке вообще невелико – 16 и 20 диссертаций соответственно), и оценки в их отношении не очень точны. Тем не менее следует отметить, что единичные диссертации с высоким уровнем некорректных заимствований встречаются почти во всех специальностях, даже в тех, в которых их общий уровень должен быть низок (физико-математические, филологические, философские и т.д.).

Наибольшее количество некорректных заимствований установлено в сельскохозяйственных науках: оно значимо выше по сравнению со всеми специальностями, за исключением юридических наук. Химики, находящиеся на втором месте, также значимо отличаются от физико-математических, исторических, философских, филологических, искусствоведческих, культурологических и геологических наук. Юристы также значимо отличаются от всех, кроме химиков, однако необходимо отметить, что юридические науки – это особый случай: «Антиплагиат» регулярно распознает выдержки из нормативно-правовых актов как некорректные заимствования.

Регрессионный анализ

На следующем этапе мы включили рассмотренные выше переменные в регрессионную модель для предсказания процента некорректных заимствований. Регрессия позволяет оценить влияние отдельных переменных при контроле по всем остальным – важная предосторожность в случае, если между переменными существует корреляция (скажем, если доля защит, приходящихся на разные дисциплины, изменялась по годам, и эффект года может быть связан с этим фактором, а не с тем, что во всех дисциплинах стали меньше заимствовать). Поскольку наши данные распределены не нормально, бета-регрессия является одним из возможных подходов. Альтернативой может служить использование линейной регрессии, но с преобразованием данных и взвешиванием, уменьшающим влияние наблюдений в «хвосте» распределения. В *таблице 2* представлены предельные эффекты, рассчитанные с помощью бета-регрессии (она показала лучшие характеристики по сравнению с линейной регрессией по предсказанным значениям и при анализе остатков).

Согласно представленной модели в ведущих университетах на 2,18% меньше некорректных заимствований по сравнению с обычными университетами при усреднении остальных характеристик. Если говорить о годах как группах, то мы видим отличие 2015 г. от 2007 г. на 5,58%, тем не менее различия между значениями по другим годам нам выявить не удалось. Для федеральных округов (в сравнении с Центральным) значимо отличаются Приволжский (2,77%) и Южный (2,37%). Что касается специальностей, то значимо большее количество некорректных заимствований по сравнению с физико-математическими науками демонстрируют химические (15,39%), биологические (7,88%), технические (5,89%), сельскохозяйственные и ветеринарные (18,23%), экономические (10,8%), юридические (19,24%), педагогические (8,03%), медицинские (9,01%) и психологические (5,46%) науки.

Таблица 2. Бета-регрессия: предельные эффекты¹¹

| Предикторы | % некорректных заимствований | |
|-------------------------------------|------------------------------|------------------------|
| | DF/dx | Стандартное отклонение |
| Мужчины | -0,71 | 0,56 |
| Другие организации | 0,51 | 0,85 |
| Институты РАН | -1,66 | 1,09 |
| Ведущие университеты | -2,18* | 1,02 |
| 2008 | -1,63 | 1,63 |
| 2009 | -1,57 | 1,6 |
| 2010 | -1,02 | 1,67 |
| 2011 | -0,98 | 1,65 |
| 2012 | -0,74 | 1,72 |
| 2013 | -1,24 | 1,74 |
| 2014 | -2,44 | 1,63 |
| 2015 | -5,58*** | 1,55 |
| Дальневосточный | -1,3 | 2,17 |
| Приволжский | 2,77** | 0,87 |
| Северо-Кавказский | 3,3 | 1,86 |
| Сибирский | 1,45 | 0,93 |
| Северо-Западный | -0,98 | 0,82 |
| Уральский | -0,85 | 1,27 |
| Южный | 2,37* | 1,12 |
| Химические | 15,39*** | 2,62 |
| Биологические | 7,88*** | 2,02 |
| Технические | 5,89*** | 1,53 |
| Сельскохозяйственные и ветеринарные | 18,23*** | 2,49 |
| Исторические | 3,51 | 1,8 |
| Экономические | 10,8*** | 1,6 |
| Философские | 0,61 | 1,86 |
| Филологические | -2,13 | 1,61 |
| Юридические | 19,24*** | 2,4 |
| Педагогические | 8,03*** | 1,83 |
| Медицинские | 9,01*** | 1,72 |
| Психологические | 5,46* | 2,55 |
| Социологические | 3,84 | 2,57 |
| Политические | 5,17 | 2,74 |
| Геологические | 1,75 | 2,24 |
| | N | 2393 |
| | Псевдо R ² | 0,1595 |

* p<0,05; ** p<0,01; *** p<0,001

¹¹ Референтные категории для пола – женский, для типа организаций – рядовой университет, для года – 2007, для округов – Центральный, для дисциплин – физико-математические науки.

Возвращаясь к нашим исходным теоретическим соображениям, мы видим, что идея альтернативных конвенций получает наибольшую поддержку. Существует значимая разница между более глобализованными институтами и остальными и более глобализованными столичными регионами и Южным, Северо-Кавказским и Приволжским округами. При этом нет различий между мужчинами и женщинами (которые предполагаются нормативно-социализационными теориями), а эффект года (предполагаемый рационально-структурными гипотезами) слаб и прослеживается только в 2015 г. Наконец, различия между дисциплинами существуют, но совсем не те, которые можно ожидать, основываясь на *Гипотезе 2*. Доля некорректных заимствований слабо связана с уровнем экспансии, иллюстрацией чего является *рисунок 2*, отражающий связь между средней долей заимствований и уровнем экспансии дисциплины, который рассчитан как отношение числа защит в 1995 г. (минимальное число защит в постсоветский период) и 2006 г. (максимальное).

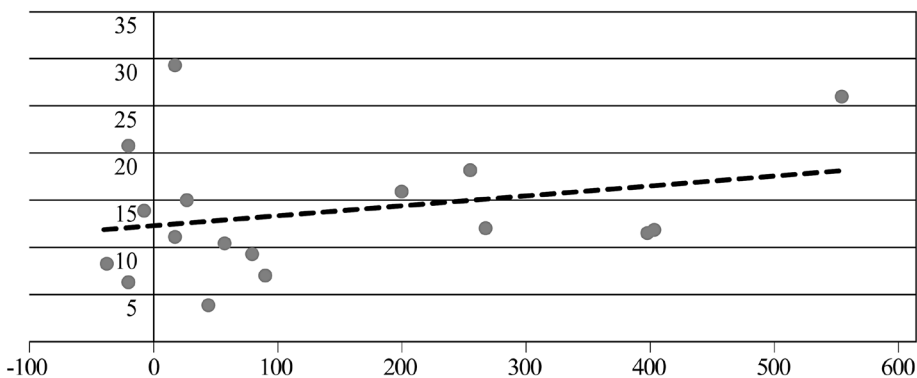


Рисунок 2. Медианный процент некорректных заимствований (ось Y) и экспансия дисциплины в 1995–2006 гг. (ось X)

Связь, если она и прослеживается, очень условна. Самые большие доли заимствований обнаруживаются в юриспруденции (но надо помнить о неточности оценок заимствований у юристов), сельскохозяйственных науках и химии, при этом среди них юриспруденция выросла сильнее всего, а химия была одной из переживших наибольшее сжатие.

Влияние известности и научной репутации

Гипотеза 3 выше гласит, что соблазн заимствовать часть чужой диссертации сильнее для тех, кто в силу общей интеллектуальной слабости не способен написать собственный текст, и, напротив, человеку, который пишет много и хорошо, проще писать самому, чем использовать копи-пэйт. Кроме того, в духе *Гипотезы 1* мы можем предполагать, что если заимствование воспринимается как разновид-

ность девиации, то те, кто не усвоил нормы научного этикета, первыми отреагируют на возросшее публикационное давление попытками обмануть систему – будут печататься в «мусорных» журналах, цитировать собственные статьи и т.д. Благодаря тому, что Российский индекс научного цитирования (РИНЦ) предоставляет широкий репертуар метрик, у нас есть возможность оценить не только количество, но и качество цитирований. К качественным показателям мы отнесли публикации и цитирования в «ядре» РИНЦ (состоящем из журналов, за которые проголосовали в ходе опроса высокоцитируемые российские ученые), импакт-факторы журналов, где были опубликованы и процитированы статьи, долю зарубежных статей и цитирований в них, а также долю статей в журналах ВАК. К имитационным показателям мы отнесли долю публикаций и цитирований, а также долю Хирш-индекса в изданиях eLibrary.ru, не вошедших в РИНЦ (категория, в которую входят в том числе исключенные из РИНЦ «мусорные» издания и сборники), долю самоцитирования и долю цитирования соавторами. Имитационные показатели были, таким образом, связаны с практиками, к которым могли бы прибегнуть беспринципные ученые, чтобы быстро улучшить свой наукометрический профиль – начать публиковаться за деньги, цитировать самого себя или обмениваться ссылками с теми коллегами, с кем сам ученый тесно связан.

Из 2468 авторов докторских диссертаций 2028 имеют профиль в платформе eLibrary.ru, что составляет 82,1% от всей выборки. При этом статистически значимой связи между отсутствием профиля и специальностью, и годом защиты диссертации выявлено не было. Из всех имеющих профиль самостоятельно зарегистрировались на сайте 82,79% докторов, профиль остальных автоматически сгенерирован системой. Для доли некорректных заимствований и показателей публикационной активности была рассчитана корреляция Спирмена. В целом сила связи между долей заимствования и показателями публикационной активности невелика. Для показателей публикаций и цитирований по «ядру» РИНЦ, так же как в переводных изданиях и доли статей в зарубежных журналах, высокие значения ассоциируются с меньшей долей некорректных заимствований в тексте диссертации, но даже для них абсолютная величина коэффициента корреляции не превышает 0,2. Таким образом, у нас нет оснований заявлять, что заимствования являются маркером «слабой науки».

На наблюдаемую связь могли, однако, повлиять различия в публикационных и цитатных практиках дисциплин. Чтобы нивелировать их, мы создали две полярные выборки докторов наук, в которые вошли 10% от представителей каждой дисциплины с минимумом заимствований и 10% тех, у кого обнаружился максимум (число представителей каждой дисциплины, таким образом, было строго пропорционально ее доле в общем числе защит). В выборках было собрано равное количество экономистов, психологов, химиков и т.д. Мы обозначили эти группы «плагиат» и «неплагиат».

Далее мы сфокусировались на публикационных показателях, которые позволяют противопоставить наши группы и при этом могут быть интерпретированы как показатели преимущественно качественной и/или имитационной деятельности ученого. Чтобы определить, существует ли разница в публикационной активности или же текст докторской диссертации не является определяющим в выбранной стратегии публикаций, мы сравнили поляризованные выборки докторов наук по указанным выше показателям (таблица 3). Статистическая значимость различий была рассчитана с помощью перестановочного теста при случайном назначении в группы «плагиат»/«неплагиат».

Таблица 3. Модельные оценки

| Переменная | Среднее | | Медиана | | Средний эффект воздействия (Δ) | Эффект воздействия по медиане (Δ) |
|-------------------------------|---------|-----------|---------|-----------|--------------------------------|-----------------------------------|
| | Плагиат | Неплагиат | Плагиат | Неплагиат | | |
| Качественные | | | | | | |
| Публикации ядро, % | 19,96 | 24,99 | 11,21 | 14,19 | -5,03* | -2,98 |
| Цитирования ядро, % | 17,79 | 24,29 | 9,09 | 9,68 | -6,50** | -0,59 |
| Импакт-фактор опубликованные | 0,37 | 0,45 | 0,28 | 0,32 | -0,07* | -0,04 |
| Импакт-фактор процитированные | 0,42 | 0,52 | 0,32 | 0,32 | -0,10* | 0 |
| Статьи зарубежные, % | 3,74 | 6,84 | 0 | 1,4 | -3,10*** | -1,4 |
| Цитирования зарубежные, % | 6,94 | 10,4 | 2,8 | 2,9 | -3,45** | -0,1 |
| Статьи ВАК, % | 50,39 | 46,51 | 50 | 47,1 | 3,88 | 2,9 |
| Имитационные | | | | | | |
| Публикации Elibrary, % | 5,40 | 4,35 | 2,26 | 2,13 | 1,05 | 0,13 |
| Цитирования Elibrary, % | 8,35 | 7,70 | 4,15 | 4,62 | 0,65 | -0,47 |
| Индекс Хирша Elibrary, % | 4,7 | 4,55 | 0 | 0 | 0,15 | 0 |
| Самоцитирования, % | 20,94 | 28,13 | 15,6 | 20,8 | -7,19*** | -5,2* |
| Цитирования соавторами, % | 36,7 | 40 | 36,7 | 40 | -3,3 | -3,3 |

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Статистически значимыми оказались различия между поляризованными выборками докторов наук по всем качественным показателям, кроме доли статей в журналах ВАК. Отсутствие различий в числе публикаций в журналах из списка ВАК, разумеется, может быть поводом усомниться в их качестве. Определенное количество статей в таких изданиях является обязательным для защиты докторской диссертации, и в этом смысле разница между более сильными и более слабыми учеными в нашей выборке может быть размыта. Группа имитационных показателей (за исключением большего самоцитирования в группе «неплагиат») в целом не показывает различий между докторами наук с высокой и низкой долей плагиата в диссертации. То, что более добросовестные ученые чаще цитируют сами себя, на первый взгляд выглядит парадоксом, но может показаться и вполне логичным: возможно, что они в принципе относятся к цитированию, включая цитирование собственных работ, аккуратнее. Это отражается и на числе ссылок на себя, и на доле некорректных заимствований. При этом если мы проводим аналогичный анализ, ориентируясь не на средние, а на медианные показатели, толь-

ко различие по показателю самоцитирования остается значимым. Этот факт свидетельствует о том, что среди собственноручно написавших диссертацию чаще встречаются ученые с исключительными высокими наукометрическими показателями, но за пределами самого верхнего слоя наличие плагиата в диссертации не позволяет определить уровень ученого.

В целом эти результаты можно считать опровержением *Гипотезы 1* и частичным подтверждением *Гипотезы 3*. До некоторой степени отсутствие заимствований объективно является сигналом, указывающим на способности индивида. В этом смысле норма текстуальной аутентичности может быть рационально обоснована со ссылкой на этот эффект: требование писать самостоятельно может способствовать отбору индивидов, которые затем будут публиковаться в лучших журналах. Эффекты, предсказанные *Гипотезой 3*, однако, оказываются очень слабыми. Контраст между полярными группами, заимствующими больше и меньше всего, достигает уровня значимости в 0,001 по единственному показателю – числу публикаций в иностранных журналах. Более того, те же наблюдения можно объяснить и с помощью альтернативных конвенций: ученые, знакомые с международными публикационными нормами с их более строгими правилами в отношении заимствований, следуют им и в публикациях на русском языке.

Выводы

В целом результаты нашего исследования лучше укладываются в представление об альтернативных конвенциях, допускающих заимствования, которые не превышают определенного предела, как приемлемые, а превышающие его – как девиацию¹². Диссертации, в которых грубо нарушаются эти нормы, хотя и встречаются, однако составляют незначительное меньшинство в общем объеме. Возможно, что в отношении этого небольшого числа заимствованных диссертаций действуют иные закономерности, отличающиеся от «нормального» плагиата, и этот факт объясняет несовпадение наших выводов с выводами «Диссернета» (например, в отношении распределения по дисциплинам).

Во-вторых, нормы, допускающие текстуальные заимствования, не изменились немедленно в результате внедрения «Антиплагиата». Это говорит против предположения о том, что заимствование чужого текста есть предмет рационального выбора стратегии нарушения норм. Если бы нарушение было делом утилитарного расчета, появление инструмента идентификации нарушителя должно было бы немедленно сократить число соответствующих случаев. Очевидно, что это не произошло.

В-третьих, объем заимствований отчетливо зависит от дисциплины, но, вопреки гипотезе о роли экспансии в разрушении норм научного этоса, предрасположенность к текстовым заимствованиям слабо коррелирует с тенденцией к расширению либо сжатию дисциплины или с положением дисциплины на шкале естественно-научный–гуманитарный (*hard-soft*).

¹² Цифра в 20% неоригинальности по «Антиплагиату» неоднократно упоминалась в разговорах авторов с секретарями диссертационных советов как тот объем, который диссовет считает безопасным. При этом, кажется, часто не проводилось различие между заимствованиями из собственных более ранних текстов и текстов других людей.

В-четвертых, на установки по отношению к заимствованиям, вероятно, повлияли тип и статус институции (ведущие университеты, институты РАН, обычные вузы) в виду контакта с «мировой культурой» с ее более жесткими стандартами, однако в целом связь является слабой, и предельные эффекты структурных переменных не превышают 5%.

В-пятых, мы нашли слабую, но значимую отрицательную связь уровня заимствований с некоторыми показателями академической силы (публикации в международных изданиях), которая сохраняется при контроле по типу институции. Эта связь может быть интерпретирована двояко: или в духе рационального выбора (те, кто может писать сам, не заимствует у других), или в духе диффузии норм (те, кто соприкасается с более строгими глобальными нормами, транслирует их дальше).

В целом гипотезы, связанные с нормативно-социализационной и рационально-структурными теориями, в ходе анализа поддержки не получили. Это может пониматься как аргумент в пользу теории альтернативных норм; было бы лучше, разумеется, если бы доводы в ее пользу не были преимущественно негативными (т.е. если бы мы обнаружили нечто, подтверждающее именно ее, а не опровергающее конкурирующие построения), однако на данный момент другого объяснения паттернам, которые мы наблюдаем, нет.

Отношение к заимствованиям частей чужого текста в российской науке, видимо, похоже на отношение к превышению скорости: большинство водителей считают, что есть допустимые превышения (в пределах 10–15 километров в час) и недопустимые (скажем, свыше 50 километров в час). Представления о допустимости «небольшого» и «технического» плагиата распространены повсеместно. Они разделяются в том числе вполне серьезными учеными, поэтому наличие заимствований в считающихся приемлемыми пределах не могут служить надежным сигналом, свидетельствующим о слабости. Этот факт в свою очередь служит стабилизации норм: видя, как уважаемые члены дисциплины позволяют себе использовать копи-пэйт, новички приходят к выводу, что и на них не падет тени позора, если их заметят в чем-то подобном. Нормы, более суровые к некорректным заимствованиям, возможно, распространяются в наиболее глобализированных учреждениях (сильных университетах и институтах РАН) в «мировых городах», однако этот процесс разворачивается очень медленно.

Литература

- Структура, объем и ключевые игроки диссеродельной индустрии в России. Доклад Вольного сетевого общества «Диссернет» (2018) // Диссернет. 17 января 2018 // https://www.dissernet.org/publications/struktura_disserodelnoy_industrii.htm
- Шахрай С.М., Аристер Н.И., Тедеев А.А. (2015) О плагиате в диссертациях на соискание ученой степени. М.: МИИ.
- Becker G.S. (1968) Crime and Punishment: An Economic Approach // The Economic Dimensions of Crime (eds. Fielding N., Clarke A., Witt R.), London: Palgrave Macmillan, pp. 13–58.
- Becker H.S. (2008) Outsiders, NY: Simon and Schuster.
- Honig B., Bedi A. (2012) The Fox in the Hen House: A Critical Examination of Plagiarism among Members of the Academy of Management // Academy of Management Learning & Education, vol. 11, no 1, pp. 101–123.
- Horbach S.S., Halfman W.W. (2017) The Extent and Causes of Academic Text Recycling or ‘Self-plagiarism’ // Research Policy, vol. 48, no 1, pp. 492–502.

- Merton R. (1973) *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago: University of Chicago Press.
- Spence M. (1973) Job Market Signaling // *The Quarterly Journal of Economics*, vol. 87, no 3, pp. 355–374.
- Stephans P. (1996) *The Economics of Science* // *Journal of Economic Literature*, vol. 34, no 3, pp. 1199–1235.
- Stiglitz J. (2000) The Contributions of the Economics of Information to Twentieth Century Economics // *The Quarterly Journal of Economics*, vol. 115, no 4, pp. 1441–1478.
- Thomas W. (1923) *The Unadjusted Girl: With Cases and Standpoint for Behavior Analysis*, Boston: Little, Brown and Company.
-
-

Plagiarism in Russian Doctoral Dissertations: How Much, Where, and in Whose Work?

A. MAKEEVA*, A. TSIVINSKAYA**, M. SOKOLOV***,
N. SOKOLOVA****, K. GUBA*****

***Alexandra Makeeva** – Assistant, Centre for Institutional Analysis of Science and Education, European University at Saint Petersburg. Address: 6/1a, Gagarinskaya St., Saint Petersburg, 191187, Russian Federation. E-mail: amakeeva@eu.spb.ru

****Angelika Tsivinskaya** – Junior Researcher, Centre for Institutional Analysis of Science and Education, European University at Saint Petersburg. Address: 6/1a, Gagarinskaya St., Saint Petersburg, 191187, Russian Federation. E-mail: atzivinskaya@eu.spb.ru

*****Mikhail Sokolov** – PhD in Sociology, Researcher, Centre for Institutional Analysis of Science and Education, European University at Saint Petersburg. Address: 6/1a, Gagarinskaya St., Saint Petersburg, 191187, Russian Federation. E-mail: msokolov@eu.spb.ru

******Nadezhda Sokolova** – Junior Researcher, Centre for Institutional Analysis of Science and Education, European University at Saint Petersburg. Address: 6/1a, Gagarinskaya St., Saint Petersburg, 191187, Russian Federation. E-mail: nsokolova@eu.spb.ru

*******Katerina Guba** – PhD in Sociology, Head, Centre for Institutional Analysis of Science and Education, European University at Saint Petersburg. Address: 6/1a, Gagarinskaya St., Saint Petersburg, 191187, Russian Federation. E-mail: kguba@eu.spb.ru

Citation: Makeeva A., Tsivinskaya A., Sokolov M., Sokolova N., Guba K. (2020) Plagiarism in Russian Doctoral Dissertations: How Much, Where, and in Whose Work? *Mir Rossii*, vol. 29, no 2, pp. 27–48 (in Russian). DOI: 10.17323/1811-038X-2020-29-2-27-48

Abstract

In recent years, plagiarism in the doctoral dissertations of politicians and other public figures has caused scandals in several countries. However, plagiarism is usually discovered by whistleblowers interested in exposing particular plagiarizers rather than seeking to uncover the whole picture of the scale and origins of plagiarism. In this paper, we fill this gap by analyzing a random sample of 2,468 Russian doctoral dissertations defended in 2006–2015 (9.3% of the total population of dissertations defended during this period) using plagiarism detection scripts. These data allow an estimation of the spread of plagiarism and test

some general hypotheses concerning the factors facilitating it. We evaluate three possible explanations: alternative conventions (groups of scholars differ in their understanding of what “plagiarism” is), anomie (plagiarism emerges from failing to embrace scientific ethos) and rational choice (plagiarism is likely when the expected costs are less than the benefits of transgressing a norm). We then evaluate the influence of a set of independent variables related to these hypotheses – discipline, defense year, region, the type of organization and the author’s bibliometric profile. According to our findings, the average and median amount of unauthorized borrowing in a doctoral dissertation is 19.1% and 13.9% respectively (with no more, than 6% of these figures attributable to “false positives” by the software mistakenly identifying plagiarism). Only a quarter of dissertations include less than 7% of incorrect borrowings. On the other hand, borrowings exceeding 50% are also relatively rare (in ca. 3,6% of doctoral dissertations). We argue that, overall, the pattern observed fits better with the theoretical model of alternative conventions. Variables associated with anomie (the rapid expansion of the discipline, evidence of imitating academic productivity by gaming the indicators) does not demonstrate significant correlations with the amount of plagiarism. Variables associated with the rational choice hypotheses (the academic fitness of the candidate and his/her institution, the spread of plagiarism-detecting software) demonstrate some, moderate, significance.

Key words: sociology of science, science in Russia, dissertations, plagiarism, academic dishonesty, academic misconduct, academic norms, automatic text analysis

References

- Becker G.S. (1968) Crime and Punishment: An Economic Approach. *The Economic Dimensions of Crime* (eds. Fielding N., Clarke A., Witt R.), London: Palgrave Macmillan, pp. 13–58.
- Becker H.S. (2008) *Outsiders*, NY: Simon and Schuster.
- Honig B., Bedi A. (2012) The Fox in the Hen House: A Critical Examination of Plagiarism among Members of the Academy of Management. *Academy of Management Learning & Education*, vol. 11, no 1, pp. 101–123.
- Horbach S.S., Halfman W.W. (2017) The Extent and Causes of Academic Text Recycling or ‘Self-plagiarism’. *Research Policy*, vol. 48, no 1, pp. 492–502.
- Merton R. (1973) *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago: University of Chicago Press.
- Spence M. (1973) Job Market Signaling. *The Quarterly Journal of Economics*, vol. 87, no 3, pp. 355–374.
- Stephans P. (1996) The Economics of Science. *Journal of Economic Literature*, vol. 34, no 3, pp. 1199–1235.
- Stiglitz J. (2000) The Contributions of the Economics of Information to Twentieth Century Economics. *The Quarterly Journal of Economics*, vol. 115, no 4, pp. 1441–1478.
- Struktura, objem i kluchevye igroki disserterdel’noj industrii v Rossii. Doklad Vol’nogo setevogo sobschestva «Dissernet» [Structure, Volume and Major Players of the Fake Dissertation Industry in Russia. Report by the Free Network Association “Dissernet”] (2018). *Dissernet*, January 17, 2018. Available at: https://www.dissernet.org/publications/struktura_disserterdelnoy_industrii.htm, accessed 14.02.2020.
- Shahraj S.M., Arister N.I., Tedeev A.A. (2015) *O plagiate v dissertatsiyakh na soiskanie uchenoj stepeni* [On Plagiarism in Dissertations], Moscow: MII.
- Thomas W. (1923) *The Unadjusted Girl. With Cases and Standpoint for Behavior Analysis*, Boston: Little, Brown and Company.