

М. Ю. КНЯЗЕВ

Институт лингвистических исследований РАН /
Санкт-Петербургский филиал Национального исследовательского университета
«Высшая школа экономики» / Санкт-Петербургский государственный университет
(Россия, Санкт-Петербург)
mknyazev@hse.ru

Е. А. РУДАЛЕВА

Санкт-Петербургский филиал Национального исследовательского университета
«Высшая школа экономики»
(Россия, Санкт-Петербург)
earudaleva@edu.hse.ru

КОРПУСНОЕ ИССЛЕДОВАНИЕ ВЛИЯНИЯ РЕГИСТРА НА ВЫБОР СПОСОБА ОФОРМЛЕНИЯ СЕНТЕНЦИАЛЬНОГО АКТАНТА В РУССКОМ ЯЗЫКЕ*

Регистр является важным фактором грамматической вариативности и может влиять на выбор одной из конструкций в ситуации конкуренции наряду с собственно грамматическими, семантико-прагматическими и прочими факторами. В данной статье на материале Национального корпуса русского языка исследуется влияние регистра на выбор стратегии оформления сентенциального актанта в индикативе (изъяснительном придаточном с союзом *что* или конструкцией с местоимением-коррелятом *то*) при глаголах, имеющих валентность на косвенное дополнение, ср. *Она настаивает (на том), что...* В качестве основной гипотезы рассматривается связь признака формальности регистра и предпочтения конструкции с коррелятом. Сравниваются пять регистров (академический, публицистический и художественный регистры письменного модуса и публицистический и непубличный регистры устного модуса) с помощью попарных плановых сравнений. Влияние регистра рассматривается тройко: а) по всей выборке (из 26 глаголов), б) по отдельным глаголам и в) по классам глаголов, выделяемых в зависимости от типа предпочтения одной (ни одной) из стратегий. Был использован «коллокационный анализ» (разработанный Ст. Грисом и А. Стефановичем), а также анализ дисперсии методом «отклонения пропорций» (разработанный Ст. Грисом). Результаты показывают небольшой рост предпочтения конструкции с коррелятом в академическом и нехудожественном регистрах (по сравнению с публицистическим и художественным регистрами) для всей выборки и для двух (из трех) классов притяжения, что в целом

* Представленное в данной статье исследование было выполнено при поддержке гранта РФФИ № 16-18-02003 «Структура значения и его отображение в системе лексических и функциональных категорий русского языка», реализованном в Московском педагогическом государственном университете (МПГУ). Авторы благодарят двух анонимных рецензентов за ценные советы и замечания.

подтверждает основную гипотезу. Влияние регистра на отдельные глаголы (операционализированное как смена класса притяжения при смене регистра), однако, незначительное. Также выявлено интересное отличие класса глаголов, не притягивающихся ни к одной из конструкций, от двух остальных классов, которое может указывать на влияние другого признака регистра, нежели формальности.

Ключевые слова: регистр, стратегии оформления сентенциального актанта, корпусные подходы, коллокационный анализ, анализ дисперсии, конкуренция.

1. Введение

В современной корпусной лингвистике принимается как данность тот факт, что тип дискурса является одним из важных параметров лингвистической вариативности [Biber 2012]. Это касается не только противопоставления между устным и письменным дискурсом, но и более тонких различий по сфере функционирования дискурса, традиционно связанных с понятием функционального стиля (ф-стиля), а также различий, обусловленных выбором речевого жанра и степенью формальности дискурса [Кибрик 2009]. В дальнейшем изложении данную группу факторов, следуя англоязычной традиции, мы будем достаточно условно называть термином «регистр» (register), используя его как наиболее общее обозначение типа дискурса, выделяемого в зависимости от конкретной ситуации его использования [Biber, Conrad 2009: 6]; для классификации дискурсов по (письменному/устному) каналу передачи информации мы будем использовать термин «модус» вслед за работой [Кибрик 2009]¹.

Хотя общепризнано, что регистр влияет на выбор лексико-грамматических средств, роль регистра в описании конкретных грамматических явлений пока представляется достаточно периферийной и уступает не только собственно грамматическим, но и семантико-прагматическим факторам, а также факторам, связанным с порождением и восприятием речи (таким как сложность обработки предложения, частотность и др.). Такое положение дел вполне оправданно при исследовании «жестких», невариативных грамматических ограничений в языке. Однако изучение более «мягких», вариативных ограничений, которые все чаще становятся предметом лингвистического описания, в большей степени требует учета роли регистра, поскольку в таких случаях последний может оказаться важным предиктором частотности/приемлемости рассматриваемых структур. Особенно важным это становится в квантитативных исследованиях (как корпусных,

¹ Регистр, как правило, подразумевает определенный модус (письменный или устный) и противопоставляется *жанру* (отражающему характеристики, связанные с типичной риторической структурой данного типа дискурса) и *стилю* (отражающему эстетические характеристики данного типа дискурса), подробнее см. [Кибрик 2009; Biber, Conrad 2009].

так и экспериментальных), где без систематического контроля регистра выявленные грамматические ограничения могут оказаться артефактом исследования, а не реальным грамматическим фактом.

Дистрибуция сентенциальных актантов в русском языке является одной из тех областей грамматики, в которой грамматические ограничения имеют нежесткий характер. Многие «запреты», приводимые в работах по русским сентенциальным актантам, либо эксплицитно помечены как неабсолютные, т. е. выявляемые только в контрасте с альтернативными стратегиями (см., например, обсуждение влияния неагентивности подлежащего на выбор сентенциального актанта [Пекелис 2014; Князев 2018]), либо оспариваются в литературе (см., в частности, обсуждение конструкции с плеонастическим *это* в [Летучий 2014], ср. [Циммерлинг 2012]). В такой ситуации регистр может легко оказаться одним из факторов вариативности, контроль которого необходим для корректной интерпретации результатов, однако для большинства исследований в этой области определение роли регистра пока остается задачей на будущее.

Прежде чем начать учитывать роль регистра в описании более сложных ограничений в области сентенциальных актантов (т. е. предполагающих взаимодействие нескольких грамматических факторов), полезно иметь представление о том, как регистр влияет на выбор актанта в более простых случаях, например в ситуации «идеальной» конкуренции двух стратегий оформления актанта. Соответственно, проблема, которую мы далее рассмотрим, будет касаться влияния типа модуса/регистра на способ оформления индикативного сентенциального актанта при глаголах речемыслительной деятельности.

Хорошо известно, что в русском языке такой сентенциальный актант может быть выражен двумя способами:

- а) непосредственно придаточным с союзом *что* (далее *что*-клаузы);
- б) конструкцией, включающей соотносительное слово (иногда называемое «коррелятом») с требуемым предложно-падежным оформлением и соответствующее придаточное (далее *то, что*-клаузы).

Каждая из этих конструкций имеет свои характерные дистрибутивные ограничения. *То, что*-клаузы в нейтральном контексте избегаются в позиции, соотносимой с позицией прямого дополнения, например при *говорить, знать* и др., как показано в (1a), см. [Шведова 1980: 483], а также при отдельных глаголах, таких как *считать, полагать, казаться* и др., подробнее см. [Кобозева 2013]². *Что*-клаузы недопустимы в позиции предглагольного подлежащего, как показано в (1b), см. [Hartman 2012], а также с глаголами, которые в силу словарных свойств требуют выра-

² В разговорной речи в нестандартных вариантах русского языка сочетание *то что* может выступать в качестве союза в этих (и других) контекстах, однако в таких случаях оно имеет особые фонетические и семантико-прагматические свойства, подробнее см. [Коротаев 2013].

женного коррелята, как показано в (1с), см. [Валгина 2003: 311; Кобозева 2013]³.

Существует и зона конкуренции, в которой возможны обе конструкции. К этой зоне, например, относится позиция послеглагольного подлежащего, как в (2а), а также позиция, соотносимая с позицией косвенного дополнения, при таких глаголах, как *надеяться*, *настаивать* и т. д., см. (2b)–(2с).

- (1) а. *Она говорит/знает (*то), что он неправ.*
 б. **(То,) что он опоздал, меня очень расстроило.*
 с. *Он начал *(с того), что поприветствовал гостей.*
- (2) а. *Меня очень расстроило (то), что он опоздал.*
 б. *Она надеется (на то), что он придет.*
 с. *Я настаиваю (на том), что они правы.*

В литературе поднимался вопрос о факторах, влияющих на предпочтение одной из конструкций в контекстах типа (2). Насколько нам известно, среди таких факторов указывались исключительно семантико-прагматические. В частности, отмечалось, что есть тенденция к выражению коррелята, если сентенциальный актант контрастивно выделен, а также если он обозначает информацию, данную в предтексте или в общем фонде знаний о мире (см. [Шведова 1980: 483; Валгина 2003: 311; Кобозева 2013]).

Вопрос о влиянии модуса/регистра на выбор типа сентенциального актанта, по нашим сведениям, ранее систематически не рассматривался. Однако есть косвенные свидетельства в пользу того, что такое влияние может иметь место. Так, одному из авторов при проведении опросов на вынесение суждений о приемлемости предложений со *что*- и *то, что*-клаузами приходилось сталкиваться с уточняющими расспросами носителей о том, в разговорной или письменной речи употреблено данное предложение. Более того, носители иногда указывали на «тяжеловесность» предложений с коррелятом при некоторых глаголах, например при *надеяться*, отмечая их «книжный» характер, ср. (2b). В других случаях, в частности при *настаивать*, носители иногда отмечали «разговорный» оттенок конструкций без коррелята, ср. (2с). Отражают ли подобные комментарии реальные различия между регистрами для соответствующих комбинаций глагола и конструкции, заранее сказать трудно. Однако мы исходили из того, что подоб-

³ *Что*-клаузы, как правило, также избегаются в позиции топики, как показано в (i), однако это ограничение кажется менее жестким и зависит от других факторов, таких как выбор глагола, тип синтаксической позиции, с которым соотносится позиция топики, наличие отрицания при главном глаголе и резюмтивного местоимения в позиции топики (подробнее см. [Кобозева 2013], а также [Knyazev 2016]).

- (i) а. *?(То,) что он опоздает, я знаю.*
 б. *??? (О том,) что мы проиграли, я не сожалею.*
 с. **Что я ему завидую, он решил.*
 д. *Что это обязательно произойдет, я *(так) не считаю.*

ные интуиции имеют под собой некую основу, и это позволяет рассматривать их в качестве рабочей гипотезы.

В качестве основного признака, различающего разные типы регистров и ассоциированных с ними дискурсов, мы взяли *формальность* как характеристику дискурса, связанную с социальной дистанцией между говорящим и слушающим [Кибрик 2009: 14]. Строго говоря, формальность является параметром, независимым от регистра, поскольку один и тот же тип регистра (по крайней мере, когда он задает максимально широкий класс дискурсов и таким образом приближается к традиционному понятию функционального стиля) может различаться по степени формальности (ср. научный регистр в рецензируемой журнальной статье и в конспекте этой же статьи). Тем не менее очевидно, что признак формальности различает не только типы дискурсов внутри данного регистра, но и регистры между собой, что позволяет говорить о более и менее формальных регистрах в целом (например, научный и публицистический регистры).

Важной характеристикой, связанной с признаком формальности дискурса, является степень фонетической редукции, к которой сильнее тяготеют менее формальные регистры [Там же]. Поскольку выбор *что*-клауз в позиции конкуренции может быть обусловлен стремлением к фонетической редукции, мы предположили, что в менее формальных регистрах, т. е. в тех, которые ближе к спонтанной устной речи, частотность *что*-клауз в сравнении с *то*, *что*-клаузами, будет выше, чем в более формальных регистрах. В обобщенном виде это можно сформулировать в (3).

(3) Основная гипотеза

В более формальных регистрах частотность *то*, *что*-клауз (в сравнении со *что*-клаузами) будет выше, чем в менее формальных регистрах.

Для проверки гипотезы (3) мы провели исследование на материале (основного подкорпуса) Национального корпуса русского языка (далее — НКРЯ) на выборке из 26 глаголов. Мы проверяли не только влияние регистра на выбор типа сентенциального актанта в целом, но и то, как регистр влияет на этот выбор для каждого конкретного глагола. Также мы проверяли, насколько влияние регистра будет зависеть от класса глагола, выделяемого в зависимости от общего предпочтения *что*- или *то*, *что*-клауз для данного глагола. Ключевым моментом для нас было выяснить не столько то, имеется ли влияние регистра вообще, сколько то, как именно влияет регистр и от каких факторов зависит само это влияние. Так, в качестве одного из важных результатов статьи мы обнаружили различия во влиянии регистра на глаголы с разным классом предпочтения относительно выбора типа клаузы.

В нашем исследовании мы опирались на различную методологию, предлагаемую в литературе для анализа влияния регистра, в первую очередь на так называемый «коллострукционный анализ» (collostructional analysis), см. [Gries, Stefanowitsch 2004; Stefanowitsch, Gries 2008; Кузнецо-

ва, Велейшикова 2010], а также на метод отклонения пропорций (Deviation of Proportions) для вычисления дисперсии [Gries 2008; Lijffijt, Gries 2012].

Статья структурирована следующим образом. В разделе 2 обсуждается выбор сравниваемых регистров и глаголов для дальнейшего анализа (раздел 2.1) и приводятся результаты критерия независимости хи-квадрат для факторов регистра и типа клаузы (раздел 2.2). В разделе 3 приводятся данные коллострукционного анализа, вначале для всей выборки глаголов (раздел 3.1), затем для глаголов по отдельности (раздел 3.2) и наконец для глаголов по классам притяжения (раздел 3.3). В разделе 4 обсуждается анализ дисперсии, уточняющий результаты, полученные в разделе 3. Раздел 5 содержит заключение.

2. Тесты на независимость: общее влияние регистра

2.1. Выбор сравниваемых регистров и глаголов для анализа

Для последующих сравнений мы использовали три основных письменных регистра (академический, публицистический и художественный), которые обычно фигурируют в подобных сравнениях. Эти регистры были отобраны с помощью соответствующей разметки основного подкорпуса НКРЯ, как показано в (4а). Также мы отобрали два устных регистра (публичный и непубличный), опираясь на разметку устного подкорпуса НКРЯ, как показано в (4б).

(4) а. Письменный модус

- i. Академический регистр (нехудожественные тексты учебно-научной сферы функционирования основного подкорпуса НКРЯ)
- ii. Публицистический регистр (нехудожественные тексты публицистической сферы функционирования основного подкорпуса НКРЯ)
- iii. Художественный регистр (художественные тексты основного подкорпуса НКРЯ)

б. Устный модус

- i. Публичный регистр (публичные тексты устного подкорпуса НКРЯ)
- ii. Непубличный регистр (непубличные тексты устного подкорпуса НКРЯ)

Поскольку наше исследование имело пилотный характер, при выборе сравниваемых регистров мы воспользовались заведомо очень грубой классификацией, не учитывая более дробные различия по типу (жанру) текста и тематике, доступные в разметке соответствующих подкорпусов НКРЯ. Также мы не учитывали диахронический фактор. Как справедливо отмечают рецензенты, оба решения могли отрицательно сказаться на результатах. В частности, публицистический регистр, операционализированный нами как в (4aii), включает тексты самых разнообразных типов и жанров, в том числе демонстрирующих определенную близость к художественной лите-

ратуре (например, мемуары). В итоге потенциальные различия между регистрами могли оказаться смазанными. В связи с этим для отбора публицистических текстов более оправдан, по-видимому, был бы выбор текстов газетного подкорпуса НКРЯ, как советует рецензент. Что касается диахронического фактора, то неучитывание года создания текста также потенциально могло исказить результаты, поскольку тексты разных регистров в НКРЯ не полностью сбалансированы по годам.

Принимая во внимание эти замечания, которые, безусловно, должны быть учтены в дальнейшем, надо отметить, что основным фокусом нашего исследования было не столько влияние регистра как таковое, сколько взаимодействие этого влияния с другими факторами — выбором глагольной лексики и класса глагола по предпочтению типа клаузы (иначе говоря, нас интересовали не только различия между регистрами, но и различия между глаголами и их классами). При такой постановке вопроса способ операционализации самого регистра может быть не столь существенным. В любом случае, мы надеемся, что недостаток контролируемости сторонних факторов в нашем исследовании не повлиял на его основные выводы⁴.

Вернемся к отобранному нами списку регистров в (4). Поскольку дальнейший анализ предполагал попарные сравнения, приведенные выше пять регистров были разбиты на четыре контрастные пары (по аналогии с методом ортогональных контрастов для регрессионного анализа, см. [Levshina 2015: 185–186]), как показано в (5), в каждой из которых был определен более формальный модус/регистр (что обозначено знаком «>»). Помимо очевидных контрастов между письменным и устным модусами в (5а) и публичным и непубличным регистрами устного модуса в (5г), были получены два других контраста путем деления письменного модуса на художественный и нехудожественный регистры (5б) и последнего на академический и публицистический (5в).

- (5) а. Письменный модус > Устный модус
- б. Нехудожественный регистр (Академический и Публицистический) > Художественный регистр
- в. Академический регистр > Публицистический регистр
- г. Публичный регистр (Устного модуса) > Непубличный регистр (Устного модуса)

Такое разбиение письменного модуса представляется наиболее естественным, поскольку оно опирается на существующую разметку регистров в НКРЯ (ср. категорию «нехудожественные тексты»), а также на представление о том, что в художественных текстах могут присутствовать элементы спонтанной устной речи за счет передачи диалогов, что сближает их с

⁴ В дальнейшем мы планируем проверить влияние регистра на выбор типа клаузы в контролируемом эксперименте. См. попытку сходного исследования в [Кибрик 2009: 16–18].

устными регистрами (см. [Biber, Conrad 2009: 135]). Мы также исходили из того, что академический регистр является более формальным, чем публицистический, т. к. именно он обычно выступает в роли прототипического формального регистра.

В качестве позиции конкуренции для *что*- и *то*, *что*-клауз была выбрана позиция сентенциального актанта глагола, соотносимая с позицией косвенного дополнения, см. (2b)–(2с). Составление выборки осуществлялось в два этапа. На первом этапе отбирались глаголы, допускающие оба типа клауз. При отборе использовались триграммы вида «V + *то* + *что*» и тетраграммы вида «V + Pr + *то* + *что*» (для глаголов с беспредложными и предложными дополнениями соответственно); далее полученные при выдаче глаголы проверялись на встречаемость со *что*-клаузой. При этом отбирались только такие глаголы, которые встречались в основном корпусе НКРЯ не менее 50 раз суммарно с обоими типами клауз. В результате было отобрано 26 глаголов, см. Таблицы 4 и 5 в разделе 3.1. Для каждого из глаголов было подсчитано количество примеров с обоими типами клауз в каждом из пяти регистров. Получившийся подкорпус использовался для всех последующих статистик.

2.2. Результаты тестов на независимость

Для первоначальной оценки влияния регистра были агрегированы данные по всем рассматриваемым глаголам и проведены тесты на независимость для факторов регистра и типа клаузы. В Таблице 1 представлена частотность каждого из двух типов клауз по пяти рассматриваемым регистрам. Как можно видеть, в среднем по выборке *что*-клаузы составляют 79% от всех употреблений, *то*, *что*-клаузы — 21%. Также обнаружилось, что частотность *что*-клауз значительно различается в зависимости от регистра ($\chi^2(1) = 1594$, $p < 0.001$). Так, максимальная частотность *что*-клауз наблюдается в художественном регистре (86%), а минимальная — в академическом (60%). Частотность *что*-клауз в остальных регистрах находится между этими двумя отметками (74–77%). Для наглядности соответствующие частотности приводятся в диаграмме на рисунке 1.

Попарные плановые сравнения выявили значимую связь между регистром и типом клаузы для контраста между художественным и нехудожественным регистрами и для контраста между публицистическим и академическим регистрами, как показано в Таблице 2 (соответствующие контрасты выделены заливкой). Эту связь можно охарактеризовать следующим образом: в нехудожественном и академическом регистрах шансы употребления *то*, *что*-клаузы больше, чем в художественном и публицистическом регистрах соответственно. На основе меры V Крамера (Cramer's V , см. [Levshina 2015: 203–210]) величину эффекта в этих случаях можно оценить в 0.14 и 0.16, что указывает на **слабый эффект**. Напротив, между устным и письменным модусами и публичным и непубличным регистрами устного модуса значимых различий не обнаружено.

Таблица 1

**Наблюдаемая (и ожидаемая) частотность *что*- и *то, что*-клауз
в зависимости от модуса/регистра**

Регистр	<i>Что</i> -Клауза		<i>То, что</i> -Клауза	
Академический	3025 (3961)	60%	2006 (1070)	40%
Публицистический	15 361 (15 651)	77%	4519 (4229)	23%
Художественный	15 209 (13 959)	86%	2522 (3772)	14%
Устный (публичный)	863 (879)	77%	253 (237)	23%
Устный (непубличный)	125 (133)	74%	44 (36)	26%
Итого	34 583	79%	9344	21%

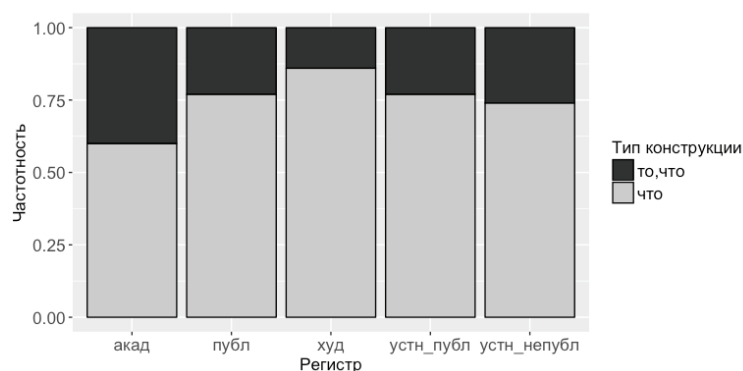
Рис. 1. Процент *что*- и *то, что*-клауз в зависимости от модуса/регистра

Таблица 2

Результаты плановых сравнений между модусами/регистрами

Устный vs. Письменный	$\chi^2 (1) = 2.57, p = 0.11$
Художественный vs. Нехудожественный	$\chi^2 (1) = 887.1, p < 0.001,$ Cramer's $V = 0.14$
Публицистический vs. Академический	$\chi^2 (1) = 609.3, p < 0.001,$ Cramer's $V = 0.16$
Устный (непубличный) vs. Устный (публичный)	$\chi^2 (1) = 0.76, p = 0.38$

Таким образом, для двух контрастов из четырех мы обнаружили связь в направлении, предсказанном гипотезой в (3). Отсутствие в результатах значимой связи для контраста между публичным и непубличным регистрами устного модуса оказалось несколько неожиданным. Однако можно предположить, что различие не было обнаружено в силу значительно меньшего количества данных для устного модуса. Что касается контраста между устным и письменным модусами, то на необнаружение связи в этом случае, по-видимому, повлияло то, что в составе обоих модусов имеются как более, так и менее формальные регистры. Кроме того, публичный и непубличный регистры устного модуса оказались по частотности *что*-клауз (77 и 77%) схожи друг с другом и одновременно с публицистическим регистром письменного модуса (77%), который, в свою очередь, схож с письменным модусом в среднем (79%)⁵.

С учетом этих соображений, полученные результаты в целом подтверждают гипотезу (3), однако не объясняют, за счет чего возникает влияние регистра на выбор типа клаузы. Так, например, теоретически наблюдаемый эффект может возникать как благодаря влиянию регистра на отдельные глаголы или группы глаголов (наиболее подверженные этому влиянию), так и благодаря влиянию на все глаголы равномерно. Для ответа на данный вопрос мы воспользовались так называемым коллострукционным анализом (точнее, его разновидностью, предлагаемой для исследования влияния регистра), см. [Stefanowitsch, Gries 2008].

3. Влияние регистра на отдельные глаголы и классы глаголов

3.1. Метод дистинктивных коллексем

Для анализа влияния регистра на отдельные глаголы была использована процедура, предлагаемая в работе [Ibid.]. Эта процедура основана на методе *дистинктивных коллексем* (разновидности коллострукционного анализа), предназначенном для семантической дифференциации конкурирующих конструкций, например «конструкции с двойным объектом» (ср. *Mary gave John a book*) и «конструкции с предложным дативом» (ср. *John gave a book to Mary*); подробнее об этом методе см. [Gries, Stefanowitsch 2004; Кузнецова, Велейшикова 2010; Levshina 2015: 223–239]⁶. Поскольку мы также имели дело с конкурирующими конструкциями (*что*- vs. *то*, *что*-клаузы), этот метод мог быть непосредственно перенесен и на наш материал. Вна-

⁵ Данный результат можно считать следствием не вполне удачного отбора регистров для сравнения. Возможно, более правильным решением для будущих исследований было бы исключение публичного регистра устного модуса из числа рассматриваемых регистров.

⁶ О современном состоянии коллострукционного анализа см. [Gries 2015a], о критике этого подхода и альтернативах см. [Schmid, Küchenhoff 2013].

чале мы рассмотрим, как применяется этот метод, а затем — как он применяется для анализа влияния регистра.

Суть метода заключается в том, чтобы на основе статистического анализа составить список (глагольных) лексем, которые встречаются с каждой из конструкций значимо чаще, чем ожидалось бы, если бы выбор лексемы и тип конструкции были независимы друг от друга (далее исходя из семантического анализа каждого из списков выводится семантический портрет соответствующей конструкции). В основе данного метода лежит привычный тест на независимость для таблиц сопряженности 2×2 , составленных для каждой из рассматриваемых лексем. В качестве одного из признаков в такой таблице выступает тип конструкции, а в качестве другого — выбор данной лексемы или всех остальных лексем. Пример соответствующей таблицы для лексемы *признаться* (одного из 26 отобранных глаголов) приведен в Таблице 3.

Таблица 3

Частотность *что-* и *то, что-*клауз для глагола *признаться*

	<i>признаться</i>	Остальные глаголы	Итого
<i>Что</i> -Клауза	3580 (2942)	31 003 (31641)	34 583
<i>То, что</i> -Клауза	157 (795)	9187 (8549)	9344
Итого	3737	40 190	

Далее для каждой таблицы вычисляется *p*-значение точного критерия Фишера. Если *p*-значение меньше 0.05, то можно говорить о (значимом) *притяжении* данной лексемы к одной из конструкций; при этом «притягивающая» конструкция определяется в зависимости от того, для какой из конструкций наблюдаемая частотность превышает ожидаемую (дана в скобках в Таблице 3). Далее *p*-значения трансформируются в отрицательный десятичный логарифм, служащий мерой притяжения лексемы к конструкции. В результате можно составить списки лексем, которые наиболее сильно притягиваются к каждой из конструкций.

Результаты применения метода дистинктивных коллексем к выборке глаголов приведены в Таблицах 4 и 5. Из 26 глаголов шесть не проявили притяжения ни к одной из конструкций, см. Таблицу 5. Отметим, что данные лексемы чаще встречались со *что*-клаузами, однако частотность последних при этих глаголах (указана в скобках после запятой) значимо не превышала среднюю частотность *что*-клауз (79%, см. раздел 2.2). Остальные 20 глаголов поделились на две группы: девять глаголов показали притяжение к *что*-клаузе и 11 глаголов — к *то, что*-клаузе, как видно из Таблицы 4. Следует также отметить, что поскольку *p*-значение зависит от размера выборки, позиция глагола в списке не отражает напрямую частотность *что*-клауз.

Таблица 4

**Коллексемы, притягивающиеся к *что*- или *то, что*-клаузам:
 частотность *что*- и *то, что*-клауз (в скобках перед запятой),
 процент *что*-клауз (в скобках после запятой), *p*-значение
 точного критерия Фишера и его отрицательный десятичный логарифм**

Притягиваются к <i>Что</i> -Клаузе		
Коллексема	p	$-\log_{10} p$
<i>надеяться</i> (10941:373, 96%)	0	Inf
<i>признаться</i> (3580:157, 96%)	2.50E-209	208.6
<i>сознаться</i> (1700:74, 96%)	2.31E-97	96.6
<i>жалеть</i> (2007:185, 92%)	1.37E-61	60.9
<i>испугаться</i> (901:55, 94%)	2.06E-41	40.7
<i>жаловаться</i> (1782:263, 87%)	1.02E-23	23
<i>рассчитывать</i> (1501:346, 81%)	6.31E-03	2.2
<i>злиться</i> (128:20, 86%)	2.03E-02	1.7
<i>сомневаться</i> (3159:815, 79%)	2.23E-02	0.7

Притягиваются к <i>То, что</i> -Клаузе		
Коллексема	p	$-\log_{10} p$
<i>гордиться</i> (462:1037, 31%)	0	Inf
<i>свидетельствовать</i> (1563:1863, 46%)	0	Inf
<i>добиться</i> (128:398, 24%)	1.32E-158	157.9
<i>настаивать</i> (432:607, 42%)	2.46E-153	152.6
<i>намекать</i> (492:471, 51%)	3.71E-82	81.4
<i>указывать</i> (2042:1123, 65%)	1.33E-81	80.9
<i>привыкнуть</i> (668:503, 57%)	3.97E-64	63.4
<i>возмущаться</i> (121:108, 53%)	3.32E-18	17.5
<i>сетовать</i> (216:142, 60%)	2.21E-15	14.7
<i>стыдиться</i> (116:82, 59%)	1.44E-10	9.8
<i>уповать</i> (114:56, 67%)	4.58E-4	3.3

Таблица 5

**Коллексемы, не притягивающиеся ни к одному из типов клауз:
частотность *что-* и *то, что*-клауз (в скобках перед запятой),
процент *что*-клауз (в скобках после запятой),
p-значение точного критерия Фишера
и его отрицательный десятичный логарифм**

Без притяжения		
Коллексема	<i>p</i>	$-\log_{10} p$
<i>радоваться</i> (1819:475, 79%)	5.12E-02	0.3
<i>извиниться</i> (192:47, 80%)	5.80E-02	0.2
<i>обидеться</i> (196:49, 80%)	6.95E-02	0.2
<i>хвастаться</i> (221:57, 79%)	8.25E-02	0.1
<i>дождаться</i> (66:24, 73%)	2.45E-02	0.6
<i>расстраиваться</i> (36:14, 72%)	2.31E-02	0.6

Результаты применения метода дистинктивных коллексем показывают, что глаголы неравномерно распределены по двум конкурирующим стратегиям оформления сентенциального актанта, поскольку часть глаголов притягивается к одной из конструкций, а часть — к другой. Таким образом, глагольная лексема — важный фактор для выбора типа клаузы. Наша задача, однако, состоит в том, чтобы понять, является ли таким фактором также регистр и, если да, то варьируется ли его влияние в зависимости от выбора лексемы. Чтобы ответить на этот вопрос, мы применили метод дистинктивных коллексем для анализа влияния регистра, см. [Stefanowitsch, Gries 2008].

3.2. Метод дистинктивных коллексем для анализа регистра

Чтобы проанализировать влияние регистра/модуса на выбор конкурирующих конструкций в работе [Stefanowitsch, Gries 2008], предлагается следующий метод. Для каждого из двух сравниваемых регистров/модусов создается таблица коллексем, притягивающихся к одной из двух конкурирующих конструкций, далее таблицы проверяются на предмет «эффектов перекреста», т. е. наличия лексем, которые в противоположных регистрах притягиваются к противоположным конструкциям.

Данный метод был последовательно применен к анализу каждого из четырех контрастов. Исходя из гипотезы в (3), мы ожидали, что направление изменения будет состоять в появлении притяжения к *то, что*-клаузе в более формальном регистре из пары. Результаты сведены в Таблицу 6, в которой для каждого из 26 глаголов показано, к какой из конструкций (или ни к одной из двух) этот глагол притягивается в целом по

корпусу и в сравниваемых регистрах внутри каждой из пар, а также позиция глагола в списке (исходные данные приведены в Приложении в Таблицах 10–13).

Глядя на Таблицу 6, можно сделать следующие выводы. Только четыре глагола из 26 показали эффекты перекреста, а именно: *рассчитывать*, *сомневаться*, *уповать*, *радоваться*. Отметим, что все эти глаголы имеют довольно низкие позиции в соответствующих списках с относительно невысокой степенью значимости притяжения к конструкциям. Кроме того, лишь один глагол (*сомневаться*) показывает перекрест в направлении, предсказанном гипотезой (3), — появление притяжения к *то, что*-клаузе при переходе от публицистического к академическому регистру. Более того, это притяжение в обоих случаях находится почти на границе уровня значимости (см. Таблицу 12). Можно также видеть, что стабильно не только притяжение глаголов к одной или другой конструкции в разных регистрах, но и степень этого притяжения (операционализованную как позицию в списке), что видно по очень слабому варьированию позиций по регистрам. На основе этих данных можно заключить, что данный тест в целом не показал влияния регистра на отдельные глаголы (ср. сходные результаты для других конструкций в [Stefanowitsch, Gries 2008]).

Таким образом, хотя и было обнаружено влияние регистра на выбор типа клаузы в целом, заметного влияния регистра на отдельные глаголы не выявлено. Данный вывод, однако, не противоречит гипотезе (3), поскольку сами авторы метода, предложенного в [Ibid.], объясняя необнаружение эффектов регистра для рассматриваемых ими конструкций, замечают, что этот метод, вероятно, предлагает излишне жесткое понимание эффекта регистра. В целом отсутствие «эффектов перекреста» в результатах согласуется со слабой величиной эффекта регистра (см. раздел 2.3).

3.3. Тесты на независимость по классам притяжения глаголов

Не обнаружив значительного влияния регистра на отдельные глаголы, мы решили проверить, будет ли регистр по-разному влиять на классы глаголов, полученные путем применения метода дистинктивных коллексем для всей выборки (см. раздел 3.1). Для этой цели мы провели стандартные тесты на независимость для каждого из трех классов глаголов, выделяемых исходя из наличия у глагола притяжения к одному из типов клауз. Результаты этих тестов приведены в Таблице 7 (см. исходные данные в Таблице 14 в Приложении). Для каждой из контрастных пар регистров/модусов в таблице указаны наблюдаемые (и ожидаемые) частотности *что*-клауз, позволяющие увидеть направление влияния регистра, а также *p*-значения и *V* Крамера, отражающие наличие значимой связи между типом регистра и типом клаузы, а также величину соответствующего эффекта.

Таблица 6

Сводная таблица притяжения глаголов к *что*- и *то*, *что*-клаузам по регистрам

	Класс	Устный	Письм.	Худож.	Нехуд.	Публиц.	Академ.	Публич.	Непубл.
<i>надеяться</i>	<i>что</i> (1)	<i>что</i> (1)	<i>что</i> (1)	<i>что</i> (1)	<i>что</i> (1)	<i>что</i> (1)	<i>что</i> (1)	<i>что</i> (1)	<i>что</i> (1)
<i>признаться</i>	<i>что</i> (2)	—	<i>что</i> (2)	<i>что</i> (2)	<i>что</i> (2)	<i>что</i> (2)	<i>что</i> (2)	—	—
<i>сознаться</i>	<i>что</i> (3)	—	<i>что</i> (3)	<i>что</i> (3)	<i>что</i> (3)	<i>что</i> (3)	<i>что</i> (3)	—	—
<i>жалеть</i>	<i>что</i> (4)	—	<i>что</i> (4)	<i>что</i> (4)	<i>что</i> (4)	<i>что</i> (4)	<i>что</i> (5)	<i>что</i> (2)	—
<i>испугаться</i>	<i>что</i> (5)	—	<i>что</i> (5)	<i>что</i> (5)	<i>что</i> (6)	<i>что</i> (5)	<i>что</i> (9)	—	—
<i>жаловаться</i>	<i>что</i> (6)	—	<i>что</i> (6)	<i>что</i> (6)	<i>что</i> (5)	<i>что</i> (6)	<i>что</i> (4)	—	—
<i>рассчитывать</i>	<i>что</i> (7)	<i>то, что</i> (6)	<i>что</i> (7)	—	<i>что</i> (7)	—	<i>что</i> (6)	—	<i>то, что</i> (6)
<i>злиться</i>	<i>что</i> (8)	—	<i>что</i> (8)	—	—	—	—	—	—
<i>сомневаться</i>	<i>что</i> (9)	—	—	—	—	<i>что</i> (7)	<i>то, что</i> (6)	—	—
<i>гордиться</i>	<i>то, что</i> (1–2)	<i>то, что</i> (3)	<i>то, что</i> (1–2)	<i>то, что</i> (1)	<i>то, что</i> (2)	<i>то, что</i> (2)	<i>то, что</i> (1)	<i>то, что</i> (1)	<i>то, что</i> (3)
<i>свидетельствовать</i>	<i>то, что</i> (1–2)	<i>то, что</i> (1)	<i>то, что</i> (1–2)	<i>то, что</i> (3)	<i>то, что</i> (1)	<i>то, что</i> (1)	<i>то, что</i> (2)	—	<i>то, что</i> (1)

Таблица 7

**Наблюдаемая (ожидаемая) частотность *что*-клауз, *p*-значение χ^2 -теста и *V* Крамера
для четырех попарных контрастов между модусами/регистрами по классам притяжения глаголов**

	Притяжение <i>Что</i>		Притяжение <i>То, что</i>		Без притяжения	
	89% (92%) — 92% (92%)	$p = 0.004$ $V = 0.02$	41% (50%) — 50% (50%)	$p = 0.001$ $V = 0.03$	88% (79%) — 79% (79%)	$p = 0.18$
Художественный Нехудожественный	94% (92%) — 90% (92%)	$p < 0.001$ $V = 0.06$	57% (50%) — 48% (50%)	$p < 0.001$ $V = 0.08$	82% (79%) — 73% (79%)	$p < 0.001$ $V = 0.1$
Публицистический Академический	91% (90%) — 87% (90%)	$p < 0.001$ $V = 0.05$	50% (48%) — 42% (48%)	$p < 0.001$ $V = 0.08$	73% (73%) — 75% (73%)	$p = 0.8$
Непубличный (Устный) Публичный (Устный)	90% (89%) — 89% (89%)	$p = 0.9$	43% (41%) — 40% (41%)	$p = 0.9$	100% (88%) — 81% (88%)	$p = 0.07^7$ $V = 0.28$

⁷ Для контраста между публичным и непубличным регистрами устного модуса был применен точный критерий Фишера, поскольку в силу небольшого числа данных не были выполнены условия для применения χ^2 -теста.

Как видно из Таблицы 7, глаголы, притягивающиеся к *что*- и *то*, *что*-клаузам, демонстрируют довольно близкое поведение. Для первых трех контрастных пар мы видим значимое влияние регистра в одном и том же направлении для обоих классов глаголов (выделено заливкой). Для контрастов между художественным vs. нехудожественным и публицистическим vs. академическим регистрами это влияние такое же, как и для тестов на независимость по всем глаголам (см. Таблицы 1–2), а именно: для более формального регистра в паре наблюдаемая частотность *что*-клауз ниже, чем ожидается. Соответственно, наблюдаемая частотность *то*, *что*-клауз в более формальных регистрах (получаемая путем вычитания первой из 100%) выше, чем ожидается, что соответствует гипотезе (3). Как видно из таблицы, этот эффект несколько сильнее для глаголов, притягивающихся к *то*, *что*-клаузе, однако эффекты различаются не очень существенно.

Для контраста между устным и письменным модусами мы также видим значимый эффект в одинаковом направлении для этих же классов глаголов. Примечательно, что данный эффект не обнаруживался при общем анализе (см. Таблицу 2). Любопытно также, что здесь мы наблюдаем уменьшение частотности *что*-клауз (и, соответственно, увеличение частотности *то*, *что*-клауз) в сравнении с ожидаемым в модусе, который, по гипотезе, *менее* формален. Этот результат на первый взгляд противоречит гипотезе (3). Однако данное противоречие, по-видимому, мнимое, так как в устном модусе подавляющее большинство примеров (см. Таблицу 1) относится к публичному регистру ($87\% (= 1116/(1116+169))$). Поэтому на месте устного модуса фактически выступает публичный регистр устного модуса, а для этого регистра вполне можно ожидать, что он превосходит по степени формальности письменный модус в целом. В свою очередь, не обнаружение указанного эффекта для всей группы глаголов (см. Таблицу 2, а также примечание 2), вероятно, можно связать с тем, что в эту группу попало большое число примеров с глаголами без притяжения, для которых описанный эффект не обнаруживается (см. Таблицу 14 в Приложении).

Что касается группы глаголов без притяжения, то их поведение менее единообразно. С одной стороны, они показывают значимый эффект регистра для сравнений художественного и нехудожественного регистров в направлении гипотезы (3). Кроме того, эти глаголы демонстрируют значимый эффект в предсказанном направлении для сравнения непубличного и публичного регистров устного модуса, который не обнаружен ни у двух других классов, ни у всех глаголов в целом; впрочем, этот эффект получен на довольно небольшой выборке (всего 39 примеров, см. Таблицу 14 в Приложении). С другой стороны, для двух других контрастов те же глаголы не обнаруживают эффекта регистра.

Данные факты указывают на различие между глаголами, притягивающимися к определенному типу клауз, и глаголами, не показывающими притяжения. Это различие означает, что гипотеза (3) в целом подтверждается для первых двух классов глаголов, но не подтверждается для глаголов без притяжения. Что могло привести к такому результату?

Как видно из Таблицы 7, те контрасты, для которых был обнаружен значимый эффект регистра для глаголов без притяжения (художественный — нехудожественный и публичный — непубличный, но не устный — письменный и не академический — публицистический) имеют нечто общее: в основе и тех и других лежит противопоставление разговорной (спонтанной) речи, в том числе переданной в диалогах в художественном тексте, и подготовленной речи. Можно предположить, что на глаголы без притяжения влияет не более общий признак формальности регистра, а более частный признак близости регистра к разговорной (спонтанной) речи. Поскольку этот признак, по-видимому, нерелевантен для контраста между академическим и публицистическим регистрами (внутри письменного модуса), мы не наблюдаем здесь эффекта регистра для глаголов без притяжения. Такая же ситуация вероятна и для контраста между устным и письменным модусами. Учитывая, что в каждом из сравниваемых модусов имеется регистр с «разговорным» компонентом (непубличный в устном и художественный в письменном модусе), различие между модусами по признаку «разговорности» может отчасти нивелироваться и, соответственно, не приводить к наблюдаемому эффекту⁸. Таким образом, возникает предположение, что для глаголов без притяжения именно степень близости регистра к разговорной речи, а не степень его формальности в целом повышает частотность *что*-клауз (и уменьшает частотность *то, что*-клауз). Однако целенаправленная проверка этой гипотезы пока не осуществлена.

4. Анализ дисперсии

В предыдущем разделе мы рассмотрели, как выбор регистра/модуса внутри контрастной пары влияет на выбор типа клаузы для глаголов трех разных классов притяжения. Мы выяснили, что признак формальности регистра, операционализированный путем использования контрастных пар, сильнее влияет на глаголы, притягивающиеся к *то, что*-клаузе, чем на глаголы, притягивающиеся к *что*-клаузе. Как объясняется такое положение дел?

Можно предположить следующее. Поскольку *что*-клаузы значительно частотнее, чем *то, что*-клаузы, они являются менее маркированным способом выражения сентенциального актанта и, соответственно, более «укорененным» (entrenched) в когнитивной системе носителя (см., например, [Divjak, Caldwell-Harris 2015]). С другой стороны, известно, что важным предиктором когнитивной укорененности, наряду с частотностью, является равномерность распределения слова/конструкции по различным частям корпуса (в частности, регистрам), что обычно обозначается как «дисперсия» (см. [Gries 2008]). Можно ожидать, что *то, что*-клаузы будут в целом

⁸ Здесь мы исходим из того, что публичный регистр устного модуса не обладает высокой степенью признака «разговорности» (в указанном смысле).

иметь большую дисперсию, чем *что*-клаузы, иначе говоря, сильнее зависеть от типа регистра. Как следствие, глаголы, притягивающиеся к *то*, *что*-клаузам, будут также иметь большую дисперсию по сравнению с глаголами, притягивающимися к *что*-клаузам. Таким образом, больший эффект регистра для глаголов, притягивающихся к *то*, *что*-клаузе, предположительно, связан с большей дисперсией *то*, *что*-клауз в целом.

Что касается глаголов без притяжения, то их поведение отличается от поведения двух других групп, поскольку, как мы видели, для этих глаголов формальность регистра не имеет последовательного влияния на тип клаузы. Поэтому не имеет смысла говорить о некоем дополнительном факторе, модулирующем степень влияния регистра. Однако можно ожидать, что эти глаголы будут отличаться от двух других групп не только по характеру влияния регистра, но и по дисперсии.

Для анализа дисперсии была использована мера *отклонение пропорций* (Deviation of Proportions, или DP), предложенная в [Gries 2008; Lijffijt, Gries 2012; см. также Levshina 2015: 80—85]. Эта мера основана на разности между ожидаемыми и наблюдаемыми пропорциями определенной конструкции в разных частях (регистрах) корпуса. Ожидаемые пропорции соответствуют тем долям, которые занимают части корпуса от суммы всех частей. Наблюдаемые пропорции соответствуют долям, которые занимает соответствующая конструкция в каждой из частей корпуса. Сама мера (DP) равна сумме модуля разностей между ожидаемыми и наблюдаемыми пропорциями, деленной на два. Эту меру также принято нормализовывать (DP_{norm}) путем деления на величину (1 — минимальная доля), в результате чего она получает значения в интервале от 0 до 1.

Для вычисления DP_{norm} мы выбрали три использованных выше письменных регистра и устный модус. Мы решили не делить последний на публичный и непубличный регистры, посчитав, что доля непубличного регистра слишком мала, чтобы существенно повлиять на результат. В Таблице 8 можно видеть размер регистра/модуса в абсолютных цифрах, а также его долю в суммарном подкорпусе.

Таблица 8

Доля регистра/модуса в подкорпусе

Регистр/модус	Количество слов	Доля
АКАДЕМИЧЕСКИЙ	33,383,793	0.12
ПУБЛИЦИСТИЧЕСКИЙ	108,127,689	0.4
ХУДОЖЕСТВЕННЫЙ	122,411,245	0.45
УСТНЫЙ	8,487,701	0.03
Итого	272,410,428	1

Далее мы получили для каждого глагола с каждым из двух типов клауз величину DP_{norm} , а затем подсчитали средние значения для классов глаголов по типу притяжения. Результаты этих подсчетов отражены в Таблице 9 (см. Таблицу 15 в Приложении с данными для отдельных глаголов) и графически представлены на диаграмме (рис. 2).

Таблица 9

Дисперсия (нормализованное отклонение пропорций) глаголов по классам притяжения и типу клаузы

Тип клаузы	Класс притяжения			
	Притяжение <i>Что</i>	Притяжение <i>То, что</i>	Без притяжения	Все глаголы
<i>Что</i>	0.06	0.18	0.23	0.05
<i>То, что</i>	0.11	0.25	0.12	0.19
среднее притяжение	0.06	0.21	0.20	0.06

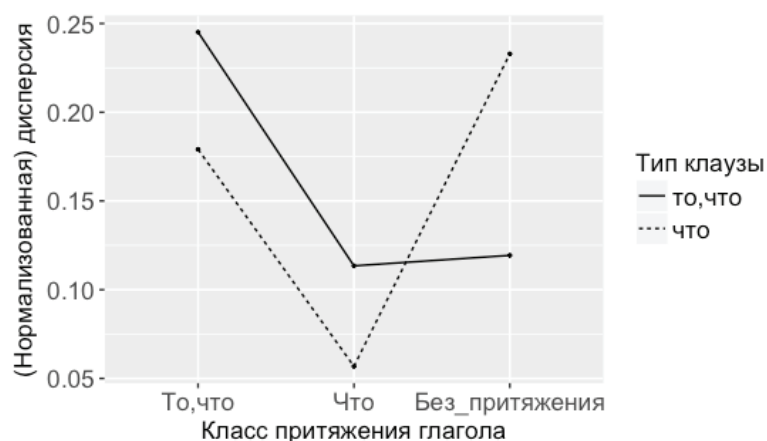


Рис. 2. Дисперсия (нормализованное отклонение пропорций) по классам притяжения глаголов

Как показывает график на рисунке 2, глаголы, притягивающиеся к *что*-клаузе, имеют значительно меньшую дисперсию, чем глаголы, притягивающиеся к *то, что*-клаузе. Кроме того, для каждого из этих классов глаголов выбор *то, что*-клаузы равномерно увеличивает величину дисперсии по сравнению со *что*-клаузой. На графике это отражено в виде двух параллельных наклонных линий. Что касается глаголов без притяжения, то они ведут себя совсем иначе. При *что*-клаузах эти глаголы имеют значительно большую дисперсию, чем при *то, что*-клаузах, вопреки ожи-

даниям, что можно описать как своеобразный эффект взаимодействия между классом глагола и типом клаузы. Как объясняются эти факты в свете различий между классами глаголов, которые мы обсуждали в разделе 3.3?

В первую очередь отметим, что наше предположение о том, что *то, что*-клаузы, а также глаголы, притягивающиеся к ним, имеют большую дисперсию (по регистрам), чем *что*-клаузы, подтвердилось. Также подтвердилась гипотеза о том, что два класса глаголов с притяжением будут вести себя иначе, чем глаголы без притяжения. Действительно, для глаголов с притяжением мы видим равномерный (линейный) эффект типа клаузы, в то время как для глаголов без притяжения мы видим эффект взаимодействия. Таким образом, результаты анализа дисперсии в целом соотносятся с результатами тестов на независимость для трех классов глаголов (см. Таблицу 7). Кроме того, большая величина дисперсии для глаголов, притягивающихся к *то, что*-клаузе, как оказалось, коррелирует с большим влиянием регистра на эти глаголы, что, очевидно, указывает на связь этих двух параметров.

Большой интерес представляет интерпретация эффекта взаимодействия для глаголов без притяжения. При анализе тестов на независимость мы предположили, что для этих глаголов релевантным фактором, повышающим предпочтение *что*-клауз, является скорее степень близости регистра к разговорной речи, нежели степень его формальности в целом. В то же время неожиданно высокая величина дисперсии *что*-клауз для данного класса глаголов указывает на то, что этот класс проявляет особую чувствительность к влиянию регистра при *что*-клаузах, иначе говоря, у таких глаголов есть некоторое свойство, вызывающее значительное предпочтение *что*-клауз в одних регистрах по отношению к другим (тогда как *то, что*-клаузы распределены по регистрам более равномерно). По-видимому, два отмеченных факта связаны следующим образом. Можно предположить, что глаголы без притяжения демонстрируют особый эффект регистра, заключающийся в том, что при этих глаголах в регистрах с большим «разговорным» компонентом будет увеличиваться предпочтение *что*-клауз. Действительно, интуитивно кажется, что при глаголах без притяжения типа *радоваться, извиняться* и др. *что*-клаузы могут иметь разговорный оттенок (ср. примеры в (6a)–(6b)), который менее заметен при глаголах других классов, — однако высказанное предположение требует дальнейшей проверки.

- (6) а. *Очень извиняюсь, что два раза наступил вам на рубашку и один раз на костюм в целом!* (Дмитрий Емец. Таня Гроттер и колодец Посейдона (2004)).
- б. *Радуйтесь, что ваш сын к этому классу не принадлежит* (И. Грекова. Перелом (1987)).

Поскольку указанный эффект «разговорности» регистра (степени его близости к разговорной речи) в общем случае не совпадает с эффектом формальности регистра, глаголы без притяжения могут быть более чувст-

вительны к первому эффекту, чем ко второму (в отличие от глаголов с притяжением, для которых ситуация может быть обратная). Это могло бы объяснить различия в характере влияния регистра на глаголы с притяжением и без притяжения (см. Таблицу 7, а также обсуждение в конце раздела 3.3).

Помимо обнаруженного эффекта взаимодействия между типом клаузы и типом притяжения глагола, любопытно также отметить, что тип клаузы оказался менее важным фактором для дисперсии, чем класс глагола, что видно по большей величине наклона параллельных линий по сравнению с расстоянием между ними (по оси y). Иначе говоря, в то время как глаголы разных классов притяжения заметно отличаются по уровню дисперсии, глаголы одного класса имеют близкую дисперсию независимо от типа клаузы, с которой они употребляются. Этот факт подтверждается наличием достаточно высокой корреляции ($r = 0.6$, $p = 0.005$) уровня дисперсии при *что*- и *то, что*-клаузах для глаголов с притяжением, что отражено на рисунке 3. Этот результат также требует дальнейшего осмысления в будущих исследованиях.

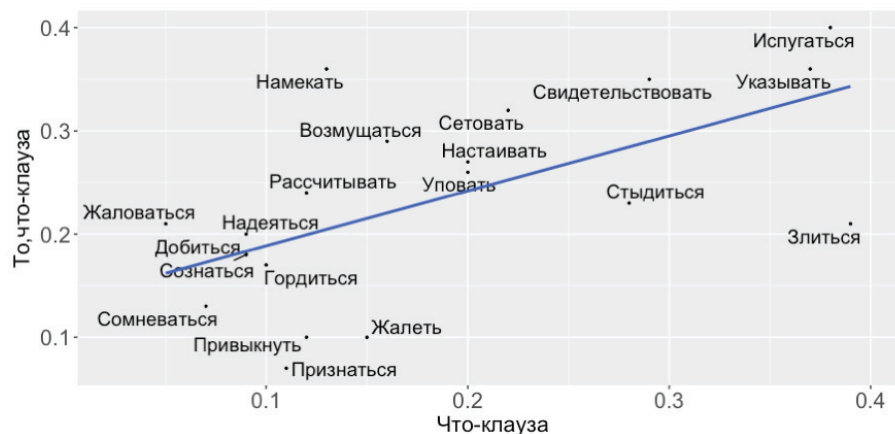


Рис. 3. Дисперсия *что*- и *то, что*-клауз для глаголов с притяжением

Подводя итог, можно сказать, что анализ дисперсии показал результаты, сопоставимые с тестами на независимость, подтвердив сходство между двумя классами глаголов с притяжением, а также больший эффект регистра для глаголов, притягивающихся к *то, что*-клаузе. Кроме того, анализ подтвердил особое поведение глаголов без притяжения (непоследовательное влияние признака формальности регистра) и позволил прояснить высказанную выше гипотезу о важности признака «разговорности» для объяснения дистрибутивного профиля данной группы глаголов.

5. Заключение

Целью настоящей статьи было рассмотреть, как регистр влияет на выбор *что*- или *то, что*-клауз в ситуации конкуренции (в позиции, соотносимой с позицией косвенного дополнения). В качестве основной гипотезы рассматривалась связь между формальностью регистра и выбором *то, что*-клаузы. Для проверки этой гипотезы были отобраны 26 глаголов и пять регистров, полученных с использованием соответствующей разметки в основном и устном подкорпусах НКРЯ, — академический, публицистический и художественный регистры письменного модуса и публичный и непубличный регистры устного модуса, — разбитые на четыре контрастные пары для плановых сравнений (устный — письменный, художественный — нехудожественный, публицистический — академический, непубличный — публичный).

Первоначальные тесты на независимость показали слабое, но значимое влияние нехудожественного и академического регистров (более формального регистра в соответствующих парах) на увеличение частотности *то, что*-клаузы, которое было интерпретировано как в целом подтверждающее основную гипотезу. Далее, для оценки влияния регистра на *отдельные* глаголы, был применен метод дистинктивных коллексем (разновидность коллокационного анализа), основанный на идее притяжения глагола к одной из конкурирующих конструкций [Gries, Stefanowitsch, 2004; Stefanowitsch, Gries 2008]. Целью этого метода являлась классификация глаголов по притяжению к *что*-клаузам vs. *то, что*-клаузам или ни к одному из типов клауз, а также идентификация глаголов, которые меняют класс притяжения (от *то, что*- к *что*-клаузе или наоборот) при изменении регистра. Этот метод показал стабильный профиль притяжения глаголов по регистрам и практически не выявил глаголов, меняющих тип притяжения, что в целом соответствует сделанному ранее выводу о *слабом* эффекте регистра. Для дальнейшего анализа были проведены тесты на независимость регистра и типа клаузы по классам притяжения, идентифицированных методом дистинктивных коллексем. Был воспроизведен эффект формальности регистра для глаголов с притяжением *что*- и *то, что*-клауз при более сильном эффекте для второй группы глаголов, при этом глаголы без притяжения показали несколько иной эффект, предположительно, связанный не с формальностью регистра как таковой, а со степенью его близости к разговорной речи. Наконец, был проведен анализ дисперсии (по регистрам) методом отклонения пропорций (см. [Gries 2008; Lijffijt, Gries 2012]) для комбинаций глаголов с двумя типами клауз. Выявлена несколько бóльшая дисперсия *то, что*-клауз, и значительно бóльшая дисперсия глаголов, притягивающихся к *то, что*-клаузе (независимо от типа клаузы), что в целом соотносится с более сильным воздействием регистра на этот класс. Кроме того, обнаружен несколько неожиданный эффект взаимодействия для глаголов без притяжения (бóльшая дисперсия *что*-клауз), что, предположи-

тельно, соотносится с их особым поведением, выявленным в тестах на независимость.

Какие выводы можно сделать из результатов проведенных тестов? В целом гипотеза о связи формальности регистра с выбором *то, что*-клауз подтвердилась; более того, оказалось, что данная связь не зависит от класса притяжения глагола (т. е. от того, с каким типом клаузы глагол встречается значимо чаще, чем «средний глагол»). Эта связь, однако, слабая и почти никогда не меняет класс притяжения глагола. Наиболее интересным результатом оказалось особое поведение глаголов без притяжения (т. е. тех, что ближе к «среднему глаголу»). Вероятно, такое поведение связано с влиянием не формальности регистра, а его «разговорности» (или некоего другого признака).

Если предложенная интерпретация в целом верна, то будущие исследования должны использовать более детальную классификацию регистров (возможно, с применением подрегистров), которая позволила бы систематически разграничить признаки формальности и «разговорности» (или какие-то другие признаки, которые окажутся релевантными для выбора типа клаузы). Продуктивным решением в этом направлении мог бы стать отказ от «априорных» делений на модусы и регистры и использование автоматической, основанной на данных (data-driven) кластеризации (например, методом главных компонент), см. [Gries 2011]. Такой метод мог бы помочь выявить релевантные признаки, по которым различаются регистры. Однако его реализация на конкретном материале пока остается задачей будущего.

Дальнейшие исследования также должны будут учесть возможное взаимодействие между классом притяжения глагола и влиянием регистра. Важно отметить, что классы притяжения, использованные в данном исследовании, отчасти являются следствием конкретной выборки глаголов (поскольку класс притяжения определялся путем сравнения с остальными глаголами из выборки); иначе говоря, другая выборка могла бы дать несколько другие классы. Возможно, для более надежного контроля этого фактора в будущем можно было бы попробовать использовать классификации глаголов на основе других мер притяжения (см., например, [Schmid, Küchenhoff 2013]), в том числе непосредственного уровня частотности каждого из типа клауз.

Наконец, еще одним направлением исследований могло бы стать использование обобщенных линейных моделей со смешанными эффектами, которые бы *одновременно* проверяли влияние различных факторов (в том числе регистра и класса глагола) на выбор типа клаузы; см. [Gries 2015b], где рекомендуется использовать именно такие модели для анализа корпусных данных, а также [Divjak 2017] в качестве примера конкретного исследования в области дистрибуции сентенциальных актантов (на материале польского языка).

Приложение

Таблица 10

Притяжения глаголов к *что-* и *то*, *что*-кляузам в письменном и устном модусах

Устный модус					
<i>Что</i> -Кляуза			<i>То</i> , <i>что</i> -Кляуза		
Коллексема	p	log ₁₀ p	Коллексема	p	log ₁₀ p
<i>надея</i> ться (590:34)	6.04E-53	52.2	<i>свидетельствовать</i> (16:55)	1.61E-23	22.8
<i>радов</i> аться (28:1)	6.77E-3	2.2	<i>настаив</i> ать (10:33)	4.69E-14	13.3
			<i>горд</i> иться (19:35)	2.54E-11	10.6
			<i>указыв</i> ать (21:28)	1.65E-07	6.8
			<i>доб</i> иться (10:16)	2.49E-05	4.6
			<i>рассчитыв</i> ать (36:23)	6.42E-03	2.2
Письменный модус					
<i>Что</i> -Кляуза			<i>То</i> , <i>что</i> -Кляуза		
Коллексема	p	log ₁₀ p	Коллексема	p	log ₁₀ p
<i>надея</i> ться (10351:339)	0	Inf	<i>горд</i> иться (443:1002)	0	Inf
<i>призн</i> аться (3554:152)	5.15E-210	209.3	<i>свидетельствовать</i> (1547:1808)	0	Inf
<i>созн</i> аться (1696:72)	4.54E-98	97.3	<i>доб</i> иться (118:382)	3.33E-155	154.5
<i>жале</i> ть (1973:179)	2.68E-61	60.6	<i>настаив</i> ать (422:574)	8.64E-142	141.1
<i>испуг</i> аться (893:53)	1.01E-41	41	<i>намека</i> ть (489:469)	1.74E-82	81.8
<i>жалов</i> аться (1748:252)	2.06E-24	23.7	<i>указыв</i> ать (2021:1095)	1.15E-77	76.9
<i>рассчитыв</i> ать (1465:323)	7.52E-04	3.1	<i>привыкн</i> уть (627:489)	5.52E-66	65.3

<i>злиться</i> (125:20)	3.20E-02	1.5	<i>возмущаться</i> (112:103)	4.56E-18	17.3
			<i>сетовать</i> (212:139)	4.92E-15	14.3
			<i>стыдиться</i> (116:82)	1.32E-10	9.9
			<i>уповать</i> (114:54)	8.85E-04	3.1

Таблица 11

Притяжения глаголов к *что-* и *то*, *что-*клаузам в художественном и нехудожественном регистрах

Художественный регистр					
<i>Что</i> -Клауза			<i>То</i> , <i>что</i> -Клауза		
Коллексема	p	log ₁₀ p	Коллексема	p	log ₁₀ p
<i>надеясь</i> (4559:111)	1.47E-210	209.8	<i>гордиться</i> (245:298)	1.11E-113	113
<i>признаваясь</i> (1529:74)	1.45E-38	37.8	<i>привыкнув</i> (360:277)	5.25E-75	74.3
<i>сознаться</i> (735:28)	9.12E-23	22	<i>свидетельствовать</i> (296:217)	5.59E-56	55.3
<i>жалеть</i> (1194:95)	4.53E-15	14.3	<i>добиться</i> (63:117)	2.01E-55	54.7
<i>испугаться</i> (733:46)	7.91E-14	13.1	<i>настаивать</i> (123:113)	1.60E-35	34.8
<i>жаловаться</i> (754:65)	2.02E-08	7.7	<i>указывать</i> (220:124)	1.57E-24	23.8
			<i>стыдиться</i> (83:55)	1.05E-13	13
			<i>намекать</i> (281:93)	3.36E-08	7.5
			<i>возмущаться</i> (46:28)	4.11E-07	6.4
			<i>радоваться</i> (1230:261)	2.29E-04	3.6
			<i>сетовать</i> (67:23)	3.80E-03	2.4
			<i>уповать</i> (33:13)	1.08E-02	2
			<i>дожидаться</i> (49:15)	4.63E-02	1.3

Нехудожественный регистр					
Что-Клауза			То, что-Клауза		
Коллексема	p	log ₁₀ p	Коллексема	p	log ₁₀ p
надеясь (5792:228)	0	Inf	свидетельствовать (1251:1591)	4.96E-284	283.3
признаться (2025:78)	2.48E-180	179.6	гордиться (198:704)	4.11E-242	241.4
сознаться (961:44)	7.450E-78	77.1	добиться (55:265)	4.31E-101	100.4
жалесть (779:84)	1.38E-34	33.9	настаивать (299:461)	1.00E-91	91
жаловаться (994:187)	3.45E-18	17.5	намекать (208:376)	1.92E-85	84.7
испугаться (160:7)	6.56E-14	13.2	указывать (1801:971)	8.91E-28	27.1
рассчитывать (881:241)	1.99E-04	3.7	привыкнуть (267:212)	9.47E-18	17
			возмущаться (66:75)	1.30E-11	10.9
			сетовать (145:116)	1.65E-10	9.8
			стыдиться (33:27)	1.74E-03	2.8

Таблица 12

Притяжения глаголов к что- и то, что-клаузам в публицистическом и академическом регистрах

Публицистический регистр					
Что-Клауза			То, что-Клауза		
Коллексема	p	log ₁₀ p	Коллексема	p	log ₁₀ p
надеясь (5000:198)	0	Inf	свидетельствовать (877:877)	9.81E-153	152
признаться (1790:70)	3.83E-126	125.4	гордиться (186:406)	6.46E-129	128.2
сознаться (817:42)	8.75E-49	48.1	добиться (47:223)	4.47E-98	97.3
жалесть (725:79)	3.30E-22	21.5	намекать (197:349)	6.52E-97	96.2

<i>испугаться</i> (149:6)	1.09E-10	10	<i>настаивать</i> (254:349)	9.70E-80	79
<i>жаловаться</i> (800:156)	5.90E-07	6.2	<i>привыкнуть</i> (244:196)	1.59E-24	23.8
<i>сомневаться</i> (1401:369)	4.98E-02	1.3	<i>указывать</i> (1102:493)	4.20E-15	14.4
			<i>возмущаться</i> (62:72)	6.67E-15	14.2
			<i>сетовать</i> (131:100)	3.58E-12	11.4
			<i>уповать</i> (51:36)	9.55E-05	4
			<i>стыдиться</i> (29:24)	3.89E-04	3.4
			<i>радоваться</i> (510:195)	1.84E-03	2.7
Академический регистр					
Что-Клауза			То, что-Клауза		
Коллексема	p	log ₁₀ p	Коллексема	p	log ₁₀ p
<i>надеяться</i> (792:30)	1.50E-151	150.8	<i>гордиться</i> (12:298)	4.00E-107	106.4
<i>признаться</i> (235:8)	1.16E-42	41.9	<i>свидетельствовать</i> (374:714)	4.99E-84	83.3
<i>сознаться</i> (144:2)	1.22E-29	28.9	<i>настаивать</i> (45:112)	7.48E-16	15.1
<i>жаловаться</i> (194:31)	2.72E-18	17.6	<i>добиться</i> (8:42)	1.52E-10	9.8
<i>жалеть</i> (54:5)	1.15E-07	6.9	<i>намекать</i> (11:27)	1.58E-04	3.8
<i>рассчитывать</i> (114:35)	2.67E-05	4.6	<i>сомневаться</i> (191:157)	4.10E-02	1.4
<i>уповать</i> (30:5)	1.50E-03	2.8			
<i>радоваться</i> (51:18)	1.86E-02	1.7			
<i>испугаться</i> (11:1)	3.47E-02	1.5			

Таблица 13

Притяжения глаголов к *что-* и *то-*, *что*-кlausам в непубличном и публичном регистрах устного модуля

Непубличный регистр (Спонтанная речь)					
<i>Что</i> -Клауза			<i>То, что</i> -Клауза		
Коллексема	p	log ₁₀ p	Коллексема	p	log ₁₀ p
<i>надеялся</i> (41:4)	2.52E-03	2.6	<i>гордиться</i> (1:34)	1.50E-25	24.8
<i>жалеть</i> (14:0)	2.20E-02	1.7			
Публичный регистр (Подготовленная речь)					
<i>Что</i> -Клауза			<i>То, что</i> -Клауза		
Коллексема	p	log ₁₀ p	Коллексема	p	log ₁₀ p
<i>надеялся</i> (549:30)	1.71E-51	50.8	<i>свидетельствовать</i> (16:55)	1.61E-23	22.8
			<i>настаивать</i> (10:33)	4.69E-14	13.3
			<i>гордиться</i> (19:35)	2.54E-11	10.6
			<i>указывать</i> (21:28)	1.65E-07	6.8
			<i>добиться</i> (10:16)	2.49E-05	4.6
			<i>рассчитывать</i> (36:23)	6.42E-03	2.2
			<i>возмущаться</i> (2:5)	8.09E-03	2.1

Таблица 14

Наблюдаемая (ожидаемая) частотность *что-* и *то, что*-клауз (в пропорциях и абсолютных цифрах)
для четырех контрастов между модусами/регистрами по классам притяжения глаголов

	Притяжение <i>Что</i>		Притяжение <i>То, что</i>		Без притяжения	
	<i>Что</i>	<i>То, что</i>	<i>Что</i>	<i>То, что</i>	<i>Что</i>	<i>То, что</i>
Устный	812 (836) 89% (92%)	98 (74) 11% (8%)	133 (163) 41% (50%)	193 (163) 59% (50%)	43 (39) 88% (79%)	6 (10) 12% (21%)
Письменный	24887 (24863) 92% (92%)	2190 (2213) 8% (8%)	6221 (6191) 50% (50%)	6197 (6227) 50% (50%)	2487 (2491) 79% (79%)	660 (656) 21% (21%)
Художеств.	11684 (11463) 94% (92%)	788 (1009) 6% (8%)	1817 (1591) 57% (50%)	1358 (1584) 43% (50%)	1708 (1647) 82% (79%)	376 (437) 18% (21%)
Нехудож.	13203 (13424) 90% (92%)	1402 (1181) 10% (8%)	4404 (4630) 48% (50%)	4839 (4613) 52% (50%)	779 (840) 73% (79%)	284 (223) 27% (21%)
Публицист.	11468 (11390) 91% (90%)	1132 (1210) 9% (10%)	3180 (3004) 50% (48%)	3125 (3300) 50% (52%)	713 (715) 73% (73%)	262 (260) 27% (27%)
Академич.	1735 (1813) 87% (90%)	270 (192) 13% (10%)	1224 (1400) 42% (48%)	1714 (1538) 58% (52%)	66 (64) 75% (73%)	22 (24) 25% (27%)
Непубличный (Устный)	81 (10) 90% (89%)	9 (10) 10% (11%)	26 (25) 43% (41%)	35 (36) 57% (59%)	18 (16) 100% (88%)	0 (2) 0% (12%)
Публичный (Устный)	731 (732) 89% (89%)	89 (88) 11% (11%)	107 (108) 40% (41%)	158 (157) 60% (59%)	25 (27) 81% (88%)	6 (4) 19% (12%)

Таблица 15

Дисперсия (нормализованное отклонение пропорций) глаголов в конструкции со *что-* и *то-* *что*-кlausами (по убыванию)

<i>Что</i> -Клауза				<i>То, что</i> -Клауза			
Глагол	Класс при- тяжения	log ₁₀ p	DP _{norm}	Глагол	Класс при- тяжения	log ₁₀ p	DP _{norm}
<i>злиться</i>	<i>что</i>	1.7	0.39	<i>намекать</i>	<i>То, что</i>	81.4	0.36
<i>испугаться</i>	<i>что</i>	40.7	0.38	<i>указывать</i>	<i>То, что</i>	80.9	0.36
<i>указывать</i>	<i>То, что</i>	80.9	0.37	<i>свидетельствовать</i>	<i>То, что</i>	Inf	0.35
<i>свидетельствовать</i>	<i>То, что</i>	Inf	0.29	<i>сетовать</i>	<i>То, что</i>	14.7	0.32
<i>стыдиться</i>	<i>То, что</i>	9.8	0.28	<i>возмущаться</i>	<i>То, что</i>	17.5	0.29
<i>радоваться</i>	—	0.3	0.23	<i>настаивать</i>	<i>То, что</i>	152.6	0.27
<i>сетовать</i>	<i>То, что</i>	14.7	0.22	<i>уповать</i>	<i>То, что</i>	3.3	0.26
<i>хвастаться</i>	—	0.1	0.22	<i>рассчитывать</i>	<i>что</i>	2.2	0.24
<i>расстраиваться</i>	—	0.6	0.19	<i>стыдиться</i>	<i>То, что</i>	9.8	0.23
<i>извиниться</i>	—	0.2	0.18	<i>жаловаться</i>	<i>что</i>	23	0.21
<i>возмущаться</i>	<i>То, что</i>	17.5	0.16	<i>злиться</i>	<i>что</i>	1.7	0.21
<i>жалеть</i>	<i>что</i>	60.9	0.15	<i>обидеться</i>	—	0.2	0.21
<i>намекать</i>	<i>То, что</i>	81.4	0.13	<i>дождаться</i>	—	0.6	0.19
<i>привыкнуть</i>	<i>То, что</i>	63.4	0.12	<i>сознаться</i>	<i>что</i>	96.6	0.18
<i>рассчитывать</i>	<i>что</i>	2.2	0.12	<i>добиться</i>	<i>То, что</i>	157.9	0.18
<i>признаваться</i>	<i>что</i>	208.6	0.11	<i>гордиться</i>	<i>То, что</i>	Inf	0.17
<i>добиться</i>	<i>То, что</i>	157.9	0.09	<i>хвастаться</i>	—	0.1	0.17
<i>надеяться</i>	<i>что</i>	Inf	0.09	<i>сомневаться</i>	<i>что</i>	0.7	0.13

сознаться	что	96.6	0.09	радоваться	—	0.3	0.12
сомневаться	что	0.7	0.07	извиниться	—	0.2	0.11
жаловаться	что	23	0.05	расстраиваться	—	0.6	0.09
дождаться	—	0.6	0.3	признаться	что	208.6	0.07
обидеться	—	0.2	0.3	испугаться	что	40.7	0.4
настаивать	То, что	152.6	0.2	надеяться	что	Inf	0.2
уповать	То, что	3.3	0.2	жалеть	что	60.9	0.1
гордиться	То, что	Inf	0.1	привыкнуть	То, что	63.4	0.1

Л и т е р а т у р а

Валгина 2003 — Н. С. Валгина. Современный русский язык. Синтаксис. 4-е изд. М., 2003.

Кибрик 2009 — А. А. Кибрик. Модус, жанр и другие параметры классификации дискурсов // Вопросы языкознания. 2009. № 2. С. 3–21.

Князев 2018 — М. Ю. Князев. Об ограничении на сентенциальный актант с союзом *что* при деагентивных употреблениях глаголов речи // Вопросы языкознания. 2018. № 3. С. 7–39.

Кобозева 2013 — И. М. Кобозева. Условия употребления «то» перед придаточным изъяснительным с союзом «что» // *Du mot au texte. Études slavo-romanes* / O. Inkova (éd.). Bern, 2013. P. 131–150.

Коротаев 2013 — Н. А. Коротаев. Полипредикативные конструкции с *то что* в непубличной устной речи // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2013» (Бекасово, 29 мая — 2 июня 2013 г.). Вып. 12 (19). М., 2013. С. 324–331.

Кузнецова, Велейшикова 2010 — Ю. Л. Кузнецова, Т. В. Велейшикова. Современные корпусные исследования языка: новые подходы // Вопросы языкознания. 2010. № 5. С. 108–124.

Летучий 2014 — А. Б. Летучий. Синтаксические свойства сентенциальных актантов при предикативах // Вестник Московского государственного гуманитарного университета им. М. А. Шолохова. Филологические науки. 2014. № 1. С. 62–84.

НКРЯ — Национальный корпус русского языка. URL: <http://www.ruscorpora.ru>.

Пекелис 2014 — О. Е. Пекелис. Инфинитив vs. придаточное с союзом *чтобы*: К вопросу о выборе способа оформления сентенциального актанта в русском языке // Вопросы языкознания. 2014. № 4. С. 13–45.

Циммерлинг 2012 — А. В. Циммерлинг. Неканонические подлежащие в русском языке // От формы к значению, от значения к форме. Сборник статей в честь 80-летия А. В. Бондарко / Отв. ред. М. Воейкова. М., 2012. С. 568–590.

Шведова 1980 — Н. Ю. Шведова (ред.). Русская грамматика. Т. 2. Синтаксис. М., 1980.

Biber 2012 — D. Biber. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*. 2012. Vol. 8. No. 1. P. 9–37.

Biber, Conrad 2009 — D. Biber, S. Conrad. Register, genre, and style. Cambridge, 2009.

Divjak 2017 — D. Divjak. The role of lexical frequency in the acceptability of syntactic variants: evidence from *that*-clauses in Polish. *Cognitive Science*. 2017. Vol. 4. No. 2. P. 354–382.

Divjak, Caldwell-Harris 2015 — D. Divjak, C. L. Caldwell-Harris. Frequency and entrenchment // *Cognitive foundations of language* / E. Dabrowska, D. Divjak (eds.). Boston, 2015. P. 53–75.

Gries 2008 — S. Th. Gries. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*. 2008. Vol. 13. No. 4. P. 403–437.

Gries 2011 — S. Th. Gries. Corpus data in usage-based linguistics: What's the right degree of granularity for the analysis of argument structure constructions? // *Cognitive linguistics: Convergence and expansion* / M. Brdar, S. Th. Gries, M. Zič-Fuchs (eds.). Amsterdam; Philadelphia, 2011. P. 237–256.

Gries 2015a — S. Th. Gries. More (old and new) misunderstandings of collocation analysis: on Schmid & Küchenhoff (2013) // *Cognitive Linguistics*. 2015. Vol. 26. No. 3. P. 505–536.

Gries 2015b — S. Th. Gries. The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models // *Corpora*. 2015. Vol. 10. No. 1. P. 95–125.

Gries, Stefanowitsch 2004 — S. Th. Gries, A. Stefanowitsch. Extending collocation analysis: A corpus-based perspective on ‘alternations’ // *International Journal of Corpus Linguistics*. 2004. Vol. 9. No. 1. P. 97–129.

Hartman 2012 — J. Hartman. Varieties of clausal complementation. PhD diss. Cambridge (MA): MIT, 2012.

Knyazev 2016 — M. Knyazev. Licensing clausal complements: The case of *što*-clauses. PhD diss. Utrecht, 2016.

Levshina 2015 — N. Levshina. How to do linguistics with R. Amsterdam, 2015.

Lijffijt, Gries 2012 — J. Lijffijt, S. Th. Gries. Correction to “Dispersions and adjusted frequencies in corpora” // *International Journal of Corpus Linguistics*. 2012. Vol. 17. No. 1. P. 147–149.

Schmid, Küchenhoff 2013 — H.-J. Schmid, H. Küchenhoff. Collocation analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings // *Cognitive Linguistics*. 2013. Vol. 24. No. 3. P. 531–577.

Stefanowitsch, Gries 2008 — A. Stefanowitsch, S. Th. Gries. Channel and constructional meaning: A collocation case study // *Cognitive sociolinguistics* / G. Kristiansen, R. Dirven (eds.). Berlin; New York, 2008. P. 129–152.

Статья получена 07.08.2018

MIKHAIL YU. KNYAZEV

Institute for Linguistic Studies of the Russian Academy of Sciences /
National Research University Higher School of Economics, St. Petersburg branch /
St. Petersburg State University
(Russia, St. Petersburg)
mknyazev@hse.ru

EKATERINA A. RUDALEVA

National Research University Higher School of Economics, St. Petersburg branch
(Russia, St. Petersburg, Moscow)
earudaleva@edu.hse.ru

REGISTER AS A PREDICTOR OF THE COMPLEMENT CLAUSE TYPE IN RUSSIAN: A CORPUS STUDY

Register is a factor in linguistic variation and may affect the choice between competing constructions, along with grammatical, semantic, pragmatic and other factors. The paper investigates the effect of register on the realization of indicative complement clauses in Russian (as a *što*-clause or as a special construction with the correlative *to*) with verbs selecting prepositional/oblique complements, cf. *Ona настаивает (na tom), što...* ‘She insists (on the fact) that...’. The study is based on the data from the Russian Na-

tional Corpus. The main hypothesis is that more formal registers tend to be associated with higher preference for the construction with the correlative. Five different registers are compared (academic, news, fiction, spontaneous and non-spontaneous speech). The influence of register was investigated in three ways: on the basis of the verb sample as a whole (26 verbs); individually for separate verbs; for attraction classes of verbs according to their choice of a certain strategy. The so-called “collostructional analysis” (for distinctive collexemes), developed by Stefan Th. Gries and Anatol Stefanowitsch, as well as the analysis of dispersion measured by the Deviation of Proportions (developed by Stefan Th. Gries) were used. The results showed a small increase in the proportion of the correlative construction in the academic and non-fiction registers (as opposed to news and fiction registers) both for the sample as a whole and for the two out of three attraction classes, which on the whole supports the main hypothesis. However, the effect of register on individual verbs (operationalized as change of attraction class accompanying change of register) proved negligible. In addition, the class of verbs neutral with respect to the attracted construction showed a profile different from the two other classes, suggesting that some property of register other than formality might influence their behavior.

Keywords: register, clausal complements, corpus approaches, collostructional analysis, analysis of dispersion, syntactic alternation.

References

- Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, 8(1), 9–37.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.
- Divjak, D. (2017). The role of lexical frequency in the acceptability of syntactic variants: evidence from that-clauses in Polish. *Cognitive Science*, 4(2), 354–382.
- Divjak, D., & Caldwell-Harris, C. L. (2015). Frequency and entrenchment. In E. Dabrowska, & D. Divjak (Eds.), *Cognitive foundations of language*. Boston: Mouton De Gruyter.
- Gries, S. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403–437.
- Gries, S. Th. (2011). Corpus data in usage-based linguistics: What's the right degree of granularity for the analysis of argument structure constructions? In M. Brdar, S. Th. Gries, & M. Žič-Fuchs (Eds.), *Cognitive linguistics: Convergence and expansion* (pp. 237–256). Amsterdam; Philadelphia: John Benjamins.
- Gries, S. Th. (2013). More (old and new) misunderstandings of collostructional analysis: on Schmid & Küchenhoff (2013). *Cognitive Linguistics*, 26(3), 505–536.
- Gries, S. Th. (2015). The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10(1), 95–125.
- Gries, S. Th., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics*, 9(1), 97–129.
- Hartman, J. (2012). *Varieties of clausal complementation* (unpublished doctoral dissertation). MIT, Cambridge (MA), USA.
- Kibrik, A. A. (2009). Modus, zhanr i drugie parametry klassifikatsii diskursov. *Voprosy jazykoznanija*, 2, 3–21.

Knyazev, M. (2016). *Licensing clausal complements: the case of čto-clauses* (unpublished doctoral dissertation). Utrecht University, Utrecht, Netherlands.

Knyazev, M. (2018). Ob ogranichenii na sententsial'nyĭ aktant s soyuzom čto pri deagentivnykh upotrebleniyakh glagolov rechi. *Voprosy jazykoznanija*, 3, 7–39.

Kobozeva, I. M. (2013). Usloviya upotrebleniya «to» pered pridatochnym iz"yasnitel'nym s soyuzom “čto”. In O. Inkova (Ed.), *Du mot au texte. Etudes slavo-romanes* (pp. 131–150). Bern: Peter Lang.

Korotaev, N. A. (2013). Polipredikativnye konstruksii s to čto v nepublichnoĭ ustnoĭ rechi. In *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po materialam mezh-dunarodnoĭ konferentsii «Dialog 2013»*. Vyp. 12 (pp. 324–331). Moscow: Izd-vo RGGU.

Kuznetsova, Yu. L., & Velejšikov, T. V. (2010). Sovremennye korpusnye issledovaniya yazyka: novye podkhody. *Voprosy jazykoznanija*, 5, 108–124.

Letuchiy, A. B. (2014). Sintaksicheskie svoystva sententsial'nykh aktantov pri predikativakh. *Vestnik Moskovskogo gosudarstvennogo gumanitarnogo universiteta im. M. A. Sholokhova. Filologicheskie nauki*, 1, 62–84.

Levshina, N. (2015). *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.

Lijijt, J., & Gries, S. Th. (2012). Correction to “Dispersions and adjusted frequencies in corpora”. *International Journal of Corpus Linguistics*, 17(1), 147–149.

Natsional'nyĭ korpus russkogo yazyka. Retrieved from <http://www.ruscorpora.ru>.

Pekelis, O. E. (2014). Infinitiv vs. pridatochnoe s soyuzom chtoby: k voprosu o vybore sposoba oformleniya sententsial'nogo aktanta v russkom yazyke. *Voprosy jazykoznanija*, 4, 13–45.

Schmid, H.-J., & Küchenhoff, H. (2013). Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics*, 24(3), 531–577.

Shvedova, N. Yu. (Ed.). (1980). *Russkaya grammatika* (Vol. 2). Moscow: Nauka.

Stefanowitsch, A., & Gries, S. Th. (2008). Channel and constructional meaning: A collostructional case study. In G. Kristiansen, & R. Dirven (Eds.), *Cognitive sociolinguistics* (pp. 129–152). Berlin; New York: Mouton de Gruyter.

Tsimerling, A. V. (2012). Nekanonicheskie podlezhashchie v russkom yazyke. In M. Voyeikova (Ed.), *Ot formy k znacheniyu, ot znacheniya k forme. Sb. stateĭ v chest' 80-letiya A. V. Bondarko* (pp. 568–590). Moscow: Yazyki slavianskikh kul'tur.

Valgina, N. S. (2000). *Sintaksis sovremennogo russkogo yazyka*. Moscow: Agar.

Received on 07.08.2018