

**БАЗА ДАННЫХ «ИНТОНАЦИЯ РУССКИХ ИНФОРМАЦИОННЫХ
ТЕКСТОВ»¹****DATABASE «INTONATION OF RUSSIAN INFORMATIONAL TEXTS»**

*Кодзасов С.В. (sankod@philol.msu.ru), Архипов А.В. (arxipov@philol.msu.ru),
Захаров Л.М. (leon@philol.msu.ru), Кривнова О.Ф. (okri@philol.msu.ru)
Московский государственный университет имени М.В. Ломоносова*

В докладе сообщается о результатах работы по созданию БД «Интонация русских информационных текстов». Этот проект продолжает многолетнее исследование, направленное на создание понятийного аппарата и компьютерного инструментария для всестороннего описания функций и средств русской интонации на основе результатов анализа представительного корпуса русских текстов (диалогических и повествовательных). В работе обсуждаются результаты начального этапа второго нового цикла наших корпусных исследований.

1. Описание проекта

Проект связан с быстро развивающейся областью отечественной лингвистики, которую можно назвать дискурсивной грамматикой русского языка. В ее задачи входит описание дискурсивной лексики, дискурсивного синтаксиса и дискурсивной интонации. До сих пор корпусные методы применялись изучалась в основном для исследования интонация диалогического дискурса. Результаты исследований разработок авторского коллектива в этой области сообщались на прошлых конференциях «Диалог» (см. список публикаций).

В осуществляемом ныне проекте впервые ставится задача создания компьютерной речевой базы для изучения интонации информационного и повествовательного дискурса. При этом просодия текстов исследуется демонстрируется и описывается в тесной связи с их иллокутивным и модальным содержанием, а также с их фазово-синтаксической структурой. Особое внимание уделяется просодическим техникам маркирования жанровых и регистровых характеристик текстов и их компонентов. Это позволит уточнить системы семантических, прагматических и стилистических дескрипторов, которые используются в текст-лингвистике в настоящее время.

Результаты проекта могут использоваться как при дальнейших лингвистических исследованиях грамматики текста, так и в прагматических целях (оптимизация учебных и инструктивных текстов, рекомендации по использованию произносительных стилей в средствах массовой информации). В число потенциальных пользователей попадают не только лингвисты, но также дикторы и составители текстов в средствах массовой информации.

Проект предполагает создание трех представительных корпусов звучащих текстов для трех базовых областей дискурса: информационные и инструктивные тексты (2007 г.), нарративные тексты разных жанров (современные литературные тексты, фольклорные тексты) (2008 г.), спонтанные нарративы (2009 г.). Характерные фрагменты для каждого типа дискурса будут введены в компьютерную базу данных. Каждый файл снабжается краткой прагматико-семантической дескрипцией и просодической транскрипцией, включающей все просодические характеристики: тональные акценты и интегральные уровни тона, выделительные акценты и громкости, долготы и темпы, фонации и просодические тембры. План работ по годам соответствует трем указанным областям дискурса.

До сих пор лингвистами с помощью корпусных методов изучалась в основном интонация диалогического дискурса. В данном проекте впервые была поставлена задача создания компьютерной базы для изучения интонации информационного и повествовательного (литературного и бытового) дискурса.

За отчетный период в целях создания корпусного материала были исследованы просодические средства оформления текстов информационных программ центрального телевидения. В качестве материала использовались записи новостных программ трех наиболее популярных телеканалов: НТВ («Сегодня»), ОРТ («Время») и РТР («Вести»), которые стилистически различны.

¹ Работа поддержана грантом РГНФ № 07-04-12160в.

База данных «Интонация русских информационных текстов»

Исследовались акцентная структура текстов (размещение семантически нагруженных выделительных акцентов), тональная структура компонентов сообщения и средства экспрессивного подчеркивания. Производился анализ функциональной структуры текста (фазовые и модальные характеристики его компонентов), что позволяет выделить средства семантического структурирования сообщения и риторического воздействия текста на слушающего.

В результате анализа было отобрано около 300 текстовых фрагментов (информационных блоков) общим временем звучания около 90 минут. Они введены в базу данных в виде звуковых файлов, снабженных несколькими типами аннотаций: простейшая функциональная (тематическая) индексация, орфографическая запись, просодическая транскрипция текстового фрагмента.

Каждый файл снабжен краткой семантико-прагматической дескрипцией и детальной просодической дескрипцией на основе разработанной авторами ранее системы интонационного описания (см. список публикаций). Эта информация представлена в виде базы данных в формате Microsoft Access. Программа позволяет прослушивать файл, видеть его осциллограмму и интонограмму и анализировать семантико-просодическую дескрипцию.

2. Пример записи информационного блока в базе данных («Новости» на НТВ)

Тема блока и темы фрагментов	Орфографическая запись фрагмента	Просодия фрагмента
БЛОК 1. ПОЖАР В ДОМЕ ПРЕСТАРЕЛЫХ		
1.1. Развитие чрезвычайной ситуации, число жертв	Число погибших при пожаре в доме престарелых под Тулой сегодня вновь возросло и достигло тридцати двух человек.	Число(\)(\)*погибших(\) при пожаре в доме *престарелых(/) под *Тулой (\) сегодня *вновь (/) *возросло(\) и *достигло(\) *тридцати(\) двух *человек (\).
1.2. Еще одна жертва	Уже после того, как основные поиски были Взавершены, на месте сгоревшего интерната кинологи обнаружили тело еще одной жертвы.	Уже после *того (\), как основные(/) *поиски(/) были *завершены(/), на *месте(/) сгоревшего(/) *интерната(/) кинологи(/) *обнаружили(\) *тело(\) *еще(\) одной *жертвы(\).
1.3. Причина большого числа жертв	В МЧС говорят, что масштабной трагедии можно было избежать. Медики сами пытались тушить огонь и долго не сообщали о возгорании по ноль один.	В *МЧС говорят(\), что *масштабной(\) *трагедии(/) можно было *избежать(\). Медики(/) *сами пытались(\) тушить *огонь(/) и долго не *сообщали(\) о *возгорании(\) по ноль *один(\).
1.4. Последствия	Сегодня в Тульской области объявлен день траура. Тем временем в городах и поселках началась проверка пожарной безопасности в старых зданиях.	*Сегодня(\) в Тульской(\) области(\) объявлен(/) день(/) *траура (\). Тем *временем(/) в городах и *поселках(\) началась *проверка(/) пожарной *безопасности(/) в старых *зданиях(\).
1.5. Текущее состояние 1	С места события передает наш специальный корреспондент Евгений Гуцал. — Все работы в доме престарелых сейчас завершаются. Режим чрезвычайной ситуации в Тульской области уже снят.	С места *события(\) передает наш *специальный(\)(\) *корреспондент(/) Евгений *Гуцал(\). Все *работы(/) в доме *престарелых(+) сейчас *завершаются(\). *Режим (/) *чрезвычайной *ситуации в *Тульской(\) *области(/) уже(/) *снят(\).
1.6. Число погибших и их похороны	Окончательное число погибших — тридцать два человека. Всех их похоронят за счет государства. У многих нет даже родственников.	Окончательное(\) число *погибших (/) — *тридцать(\) *два(\) человека. Всех(\) их похоронят(\) за счет *государства(\). У многих(/) нет даже *родственников (\).
1.7. Текущее состояние 2	Сегодня утром составлены списки: кто из пациентов сейчас в больнице, кого отправили в ближайшие интернаты. К месту трагедии съезжаются близкие пострадавших.	*Сегодня(/) *утром(/) *составлены(\)(\) списки: кто из *пациентов(\) *сейчас(\) в *больнице(/), *кого(/) отправили в *ближайшие(\) *интернаты(\)(\). К месту *трагедии(/) *съезжаются(\)(\) *близкие(\) пострадавших.

Тема блока и темы фрагментов	Орфографическая запись фрагмента	Просодия фрагмента
Г.8. Подтверждение прежней версии причины события	Продолжают работать следователи. Официальной версией пока так и остается возгорание проводки.	*Продолжают(/) работать(/) *следователи(\). Официальной *версией(/) пока так и *остае-ся(\) возгорание *проводки.
Г.9. Последние дан-ные с места события	О том, что найдено тело тридцать *вто-рого постояльца, стало известно во время заседания специальной комиссии по расследованию ситуации.	О том, что найдено тело тридцать *второго(+) постояльца(\), стало известно(/) во время засе-дания специальной *комиссии(/) по расследова-нию(/) ситуации(\).
Г.10. Дальнейшая судьба дома престарелых	Сотрудники МЧС заявили, что здание дома престарелых восстановлению не подлежит. В области сейчас решается вопрос о строительстве нового интерната.	Сотрудники *МЧС *заявили(/), что здание дома *престарелых (/) восстановлению(/) не *подле-жит(/). В области сейчас решается *вопрос(/) о строительстве *нового(\) интерната.
Г.11а. Сопутствующие гибели людей мероприятия	Похороны погибших состоятся завтра.	Похороны(/) *погибших(/) состоятся(\) *зав-тра(\).
Г.11б. Объявление траура	Сегодня в Тульской области объявлен траур.	*Сегодня(/) в Тульской *области(/) объявлен *траур(\).

3. Просодическая специфика информационных текстов (предварительные наблюдения)

Материал нашей базы данных обнаруживает немало отклонений просодии телепрограмм от типичных характеристик диалогического текста, что видно и по просодической разметке приведенного выше фрагмента. Укажем наиболее важные из них.

Прежде всего, обращает на себя внимание гораздо большая степень акцентированности текста в целом. Это характерно не только для новостных программ НТВ, но и для новостей на других каналах. Нередки случаи выделения громкостными или тональными акцентами всех слов предложения (своего рода «акцентное скандирование» текста). Другая особенность: наличие многочисленных случаев несовпадения выделительных и тональных акцентов, что для реплик диалога нехарактерно. Немало также случаев двойных тональных акцентов внутри слова. Все это говорит о том, что в текстах рассматриваемого вида используется особая акцентная подсистема, ориентированная на высокую степень риторического воздействия на слушателя.

Другой особенностью проанализированных текстов является довольно частое использование восходящего тона в качестве финального акцента неконечных предложений внутри фрагмента сообщения. Риторическая функция этой техники очевидна: диктор указывает на незавершенность текста, мобилизуя внимание слушающего на его продолжение.

Мы приводим в Приложении расшифровку знаков используемой в данной работе просодической транскрипции, поскольку она не общепринята и может вызвать затруднения у читателя. Эта же система использовалась нами при создании баз данных по интонации русских диалогических реплик (см. список литературы).

Приложение. Знаки интонационной транскрипции

Знаки для тональных акцентов

- \ - нисходящий нейтральный (4-7 полутонов)
- \' - нисходящий малый (3-4 полутона)
- \'' - нисходящий сверх-малый (менее 3 полутонов)
- \\ - нисходящий увеличенного интервала
- \в - нисходящий в высоком регистре
- \н - нисходящий в низком регистре
- \с - нисходящее движение происходит на начальном согласном слога
- \~ - нисходящий тон растянут на слово
- / - восходящий нейтральный
- /' - восходящий малый
- /''' - восходящий сверх-малый
- // - восходящий увелич. интервала
- /в - восходящий в высоком регистре

База данных «Интонация русских информационных текстов»

/н – восходящий в низком регистре
 /с – восходящее движение происходит на начальном согласном слога
 /~ - восходящий тон растянут на слово
 /\ - восходящий – ровный – нисходящий
 ^ - восходяще-нисходящий тон внутри гласного
 /с\г – восходящий тон на согласном и нисходящий на гласном
 v - нисходяще-восходящий тон внутри гласного
 /- - восходящий плюс ровный без падения

Знаки для нетональных акцентов

: - долготный акцент
 * – громкостный акцент на подчеркнутом сегменте слова

Знаки для синтагменных просодий

инк – инклинация тона
 дек – деклинация тона
 ТИ – низкая громкость (тихо)
 ГР – высокая громкость (громко)
 Н – низкий регистр
 В – высокий регистр
 Б – быстрый темп
 НПР – напряженная фонация
 ПДХ – придыхательная фонация
 ФЦТ – фальцетный регистр

Список литературы

1. Кодзасов С.В., Бонч-Осмоловская А.А., Захаров Л.М., Кобозева И.М., Кривнова О.Ф. База данных «Интонация русского диалога»: вопросительные реплики // Труды международной конференции Диалог 2005 г.
2. Кодзасов С.В., Архипов А.В., Бонч-Осмоловская А.А., Захаров Л.М., Кривнова О.Ф. База данных «Интонация русского диалога»: побудительные реплики. // Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции Диалог 2006 г.
3. Кодзасов С.В., Архипов А.В., Захаров Л.М., Кривнова О.Ф. База данных «Интонация русского диалога»: реплики-сообщения. Компьютерная лингвистика и интеллектуальные технологии // Труды международной конференции Диалог 2007 г.