

## **КОРПУСНЫЕ МЕТОДЫ В ИССЛЕДОВАНИИ РЕЧЕВЫХ ЖАНРОВ: ПРОБЛЕМА КЛЮЧЕВЫХ ФРАЗ**

**В.В. Дементьев, Н.Б. Степанова**

Саратовский государственный университет им. Н.Г. Чернышевского  
410012, Россия, г. Саратов, ул. Астраханская, д. 83

В статье обсуждаются теоретические и практические проблемы, связанные с изучением речевых жанров на основе корпусных баз данных и в целом при помощи методов корпусной лингвистики. Богатство, разнообразие и несомненная эффективность корпусных методов для решения различных лингвистических задач убеждают в том, что и при лингвистическом изучении речевых жанров такие методы должны быть задействованы. В то же время разработанной корпусной методике для решения речежанровых проблем лингвистики пока нет, и в целом не ясны ни роль, какую должны сыграть корпусные методы в общем изучении РЖ, ни — главное — отношение тех единиц, которые используются/доступны оперативному анализу в корпусах, к жанрам речи: о методах корпусной лингвистики можно говорить только в тех случаях, когда лингвистические задачи решаются количественно и с использованием технических возможностей корпусно-поисковых систем.

На первый план выходит проблема «ключевых фраз», или текстовых маркеров РЖ, по которым можно осуществлять поиск в корпусах. По мнению авторов, ключевая фраза речевого жанра представляет собой его существенную характеристику, которая присутствует в речежанровой компетенции носителя языка и должна быть включена в соответствующие жанроведческие описания в качестве дополнительного параметра. Характеристики жанра такого рода в будущих «энциклопедиях речевых жанров», добавляемые к характеристикам, сформулированным в результате других, более традиционных жанроведческих исследований, могли бы выглядеть так: жанр — наличие ключевой леммы — ключевая фраза / ключевые фразы для идентификации жанра — количество вхождений в корпус (какой) — ключевые фразы для исследования контекста — количество вхождений в корпус.

Формулируются главные требования, предъявляемые к ключевой фразе: 1) ключевая фраза должна максимально соответствовать семантике жанра; 2) в системе по этому критерию можно реально искать и находить большое количество примеров.

Предлагается начальная классификация ключевых фраз речевых жанров, которые противопоставляются: 1) по природе самих речевых жанров различных типов (особенно значимым является противопоставление стандартизованных и свободных, прямых и косвенных жанров); 2) по техническим возможностям корпусов; 3) по задачам конкретного жанроведческого исследования (требуется ли идентифицировать жанры речи в речевом потоке или же осуществлять тот или иной количественный анализ текста в рамках уже идентифицированных речевых жанров).

**Ключевые слова:** изучение речевых жанров, корпусная лингвистика, ключевые фразы речевых жанров

### **1. Введение** **Проблематика корпусной генристики**

В статье, во-первых, обсуждается пересечение теории речевых жанров и корпусной лингвистики: корпусная лингвистика участвует в данном пересечении не только своей практической частью, т.е. собственно корпусными методами, — они, как известно, довольно легко заимствуются различными направлениями традиционной лингвистики для решения их специфических задач, — но и теоретиче-

ской частью. О теоретических проблемах корпусной лингвистики, включая ее «философию», см., например: (Баранов, 2003; Копотев, Мустайоки, 2008; Пешё, 1999; Плунгян, 2005; Почепцов, 2001; Bhatia, 2002; Cheng, 2011; Facchinetti, 2007; Lüdeling, Kytö, 2009; McEnery, Hardie, 2012; O'Keeffe, McCarthy, 2010; Rühlemann, 2014; Stenstrom, 2002).

Во-вторых, в статье высказывается несколько предложений методологического характера. Наш опыт работы с жанрами речи, включая изучение их при помощи корпусов, позволяет выделить некоторую структурную основу такой методологии.

С появлением корпусной лингвистики как будто бы сбылась давняя мечта жанроведа: анализировать РЖ количественно, — ибо, в общем, для решения этой задачи недоставало именно того, что и дают сегодня корпуса: главное, появляется надежда получить достаточно материала для всестороннего освещения каждого жанра со всеми значимыми вариациями — а до сих пор практически все «полевые» исследователи соглашались, что на уровне РЖ это невозможно (например: [Седов, 2006:85—97]).

То, что корпуса и корпусные методы должны быть использованы во всех магистральных направлениях лингвистического изучения жанров — текстовом, социокогнитивном и социокритическом, — подчеркивает В.К. Бхатъя: «Количественный лингвистический анализ часто дает возможности выявить закономерности в области формы и функционирования конкретных жанров и сферах использования языка, которые являются неожиданными для исследователя и трудно обнаруживаемыми другими способами» (Bhatia, 2002:13). К сожалению, конкретных способов изучения жанров при помощи корпусов и корпусных методов Бхатъя не предлагает.

В исследованиях речевых жанров: в диссертациях, монографиях, статьях, посвященных различным аспектам и проблемам теории речевых жанров (Дехган, 2015; Колоян, 2006), — совсем не редко используются те или иные возможности корпусной лингвистики, при этом, как представляется, цели и способы обращения к данным возможностям могут быть сведены к нескольким основным типам (к сожалению, среди них лишь небольшое количество отличаются настоящей оригинальностью).

Так, при изучении речевых жанров корпуса и формируемые в их форматах пользовательские подкорпуса могут выступать как база/фонд материала исследования. Используя хранимые и размеченные в корпусах тексты различных жанров, можно изучать соответствующие РЖ (*рассказ, роман, очерк, интервью, фелетон* и мн. др.). Например, в художественном подкорпусе НКРЯ выделяются «жанры текста»: *нежанровая проза, детектив, боевик, детская, историческая проза, приключения, фантастика, любовная история, юмор и сатира, документальная проза, драматургия, перевод* (о некоторых принципах жанрового членения корпуса письменных текстов XIX в. НКРЯ см.: [Оскольская, 2009:46—51]). Исследователи, изучающие на корпусном материале *религиозный* дискурс (например, Балашова, 2015), опираются на жанровое членение, осуществленное составителями корпуса (для НКРЯ — «церковно-богословские жанры»: *беседа, житие, катехизис, молитва, отчетник, описание, паломничества, послание, поучение,*

проповедь, Священное писание, слово), что, конечно, удобно, однако само по себе данное членение не является результатом корпусного анализа речевых жанров. Следует отметить, что этот методологически небезупречный, но соблазнительный в силу своей «уже данности» жанрового членения текстов прием нередко используется при количественном анализе жанров в других типах дискурсов, за пределами собственно корпусов. Например, количественное исследование жанров *внешнеполитического* дискурса (Ярославцева, Дубровская, 2015:182—191) исходит из жанрового членения данного дискурса (*выступления, ответы на вопросы СМИ, интервью, комментарии, remarks* ‘выступления’, *press availability* ‘пресс-конференции’, *interview* ‘интервью’, *press statements* ‘заявления для прессы’) на том основании, что соответствующие рубрики выделены на сайтах МИД РФ ([www.mid.ru](http://www.mid.ru)) и Госдепартамента США ([www.state.gov](http://www.state.gov)).

Кроме того, используя этот же материал, хранящийся в корпусах, можно изучать «жанрово-ролевые сценки», т.е. изображенные в художественных произведениях диалоги персонажей различных жанров (*беседа, болтовня, разговор по душам, комплимент, признание в любви, флирт, ссора, угроза, оскорбление* и мн. др.) (Дементьев, 2008:95—109).

Как видим, названные направления изучения РЖ весьма разнообразны, и вряд ли правомерно и целесообразно говорить, что во всех «используются методы корпусной лингвистики». В этом отношении само по себе обращение к корпусам в качестве источника материала малопоказательно. Так, например, статьи, диссертации, дипломы, в которых во введении имеет место формулировка «на материале НКРЯ» (а это, без преувеличения, сотни работ), при ближайшем рассмотрении оказываются чрезвычайно разнородны — как чрезвычайно разнородны тексты, хранимые в НКРЯ: как известно, это и тексты XX в., и XIX (отчасти и XVIII), и тексты СМИ, и разговорной речи, и диалектной речи.

С нашей точки зрения, о методах корпусной лингвистики как таковых можно говорить только в тех случаях, когда лингвистические задачи решаются количественно и с использованием технических возможностей корпусно-поисковых систем, при этом на первый план выходит проблема «ключевых фраз».

## **2. Поиск ключевых фраз речевых жанров как теоретическая и практическая проблема теории речевых жанров и корпусной лингвистики**

Важность этой проблемы, вынесенной в название статьи, несомненна. О значимости ключевых фраз для ТРЖ писали некоторые исследователи еще до появления корпусов (Шмелева, Шмелев, 2002: 144; Шмелева 1997:88—98).

Действительно, при работе с корпусами проблема таких ключевых фраз, или текстовых маркеров РЖ, по которым одним только и можно осуществлять поиск в корпусах, становится не просто важной, но критической, а главное — исследователю приходится решать данную проблему фактически с нуля, ибо раньше поиском ключевых фраз для конкретных РЖ жанроведы фактически не занимались. В результате, конечно, в анализ будет привнесена доля субъективизма, что в известной степени обесценивает достоинства корпусного подхода.

Приведем пример.

Допустим, предполагается исследование РЖ *роман* на основе текстового корпуса/корпусов. Возможные направления анализа, кажется, ясны: в базах данных различных корпусов хранится (доступно количественному анализу и, как правило, подвергнуто различным типам разметки) большое количество разнообразных романов, исследователь может достаточно легко осуществить целый ряд количественных исследований — например, сравнить несколько романов по ряду показателей, начиная с частотности словоупотребления, формоупотребления и т.д. и заканчивая выявлением авторских и жанровых (*идеологический роман, психологический роман, философский роман*) особенностей стиля.

Задача оказывается гораздо сложнее, если предполагается исследование устных жанров, например *светской беседы* или *разговора по душам*.

Сложности начинаются, собственно, с того, что следует искать в корпусах: понятно, что лексемы и идиомы, наиболее (и почти единственно по-настоящему) доступные обнаружению, выделению и исчислению в корпусах (такие как «*светская беседа*», «*разговор по душам*», а также «*светский*», «*душа*») в этой роли не годятся: они могут вообще не встретиться в даже весьма длинных текстах (записях) соответствующих жанров.

В нашем исследовании (Дементьев, 2010; Дементьев, 2015: 78—107) было показано, как лексемы - названия трех РЖ (*светский, гламурный и по душам*) позволяют найти в корпусах довольно обширную рефлексию (в нашем случае — современных российских журналистов) о данных жанрах, анализ которой позволяет уточнить некоторые аспекты в них.

Но речежанровое исследование, претендующее на адекватность, не может ограничиваться изучением рефлексии над РЖ. Если речь идет о доброкачественном анализе с использованием корпусных методов, на первый план выходит поиск не ключевых лексем, а некоторых ключевых коммуникативно-речевых единиц для конкретных жанров, по которым можно осуществлять формальный поиск и подсчет.

(Собственно, лексема «*роман*» тоже не годится для данной цели, если анализировать некие неизвестные и жанрово немаркированные тексты, еще не размеченные в корпусах. Но, конечно, практически все существующие *романы*, как и другие художественные произведения, уже жанрово маркированы (прежде всего авторами, но нередко и критиками, которым, как известно, иногда «виднее») до размещения в корпусах, и задача жанровой идентификации просто не стоит — стоят совершенно другие задачи «уточнения и прояснения», более традиционные для корпусной лингвистики.)

Как идентифицировать, анализировать и систематизировать («что искать?») *светскую беседу*, пока совершенно неясно. Эта трудность, которую еще только предстоит преодолеть, вполне понятна и даже очевидна на практике. Как это обычно бывает, решение практической задачи сильно облегчается осмыслением изучаемого феномена в теоретическом плане — и здесь несомненно, что у обсуждаемых ключевых фраз РЖ есть общее с речевыми стереотипами, формулами,

клише, коллокациями и т.д., иногда фиксируемыми в фразеологических, идиоматических и т.п. словарях (например: Баранов, Добровольский, 2007; Меликян, 2001; Балакай, 2001), но они далеко не сводятся к ним.

Сложность теоретического характера усиливается тем, что сама постановка данной задачи в жанроведении предполагает определенное противоречие: исследователю предстоит искать конкретные, «материальные» языковые или речевые единицы, которые присущи скорее прецедентным феноменам, чем речевым жанрам, как известно, представляющим собой не сами фразы/тексты, а только их формы (Бахтин, 1996:192); так осуществляется и распознавание жанров в речи (Шмелева, Шмелев, 2002:18; Рогачева, 2011:5). В сущности, это не что иное, как попытка соединить *типологический* подход (чистые структуры, свобода от «материальных» единиц и их конкретных формы-содержания), присущий, как уже неоднократно отмечалось, изучению речевых жанров в целом (Шмелева, 1997: 88—98; Дементьев, 2010), и в большей степени *субстанциональный* подход, присущий, например, сравнительно-историческому языкознанию (как известно, в истории лингвистической науки, в частности лингвистической типологии, таких попыток соединения было немало, положительных результатов же — мало: типологическое изучение отдельных языков и даже отдельных грамматических категорий в них, стадильная типология и т.п.; при этом большинством современников они воспринимались обычно как лишь факультативные «отклонения» см. [Алпатов, 2001:92]). В этой связи следует подчеркнуть, что, хотя изначально в ТРЖ и не предполагался поиск и систематизация подобных «конкретно-материальных» единиц, в современной активно развивающейся ТРЖ уже появились некоторые новые направления, реализующие эту «нетипологическую типологию»: например, изучение жанра как исторического феномена (Шеффер, 2010), противопоставление первичных и вторичных жанров религиозного дискурса на основе тематического ряда (Бобырева, 2007) и т.п. Кроме того, в теоретическом плане «жанровая типология» и «текстовая субстанциональность» осмысляются как два разных *интертекстуальных* аспекта РЖ, а именно: противопоставляются жанровые связи на уровне структур, реализуемые во *вторичных речевых жанрах*, и прецедентные связи на уровне самой речи/текста, реализуемые во *вторичных текстах* (Дементьев, 2015:78—107).

Этим объективным сложностям корпусной генристики как будто бы противоречит тот факт, что существует целый ряд жанров, вопросы идентификации которых и «до корпусов» решались относительно просто, и проблем с поиском материала не возникало (это не значит, конечно, что тут совсем не нужны или бесполезны окажутся корпусные методы): это жанры, идентификация которых осуществляется по *простой языковой (грамматической) форме*: императива (*просьба*), перформатива (*приказ*), инфинитива (*команда*) или, в случае *комплимента* (или, наоборот, *инвективы*), использования характеризующей конструкции с местоименным именованьем адресата речи в функции подлежащего и составным именным сказуемым с предикатной (положительно- или отрицательно-оценочной) лексемой в именной части («*Ты — X*»).

Сюда же следует отнести так называемые *этикетные* РЖ, осуществление которых часто равно произнесению соответствующих формул речевого этикета: *приветствие* («Здравствуйте!»), *благодарность* («Спасибо!») и т.п.

(Здесь следует вспомнить интересное, хотя и небесспорное предложение А.Д. и Е.Я. Шмелевых распознавать жанр *рассказывания анекдота* по ряду формальных признаков: метатекстовый ввод; изобразительные/иконические элементы вербальной и невербальной коммуникации; зачиновый глагол в форме наст. врем., несов. вида (Шмелева, Шмелев, 2002:29—30).

Такие случаи нами не рассматриваются, поскольку, собственно говоря, нет нерешенной проблемы корпусной лингвистики (хотя, например, в случае жанра *комплимента* могут быть различные ключевые фразы типа: «*Любуюсь тобой*», жанра *похвалы* — «*Горжусь тобой*», жанра *просьбы* — «*Не могли бы вы...?*» и «*Можно вас попросить?*» (по форме и функции иногда тоже близкие к этикетным формулам)).

Гораздо больше в данном случае нас интересуют содержательно и/или формально *сложные* жанры (предполагающие варьирование, иногда косвенные), которые — задолго до появления корпусов — обычно и вызывали больше всего сложностей при анализе; до сих пор методика их описания и систематизации далеко не достаточно разработана. Исследователи (Седов, 2006; Седов, 2007; Фенина, 2005) обычно исходили из того, что попытки выявить четкие формальные показатели данных жанров (как и для других подобных «нежестких» жанров), а также сформулировать принципы безэквивалентной семантики данных явлений (например, в терминах универсальных семантических метаязыков) бесперспективны. (Ср. в этой связи весьма интересное исследование Т.В. Лариной и Е.Б. Щелчковой (Ларина, Щелчкова, 2013:73—79), в котором детально разбираются и анализируются случаи неправильного понимания *приглашений*, проистекающие из межкультурных различий.) Понятно, что «охватить» содержание (и форму) данных жанров через одну или несколько счастливо найденных ключевых фраз несколько не менее сложно.

Следует отметить, что в реальной речи таких жанров не просто много — их **б о л ь ш и н с т в о**: содержательно сложных, многомерных, не предполагающих единственно верной интерпретации (причем ее может не быть не только у адресата, но и у адресанта высказывания), что характерно для широко распространенной в реальной речи непланируемой непрямой коммуникации (Дементьев, 2006).

Существуют ли вообще у таких жанров ключевые фразы (в нашем понимании: фиксируемые и пригодные для количественного анализа в корпусах)? И если да, могут ли они быть исчислены и систематизированы, а тем самым — представлены как особая формальная характеристика данных содержательно сложных жанров в описании?

Мы полагаем, что ответ должен быть положительным, а проблема поиска ключевых фраз для идентификации речевых жанров непланируемой непрямой коммуникации (конечно, не только для них) далеко не сводится к методологической проблеме грамотного пользования корпусами. Данная проблема регулярно ставится и решается на практическом уровне за пределами корпусов (и, так ска-

зять, корпусной генристики) при работе с жанрами: как специалистами-жанроведами, имеющими дело с реальным речевым материалом, так и неспециалистами — носителями языка в «живой жизни»; до теоретического же осмысления данной проблемы (которое, как уже было сказано, могло бы сильно помочь практике) дело, как правило, не доходило.

Так, например, очевидно сходство ключевых фраз РЖ с тем, что ищут (и, так сказать, «за что судят») в *лингвистической экспертизе*: эксперты-жанроведы (например: Шарифуллин, 2002:266—276) неоднократно указывали на то, что ТРЖ должна разработать модель опознавания РЖ (инвективных) по формальным признакам, которые только и можно учитывать в экспертизе. Увы, современной ТРЖ до этого еще далеко.

Иллюстрацией второго «неспециального» подхода могут послужить списки «фраз для начала разговора во флирте» (такие рекомендации размещают на страницах изданий разной степени «несерьезности», им посвящают специальные издания, сайты такой направленности множатся в Интернете) наподобие следующих:

**Ср.: Топ 10 флирт фраз**

1. — *Привет, как мы давно с тобой не виделись! — говорите вы прекрасной девушке.*

— *Мы с вами знакомы? — отвечает она.*

— *Да, я тебя знаю, ты «моя мечта!» (Как же девушка может отказаться от такого изумительного имени?)*

2. *Мне так жалко всех этих людей, ведь твоя красота делает их серой массой.*

3. *Здравствуйте, девушка! Скажите, пожалуйста, как вы себя чувствуете, ведь вы — самая красивая девушка в этом городе?*

5. *Догоните понравившуюся вам девушку, по ходу купив цветок, и нежно скажите:*

— *Извините, что преподножу вам только один цветок, но я исправлюсь. В нашу следующую встречу, я подарю букет ваших любимых цветов.*

6. *Девушка, помогите мне! У меня в кармане две путевки на Бали. Я хочу предложить вам совместную поездку, ведь вы так прекрасны.*

7. *Простите, у меня был трудный день, но вдруг я увидел подарок судьбы — ваше лицо. Можно ли полюбоваться вами? (Срабатывает всегда).*

8. *Вы взяли меня в полон своей красотой, и я был бы очень счастлив отбывать пожизненное наказание в плену Вашего сердца.*

9. *Вызовите, пожалуйста, мне скорую помощь. Меня только что подстрелил Амур.*

10. *Вы красивы, как звезда! И освещаете собой серые улицы.*

(<http://www.best-woman.ru/tehnika-flirta-flirt-flirtum-pravilno-top-10-flirt-fraz/>)

Понятно, что фразы эти близки к обсуждаемому нами феномену ключевых фраз речевых жанров. Понятно также, что осмысление данных списков фраз (как и сравнение разных списков) могло бы быть полезно для осмысления речевых особенностей *флирта* — см., например, обзор работ в (Колоян, 2006). Но такое осмысление — дело будущего. (Дополнительная трудность состоит в том, что успешность данных фраз на практике определяется, вероятно, не только и не столько их распространенностью, частотностью, соответствием каким-то нормам, сколько оригинальностью, а значит, скорее, наоборот, редкостью.)

### 3. Классификация ключевых фраз речевых жанров

Сложная лингвистическая и коммуникативная природа феномена столь высокого уровня, как РЖ, не может не породить очень большого количества и очень большого многообразия как отдельных характеристик жанров (это положение давно стало общепризнанным в жанроведении), так и их маркеров, к которым относятся обсуждаемые здесь ключевые фразы.

Рассмотрим данное многообразие и его возможные границы — возможные повторяющиеся типы, а значит, классификацию. В основе разных типов ключевых фраз лежат разные пути выявления таких фраз, требования, предъявляемые к ним: данные требования проистекают, *во-первых*, из природы самих речевых жанров различных типов — здесь особенно значимым является противопоставление стандартизованных и свободных, прямых и косвенных (интерпретативно сложных) жанров. Данное основание обуславливает особые технические характеристики ключевых фраз, выделяемых в корпусах: следует ли осуществлять поиск неизменной при любых обстоятельствах фразы (типа упомянутой идиомы *по душам*) или поиск может/должен быть более гибким и включать варианты, формы слов и фраз (типа *гламур/гламурный* + полные падежные парадигмы данных имен существительного и прилагательного), вплоть до переменных лексем. Но, как это обычно бывает, допущение варьирования формы в итоге приводит к допущению содержательного варьирования.

*Во-вторых*, данные требования проистекают из технических возможностей корпусов, *в-третьих* — из задач конкретного жанроведческого исследования: требуется ли, прежде всего, идентифицировать жанры речи в тексте / речевом потоке или же осуществлять тот или иной количественный анализ текста в рамках уже идентифицированных речевых жанров.

Данные три типа требований, очевидно, порождают три типа ключевых фраз. Однако даже начальный анализ показывает, что данные типы имеют самый общий характер, внутри всех возможно очень большое количество конкретных вполне пригодных для того или иного жанроведческого анализа критериев, соответственно — типов ключевых фраз.

Особенно велико разнообразие последнего типа (два первых, в общем, сводятся к противопоставлению, исходит ли исследователь из особенностей жанра и его природы или из особенностей корпуса): цели и задачи конкретного жанроведческого исследования, начиная с уже названного противопоставления задачи начальной идентификации жанра и «всего остального».

Особенно велико разнообразие этого самого «всего остального»: в рамках уже идентифицированных жанров может быть выделено множество различных дополнительных аспектов, обладающих свойством исчислимости, список которых (особенно у упоминавшихся сложных и косвенных жанров) кажется поистине безграничным, причем, кажется, каждый из таких аспектов способен дать при обращении к корпусам какую-то ценную информацию речевжанрового характера.

Например, для жанра *романа*:

- ◆ количество/процент лексем «X» (исчислению может быть подвергнут текст всего романа и отдельных его частей; текст романа может быть сопоставлен с другими текстами этого же автора разных периодов и текстами других авторов и т.д.; все это касается и следующих аспектов);

- ◆ морфологических форм «Y» (в т.ч. — частей речи);

- ◆ синтаксических конструкций того или иного типа «Z»;

- ◆ стилистически маркированных лексем или идиом и т.д. (Mustajoki, 2006; Vamford, Cavalieri, Diani, 2013; Rühlemann, 2014).

Много в этом плане, как и следовало ожидать, может дать относительно хорошо разработанный в прикладной лингвистике **контент-анализ** (Пешё, 1999; Почепцов, 2001; Шалак, 2004; Krippendorff, 1980, Rosenberg, Schnurr, Oxman, 1990; Weber, 1990). В частности, в контент-анализе традиционно много внимания уделялось «единицам счета», которые могут совпадать со смысловыми единицами или носить специфический характер (на основе анализируемого материала и целей исследования исследователями выдвигались такие единицы счета, как физическая протяженность текстов, площадь текста, заполненная смысловыми единицами, число строк (абзацев, знаков, колонок текста), длительность трансляции по радио или ТВ, метраж пленки при аудио- и видеозаписях, количество рисунков с определенным содержанием, сюжетом и т.д.). Несомненно значим для жанров речи (ср. понятие «темы жанра» М.М. Бахтина) «качественный контент-анализ»: исследователя интересует не столько ЧТО говорится, сколько КАК говорится.

В качестве «единиц счета» могут выступать ключевые фразы речевых жанров, есть возможность и выявить разную роль разных ключевых фраз в разных вариантах и разновидностях одного жанра, однако установить, какие ключевые фразы соответствуют конкретным речевым жанрам, выявить список ключевых фраз контент-анализ, видимо, не помогает (насколько нам известно, исследователи, пользующиеся контент-анализом и разрабатывающие его, вообще практически не занимались специально речежанровой проблематикой (определение жанра или жанровой разновидности, поиск жанра в массиве текстов)).

Таким образом, на основе корпусов может быть осуществлено множество доброкачественных лингвистических (подчеркнем: именно лингвистических!) исследований.

Труднее ответить на вопрос, в какой степени эти возможные исследования будут собственно речежанровыми. Подобные критерии в самых разных вариантах неоднократно использовались при решении неречежанровых задач (безусловно, мы не собираемся отрицать, что во многих случаях они обогащали речежанровую теорию и наше совокупное знание о данном жанре или системе жанров). В то же время ни постановку задач, ни их результаты еще не пытались осмыслить в единой речежанровой теоретической плоскости. Осуществленные исследования и их результаты представляются, с точки зрения данной теории (тем более — с точки зрения будущей «корпусной генристики»), во многом разрозненными и бессистемными. Кроме того, данный подход исходит из аксиоматического признания данности нам жанра, что во многих случаях небесспорно.

Поэтому предлагаем вернуться к критериям идентификации жанров речи в тексте / речевом потоке: собственно, их осмысление и есть то **новое**, чем мы надеемся обогатить генристическую теорию и методику.

Для начала охарактеризуем их в соответствии с обсуждавшимися выше наиболее общими путями выделения ключевых фраз РЖ: исходит ли исследователь из особенностей жанра и его природы или из особенностей корпуса. (На самом деле, конечно, и первое, и второе могут быть выявлены только в результате исследований, в том числе на основе корпусов.)

С этой точки зрения, во-первых, ключевая фраза должна отвечать *семантике* (точнее — *главной семантике*) жанра (поэтому, например, *светский* для СБ не годится, а *по душам* для РпД годится); во-вторых, в системе по этому критерию можно реально искать и находить *большое количество примеров* (так, для жанра *просьбы* в качестве ключевых фраз для идентификации жанра, вероятно, могут быть выделены две формулы: «*Не могли бы вы...?*» и «*Можно вас попросить?*», но при этом первая дает 306 вхождений в НКРЯ, а вторая — всего 10).

Задачи такого рода еще далеко не решены (они пока, в большинстве случаев, еще даже не ставились), а многочисленные сложности, связанные с их решением, еще только предстоит преодолеть, — мы здесь обрисовали лишь предварительный список таких сложностей; понятно, что в действительности их намного больше.

В случае же успешного решения поставленных задач мы получим *существенную характеристику жанра*, которая присутствует в речезанровой компетенции носителя языка и должна быть включена в соответствующие жанроведческие описания. С точки зрения данного описания наличие той или иной ключевой фразы/фраз представляет собой что-то вроде дополнительного параметра к разнообразным моделям РЖ типа известной «анкеты РЖ» Т.В. Шмелевой (Шмелева, 1997).

(Возвращаясь же к общему определению речезанрового подхода как типологического, обсуждаемая характеристика — наличие той или иной ключевой фразы/фраз речевого жанра — сродни известной типологической идее Владимира Скалички (Скаличка, 1996:22—30): для изучения отдельного языка использовать типологические характеристики, разработанные и традиционно используемые (поскольку они типологические) принципиально НЕ для отдельных языков.)

Характеристики жанра такого рода (например, в будущих «энциклопедиях речевых жанров», о которых говорит Т.В. Шмелева), добавляемые к характеристикам, сформулированным в результате других, более традиционных жанроведческих исследований, могли бы выглядеть примерно так (табл. 1):

Таблица 1

**Ключевые фразы речевых жанров**  
**Key phrases of speech genres**

Жанр	Наличие ключевой лексики	Ключевая фраза / ключевые фразы (для идентификации жанра)	Количество вхождений в корпус (какой)	Ключевые фразы (для исследования контекста)	Количество вхождений в корпус (какой)
X	A <sub>1</sub>	B <sub>1</sub>	n <sub>1</sub>	C <sub>1</sub>	N <sub>1</sub>
	A <sub>2</sub>	B <sub>2</sub>	n <sub>2</sub>	C <sub>2</sub>	N <sub>2</sub>
	...	B <sub>3</sub>	n <sub>3</sub>	C <sub>3</sub>	N <sub>3</sub>
		...	n...	...	N...

Например, для РЖ *просьба* (табл. 2):

Таблица 2

**Ключевые фразы для речевого жанра просьба**  
**Key phrases of the request speech genre**

Жанр	Наличие ключевой лексемы	Ключевая фраза / ключевые фразы (для идентификации жанра)	Количество вхождений в корпус (НКРЯ)	Ключевые фразы (для исследования контекста)	Количество вхождений в корпус (НКРЯ)
Просьба	<i>Прошу</i>	<i>Не могли бы вы</i>	306	<i>Буду очень благодарен</i>	32
	<i>Пожалуйста</i>	<i>Можно вас попросить</i>	10	<i>Для меня это очень важно</i>	15
	<i>будьте добры</i>				

Заявленное нами направление поиска, естественно, пересекается с традиционной для лингвистики теоретической и практической проблемой (для многих устных, еще больше — письменных жанров уже неоднократно ставившейся и решавшейся) конкретного *лексического и грамматического наполнения* текстов в рамках тех или иных жанров. Здесь, с одной стороны, открывается особенно обширное поле деятельности для исследователя-лингвиста, с другой — обширное поле возможностей для классификации (гораздо более, так сказать, лингвистической, чем многие традиционные речежанровые классификации).

Например:

- ◆ *вы не могли бы...?* — вежливая просьба,
- ◆ *я не могу жить без тебя!* — признание,
- ◆ *все будет хорошо* — утешение,
- ◆ *ты как всегда начинаешь... / хочешь... / забыл...* — упрек,
- ◆ *слушай я вот подумал...* — сентенция (?),
- ◆ *посмотрим, до чего ты так дойдешь / дождешься / договоришься* — ворчание,
- ◆ *я ведь тебя предупреждал(а)...!* — (?).

Большое место, вполне логично, здесь будет принадлежать *императивам* (при этом семантическое разнообразие и разнотипность данных формул и стоящих за ними типичных ситуаций очевидны уже из этих небольших предварительных списков): *смотри, представь (себе), подумай (о...), ответь, подожди, ступай, перестань, отстань, заткнись, успокойся, расслабься, согласись, поклянись, одумайся*, при этом показательно большое количество с «не» среди них: *не сердись, не переживай, не расстраивайся, не торопись, не ври, не хами, не перебивай, не спорь, не хвастайся, не спи, не вредничай*.

Одно из наиболее естественных оснований для классификации ключевых фраз речевых жанров — *сферы коммуникации*.

Данное основание классификации вполне очевидно; главная же и тоже очевидная трудность связана с тем противоречием, что большинство найденных нами ключевых фраз, как видим, относятся к разговорной устной сфере, которая хуже всего представлена в корпусах.

При работе же с письменными текстами (и жанрами) сама необходимость поиска ключевых фраз представляется на порядок менее критичной, поскольку гораздо лучше разработана (хотя и такая необходимость есть).

Так, в докторской диссертации С.П. Кушнерука (Кушнерук, 2008), посвященной лингвистическому изучению документного текста, рассчитывается целый ряд коэффициентов — близость их к характеристикам высокоформализованных речевых жанров (официально-деловых, технических и подоб.), которые могут быть выявлены количественными методами, в том числе с помощью корпусов, очевидна, особенно (4) и (7):

1) коэффициенты лексического разнообразия и единичной встречаемости для документных и недокументных текстов (Кушнерук, 2008:119),

2) отношение числа единиц, характеризующих рост словаря документных текстов, к размеру документного текста (Там же, 2008:127—129),

3) коэффициент терминированности текста (Там же, 2008:137),

4) доля разных единиц (словосочетание, предложение, абзац, текст) в качестве документных формул (Там же, 2008:157—158),

5) распределительные параметры терминопотреблений для терминов различной компонентной сложности (Там же, 2008:181),

6) доля существительных в документных текстах (соотносимая с их уровнем предметности), глаголов (соотносимая с их степенью процессуальности), прилагательных (признаковость) (Там же, 2008:187—188),

7) степень синтаксической устойчивости и однородности документных текстов с точки зрения реализуемости в них основных видов предложений (Кушнерук, 2008:196—197),

8) степень унифицированности документных текстов на лексико-фразеологическом уровне (Там же, 2008:207).

В итоге внимание исследователя с неизбежностью сосредоточивается: 1) на художественных, прежде всего драматических текстах, наиболее подробно воспроизводящих разговорную речь, но тем самым — и искусственных, стилизованных (воспроизведение ими разговорности весьма специфично); 2) на текстах *электронной коммуникации*, в отличие от художественных, реальных, а не искусственных, но совершенно очевидно имеющих свою очень существенную специфику (в сравнении с устной, и с письменной неинтернет-коммуникацией) (ср. наше исследование англицизмов — имен жанров интернет-коммуникации [Дементьев, 2016: 7—24]). Следовательно, необходимы поиски и специфических методик — но это скорее плюс, ибо ставит перспективы.

#### 4. Заключение

Как видим, поставленная в начале статьи проблема жанроведения, изначально возникшая из чисто практических потребностей (фактически это были технические параметры поиска, осуществляемого человеком-исследователем при работе с механическим онлайн-корпусом), имеет и важный теоретический аспект, пересекающийся с общей теоретической проблемой природы РЖ, механизмов их обнаружения и порождения (см. приводившиеся ключевые фразы *флирта*), условий бытования в реальной коммуникации, особенно в письменной официально-деловой: там не только идентификация жанра, но и само его существование часто определяются наличием/отсутствием ряда шаблонных фраз (Кушнерук, 2008).

Дальнейшие направления исследований могут быть связаны, с одной стороны, с выявлением конкретных ключевых фраз для конкретных жанров различных типов и их апробированием на материале корпусов. Это направление наиболее очевидно, два его главных практических результата (конечно, в идеале) — систематизация корпусных данных и дополнение данным (важным) пунктом соответствующих описаний жанров в энциклопедиях РЖ (тоже будущих).

С другой стороны, последовательное решение поставленных задач и практического, и теоретического характера предполагает выявление в жанровом и околожанровом пространстве других феноменов, которые имеют онтологическую близость с ключевыми фразами (и корпусами), соответственно работа в обозначенном направлении может помочь пролить некоторый свет на данные феномены — например, вкратце рассмотренные здесь шутки и анекдоты, обыгрывающие неправильную интерпретацию коммуникативной тональности.

© Дементьев В.В., Степанова Н.Б., 2016

### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- Алпатов В.М. *История лингвистических учений*. М.: Языки славянской культуры, 2001. 251 с.
- Антология речевых жанров: повседневная коммуникация* / Под ред. К.Ф. Седова. М.: Лабиринт, 2007. 320 с.
- Балакай А.Г. *Словарь русского речевого этикета*. М.: АСТ-ПРЕСС, 2001. 672 с.
- Балашова Е.Ю. *Роль телеономных концептуальных полей в организации религиозного дискурса*: Автореф. дис. ... докт. филол. наук. Саратов, 2015.
- Баранов А.Н. Корпусная лингвистика // Баранов А.Н. *Введение в прикладную лингвистику*. М.: Эдиториал УРСС, 2003. С. 112—137.
- Баранов А.Н., Добровольский Д.О. (ред.) *Словарь-тезаурус современной русской идиоматики* / Под ред. А.Н. Баранова, Д.О. Добровольского. М.: Мир энциклопедий Аванта+, 2007. 1135 с.
- Бахтин М.М. Проблема речевых жанров. Из архивных записей к работе «*Проблема речевых жанров*». Проблема текста // Бахтин М.М. Собр. соч.: В 5 т. М.: Языки русской культуры, 1996. Т. 5. Работы 1940-х — начала 1960-х годов. С. 159—206.
- Бобырева Е.В. *Религиозный дискурс: ценности, жанры, стратегии (на материале православного вероучения)*. Волгоград: Перемена, 2007.
- Дементьев В.В. *Непрямая коммуникация*. М.: Гнозис, 2006. 376 с.
- Дементьев В.В. Аспекты проблемы «жанр и культура» // *Жанры речи*. Саратов: ИЦ «Наука», 2007. Вып. 5. Жанр и культура. 440 с. С. 5—20.
- Дементьев В.В. Жанрово-ролевые сценки: обоснование подхода к составлению энциклопедии речевых жанров // *Жанры и типы текста в научном и медийном дискурсе*. Орел: Орловский гос. ин-т искусств и культуры, 2008. № 6. С. 95—109.
- Дементьев В.В. *Теория речевых жанров*. М.: Знак, 2010. 600 с.
- Дементьев В.В. Теория речевых жанров и актуальные процессы современной речи // *Вопросы языкознания*. 2015. № 6. С. 78—107.
- Дементьев В.В. Интертекстуальный аспект речевых жанров // *Жанры речи*. 2015. Т. 12. № 2. С. 9—26.
- Дементьев В.В. Некоторые методологические принципы изучения речжанровых заимствований // *Вестник Российского университета дружбы народов. Серия: Лингвистика*. 2016. Т. 20. № 1. С. 7—24.

- Дехган Халили Можде. Речевой жанр «просьба» в современной русской коммуникации: Автореф. дис. ... канд. филол. наук. М., 2015.
- Колоян Д.Л. Ухаживание как тип коммуникативного поведения: Дис. ... канд. филол. наук. Волгоград, 2006.
- Копотев М., Мустайоки А. Современная корпусная русистика // *Инструментарий русистики: корпусные подходы*. Хельсинки: Helsinki University Press, 2008. (Slavica Helsingiensia. № 34). С. 7—24.
- Кремшюкалова М.Ч. Когнитивно-дискурсивная модель малых жанров устной речи (на материале благопожеланий и проклятий): Автореф. дис. ... докт. филол. наук. Нальчик, 2014.
- Кушнерук С.П. Теория современного документного текста: Дис. ... докт. филол. наук. Волгоград, 2008.
- Ларина Т.В., Щелчкова Е.Б. Речевой акт *Приглашение* и проблемы понимания: межкультурный аспект // *Вестник Новосибирского государственного университета*. Сер.: Лингвистика и межкультурные коммуникации. 2013. Т. 11. № 2. С. 73—79.
- Меликян В.Ю. Словарь: Эмоционально-экспрессивные обороты живой речи. М.: Флинта; Наука, 2001. 240 с.
- Оскольская С.А. Корпус письменных текстов XIX века: сферы употребления и жанровое разнообразие // *Национальный корпус русского языка: 2006—2008*. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 46—51.
- Пешё М. *Контент-анализ и теория дискурса* // Квадратура смысла. М.: Прогресс, 1999. С. 302—336.
- Плунгян В.А. Зачем нужен Национальный корпус русского языка? Неформальное введение // *Национальный корпус русского языка: 2003—2005*. М.: Индрик, 2005.
- Плунгян В.А., Сичинава Д.В. Национальный корпус русского языка: опыт создания корпусов текстов современного русского языка // Л.Н. Беляева и др. (ред.). *Труды Межд. конференции «Корпусная лингвистика-2004»*. СПб.: СПбГУ, 2004. С. 216—238.
- Почепцов Г.Г. *Теория коммуникации*. М.: Рефл-бук, 2001. 656 с.
- Рахилина Е.В. Корпус как творческий проект // *Национальный корпус русского языка: 2006—2008*. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 7—26.
- Рогачева Н.Б. Структура и функционирование вторичных речевых жанров интернет-коммуникации (на материале русского и английского языков): Автореф. дис. ... канд. филол. наук. Саратов, 2011.
- Седов К.Ф. Жанры «праздноречевой» коммуникации: *болтовня, светская беседа, разговор по душам* // *Проблемы речевой коммуникации*. Саратов: Изд-во Сарат. ун-та, 2006. № 6. С. 85—97.
- Скаличка В. К вопросу о типологии // *Вопросы языкознания*. 1966. № 4. С. 22—30.
- Федосюк М.Ю. Нерешенные вопросы теории речевых жанров // *Вопросы языкознания*. 1997. № 5. С. 102—120.
- Фенина В.В. Речевые жанры *small talk* и *светская беседа* в англо-американской и русской культурах: Автореф. дис. ... канд. филол. наук. Саратов, 2005.
- Шалак В.И. *Современный контент-анализ*. М.: Омега-А, 2004. 272 с.
- Шарифуллин Б.Я. Коммуникативные стратегии и тактики инвективы: семейный скандал // *Языковая ситуация в России начала XXI века: Материалы Международной научной конференции* (КемГУ). Кемерово: КемГУ, 2002. Т. 2. С. 266—276.
- Шеффер Ж.-М. *Что такое литературный жанр?* М.: Едиториал УРСС, 2010. 192 с.

- Шмелева Е.Я., Шмелев А.Д. *Русский анекдот: текст и речевой жанр*. М.: Языки славянской культуры, 2002. 144 с.
- Шмелева Т.В. Модель речевого жанра // *Жанры речи*. Саратов: Колледж, 1997. № 1. С. 88—98.
- Ярославцева Я.А., Дубровская Т.В. Конструирование межнациональных отношений в жанрах внешнеполитического дискурса (на примере России и США) // *Политическая лингвистика*. 2015. Т. 53. № 3. С. 182—191.
- Bamford, J., Cavalieri, S., Diani, G. (2013). *Variation and Change in Spoken and Written Discourse: Perspectives from corpus linguistics*. John Benjamins Publishing Company.
- Bhatia, Vijay K. (2011). Applied genre analysis: a multi-perspective model, *Ibérica*, 4, 3—19.
- Cheng, W. (2012). *Exploring Corpus Linguistics: Language in Action*. London: Routledge.
- Facchinetti, R. (2007). *Corpus Linguistics 25 Years on*. New York, Amsterdam: Rodopi.
- Integrum: точные методы и гуманитарные науки / Г. Никипорец-Такигава. М.: Летний сад, 2006. С. 50—76.
- Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Newbury Park, CA: Sage.
- Lüdeling, A., Kytö, M. (2009). *Corpus Linguistics. An International Handbook*. Walter de Gruyter.
- McEnery, T., Hardie, A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press, 2012. 312 p. (Cambridge Textbooks in Linguistics).
- Mustajoki, A. The Integrum Database as a Powerful Tool in Research on Contemporary Russian // *Integrum: точные методы и гуманитарные науки* / Г. Никипорец-Такигава. М.: Летний сад, 2006. Р. 50—76.
- O'Keefe, A., McCarthy, M. (2010). *The Routledge Handbook of Corpus Linguistics*. London: Routledge.
- Rosenberg, S.D., Schnurr, P.P., & Oxman, T.E. (1990). Content analysis: A comparison of manual and computerized systems. *Journal of Personality Assessment*. 54, (1 & 2), 298—310.
- Rühlemann, C. (2014). *Narrative in English Conversation: A Corpus Analysis of Storytelling*. Cambridge University Press.
- Stenstrom, A.-B. et al. (2002). *Trends in Teenage Talk: Corpus Compilation, Analysis and Findings*. John Benjamins
- Weber, R.P. (1990). *Basic Content Analysis* (2d ed.). Newbury Park: CA.

#### **История статьи:**

Дата поступления в редакцию: 15 мая 2016

Дата принятия к печати: 01 июня 2016

#### **Для цитирования:**

**Дементьев В.В., Степанова Н.Б. Корпусные методы в исследовании речевых жанров: проблема ключевых фраз // Вестник Российского университета дружбы народов. Серия: Лингвистика. 2016. Т. 20. № 3. С. 57—76.**

#### **Сведения об авторах:**

*Дементьев Вадим Викторович*, доктор филол. наук, профессор, профессор кафедры теории, истории языка и прикладной лингвистики Саратовского национального исследовательского государственного университета имени Н.Г. Чернышевского. *Сфера научных интересов*: теория дискурса, коммуникативная аксиология, теория речевых жанров. Автор более 106 публикаций, среди которых монографии (6), учебники, научные статьи в российских и зарубежных изданиях. *Контактная информация*: e-mail: dementevvv@yandex.ru.

Степанова (Рогачева) Наталья Борисовна, кандидат филол. наук, доцент кафедры теории, истории языка и прикладной лингвистики Саратовского национального исследовательского государственного университета имени Н.Г. Чернышевского. *Сфера научных интересов*: теория речевых жанров, интернет-коммуникация, языковая личность. Автор ряда научных статей. *Контактная информация*: e-mail: gaerven@yahoo.com.

## **CORPUS METHODS IN THE STUDY OF SPEECH GENRES: A PROBLEM OF KEY PHRASES**

**Vadim V. Dementyev, Natalia B. Stepanova**

Saratov State University

83, Astrakhanskaya Str., 410012, Saratov, Russia

The article discusses the theoretical and practical problems related to the study of speech genres on the basis of the material of databases, and generally using the methods of corpus linguistics. The richness, diversity and undoubted effectiveness of corpus methods for solving various linguistic tasks convince that such methods should be also used in linguistic studies of speech genres. At the same time, there is no developed corpus techniques for solving genre problems in linguistics.

The role that should corpus methods play in general studies of speech genres is not clear. Also, most importantly, the relation between the units that are used / available in corpus databases, and speech genre is not clear. Methods of corpus linguistics exist only in cases where linguistic problems are solved with the use of quantitative and technical capabilities of corpus systems.

In the foreground there is the problem of “key phrases”, or text markers of speech genres, which can be searched in the case. According to the authors, the key phrase of speech genre is its essential characteristic, which is present in speech genre competence of a native speaker and should be included in the relevant speech genre description as an additional parameter. Genre characteristics of this kind, in the future “encyclopedias of speech genres”, may be added to the characteristics set forth as a result of other, more traditional speech genre research. They might look like this: the genre — the presence of the key lexeme — key phrase / key phrases (for the identification of the genre) — the number of accesses to corpus — the key phrase (for the research of the context) — the number of accesses to corpus.

The main requirements for key phrases are formulated: 1) a key phrase must have the main semantics of the genre; 2) in the corpus database system according to this criterion, you can actually search for and find many examples.

An initial classification of the key phrases of speech genres has been proposed, that are opposed to themselves by 1) the nature of speech genres themselves of various types (particularly meaningful is the juxtaposition of standardized and free, direct and indirect genres); 2) the technical capabilities of buildings; 3) specific tasks of a speech genre study (whether to identify the genre of speech in the speech flow, or to carry out one or other quantitative analysis of the text within the already identified speech genres).

**Key words:** studying speech genres, corpus linguistics, key phrases of speech genres

### **REFERENCES**

- Alpatov, V.M. (2001). *Istoriya lingvisticheskikh uchenii*. [History of Linguistic Studies]. Moscow: Yazyki Slavianskoy Kul'tur (In Russian).
- Antologiya rechevykh zhanrov: povsednevnaya kommunikatsiya* (2007) [Anthology of Speech Genres]. In K.F. Sedov (ed.). Moscow: Labirint (In Russian).
- Balakai, A.G. *Slovar' russkogo rechevogo etiketa*. (2001). [Russian Speech Courtesy Dictionary]. Moscow: AST-PRESS. (In Russian).

- Balashova, Ye.Yu. (2015). *Rol' teleonomnykh kontseptual'nykh poley v organizatsii religioznogo diskursa* [Role of Teleonomic Conceptual fields in Religious Discourse]. Saratov (In Russian).
- Baranov, A.N. (2003). *Korpusnaya lingvistika* [Corpus Linguistics]. In A.N. Baranov (ed.). *Vvedeniye v prikladnuyu lingvistiku* [Introduction to Applied Linguistics] (pp. 112—173). Moscow: Editorial URSS. (In Russian).
- Baranov, A.N., Dobrovolskii, D.O. (2007). *Slovar'-tezaurus sovremennoy russkoi idiomatiki* [Dictionary-Thesaurus of Modern Russian Idioms]. In A.N. Baranova, D.O. Dobrovolskogo (Eds.). Moscow: Mir entsiklopediy Avanta+, 2007. (In Russian).
- Bakhtin, M.M. (1996). *Problema rechevykh zhanrov* [Problems of Speech Signs] V. 5 (pp. 159—206). Moscow: Yazyki russkoi kul'tury. (In Russian).
- Bamford, J., Cavalieri, S., Diani, G. (2013). *Variation and Change in Spoken and Written Discourse: Perspectives from corpus linguistics*. John Benjamins Publishing Company.
- Bhatia, V. K. (2011). Applied genre analysis: a multi-perspective model, *Ibérica*, 4, 3—19.
- Cheng, W. (2012). *Exploring Corpus Linguistics: Language in Action*. London: Routledge.
- Bobyreva, Ye.V. *Religioznyy diskurs: tsennosti, zhanry, strategii (na materiale pravoslavnogo veroucheniya)* [Religious Discourse: values, genres, strategies in Orthodoxy]. Volgograd: Peremena. (In Russian).
- Dement'ev, V.V. (2006). *Nepryamaya kommunikatsiya* [Indirect Communication]. Moscow: Gnozis. (In Russian).
- Dement'ev, V.V. (2007). Aspekty problemy «zhanr i kul'tura» [Aspect of a Problem “Genre and Culture”]. *Zhanry rechi* [Speech Genres]. Saratov: ITS «Nauka», 5, 5—20. (In Russian).
- Dement'ev, V.V. (2008). Zhanrovo-rol'nyye stsenki: obosnovaniye podkhoda k sostavleniyu entsiklopedii rechevykh zhanrov [Genre-role Scenes: approach to compilation of Speech Signs encyclopedia]. *Zhanry i tipy teksta v nauchnom i mediyonom diskurse*. Orol: Orlovskiy gos. in-t iskusstv i kul'tury, 6, 95—109. (In Russian).
- Dement'ev V.V. (2010). *Teoriya rechevykh zhanrov* [Theory of Speech Signs]. Moscow: Znak. (In Russian).
- Dement'ev V.V. (2015). *Teoriya rechevykh zhanrov i aktual'nye protsessy sovremennoy rechi* [Theory of Speech Signs and Actual Processes of Modern Speech]. *Voprosy yazykoznaneya* [Questions of Linguistics], 6, 78—107. (In Russian).
- Dement'yev, V.V. (2015). Intertekstual'nyi aspekt rechevykh zhanrov [Intertextual Aspect of Speech Genres]. *Zhanry rechi* [Speech Genres]. 2(12), 9—26. (In Russian).
- Dement'yev, V.V. (2016). Nekotorye metodologicheskie printsipy izucheniya rechezhanrovnykh zaimstvovaniy [Some Methodological Principles of Study on Borrowing Speech Genres]. *Russian Journal of Linguistics*, 20 (1), 7—24. (In Russian).
- Dekhgan, K. M. (2015). *Rechevoy zhanr «pros'ba» v sovremennoi russkoi kommunikatsii* [Request as Speech Genre in Modern Russian Communication]. Moscow. (In Russian).
- Facchinetti, R. (2007). *Corpus Linguistics 25 Years on*. New York, Amsterdam: Rodopi.
- Fedosyuk, M.Yu. (1997). *Nereshennyye voprosy teorii rechevykh zhanrov* [Unresolved Questions of Speech Genres Theory]. *Voprosy yazykoznaneya*, 5, 102—120. (In Russian).
- Fenina, V.V. (2005). *Rechevyye zhanry small talk i svetskaya beseda v anglo-amerikanskoi i russkoi kul'turakh* [Speech Genre of small talk in Anglo-american and Russian Culture]. Saratov. (In Russian).
- Koloyan, D.L. (2006). *Ukhazhivaniye kak tip kommunikativnogo povedeniya* [Courtship as type of Communication Behaviour]. Volgograd. (In Russian).

- Kopotev, M., Mustayoki, A. (2008). *Sovremennaya korpusnaya rusistika* [Modern Corpus Russian Studies]. *Instrumentariy rusistiki: korpusnyye podkhody*. Khel'sinki: Helsinki University Press, Slavica Helsingiensia. 34, 7—24. (In Russian).
- Kremshokalova, M.Ch. (2014). *Kognitivno-diskursivnaya model' malykh zhanrov ustnoi rechi (na materiale blagopozhelanii i proklyatii* [Cognitive and Discourse Model of Small Oral Speech Genres on the material of Wishes and Damnations]. *Nal'chik*. (In Russian).
- Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Newbury Park, CA: Sage.
- Kushneruk, S.P. (2008). *Teoriya sovremennogo dokumentnogo teksta* [Theory of Modern Document Text]. Volgograd. (In Russian).
- Larina T.V., Shchelchkova Ye.B. (2013). Rechevoy akt *Priglaseniye* i problemy ponimaniya: mezhkul'turnyi aspekt [Speech Act „Invitation” and Problem of Intercultural Aspect]. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Ser.: Lingvistika i mezhkul'turnye kommunikatsii*. 11 (2), 73—79. (In Russian).
- Lüdeling, A., Kytö, M. (2009). *Corpus Linguistics. An International Handbook*. Walter de Gruyter.
- McEnery, T., Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- Melikyan, V.Yu. (2001). *Slovar': Emotsional'no-ekspressivnyye oboroty zhivoi rechi* [Dictionary of Emotionally-expressive Speech Turns]. Moscow: Flinta; Nauka. (In Russian).
- Mustajoki, A. (2006). The Integrum Database as a Powerful Tool in Research on Contemporary Russian. *Integrum: exact methods and humanities*. In G. Nikiporets-Takigava (ed.). Moscow: Letniy Sad, 50—76.
- O'Keeffe, A., McCarthy, M. (2010). *The Routledge Handbook of Corpus Linguistics*. London: Routledge.
- Oskol'skaya, S.A. (2009). Korpus pis'mennykh tekstov XIX veka: sfery upotrebleniya i zhanrovoye raznoobrazie [Corpus of Written Texts of XIX century: performance field and genres diversity]. *Natsional'nyi korpus russkogo yazyka: 2006-2008. Novyye rezultaty i perspektivy*. SPb.: Nestor-Istoriya, 46—51. (In Russian).
- Pesho, M. (1999). Kontent-analiz i teoriya diskursa [Content Analysis and Discourse Theory]. *Kvadratura smysla*. Moscow: Progress, 302—336. (In Russian).
- Plungyan, V.A. (2005). Zachem nuzhen Natsional'nyi korpus russkogo yazyka? Neformal'noye vvedenie [Why do we need National Russian Language Corpus? Informal Introduction]. *Natsional'nyy korpus russkogo yazyka: 2003—2005*. Moscow: Indrik. (In Russian).
- Plungyan V.A., Sichinava D.V. (2004). Natsional'nyy korpus russkogo yazyka: opyt sozdaniya korpusov tekstov sovremennogo russkogo yazyka [National Russian Language Corpus: Experience of Corpus Text Creation in Modern Russian]. In L.N. Belyayeva (ed.). *Korpusnaya lingvistika-2004* [Corpus Linguistics]. SPb: SPbGU, 216—238. (In Russian).
- Pochepstov G. G. (2001). *Teoriya kommunikatsii* [Theory of Communication]. Moscow: Refl-buk. (In Russian).
- Rakhilina, Ye.V. (2009). Korpus kak tvorcheskiy proyekt [Corpus as Creative Project]. *Natsional'nyy korpus russkogo yazyka: 2006-2008. Novyye rezultaty i perspektivy*. SPb.: Nestor-Istoriya, 7—26. (In Russian).
- Rogachova, N.B. (2011). *Struktura i funkcionirovaniye vtorichnykh rechevykh zhanrov internet-kommunikatsii (na materiale russkogo i angliyskogo yazykov)* [Structure and Functioning of Secondary Speech Genres in English and Russian Internet Communication]. Saratov. In Russian.

- Rosenberg, S.D., Schnurr, P.P., & Oxman, T.E. (1990). Content analysis: A comparison of manual and computerized systems. *Journal of Personality Assessment*. 54, (1 & 2), 298—310.
- Rühlemann, C. (2014). *Narrative in English Conversation: A Corpus Analysis of Storytelling*. Cambridge University Press.
- Sedov, K.F. (2006). Zhanry «prazdnorechevoy» kommunikatsii: boltovnya, svetskaya beseda, razgovor po dusham [Genre of Everyday Communication: idle talk, small talk, intimate conversation]. *Problemy rechevoy kommunikatsii* [Problems of Speech Communication]. Saratov: Izd-vo Sarat. un-ta. 6, 85—97. (In Russian).
- Shalak, V.I. (2004). *Sovremennyy kontent-analiz* [Modern Content Analysis]. Moscow: Omega-A. (In Russian).
- Sharifullin, B.Ya. (2002). Kommunikativnye strategii i taktiki invektivny: semeinyi skandal [Communicative Strategies and Invektive tactics: family scandal]. *Yazykovaya situatsiya v Rossii nachala XXI veka*. Kemerovo: KemGU. V. 2, 266—276. (In Russian).
- Sheffer, Zh.-M. (2010). Chto takoye literaturnyi zhanr? [What is Literature Genre?]. Moscow: Editorial URSS. (In Russian).
- Shmeleva, Ye.Ya., Shmelev, A.D. (2002). Russkiy anekdot: tekst i rechevoy zhanr [Russian Anecdote: text and speech genre]. Moscow: Yazyki slavyanskoi kul'tury. (In Russian).
- Shmeleva, T.V. Model' rechevogo zhanra [Speech Genre Model]. *Zhanry rechi*. Saratov: Kolledzh, 1, 88—98. (In Russian).
- Skalička, V. (1966). K voprosu o tipologii [To the Problem of Typology]. *Voprosy yazykoznaniya*. 4, 22—30. (In Russian).
- Stenstrom, A.-B. et al. (2002). *Trends in Teenage Talk: Corpus Compilation, Analysis and Findings*. John Benjamins
- Weber, R.P. (1990). *Basic Content Analysis* (2d ed.). Newbury Park: CA.
- Yaroslavtseva Ya.A., Dubrovskaya T.V. Konstruirovaniye mezhnatsional'nykh otnoshenii v zhanrakh vneshnepoliticheskogo diskursa (na primere Rossii i USA) [Construction of International Relationship in Political Discourse Genres]. *Politicheskaya lingvistika*. [Political Linguistics]. 53 (3), 182—191.

**Article history:**

Received: 15 May 2016

Revised: 25 May 2016

Accepted: 1 June 2016

**Благодарности и финансирование:**

Работа выполнена при финансовой поддержке Минобрнауки России в рамках базовой части государственного задания в сфере научной деятельности по заданию № 2014/203, код проекта 1549.

**Acknowledgements:**

The article is published with financial support from Ministry of Education and Science of the Russian Federation within base part of state order in research field according to the Order N. 2014/203, project code 1549.

**For citation:**

Demytyev, V., Stepanova, N. (2016). Corpus Methods in the Study of Speech Genres: A Problem of Key Phrases. *Russian Journal of Linguistics*, 20 (3), 57—76.

**Bio Note:**

*Dementyev Vadim Viktorovich*, Doctor of Philology, Professor of the Department of Language Theory and History, and Applied Linguistics of Chernyshevsky Saratov State University, member of the editorial board of Russian journal of linguistics — *Vestnik RUDN* (Peoples' Friendship University of Russia, Moscow). *Research interests*: the theory of discourse, communicative axiology, theory of speech genres. *Contact information*: e-mail: dementevvv@yandex.ru.

*Stepanova (Rogachova) Natalya Borisovna*, Candidate of Philology, Associate Professor of the Department of Language Theory and History, and Applied Linguistics of Chernyshevsky Saratov State University. *Research interests*: the theory of speech genres, online communication, language personality. *Contact information*: e-mail: gaerven@yahoo.com.