

База данных «интонация русских повествовательных текстов»¹

The database on intonation of russian narrative texts

Кодзасов С. В. (sankod@philol.msu.ru),

Архипов А. В. (arhipov@philol.msu.ru),

Захаров Л. М. (leon@philol.msu.ru),

Кривнова О. Ф. (okri@philol.msu.ru)

Московский государственный университет им. М. В. Ломоносова,
Москва, Россия

В докладе сообщается о результатах 2-го этапа работы по созданию БД «Интонация русских информационных и повествовательных текстов». Этот этап открывал 2-й трехлетний цикл исследований русской интонации.

Напомним, что **1-й трехлетний цикл** работ нашей группы (2004–2006 гг.) был связан с интонацией **диалогического** дискурса (интонация вопросов, побуждений, сообщений и реакций на них). Результаты исследований авторского коллектива в этой области сообщались на прошлых конференциях «Диалог» (см. список публикаций).

2-й трехлетний цикл посвящен интонации **недиалогического** дискурса. Он также состоит из трех этапов:

2007 г. — интонация информационных текстов (на материале телевизионных новостей);

2008 г. — интонация литературных повествовательных текстов (на материале записей современной прозы: Булгаков, Пелевин, Маринина), а также интонация фольклорных повествовательных текстов в исполнении профессиональных чтецов;

2009 г. — интонация спонтанных нарративов (рассказы о случаях из жизни).

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.

Как было сказано выше, до сих пор нашей группой изучалась в основном интонация диалогического дискурса. В осуществляемом ныне проекте ставится задача создания компьютерной базы для изучения интонационной специфики информационного и повествовательного дискурса. При этом просодия текстов исследуется в тесной связи

с их иллокутивным и модальным содержанием, а также с их фазовой структурой. Особое внимание уделяется просодическим техникам маркировки жанровых и регистровых характеристик текстов и их компонентов. Это позволит уточнить системы семантических, прагматических и стилистических дескрипторов, которые используются в отечественной 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 разных произносительных техник). В число потенциальных пользователей таким образом попадают не только лингвисты, но также дикторы и составители текстов в средах массовой информации.

Проект предполагает создание трех представительных корпусов звучащих текстов для трех базовых недиалогических областей дискурса: информационные тексты (2007 г.), нарративные тексты разных жанров (современные литературные тексты, фольклорные тексты) (2008 г.), спонтанные нарративы (2009 г.). Характерные фрагменты для каждого типа дискурса вводятся в компьютерную базу данных. Каждый файл снабжается 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.

Ниже приводятся фрагменты базы данных для фольклорных и нарративных текстов, прочитанных профессиональными чтецами.

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

1. Сказка. *Кот, козёл да баран*

Код	Орфография	Просодия	Комментарии
КОТ.0.			
КОТ.0.1.	Кот, козел да баран.		
КОТ.1.			
КОТ.1.1.	Жили-были в одном дворе кот, козел да баран; жили они дружно: сена клок и тот пополам.	*Жили(\)\-*были(/) в *одном(\) дворе(/) *кот(-/), *козел(-/ да(/-) *баран(-\); *жили(-/) они *дружно(-\): *сена(\) *клок(\) и *тот(\) пополам(/~).	Пословная акцентуация интродуктивного предложения — типичная фигура при чтении сказок.
КОТ.1.2.	А коли вилы в бок — так одному коту Ваське!	А коли *вилы(/) в бок — так одному коту *Ваське(\)!	
КОТ.1.3.	Он такой вор и разбойник, каждый час на промысле, и где что плохо лежит, туда и глядит.	*Он *такой(/) *вор(/(:) и *разбойник(\):), *каждый(/) *час(\) на промысле, и где что плохо *лежит(/), *туда(\н) и *глядит(\в).	Экспрессивное удлинение гласных: <i>вор</i> (:), <i>разбойник</i> (:)
КОТ.2.			
КОТ.2.1.	Вот однажды лежат себе козел и баран и *разговаривают промеж собой.	*Вот(/) *однажды(/) *лежат(/-) себе козел и *баран(/-) *и *разговаривают(/)\) промеж *собой(\).	Лексическая безакцентность слова <i>козел</i> . Слово <i>коза</i> в этом контексте скорее имело бы выделительный акцент.
КОТ.2.2.	Откуда ни возьмись — котикшко-мурлышко, серый лобишко, идет да таково жалостно плачет.	*Откуда ни *возьмись — *котишко(/)-*мурлышко(\), *серый(\) *лобишко(\), *идет(\) да *таково(\) *жалостно(:) *плачет(:).	Иконические долготы в словах <i>жалостно плачет</i> .
КОТ.3.			
КОТ.3.1.	Козел да баран спрашивают:	*Козел(/) да *баран(/) *спрашивают(\):	
КОТ.3.2.	Кот-коток, серенький лобок, о чем плачешь, на трех ногах скачешь?	*Кот-коток(\ — \), *серенький(\) *лобок(/)\), о *чем(/) *плачешь(-), на *трех(/) *ногах(\) *скачешь(\)?	

2. Литературный текст в жанре детектива (А. Маринина)

Код	Орфография	Просодия	Комментарии
МАР1.1.			
МАР1.1.1.	Субботнее утро Настя Каменская провела за своим любимым занятием.	*Субботнее(/) *утро(/) *Настя(/) *Каменская(/) *провела(/) за *своим(/) *любимым(/) *занятием(/).	Пословная акцентуация, характерная для экспозиции повествования.
МАР1.1.2.	Она ленилась.	Она *ленилась(/).	Восходящий тон на сонорном. Направление тона трудно объяснить иллюкативной семантикой.
МАР1.1.3.	Еще вчера вечером на вопрос мужа: «Чем собираешься завтра заниматься?» — она честно ответила: «Буду лениться».	*Еще(/)\) вчера *вечером(/) на вопрос *мужа(\): «*Чем(\) *собираешься(\) *завтра(\) *заниматься(/)?» — она *честно(\) *ответила(/): «*Буду(/) *лениться(\)\)».	Специфическая тональная огласовка слова <i>еще</i> нуждается в интерпретации.

Код	Орфография	Просодия	Комментарии
МАР1.2.			
МАР1.2.1.	И вот теперь она валялась в постели, прихлебывая крепкий горячий кофе, слушала музыку и предавалась неспешным размышлениям.	И *вот(\\) *теперь(\\) она *ва- лялась(\\) в *постели(/), прихлебы- вая *крепкий(/) горячий *кофе, *слу- шала(\\) *музыку(/) и *предавалась *неспешным(/) *размышлениям(\\).	Непонятно отсутствие тонального акцента на слове кофе. Тре- бует истолкования восходяще-нисходящий тон на слове слушала.
МАР1.2.2.	Правда, надо отдать ей долж- ное — размышления были все-таки связаны с работой.	[*Правда(/), *надо(\\) *отдать(\\) ей *должное(\\) — *размышле- ния(\\) *были(\\) все-таки *связаны(\\) с *работой(\\)](Н).	Низкий регистр тона всего фрагмента.
МАР1.2.3.	Во-первых, она думала об исчезновении веществен- ных доказательств по делу об убийстве пятнадцатилет- него подростка.	*Во-первых(\\), она *думала(\\) об ис- чезновении(\\) вещественных *до- казательств(\\) по *делу(\\) об *убий- стве(\\) *пятнадцатилетнего(\\) *под- ростка(\\)	Цепочка нисходящих акцентов, в том числе и в позициях незавер- шенности высказыва- ния.
МАР1.2.4.	Этим убийством их отдел занимался вот уже четыре месяца.	Этим *убийством(/) их *отдел(/) *занимался(/) *вот(\\) *уже(\\) *четы- ре(\\) *месяца(\\).	Цепочка восходящих тонов сменяется цепоч- кой нисходящих. Можно проинтерпретировать как различие тематиче- ского и рематического компонента высказы- вания.
МАР1.2.5.	Во-вторых, она думала о свалившемся на них два дня назад убийстве пятерых человек — целой семьи из- вестного московского худож- ника портретиста.	*Во-вторых(/), она *думала(/) о *свалившемся(/) на них *два *дня *назад *убийстве(\\) *пятерых(\\) \\) *человек(\\) /) — *целой(\\) *се- мьи(\\) /) *известного(/) *моск ^о вско- го(\\) *художника - *портретиста(\\).	Мотивация направле- ния тона для большин- ства слов неочевидна. Простую интерпрета- цию допускают лишь начальный восходящий и конечный нисходя- щий акценты.
МАР1.2.6.	В-третьих, Анастасия Камен- ская с раздражением думала о том, что нужно получить новый комплект форменной одежды, а для этого нужно найти старые ордера, по ко- торым она так и не получила форму в прошлый раз.	*В-третьих(\\) /), *Анастасия(\\) *Каменская(/) с *раздражением(\\) (/) думала о *том(/), что *нужно(/) *получать(\\) *новый(/) *комплект(\\) форменной *одежды(/) /) , а для *этого(/) *нужно(/) найти старые *ордера(/), по *которым(\\) она *так(/) и *не(\\) *получила(\\ \\) *фор- му(/) в *прошлый(/) *раз(\\).	Также и здесь почти все слова внутри фрагмента могут быть огласованы разными тональными акцентами. Кроме того, тональные акценты для большинства слов необязательны.
МАР1.2.7.	Куда она эти ордера засуну- ла, Настя вспомнить не мог- ла, стало быть, придется сочинять покаянный рапорт об их утере.	*Куда(/) она *эти(\\) ордера засунула, *Настя (/) вспомнить не *могла(-\\), стало быть, *придется (\\) сочинять покаянный *рапорт(/) об их *утере. (\\)	Также и этот фрагмент допускает разные про- чтения. Например, возможен восходящий акцент на слове засу- нула. Также, возможен выделительный акцент на слове сочинять.

Как и результаты предшествующего этапа работы, полученная информация представлена в виде базы данных в формате MICROSOFT ACCESS. Программа позволяет прослушивать файл, видеть его осциллограмму и интонограмму, а также анализировать его просодическую дескрипцию. Какие интонационные характеристики фиксирует эта дескрипция? Прежде всего, это акцентная структура текстов (размещение семантически нагруженных выделительных акцентов — они обозначены звездочками). Далее, это тональная структура компонентов сообщения (тоны обозначены слешами) и средства их экспрессивного подчеркивания. Производился также упрощенный анализ фазовых и тематических компонентов текста, что облегчает пользователю анализ функций просодических средств его оформления.

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

Приложение

Интонационная транскрипция:

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
- /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
- \/ — нисходяще-восходящий тон внутри гласного
- /- — восходящий плюс ровный без падения

Нетональные акценты

- : — долготный акцент на гласном
- * — громкостный акцент на подчеркнутом гласном 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

Литература

1. Кодзасов С. В., Бонч-Осмоловская А. А., Захаров Л. М., Кобозева И. М., Кривнова О. Ф. База данных «Интонация русского диалога»: вопросительные реплики // Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции «Диалог 2005». М., 2005. С. 245–249.
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». М., 2006. С. 236–242.
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». М., 2007. С. 269–277.
4. Кодзасов С. В., Архипов А. В., Захаров Л. М., Кривнова О. Ф. База данных «Интонация русских информационных текстов» // Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции «Диалог '2008». М., 2008. С. 206–209.