

Тексты ограниченного обращения в составе Национального корпуса русского языка

Обсуждаются вопросы, связанные с классификацией и описанием в Национальном корпусе русского языка текстов, создаваемых и используемых людьми в их повседневной жизни, которые обычно нерегулярно представлены в больших корпусах, — документация, листовки, буклеты и т. п. (ephemera), корреспонденция, рекламно-информационные и методические тексты. Описывается текущее состояние раздела корпуса, содержащего эти виды текстов — принципы выделения, источники текстов, наполнение, разнообразие типов текстов, принятая в настоящее время классификация или признаки метаописания, проблемы описания, актуальные задачи.

1. ВВЕДЕНИЕ

Для описания текстов в Национальном корпусе русского языка (НКРЯ) мы опираемся на две классификации: современную европейскую EAGLES [1], которая далее называется “формальной классификацией”, и традиционную отечественную типологию текстов, которая далее называется “функциональной классификацией”. Формальная классификация приложима к корпусу в целом и позволяет описывать большие группы и жанры текстов [2]. С ее помощью можно обозначить свойства больших групп текстов, в частности группы текстов, создаваемых и используемых людьми в их повседневной жизни, отличных от книг и периодических изданий, которые составляют основу национальных корпусов. Эту группу мы далее называем “Текстами ОГраниченного обращения” (ТОГ). Верхний уровень формальной классификации представлен четырьмя видами коммуникации: устная (spoken), письменная (written), письменная для устного произнесения (written-to-be-spoken) и электронная (electronic). Письменная коммуникация осуществляется в следующих формах: печатные издания (printed material), переписка (correspondence), машинописные материалы (typed material) и рукописные тексты (manuscripts). Среди печатных изданий различаются такие виды, как книги, газеты, журналы и “прочее” (ephemera), куда включаются брошюры, листовки, буклеты, вкладыши, проспекты, памятки и под.

Объектом рассмотрения в настоящей статье является совокупность письменных текстов, включающая “прочие” виды печатной продукции (ephemera), а также тексты, относящиеся к письменной коммуникации, но не являющиеся печатными изданиями. Это машинописные и рукописные тексты, а также неопубликованная переписка, в частности, и в электронной форме. В этот подкорпус включаются также письменные тексты для устного произнесения (например, “Клятва юного пионера Советского Союза”, печатавшаяся на последней странице обложек школьных тетрадей в советский период).

Формальная классификация имеет недостатки, которые связаны с тем, что за внешней точностью

кроется большая доля условности при присыпывании признака. Например, такой признак, как “цель текста” (discussion, recommendation, information, instruction), отражает важнейшее свойство текста — коммуникативную задачу — только у определенных типов текстов. Присыпывание признака цели текста личному письму вызывает затруднение. Признак “размер аудитории” (small, large, etc) фактически является производным от других свойств. Например, личное письмо обычателья имеет единичную аудиторию, а опубликованное личное письмо писателя — очень большую аудиторию, хотя также адресовано одному человеку и, скорее всего, и не предназначалось для большей аудитории. Признак “форма текста” в данной классификации (spoken, written, written-to-be-spoken, electronic) фактически не является формальным признаком, а связывается с группами текстов, функционирующими в определенной сфере и обладающими определенными качествами. Текст, созданный в электронной форме и не существующий в других формах, может быть отнесен к печатным текстам, если он является художественным произведением; документация имеет признак “машинописная”, хотя сам текст-источник может иметь электронную, машинописную или рукописную форму; анкета обычно имеет печатную форму, но заполняется на компьютере или от руки, деловое письмо часто представляет собой документ, посланный по электронной почте, и т. д.

Функциональная классификация позволяет обсуждать свойства индивидуальных текстов и групп, включает традиционные самоназвания жанров и типов текстов; параметры этой классификации гораздо понятнее широкому кругу собирателей текстов.

В разделе 2 рассматриваются свойства, позволяющие объединить ТОГ в один подкорпус. В разделе 3 определяются признаки, которые используются в настоящее время для описания этих текстов, в частности функциональные признаки. В разделе 4 описываются группы ТОГ, вошедшие в состав НКРЯ к настоящему моменту. В Заключении подводятся предварительные итоги наполнения и классификации ТОГ, а также намечаются задачи, которые позволят завершить создание этого подкорпуса.

2. ТЕКСТЫ ОГРАНИЧЕННОГО ОБРАЩЕНИЯ КАК ПОДКОРПУС И КАК ОБЪЕКТ ИССЛЕДОВАНИЯ

Основной объем текстов в НКРЯ, как и в других национальных корпусах, составляют художественные и публицистические тексты. Например, в Британском национальном корпусе (BNC) 27% составляет художественная литература и 29,7% — публицистические тексты, а в Чешском национальном корпусе (CNC) — 15% и 60% соответственно. Доля эфемерных текстов (*miscellanea*) составляет в BNC 8%. Из них 4,38% — это печатные тексты (*miscellanea published*): брошюры, рекламные листовки; 4% — рукописные и машинописные тексты (*miscellanea unpublished*): личные письма, дневники, школьные и университетские сочинения (*essays*), заметки для памяти (*memoranda*), административные документы. Разработчики Американского национального корпуса (ANC), которые используют архитектуру BNC как модель, собираются расширить этот список за счет электронной переписки, внутренних отчетов и др. В Чешском корпусе доля административных документов составляет 0,49%. В Польском корпусе тексты, характеризуемые как *ephemera*, занимают 2%.

Национальный корпус русского языка рассматривается его создателями как отражение реального употребления языка данной эпохи во всем его функциональном и жанровом разнообразии, в котором ТОГ играют существенную роль. В связи с задачей максимально полного отражения современного языкового употребления объем этой группы планируется довести до 1 млн словоупотреблений, что составит приблизительно 0,01 общего объема Корпуса.

В подкорпус ТОГ включаются оригинальные (не переводы, не компиляции) монологические, прозаические тексты на русском языке. Эти ограничения отсекают некоторые виды ТОГ, в частности школьные сочинения по литературе (которые, как правило, представляют собой компиляции¹), инструкции к импортной технике (которые являются переводами), тексты поздравительных открыток, написанные в стихах, статьи по лингвистике, содержащие примеры словосочетаний, в частности и неправильные, и т. п. Есть также приблизительное ограничение на объем текста — не менее 50 словоупотреблений. Однако в массиве ТОГ в порядке исключения иногда представлены и более короткие тексты, например, надписи, SMS-сообщения.

Основным критерием включения текстов в подкорпус ТОГ является материальная форма текстов, которая имплицитно подразумевает такую характеристику, как размер аудитории. У ТОГ-текстов, изначально не предназначенные для публикации, аудитория невелика: машинописные материалы (независимо от того, отпечатаны ли они на пишущей машинке или на принтере, т. е. существуют и в электронной форме) создаются в небольшом количестве экземпляров и предназначены для ограниченного числа пользователей, например, документация обычно создается и используется в одной организации или в группе организаций. Еще более ограничена аудитория рукописных текстов,

в частности, переписка предполагает, как правило, единичного адресата. ТОГ, существующие в печатной форме и представленные такими видами печатной продукции, как листовки, вкладыши, памятки, буклеты, проспекты, брошюры, также, как правило, издаются небольшим тиражом для ограниченного круга адресатов, которыми являются, например, посетители выставки, музея, работники предприятия и т. п.

ТОГ ограничены и в плане тематики. Они используются в бытовой и производственной сфере и связаны с удовлетворением практических потребностей. Поэтому тематика этих текстов может быть разнообразна, но она не имеет ни социальной, ни эстетической значимости в отличие от тематики большинства печатных текстов.

Ограничность ТОГ проявляется и в отношении профессионализма и свободы языкового выражения. Тексты опубликованные создаются, в основном, профессионалами: писателями (художественная литература), научной и технической интеллигенцией (научная, учебно-научная, научно-справочная, научно-популярная литература), политиками и журналистами (публицистика), духовными мыслителями и церковными деятелями (религиозно-философская литература), юристами (законодательные тексты) и т. д. Авторы ТОГ создают свои тексты по случаю для достижения практических целей (объявление, письмо, методичка, анкета, документ и т. п.), хотя в некоторых сферах создание текстов, решаяющих практические задачи, становится профессией, например, технический писатель (создание пользовательских инструкций), специалист по рекламе и пропаганде (создание рекламных изданий). ТОГ, как правило, нейтральны с точки зрения стиля, который определяется скорее функциональной сферой; многие типы текстов создаются по образцу, модели и даже трафарету. Иногда это объясняется непрофессионализмом автора, боящегося отступить от образца, чаще (в деловых документах) следование шаблонам и трафаретам диктуется целесообразностью.

ТОГ собираются из различных источников: Интернет, рабочий стол, семейный архив, объявление на улице, в транспорте, на доске объявлений, рекламные листовки и т. п. Поскольку для большинства таких текстов отсутствует постоянно существующий и общедоступный источник, ответственность за аутентичность текстов лежит на собирателе текстов. В подкорпус ТОГ не включались тексты из регулярных источников, например, новостные сообщения, представленные на поисковых сайтах в Интернете, кулинарные рецепты, гороскопы, анекдоты, печатающиеся в журналах и газетах в специальных рубриках (в частности, некрологи) и издающиеся в виде книжек и брошюр. Объявления и реклама представлены электронными объявлениями (спам) и печатными или "машинописными" и рукописными объявлениями с досок объявлений, столбов, остановок транспорта и т. п., листовками, буклетами. Тексты для подкорпуса ТОГ собирались составителями подкорпуса с участием студентов двух вузов Москвы —

¹ Многие школьные нелитературные сочинения тоже представляют собой плагиат, например, одна девочка использовала для сочинения из жизни пчелок текст о пчелах, заменив третье лицо (пчелы) на первое лицо множественного числа, например, *мы строим соты из воска* и т. п.

РГГУ (Российский государственный гуманитарный университет) и МГППУ (Московский городской психолого-педагогический университет), а также добровольцев².

Подкорпус ТОГ в настоящее время представляет наиболее доступные типы текстов. Таким образом, можно предполагать, что собранная часть подкорпуса в некотором приблизительном смысле отражает подобие реального баланса такого типа текстов в обществе.

Типы текстов в корпусе представлены обобщенно, без излишней детализации. Использовать детализированные самоназвания текстов в качестве названия типов было бы неудобно, в особенности это относится к многочисленным формам деловых документов: так, одних деловых писем насчитывается несколько десятков типов.

Тексты в подкорпусе ТОГ деперсонифицируются, т. е. идентификаторы личности и учреждений в корреспонденции, документации и других видах текстов, содержащих идентификаторы, обозначаются как XXXX, 0000 или искажаются (например, фамилия может заменяться на фамилию *Иванов, Петров* в тех случаях, когда собирателю хочется сохранить естественный вид текста). Имена также изменяются по желанию дарителя. Сохраняются имена, фамилии, названия в тех случаях, когда текст был распространён авторами и содержание текста связано с выпустившей его организацией, например, в предвыборных листовках партий и кандидатов на выборные должности. В рекламных текстах названия и идентификаторы не сохраняются. Принципиальная позиция составителей корпуса состоит в том, чтобы он был источником только лингвистических аспектов употребления языка.

Тексты, выделенные нами в подкорпус ТОГ, лингвистически недостаточно изучены, хотя в последнее время ситуация меняется [3, 4]. Многие виды этих текстов становятся объектом изучения, например, деловая переписка — в связи с растущей значимостью коммерческой грамотности [5]; отчеты, описания, инструкции — в связи с созданием систем автоматической генерации текстов; эпистолярные жанры, письменные жанры повседневного городского общения (объявления, вывески, плакаты, граффити) — в связи с исследованием стилистической системы современного русского языка [6]. Предлагается выделить в качестве объекта самостоятельного изучения разнородный письменно-речевой материал, объединенный на основе письменной формы выражения и непрофессиональности изложения и представленный такими жанрами, как записка, открытка, письмо, телеграмма, личные записи (в дневниках, еженедельниках, записных книжках), учебные записи (на полях, на доске, черновики, шпаргалки, конспекты, изложения, сочинения), деловые записи (книги отзывов, дневники дежурств), официальные тексты (автобиографии, протоколы, докладные, объяснительные записки) и т. д. [7]. В то же время обсуждается вопрос о создании интегративной дисциплины — теории специального текста, в рамках которой наряду с хорошо изученными специальными текстами (научными, техническими, юридическими) могут описываться такие малоизученные, несмотря

на их повсеместную распространённость, тексты, как указания по выращиванию комнатных растений или уходу за домашними животными, инструкции по пользованию бытовой техникой и т. д. [8]. Интерес филологов вызывают тексты, циркулирующие в электронной коммуникации [9, 10].

Структура функциональных типов текстов становится объектом исследования в связи с расширением использования информационных технологий. В трудах междисциплинарной конференции Диалог 2001–2004 гг. представлены работы по автоматическому исследованию структуры научного текста (дискурса). Конкретные функциональные типы текстов становятся объектом автоматического распознавания и моделирования — деловые письма, инструкции, характеристики.

3. КЛАССИФИКАЦИЯ ТЕКСТОВ

Обращение пользователя к Национальному корпусу текстов предполагает наличие заданных параметров, по которым им может быть выбран подкорпус интересующих его текстов. В этом разделе рассматриваются признаки, используемые для классификации текстов и связанные с традиционной, функциональной, классификацией текстов. Формальная классификация в перспективе приложения к русскому материалу подробно описана в [9].

В отечественной типологии текстов принята характеристика текстов по их принадлежности к определенной функциональной сфере. Под функциональной сферой понимается социально значимая область речевой практики, которая объединяет тексты определенного содержания и целевого назначения и которая вследствие этого связана с определенными разновидностями языка (стилями). Для описания текстов Национального корпуса выделено восемь функциональных сфер: бытовая (обычно-повседневная), художественная, учебно-научная, производственно-техническая, официально-деловая, сфера публицистики и массовой информации, рекламы, церковно-богословская. Две сферы — бытовая и художественная — каждая по-своему противопоставлены всем остальным, которые, напротив, имеют между собой много общего (например, [10, 11]).

Наиболее широкой, с точки зрения коммуникативных задач, является бытовая сфера. Это сфера повседневного, непринужденного, неформального общения в кругу лиц, объединенных неофициальными отношениями (родственников, друзей, коллег по работе, учебе и под.). В бытовой сфере реализуется функция общения, в которой могут быть выделены более конкретные функции (функциональная, информационная, воздействия и т. д. и их сочетания), которые неразделимо сочетаются в спонтанной речи. Исконной формой этих текстов является устная форма. Традиционные письменные формы — личная переписка, записки, дневники, поздравления и др. Последние годы функция общения в письменной форме реализуется также в электронной коммуникации. Причем возникают новые формы спонтанной коммуникации — электронная переписка, чаты, форумы и т. д.

² Небольшая коллекция SMS-сообщений, положившая начало этому типу текстов в подкорпусе, была собрана учениками лицея N1511 при МИФИ под руководством С. В. Абрамовой. Сочинения детей по актерскому мастерству собраны П. Л. Печенкиной.

Художественная сфера — это сфера словесного художественного творчества, объединяющая тексты, в которых в наглядно-образной форме воплощается созданный воображением автора мир. Главной отличительной особенностью языка художественной литературы является его особая, по сравнению со всеми другими разновидностями, предназначность. Организация языковых средств в художественной литературе подчинена не просто передаче содержания, а передаче содержания художественными средствами. В художественной сфере язык выступает в эстетической функции и, опосредованно, в функции воздействия.

Учебно-научная сфера объединяет тексты научного и научно-методического содержания, относящиеся к различным областям науки и образования, целью которых является описание, объяснение и прогнозирование процессов и явлений действительности в логико-понятийной форме. Основная функция языка в учебно-научной сфере — информативная.

Производственно-техническая сфера порождает описания технических устройств и производственных процессов, предписания, регулирующие профессиональные действия человека в среде искусственных объектов. В производственно-технической сфере реализуется информативная функция языка и функция воздействия (в инструкциях).

В официально-деловой сфере функционируют тексты, основное назначение которых состоит в регламентации отношений между государством и его гражданами, организациями и другими государствами, между организациями и внутри них, между организациями и частными лицами в процессе производственной, хозяйственной и юридической деятельности. Функция воздействия является основной наряду с информативной.

Сфера публистики объединяет тексты, назначение которых состоит в формировании населения и формировании общественного мнения по вопросам общественной значимости в области политики, экономики, искусства, науки, морали и пр. В этой сфере реализуются функции: информативная, воздействия и отчасти эстетическая.

В сфере рекламы функционируют тексты, нацеленные на формирование потребностей, главным образом, материальных; их задача — заинтересовать, привлечь внимание потенциального покупателя, заказчика, партнера к предмету рекламы. Сфера рекламы смыкается с деловой сферой в области торговой рекламы и с публистикой в области социальной и политической рекламы. В ней реализуются функции: информативная и воздействия.

Церковно-богословская сфера объединяет тексты религиозного содержания, которые представляют религиозную картину мира и обслуживают различные стороны религиозной жизни индивида (молитвы, церковные обряды, исповедь, проповедь и пр.). Функции языка в этой сфере — информативная и воздействия.

Необходимым признаком описания текстов является их жанрово-стилистическая принадлежность. Для характеристики текстов корпуса используется параметр “тип текста”, набор значений

которого основывается на самоидентификации текста. Список типов текстов принципиально открыт и пополняется по мере пополнения корпуса.

Каждая функциональная сфера связана с определенным набором типов текстов. Например, в сфере публистики функционируют такие типы текстов, как заметка, репортаж, интервью, статья, комментарий, фельетон, памфлет и др. Однако эта связь не жесткая. Так, существуют пограничные типы, например, очерк; типы, общие для разных функциональных сфер, например, отчет (научный, деловой и газетный); типы, используемые в разных функциональных сферах с разной целевостановкой (например, реклама заимствует типы газетных, научных текстов, наполняя их новым содержанием).

Сфера функционирования текста связана с его тематикой. Для одних сфер (научной, производственно-технической, официально-деловой, религиозной) характерна сравнительно однородная тематика текстов, для других — более разнообразная (например, реклама может быть бытова, производственная, социальная, политическая); в публистической, бытовой и художественной сферах набор тем практически не ограничен. Хотя тематическая характеристика текста — наиболее субъективная составляющая всей классификационной схемы, тем не менее набор предметных областей в разных корпусах оказывается приблизительно одинаковым (ср., например, список тематических рубрик в рекомендациях EAGLES и в НКРЯ).

Предложенная классификация по Синклеру удобна для отбора и балансировки текстов в составе корпуса, но менее удобна для практической работы с корпусом: и разработчику, выполняющему описание текстов, и рядовому пользователю, осуществляющему поиск по нему, привычнее пользоваться традиционными для данной культуры классификационными признаками, каковыми в нашем случае являются сфера функционирования и тип текста.

Описание ТОГ строится в той же системе метатекстовой разметки, которая используется и для характеристики всех остальных текстов корпуса, однако имеет некоторые особенности по сравнению, скажем, с описанием художественных и устных текстов. Для характеристики ТОГ в базе данных корпуса используется 16 параметров, или атрибутов. Семь из них относятся к характеристике самого текста. Это название текста, дата его создания, объем (в словоупотреблениях), сфера функционирования, тип текста, тема, стиль. Поскольку, как правило, ТОГ не имеют названий, в качестве названия текста при его описании используется уточненный тип текста — приказ, квитанция, лотерейный билет, личное письмо, и т. п. Для заполнения атрибута “тип текста” используются общепринятые названия соответствующих жанров, например: письмо, записка, приказ, комментарий, лабораторная работа и т. д. Уточнение типа каждого описанного текста происходит за счет указания значений других признаков-атрибутов — функциональной сферы, размера аудитории и т. д.

Три параметра характеризуют автора (имя автора, пол, дата рождения) и возможную аудиторию (возраст аудитории, размер, уровень подготовленности), остальные параметры (название файла, размер, источник получения и др.) имеют служебный характер.

4. ГРУППЫ ТОГ В КОРПУСЕ

Общее свойство текстов подкорпуса ТОГ состоит в том, что это небольшие тексты приблизительно от 10 (надпись, SMS-сообщение) до 10 тыс. (сочинение, курсовая работа) словоупотреблений. Основная масса текстов имеет размер от одной до нескольких сотен словоупотреблений — документация, корреспонденция, инструкции и т. д. В настоящее время объем обработанных ТОГ в НКРЯ составляет 0,5 млн словоупотреблений. Поскольку количество и разнообразие типов в планируемом миллионном подкорпусе велико, для обработки они разделяются на шесть групп. Основанием для образования групп ТОГ является структурная и качественная общность текстов, которой иногда сопутствует также общность значений некоторых формальных или функциональных признаков. Например, документация принадлежит функционально-деловой сфере, но обратное не верно — не всякий текст официально-деловой сферы является документацией. Группировка ТОГ скрыта от пользователя текстов, имеющего возможность получать из него фрагменты размером с предложение, но, вообще говоря, может представлять самостоятельный интерес для исследователей, занимающихся структурой текста. Например, этот аспект интересует такое прикладное направление лингвистики, как генерация текстов на естественном языке. Выделение групп коррелирует с понятием жанра текста в его традиционном неформальном употреблении. Так, рекламный текст, написанный в форме письма “подруги” или “своего парня”, не является корреспонденцией, но использует и отражает законы жанра “письма” и поэтому попадает в группу корреспонденций.

(1) **Корреспонденция.** Группа выделяется как вид информационного и фактического взаимодействия между людьми. Признаки этой группы — **характер участников акта коммуникации, принципиальная возможность ответной реакции.** Автор и аудитория являются либо частными лицами (один автор или коллективный автор, аудитория частная — один человек или группа в бытовой сфере); либо представителями общественной или государственной структуры (обобщенный автор — чиновники, партийные функционеры, аудитория — группа в официально-деловой сфере). Промежуточным типом является, например, официальное поздравление юбиляра. Автор — обобщенный, но адресатом является конкретный человек, и содержание текста касается его частной жизни и личности. Обратная ситуация — письмо обычавшего должностному лицу, в котором автором является частное лицо с его стилем, верованиями и проблемами, а получателем — группа чиновников. Типы текстов в этой группе — письмо, записка, SMS сообщение.

(2) **Документация.** Группа выделяется по признаку функциональной сферы — **официально-деловая.** Автор — обобщенный, аудитория — группа, т. е. человек как занимающий

или готовый занять (заявление о приеме на работу) некоторую должность. Типы текстов документов соответствуют самоназваниям документов — договоры, приказы, заявления и т. д. К группе Корреспонденция и к группе Документация относятся письма и записки в официально-деловой сфере (служебное письмо, докладная записка и др.). Автор и адресат текста, как правило, различны по положению (выше/ниже), для документации характерно также наличие отчетливой цели создания текста.

(3) **Рекламно-информационные тексты.** Группа выделяется по признаку функциональной сферы — **реклама.** Автор — обобщенный или неизвестный, а аудитория всегда публичная, малая публичная для объявлений в учреждении, средняя или большая — в других случаях. Типы текстов — объявление, листовка, буклет. Реклама “паразитирует” и на других жанрах, например, принимает форму корреспонденции — спам-реклама. В отличие от корреспонденции, реклама социально значима. Ее аудиторией становятся определенные социальные группы людей. В отличие от корреспонденции и документации, в которых коммуникантов объединяет личная или производственная (в самом широком смысле) сфера, разделенная обобщими участниками, рекламно-информационные тексты преследуют цель решения материальных проблем автора. Типы рекламных текстов обсуждаются также в [12].

(4) **Методические тексты.** Группа выделяется по формальному признаку **цели** создания текста (инструктирование, рекомендация), который в случае методических текстов отражается и в структуре текста. Типы текстов — инструкции, правила, распорядки, положения, учебные тексты, методические издания для внутреннего пользования, рекомендации. К этой же группе относятся надписи на дверях, приемных окошках мастерских, в проходах, в транспорте и т. д., например, *Проход в зону лифта через 4-ый этаж, Вход с домашними животными категорически воспрещен*. Надписи находятся в корпусе в виде коллекции текстов, т. е. списка надписей.

Следующие две группы текстов не объединяются каким-либо постоянным значением формального или функционального признака, а выделяются по характеру и способу организации содержания. Первая из них (5), противопоставлена второй как **довольно жесткое по форме отражение в виде текста свойств некоторой конкретной и скорее статичной реальности — предмета, человека, способа, в отличие от второй (6), допускающей динамику, фантазирование, различные стратегии и формы изложения содержания.**

(5) **Описания.** Группа объединяет тексты, являющиеся описанием предмета (например, табличка в ботаническом саду), источника информации (например, рецензия на книгу или спектакль), способа решения проблемы (патент), лица, его жизненного пути (характеристика, биография и т. д.), его подтвержденных тестированием или авторитетом автора физиологических или душевных свойств — тест (результаты тестирования), характеристика. Такой тип текста, как “**должностная инструкция**”, в принципе, может относиться и к группе Описания, и к группе Методические тексты. Мы относим его ко второй.

(6) **Сочинения**. Группа объединяет различные тексты, имеющие или могущие иметь эмоциональные составляющие, различную форму, мнение автора, не подтвержденное фактическими данными. Типы текстов — дневник, комментарий, сочинение (школьное), курсовая работа (студент), проба пера и тексты субкультуры (например, “растаманские сказки” — плоды творчества любителей конопли), а также обзоры и отчеты.

Ниже более подробно обсуждаются типы текстов, представленные в Корпусе, по выделенным шести группам. Числом в скобках обозначено приблизительное количество текстов данной группы по состоянию на декабрь 2004 г.

4.1. Корреспонденция (270)

В зависимости от значений формальных и функциональных признаков в метаописании текста пользователь может выделить для себя несколько разновидностей корреспонденции. Личная корреспонденция — личные письма, монотематические письма (приблизительно 200, не считая SMS-сообщений). Объем SMS-сообщений в словоупотреблениях составляет приблизительно 1,5 тыс. Деловая корреспонденция представлена служебными письмами и записками (70).

Письма:

(1) личное письмо (120). Характеризуется частным автором и адресатом (аудиторией), разнообразием тематики, смешанной в рамках одного документа, которая в целом характеризуется как “частная жизнь”, индивидуальным стилем, отражающим, частности, возрастной, половой, региональный и образовательный статус автора. Автор, как правило, “неизвестный”, хотя, в принципе, авторство может быть сохранено в метаописании текста по желанию дарителя аналогично письмам писателей;

(2) деловое, служебное письмо (50). Функционирует в официально-деловой сфере. Характеризуется обобщенностью как автора, так и аудитории, которая, как правило, описывается как “группа”. Тематика определяется в зависимости от направления деятельности организаций и затрагиваемых вопросов — “производство”, “бизнес, коммерция, экономика, финансы”, “администрация и управление”, “образование”, “спорт”;

(3) монотематическое письмо (50) — письмо, автором и адресатом которого является конкретное лицо, но аудитория может быть шире, в том числе, открытое письмо. Здесь имеются две разновидности: письма по работе — это короткие письма между коллегами, уточняющими определенные вопросы совместно выполняемой работы, и письма-мнения в форумах и на сайтах в Интернете, содержащие обмен мнениями на определенную тему. Тематика писем по работе связана с характером работы: “производство”, “искусство и культура”, “бизнес, коммерция, экономика, финансы”; тематика писем-мнений — “политика и общественная жизнь”, “религия” и др. По размеру первые — короткие, вторые обычно длинные. Язык обеих разновидностей сохраняет индивидуальные особенности речи автора.

(4) официальное письмо (50) — письмо, один из коммуникантов которого является обобщенным или группой, а второй — частным лицом. Выделяются две разновидности. Одна — поздравления

с юбилеями и всенародными праздниками, например, работника предприятия, члена Совета ветеранов, отражающие личные заслуги адресата. Другая разновидность представлена письмами формально с конкретным адресатом, но с публичной аудиторией, например, предвыборное письмо кандидата в депутаты, адресованное избирателю, с разъяснениями политики депутата, извещение налоговой или коммунальной службы об изменении порядка оформления платежей и т. п. В том случае, если тематика письма касается деятельности адресата, указывается два тематических признака: “частная жизнь” и признак, отражающий тематику лексики письма — “производство”, “армия и вооруженные конфликты”, “спорт” и т. п.

Записка (10) — короткий текст, в зависимости от сферы функционирования представлен следующими разновидностями: бытовая записка и служебная записка в официально-деловой сфере. В самоназваниях служебных записок представлено четыре разновидности: служебная или деловая, докладная, пояснительная и объяснительная. В настоящее время в коллекции записок в основном представлена официально-деловая сфера.

SMS-сообщения, так же, как надписи, объединяются в коллекции, а также размечаются по полу и возрасту автора. Общий объем подготовленных сообщений — более тысячи словоупотреблений.

4.2. Документация (без деловой переписки 100)

Документацию можно разделить на две разновидности — деловая и судебно-правовая. В настоящее время эти разновидности выделяются признаком “тематика”. Тематика деловой документации представлена одним признаком — “производство”, “администрация и управление”, “образование”, но могут быть и другие, например, “дом и домашнее хозяйство” — договор о гарантитном обслуживании бытовой техники. Тематика правовой документации представлена или одним признаком “право” или его сочетанием с другим признаком, например, комбинацией признаков “право | дом и домашнее хозяйство” (например, “Договор о правовом режиме имущества супругов”) или “право | путешествия” (например, “Уведомление об ограничении ответственности за багаж”). Типы деловой документации, представленные в корпусе: договор (47), приказ (22), заявление (7), протокол, выписка из протокола (7), акт (5), распоряжение (5), справка (5), предложение (4), соглашение (4), уведомление (3), постановление (2), решение (2), квитанция, контракт, расписка, резолюция, указ, ходатайство (1). Типы правовой документации в Корпусе: доверенность (8), заявление (3), свидетельство, удостоверение (2), диплом, договор, завещание, заключение, постановление, решение, уведомление (1). Общий объем документации правовой тематики — около 30 документов. Типы документации соответствуют распространенным самоназваниям соответствующих документов. Есть типы, характерные для деловой (например, приказ, квитанция и др.) и судебно-правовой разновидности (доверенность, завещание и др.), а есть — общие для этих двух разновидностей документации — договор, заявление, постановление, решение, уведомление и др.).

В основном в подкорпусе представлена документация уровня предприятий. В подкорпусе имеются единичные документы уровня городской и государственной администрации, касающиеся образования и общегородских мероприятий с типами текста — “указ” и “распоряжение”, — а также несколько международных договоров.

4.3. Рекламно-информационные тексты (210)

Рекламно-информационные тексты представлены объявлениями, листовками, буклетами, вкладышами и надписями на упаковке товаров. Основу подкорпуса составляют объявления. По функциональной сфере объявления можно разделить на две разновидности: объявления, которые имеют функциональную сферу “бытовая”, “производственная” и другие (40), и рекламные объявления, которые имеют функциональную сферу “реклама” (140). Первые имеют целью решить какую-либо некоммерческую проблему, например, найти потерявшуюся собаку, пристроить ставший не нужным учебник, котят и т. п. Для организаций нерекламные объявления — это, например, приглашение на должность. Цель рекламных объявлений — продать предмет или услугу, получив коммерческую выгоду. Большая часть рекламных объявлений в подкорпусе поступает через Интернет (спам), остальные — с досок объявлений, с улицы, и т. п. Как правило, рекламное объявление имеет рекламную и информационную составляющие. Структура текста рекламного объявления изменяется от присутствия полноценной информационной составляющей до фирменного логотипа. Последняя разновидность похожа скорее на лозунг, внушающий зрителю “веру в себя” в контексте фирменного логотипа или картинки рекламируемого предмета, например, *Поверь в себя, ты можешь больше!* (на рекламе отечественного напитка джин-тоник “Салют, Златоглавая” в метро).

Листовки в основном относятся к функциональной сфере “публистика” (20). В основном это партийные предвыборные листовки. Есть единичные экземпляры других направлений, например, листовка антиглобалистов и т. п. Встречаются также печатные тексты рекламной сферы, которые по структуре содержания ближе к листовкам, чем к рекламным объявлениям, например:

Время менять шнурки
Это удобно
Больше не надо завязывать свои шнурки!
Это модно
...
При стирке не теряют качества.
Пружинные шнурки с самофиксацией Шнуркс.

Листовками в подкорпусе ТОГ также называются короткие (не более страницы) информационно-рекламные описания на бумажном носителе, которые бросают в почтовые ящики, раскладывают на столах в аптеках и других торговых и коммунальных учреждениях.

Между объявлениями и листовками нет четкой грани, один и тот же текст может быть разослан по

электронной почте, распространен в виде листовок и расклеен в виде объявления на доске объявлений. Также нет четких языковых различий между листовками, вкладышами (сведения о товаре на бумажном носителе, вкладываемые внутрь упаковки товара) и надписями на упаковках. Особенность вкладышей и надписей на упаковке состоит в том, что они часто содержат элементы рекомендательного характера об использовании товара, например, формы для выпечки кекса, ножниц и т. п., что приближает их к методическим текстам. Буклеты (6) и брошюры отличаются от листовок размерами и более подробной информацией. Они, как правило, представляют не один конкретный вид товара, а фирму, например, поставщика кухонных гарнитуров мебели, или вид услуг, например, валютные вклады в сберкассе. Специфическим типом текстов, представляющим информационные объекты и мероприятия, является анонс (7).

4.4. Методические тексты (100)

Типичным видом методических текстов является инструкция, описывающая использование предмета. В корпусе они представлены инструкциями и руководствами к бытовым приборам — холодильнику, стиральной машине, огнетушителю и т. п. (30). Близок к инструкциям тип Памятка (3). Распространенный вид методических текстов в производственной и учебно-научной сферах — должностные инструкции: врача, грузчика, менеджера, шофера и т. д. (30). Манера написания должностных инструкций может быть похожа как на инструкцию, т. е. содержать побудительные формы, так и на описание. Похожи на описания и регламентирующие тексты — правила (13), положение (10), а также порядок (3) и устав (1). В учебно-научной сфере выделяется тип Лекция (2), и Лабораторная работа (1), а в бытовой — надписи (имеется небольшая коллекция).

4.5. Описания (60)

Под описаниями понимаются тексты, которые специально созданы, чтобы оперативно предоставлять человеку информацию в повседневной жизни — на работе, при посещении музея, зоопарка, театрального представления, просто на отдыхе. Это — не справочная информация, которая содержится в специальных изданиях, а документы: описания к покупкам, маленькие тексты, которые делают нашу жизнь удобнее и интереснее, а иногда просто развлекают. Они имеют достаточно определенную форму в зависимости от объекта описания и аспекта описания и всегда основаны на фактической информации, конкретном предмете. В подкорпусе имеются следующие типы текстов: описание человека — резюме (20), заявка-вакансия (4), тест (результат тестирования) (1), биография (1), автобиография (1), характеристика (2), анкета (4); описания предметов — заявка (1), табличка (2), паспорт прибора (1), вкладыш (6), лотерейный билет, например, серия лотерейных билетов “Острова”, (1); описание носителя информации — книги, диссертации, спектакля, сайта — отзыв (9), рецензия (3), аннотация (2), обзор сайта (2).

4.6. Сочинения (60)

Сочинения — это разнообразные тексты, не обязательно основанные на фактической информации и существующие в различных сферах. На фактической информации основаны отчеты (25) различных сфер и тематик, комментарии к футбольному матчу (1), заметки (2), дневники (отрывки из дневников) (3). На наблюдениях — контрольные и курсовые работы студентов (1). Имеются также тексты электронной коммуникации, которые трудно отнести к одному из традиционных типов текста (10). Сочинения типа “проба пера” могут быть близки к фактической жизни, например, рассказ о дне рождения, а могут намеренно удаляться в область фантазии, например, сочинение — отчет в письмах о некоторых событиях ролевой игры в фантастическом мире (2). Из школьных нелитературных сочинений есть детские сочинения по актерскому мастерству (12). В группу включаются также некоторые тексты субкультур, в частности, несколько растаманских сказок — творчество любителей конопли (6).

5. ЗАКЛЮЧЕНИЕ

На первых двух этапах проекта (2003–2004 гг.) решались задачи сбора письменных текстов — ТОГ, которые используются людьми в их повседневной жизни. Состав подкорпуса включает в основном короткие тексты от десятка до примерно 10 тыс. словоупотреблений (в среднем несколько сотен словоупотреблений) и на момент написания статьи в целом составлял 0,5 млн словоупотреблений. По содержанию он определялся индуктивно, т. е. в него попадали наиболее доступные тексты, собранные в основном студентами двух вузов Москвы. Классификация и описание текстов опиралась на формализованную европейскую традицию описания текстов в больших корпусах и на неформализованную отечественную классификацию. Сложившаяся в результате работы форма метаописания, принятая в НКРЯ, включающая формальные и функциональные признаки, удобна для собирателя текстов и пользователя корпуса, так как позволяет им оперировать более привычными понятиями, чем формальные признаки, такими, как Тип текста и Функциональная сфера. Некоторые формальные признаки, характеризующие отдельные типы текстов, частично отражены функциональной классификацией и, поэтому, остались вне внимания в метаописании НКРЯ. Однако они полезны для описания некоторых типов текстов, в частности, признак “форма текста”, например, электронная, рукописная, полезно было бы отмечать для личной корреспонденции, объявлений; “цель создания текста” может выделить определенные разновидности текстов в документации и деловой корреспонденции. Функциональные признаки также могут быть расширены. Так, судебно-правовые тексты, которые сейчас относятся к формально-деловой сфере, хотя они обладают явным своеобразием языка и стиля — консервативный стиль, точность формулировок, перлокутивные речевые акты, наличие специфических типов текстов, например, закон, завещание, требуют изучения свойства электронной коммуникации. Кроме явно речевого, спонтанного характера этих текстов

и групповой аудитории они отличаются еще и тем, что отнесение этих текстов к традиционным типам вызывает протест собирателя, например, впечатления участников о летней школе на форуме или в групповой рассылке нельзя назвать ни отзывом, ни отчетом, ни сочинением, ни письмом, ни дневником. Специальные типы электронных текстов, например, “сетевые дневники”, сильно отличаются от традиционных — содержание слишком объемное и размытое для традиционного дневника, конкретных событий, их временных последовательностей такие тексты содержат мало и т. д.

Для работы с большим количеством и разнообразием небольших текстов потребовалась дополнительная группировка текстов, которая представлена в статье в виде шести групп, имеющих следующие условные названия: корреспонденция, документация, рекламно-информационные тексты, методические тексты, описания и сочинения. Она удобна для обработки массивов, а также, возможно, полезна для приблизительной оценки баланса различных видов текстов повседневного употребления.

Подкорпус ТОГ должен быть дополнен текстами, менее доступными в повседневной жизни, например, патентами, некоторыми видами печатных изданий (ephemera), в частности, либретто, буклетами и брошюрами. Последние обычно относятся к рекламно-информационной или методической группе текстов, например, благотворительный фонд “Страна живых”, Методическое пособие для ресторанных работников, и т. п.

* * *

Авторы благодарны собирателям и дарителям, проявившим интерес и внесшим свой фактический вклад в начало создания подкорпуса ТОГ — без их участия задача была бы невыполнима. Мы и в дальнейшем будем рады любым вкладам от дарителей, понимающих важность того, чтобы наш повседневный русский язык был представлен в НКРЯ наиболее полно и разнообразно, т. е. так, как мы им пользуемся.

СПИСОК ЛИТЕРАТУРЫ

1. Sinclair J. Preliminary recommendations on text typology // EAGLES Document EAG-TCWG-TTYP/P, 1996. <http://www.ilc.pi.cnr.it/EAGLES96/texttyp/texttyp.html>
2. Шаров С. А., Савчук С. О. Типология текстов для представительного корпуса.— В печати.
3. Шмелева Т. В. Повседневная речь как лингвистический объект // Русистика сегодня: функционирование языка: лексика и грамматика.— М., 1993.
4. Современный русский язык: Социальная и функциональная дифференциация.— М., 2003.
5. Штрекер Н. Ю. Русский язык и культура речи: Учеб. пособие для вузов.— М., 2003.
6. Китайгородская М. В., Розанова Н. Н. Современное городское общение: типы коммуникативных ситуаций и их жанровая реализация (на примере Москвы) // Современный русский язык. Социальная и функциональная дифференциация.— М., 2003.
7. Лебедева Н. Б. Естественная письменная русская речь: проблемы изучения // Русский язык: исторические судьбы и современность.— М., 2001.— С. 260–261.