

# Звуковой корпус русского языка «один речевой день»: пути пополнения и первые результаты исследования<sup>1</sup>

## The ord speech corpus of Russian everyday communication: ways of replenishment and first results of analysis

**Богданова Н. В.** (nvbogdanova\_2005@mail.ru),  
**Асиновский А. С.** (a.s.asinovsky@gmail.com),  
**Маркасова Е. В.** (markasovaelena@yandex.ru),  
**Степанова С. Б.** (stsvet\_2002@mail.ru),  
**Супрунова А. В.** (nsuprunova@mail.ru),  
**Шерстинова Т. Ю.** (sherstinova@gmail.com)

Факультет филологии и искусств Санкт-Петербургского государственного университета, Санкт-Петербург, Россия

В докладе представлены некоторые конкретные результаты анализа материалов звукового корпуса «Один речевой день», а также пути его пополнения и дальнейшего развития.

Корпус «Один речевой день» (ОРД) представляет собой один из блоков Звукового корпуса русского языка (ЗКРЯ), работа над которым ведется на факультете филологии и искусств СПбГУ. Принцип, положенный в основу организации этой части корпуса, условно можно назвать «*принципом невода*»: забрасываем широкую сеть в средуносителей языка, вытягиваем все, что в нее попало, и делаем это объектом многоуровневого анализа. Предложенная в определении принципа метафора легко поддерживается высказываниями авторитетных лингвистов, которые сравнивают народный разговорный язык (т. е. речь) с широкой и мощной рекой, ледяной покров которой — это письменность (см.: Вандриес 1937: 253; Bally 1952: 13). Еще Л. В. Щерба писал о необходимости «эмансипации от письменного языка и обращения к живой речи» (Щерба 1974: 144). Именно в целях такого обращения к живой речи в ее свободном функционировании, не ограниченном ни лабораторными условиями записи, ни конкретными речевыми заданиями для информантов, и была применена методика 24-часовой

записи (подробнее о корпусе ОРД см.: Степанова и др. 2008; Богданова и др. 2009). *Принцип невода* позволяет увидеть реальную, естественную, а не искусственно созданную в лабораторных условиях, жизнь, отраженную в речи.

Работа над этой частью корпуса продолжается в настоящее время по двум направлениям: пополнение его новыми записями и анализ существующего материала. Целью настоящего сообщения является как раз представление новых поступлений в корпус, первые конкретные результаты лингвистического анализа материала на разных уровнях и определение перспектив дальнейших исследований.

Новые записи «Одного речевого дня» были осуществлены с целью сбалансировать гендерный состав информантов. Так, первый «заброшенный невод» принес вполне ожидаемый в этом отношении «улов»: из 30 информантов — участников пилотного эксперимента — оказалось 11 мужчин и 19 женщины, что вполне отражает гендерную ситуацию в нашем обществе, однако снижает возможности сопоставительного описания речи мужчин и женщин в сходных

<sup>1</sup> Исследование выполнено при поддержке гранта РГНФ «Разработка информационной среды для мониторинга устной русской речи» (09-04-12115в).

коммуникативных ситуациях. Для выравнивания (балансировки) этого состава были записаны еще 9 информантов-мужчин, в основном старшего возраста, и 1 информант-женщина. Некоторое отступление от выдвинутого принципа невода сполна окупается тем, что во всех остальных отношениях он остается ведущим при формировании ОРД. В наибольшей степени принцип невода работает, когда речь идет не о самих информантах, а об их коммуникантах, число которых достигает в настоящее время 600 и состав которых — не только гендерный, но также социологический и психологический — не поддается никакой балансировке. В результате в зоне внимания исследователей, авторов и разработчиков этого проекта, оказываются и записи речевого материала (также далекие от всякой балансировки — тематической, ситуативной и т. п.), и все характеристики всех информантов и их коммуникантов, что открывает почти безграничные возможности для лингвистического (в том числе социо- и психолингвистического) анализа современной русской речи.

Другой путь пополнения ОРД и в целом развития исследований в этом направлении — это некоторое пересечение материала двух блоков ЗКРЯ: данного корпуса повседневной речи носителей языка и сбалансированной части корпуса, устроенной, в отличие от первой, по «принципу ковчега» (см. о ней подробнее: Богданова и др. 2008, а также статью Н. В. Богдановой в настоящем сборнике). Так, одним из новых участников эксперимента стал информант (девушка-студентка, типичный экстраверт), речь которого до этого записывалась многократно в течение трех лет по одной и той же лингвистической программе (чтение, пересказ, описание изображения, свободный рассказ на заданную тему). Каждый речевой сценарий повторялся 4 раза, с интервалом в 2 месяца, что позволило выявить особенности не конкретного речевого акта, а речи данной языковой личности в целом (создать речевой портрет человека). Даже с учетом того факта, что и на этот раз информант знал о проводящейся записи, его речь следует признать значительно более естественной по сравнению с речью в микрофон, в присутствии исследователя. Таким образом появилась возможность сравнить речь одного говорящего в разных условиях записи: в перспективах этого направления исследования — повторные (как минимум, еще одна, через год) записи речевого дня данного информанта, для выявления всех особенностей его речевого поведения и уточнения его речевого портрета.

Любопытным представляется и опыт записи ОРД от информанта-иностранца, живущего в данное время в России. Таким информантом стала студентка из Швейцарии, приехавшая учиться в СПбГУ. Первая запись ее ОРД была осуществлена в ноябре 2009 г., в первые месяцы пребывания в стране изучаемого языка. Эта запись продемонстрировала не только определенный уровень владения инфор-

мантом русским языком, а также смену языков (русский, немецкий, французский) в различных коммуникативных ситуациях, но и специфику обращенной к нему речи его коммуникантов, носителей русского языка. Повторные записи ОРД от того же информанта, которые планируются на ноябрь 2010 и 2011 гг., позволят увидеть не только динамику овладения иностранцем русской речью (на всех уровнях), но и изменение особенностей речи его коммуникантов, которые невольно реагируют на степень интерферированности речи собеседника.

В целом корпус ОРД характеризуется в настоящее время такими количественными показателями: 320 часов звучания, полученные от 40 информантов (20 мужчин и 20 женщин). Звукозаписи переформатированы, убраны длительные (больше 5 минут) шумовые фрагменты, не содержащие речи. Звукозаписи разрезаны на коммуникативные эпизоды по принципу общих условий коммуникации и качества звукозаписи (см.: Asinovsky et al. 2009). В результате было получено 994 файла-эпизода общей продолжительностью 268 часов. Подготовлены методики многоуровневого аннотирования данных на лингвистическом и паралингвистическом уровнях (см.: Шерстинова и др. 2009). Общая структура корпуса «Один речевой день» описана в работе: Sherstinova 2009a.

Осуществлена расшифровка 34 часов звукозаписей для 40 информантов и их многоуровневое аннотирование в профессиональной программе ELAN. В результате расшифровки получены тексты общим объемом в 244 075 словоупотреблений на уровне Frase (реплики говорящих), которые относятся к 125 эпизодам и соответствуют 33,87 часам непрерывной звучащей речи. Для 20 информантов-мужчин получено 106 109 словоупотреблений на уровне реплик, что соответствует 15,10 часам непрерывной речи, для 20 информантов-женщин — 137 966 словоупотреблений, что соответствует 18,77 часам.

Аннотирование осуществлялось по следующим уровням:

- Phrase (реплики говорящих),
- Speaker (код говорящего),
- Events (невербальные аудиосообщения),
- Voice (качество голоса говорящего),
- PhonetCom (фонетический комментарий),
- PhraseComment (фразовый комментарий),
- Notes (общий комментарий),
- Episode (мини-эпизод речевой коммуникации).

Реплики говорящих содержат синтагматическое и фразовое членение.

По материалам расшифрованных записей 40 информантов был получен частотный словарь всех использованных в речи словоформ и выполнен первичный анализ этих данных (Sherstinova 2009b). Полученные количественные данные о частоте употребления каждой словоформы были перенесены в программу MS Excel и пропущены через сортировку от максимального к минимальному показателю упо-

требительности. Единицей описания в настоящем исследовании признается не только графическое слово («от пробела до пробела»), с которым работают многие морфологические анализаторы и программы по созданию конкорданса, но и знаки сегментирования, а также символы, обозначающие паралингвистические явления, сопровождающие естественную речь (<смех>, <кашель>, <вдох> и под.).

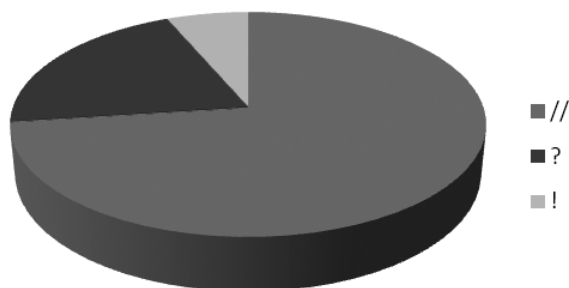
Параллельно с созданием корпуса начаты исследования фонетики и грамматики русской повседневной речи, а также особенностей повседневной коммуникации (см.: Богданова 2008; Королева 2009; Маркасова 2009а; Маркасова, Корякова 2008; Асиновский 2010; Королева, Журавлева 2010; Рыко, Степанова 2010).

Расшифрованный материал корпуса включает более чем 32 000 фраз. Подавляющее большинство высказываний в материале исследования оказались утвердительными (23 494), восклицательных и вопросительных высказываний встретилось около десяти тысяч (см. табл. 1. и рис. 1).

**Таблица 1.** Абсолютные данные сегментирования в материале ОРД

Значение символа	Знак сегментирования	Абсолютное количество
Знак членения на повествовательные высказывания	//	23 494
Знаки для обозначения восклицательных и вопросительных высказываний	?	6627
	!	2080
ИТОГО		32 201

Как хорошо видно на диаграмме (см. рис. 1), говорящий в течение своего речевого дня гораздо чаще нечто сообщает собеседникам, лишь четвертая часть всех высказываний в материале корпуса — вопросительные и восклицательные конструкции.



**Рисунок 1.** Соотношение различного рода высказываний в материале ОРД

Начато сегментирование речевого материала на лексическом уровне. Осуществлена сегментация 14 884 слов для 5 информантов (S01, S19, S24, S35, S37) (147 минут звучания).

Получена выборочная реальная транскрипция для 1000 словоизменяемых морфем из речи 4 информантов. Морфемное аннотирование включает три уровня — орфографическую запись морфемы, её функциональный тип и реальную фонетическую транскрипцию.

Что касается конкретных результатов исследования материалов этой части ЗКРЯ, то они связаны с разными уровнями анализа и частично представлены в настоящем сборнике (см. доклад Е. В. Маркасовой и С. А. Воробьёвой о функционировании в корпусе ОРД дискурсивного слова *конечно*, а также доклад А. С. Асиновского и др. «Звуковая реальность словоизменяемых аффиксов (по данным Звукового корпуса русского языка)»).

Весьма перспективным представляется анализ лексики на материале ОРД, сравнение её частотности и контекстуальных значений с тем, что представлено в Национальном корпусе русского языка (НКРЯ) и словарях, построенных на материале письменной речи. Так, анализ дискурсивной лексики в ОРД позволил констатировать постепенный «уход» из повседневной речи таких вводных слов, выражающих (по классическим их определениям) оценку говорящим степени достоверности сообщаемого, как *несомненно*, *бесспорно*, *вне всякого сомнения*, а встретившееся *безусловно* можно рассматривать, скорее, как маркер манипулятивного речевого поведения личности, имеющей иллюзивную установку доминирования (см. Маркасова 2009а).

Аналогичная ситуация с «маркерами искренности»: такие конструкции, как *говоря по совести*, *чего скрывать*, *положа руку на сердце* и под., не встретились в корпусе ОРД ни разу. Конструкции же, достаточно частотные в повседневной речи — *честно говоря* и её вариант (*если*) *честно*, — в языке повседневного общения используются не в прямом значении, то есть не в качестве маркера искренности в речевом акте признания, а в качестве контактоустанавливающего элемента, вносящего в высказывание семантику солидаризации или интимизации общения (см. примеры 1, 2, 3), или как заполнитель паузы гезитации (примеры 4, 5) (см. Маркасова 2009б):

- (1) у меня вот **честно говоря** / желание / только <...> / пойти (...) влить / в себя что-нибудь ... / чтоб повеселело;
- (2) я тоже **честно говоря** с удовольствием кофе бы попил;
- (3) по-моему попьём и не умрем // **честно говоря** / я дома тут пристрастился тоже Нурю пить;
- (4) да ну не знаю / **честно говоря** не помню;
- (5) я **честно говоря** // во-первых ну я могу конечно показать вам эти анкеты.

Проведенный анализ фразеологических оборотов (ФО), используемых в повседневной речи носителей русского языка, также выявил ряд интересных особенностей их функционирования (см. подробнее: Супрунова 2010). Количество употребляемых фразеологизмов, как показало исследование, напрямую связано с их лингвистическими характеристиками. Можно говорить о достаточно высокой частоте употребления ФО в живой речи, однако это относится лишь к определенным их типам.

Так, книжные фразеологизмы в спонтанной речи (СР) почти не употребляются (2 % от общего количества ФО в расшифровках ОРД) (*завершающий аккорд; притча во языцех; пожинать плоды; камень преткновения*), тогда как разговорных и просторечных ФО достаточно много (68 %) (*челюсть отвисла; крышник едет; вынь да положь; поджилки трясутся; дать на лапу; галопом по Европам; ни в одном глазу*).

В живой речи практически отсутствуют фразеологические сочетания (0,3 %) (*В кромешной тьме*) и мало фразеологических сращений (14 %) (*зуба не дура; дойти до ручки; ни в одном глазу; плести всякую ахинею*), при большом количестве фразеологических единств (35 %) (*стоять над душой; раз плюнуть; ставить на карту*) и фразеологических выражений (51 %) (*в крайнем случае; будь другом; время от времени; изо всех сил; посреди белого дня*).

Носители языка в целом предпочитают с помощью ФО характеризовать в своей речи действия, поступки человека (17 %), а также выражать эмоции (14 %), что обуславливает преобладание в СР фразеологических оборотов следующих лексико-грамматических разрядов:

- адвербиальные (27 %) (*задним числом; всеми силами души; кровь из носу*),
- глагольные (18 %) (*делать крюк; дать сдачи; убить время; лезть в дебри*);
- междометные (21 %) (*не дай бог; с ума сойти!; ничего себе!; слава богу; мать моя женщина*).

Что касается эмоционально-оценочной коннотации, то в живой речи используется небольшое количество ФО с положительной оценкой (12 %) (*первый парень на деревне; голубая кровь; не из робкого десятка*) и очень мало ФО с эмоционально-экспрессивной окраской (ЭЭО):

- осуждение (0,6 %) (*на иглу сажали; денег некуда девать*),
- восхищение (1,2 %) (*владеет словом; как игрушка*).

Носители языка стремятся быть достаточно сдержанными в своей речи, поэтому в ней доминируют фразеологические обороты с нейтральной коннотацией (57 %), без ЭЭО (70 %) (*приходить в голову; асфальтовая болезнь; изо всех сил; один к одному; держать на замке*).

Тем не менее, в речи информантов встретилось большое количество ФО с отрицательной коннотацией (28 %) (*понты раскинуть; горло рвать; поте-*

*рять себя; глухая тетеря; дошла до ручки; на ладан дышать*). Из всех фразеологических оборотов с ЭЭО количественно преобладают единицы с не столь резкими, категоричными, в сравнении с другими, типами окраски:

- шутливость (10 %) (*на халяву и уксус сладкий; трудовая мозоль; глухая тетеря; два брата-акробата*),
- неодобрение (5 %) (*потерять себя; убить время; сидеть на шее*).

Не последнее место среди возможностей использования материалов ЗКРЯ занимает *создание лексикографического описания бытовой спонтанной звучащей речи*, что является еще одной перспективной задачей настоящего проекта. Думается, что такое описание может вестись, как минимум, в трех направлениях.

1. *Словарь русской бытовой разговорной речи*. Возможности для создания такого словаря дает электронная картотека *E-Kar*, одно из программных средств, использованных при построении Звукового корпуса. Эта программа автоматически создает конкорданс по выбранным текстам и позволяет решать многообразные задачи классификации и описания языковых единиц. В частности, она позволяет собрать по тексту все словоформы, в нем встречающиеся, посчитать их частоты, предъявить для этих словоформ любое лингвистическое расширение по тексту или группе текстов, имеющихся в электронной коллекции. Результаты автоматической работы программы (конкорданса) и последующей работы эксперта дают возможность новых содержательных форм интерпретации лингвистического материала. Необходимость и целесообразность создания такого словаря можно проиллюстрировать простым примером. В существующих к настоящему времени расшифровках материалов ЗКРЯ обнаружилось всего два употребления слова *хлеб*, оба — в значении 'кусоч хлеба' в контекстах одинакового типа: *намазать хлеб маслом*. Проверка по материалам МАС показала, что подобное — весьма и весьма распространенное — значение данного (очень частотного) слова в Академическом словаре попросту отсутствует. Электронная версия словаря русской бытовой разговорной речи позволит не только увидеть все новые значения привычных слов и получить их контексты, но и — при необходимости — услышать их звучание в реальной речи, что обеспечивается другими программными средствами, использованными при создании ЗКРЯ.

2. *Словарь контекстных экспрессивных русской разговорной речи*. Уже первые шаги в реализации этого проекта убеждают в том, насколько он может быть интересен и полезен самым разным специалистам по изучению русской речи.

Вот, например, как выглядят фрагменты корпуса, размеченные с лексикографической точки зрения:

- она привозила с собой этого немца / который () вот **явный немец** / а **косил под русского Ваньку** // ему первые два дня было запрещено говорить // потом (...) он... **под дурачка косил** ? он был **первым парнем на деревне** // там к сему (...) / к нему все девки (...) пока не проговорился к нему все девки приставали // пока не проговорился / да / по-немецки / пока молчал // угу // это я на колонке угостил (э-э) немца // салом // выпьет водки // он долго не хотел / а потом как **подсел на это сало** / короче // и давай **один бутерброд за другим наворачивать** // просто за друг... (...) перед другими людьми уже неудобно было // так **попёрло его**;
- Дора / **типа** привет / у меня к тебе **шкурная тема** Коля / и какая у тебя **шкурная тема** ко мне? ну **типа** / причем разговаривает в каком-то таком стиле // **красавица с накладными мозгами**;
- они (э) дела... стараются делать для всех / то есть / невзирая на личности / да ? угу просто / что видит человека / вот **из него можно доить денежку** // вот например торговый представитель / определённый // **с него можно тянуть денежку** / определённую // вот он сейчас... / его запустить угу / **выжать из него все соки** / получить с него максимум / (а...) прибыли / и его можно выкинуть.

Видно, как много материала для лексикографического описания экспрессивной лексики содержат материалы ЗКРЯ.

3. Наконец, третье направление лексикографического описания материалов ЗКРЯ может быть связано с *корпусом редуцированных форм* (РФ) русской речи (*рупь, дянька, пийсят, прально, тода, то ись, кагоритсы, собсно гря* и под.). В известном смысле это направление пересекается со словарем русской бытовой разговорной речи (см. выше п. 1), поскольку электронная версия такого словаря предполагает возможность прослушать любое слово в любом контексте:

- *просто не горит* // а чек ты в сумку не **хошь** положить?
- **здрасть** / отдел кадров уже закрыт? ага / **сён-ня** же пятница;
- а-а вот мы / живем / на Смолячкова / **ничо** не знаем.
- мы как раз здесь побудем / за час уже выболтается всё // а он же там **буит**... это;
- м-м хорошо / давай // ты во **скоко** будешь дома?

Наличие словаря РФ, словник которого построен по алфавиту исходных словарных форм, позволит целенаправленно искать ту или иную единицу в материале корпуса и получать, например, статистику не только употребительности того или иного слова или словосочетания, но и представленности в нашей повседневной речи всех реальных форм языковых единиц. Образец словарной статьи подобного словаря см. в табл. 2.

Таблица 2. Структура словарной статьи словаря редуцированных форм русской речи

Лексема	Частеречное значение	Варианты написания	Примеры из НКРЯ и др. письменных источников	Частотность (если возможно)	Варианты произношения	Примеры из ЗКРЯ	Частотность	Специфика функционирования	Специфика значения	Примечания

Такой словарь может не только быть полезен специалистам по устной спонтанной речи, но и стать неоценимым помощником преподавателя русского языка в любой аудитории, ориентированной на знакомство с современной русской речью (см. о нем подробнее: Богданова, Пальшина 2010).

По мере сбора и обработки материалы ЗКРЯ традиционно передаются в устный подкорпус Национального корпуса русского языка («Из материалов корпуса “Один речевой день”, подготовленного группой А. С. Асиновского»), что делает их доступными для самого широкого круга пользователей.

## Литература

1. Асиновский А. С. Как звучат русские флексии // Материалы XXXVIII международной филологической конференции. Выпуск 22. Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. 16–20 марта 2009 г. Санкт-Петербург / Отв. ред. А. С. Асиновский, науч. ред. Н. В. Богданова. СПб., 2010. С. 4–28.
2. Богданова Н. В. Аллегорные формы русской речи: от произносительной редукции к письменной фиксации и лексикализации в языке // Материалы XXXVII международной филологической конференции. Фонетика. 10–15 марта 2008 года / Отв. ред. Н. Д. Светозарова. СПб., 2008. С. 31–44.
3. Богданова Н. В., Асиновский А. С., Русакова М. В., Рыко А. И., Степанова С. Б., Шерстинова Т. Ю. Звуковой корпус как способ монито-

- ринга и фиксации разных форм естественного языка // Компьютерная лингвистика и интеллектуальные технологии. Выпуск 8 (15). По материалам ежегодной международной конференции «Диалог» (2009) / Гл. ред. А. Е. Кибрик. М., 2009. С. 38–44.
4. Богданова Н. В., Бродт И. С., Куканова В. В., Павлова О. В., Сапунова Е. М., Филиппова Н. С. О «корпусе» текстов живой речи: принципы формирования и возможности описания // Компьютерная лингвистика и интеллектуальные технологии. Выпуск 7 (14). По материалам ежегодной международной конференции «Диалог» (2008) / Гл. ред. А. Е. Кибрик. М., 2008. С. 57–61.
  5. Богданова Н. В., Пальшина Д. А. Редуцированные формы русской речи (опыт лексикографического описания) // Материалы ежегодной Всероссийской научной конференции «Слово. Словарь. Словесность: Текст словаря и контекст лексикографии». Санкт-Петербург, 11–13 ноября 2009 г. СПб., 2010 (в печати).
  6. Вандриес Ж. Язык. М., 1937.
  7. Королева И. В. Коммуникативное взаимодействие: смена стратегии в зависимости от психо-социальных особенностей говорящих (на материале Звукового корпуса русского языка «Один речевой день») // Речевая коммуникация в современной России: Материалы I международной научной конференции (Омск, 27–29 апреля 2009 г.) / Под ред. О. С. Иссерс, Н. А. Кузьминой. Омск, 2009. С. 155–162.
  8. Королева И. В., Журавлева А. А. Личностные характеристики говорящих и их влияние на коммуникативное взаимодействие (на материале Звукового корпуса русского языка «Один речевой день») // Материалы XXXVIII международной филологической конференции. Выпуск 22. Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. 16–20 марта 2009 г. Санкт-Петербург / Отв. ред. А. С. Асиновский, науч. ред. Н. В. Богданова. СПб., 2010. С. 76–89.
  9. Маркасова Е. Слово-резонанс «безусловно» // *Auspicia* № 2, 2009а.
  10. Маркасова Е. В. Маркеры искренности в языке повседневности (признаться сказать, говоря по совести, по чести говоря, честно говоря) // *Rossica Olomucensia* 2009б. Vol. XLVIII. № 2. S. 149–155.
  11. Маркасова Е. В., Корякова Е. В. Речевая агрессия и диагностика личностной акцентуации (по данным корпуса «Один речевой день») // Материалы XXXVII международной филологической конференции. Выпуск 21. Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. 10–15 марта 2008 года / Отв. ред. А. С. Асиновский, науч. ред. Н. В. Богданова. СПб., 2008. С. 62–70.
  12. Рыко А. И., Степанова С. Б. Индивидуальные стратегии членения спонтанной речи на предложения (по результатам перцептивного эксперимента) // Материалы XXXVIII международной филологической конференции. Выпуск 22. Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. 16–20 марта 2009 г. Санкт-Петербург / Отв. ред. А. С. Асиновский, науч. ред. Н. В. Богданова. СПб., 2010. С. 128–137.
  13. Степанова С. Б., Асиновский А. С., Богданова Н. В., Русакова М. В., Шерстинова Т. Ю. Звуковой корпус русского языка повседневного общения «Один речевой день»: концепция и состояние формирования // Компьютерная лингвистика и интеллектуальные технологии. Выпуск 7 (14). По материалам ежегодной международной конференции «Диалог» (2008) / Гл. ред. А. Е. Кибрик. М., 2008. С. 488–494.
  14. Супрунова А. В. Закономерности употребления фразеологических оборотов в устной речи (на материале Звукового корпуса русского языка) // Материалы XXXVIII международной филологической конференции. Выпуск 22. Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. 16–20 марта 2009 г. Санкт-Петербург / Отв. ред. А. С. Асиновский, науч. ред. Н. В. Богданова. СПб., 2010. С. 147–155.
  15. Шерстинова Т. Ю., Степанова С. Б., Рыко А. И. Система аннотирования в звуковом корпусе русского языка «Один речевой день» // Материалы XXXVIII международной филологической конференции. Формальные методы анализа русской речи. 16–20 марта 2009 г. Санкт-Петербург. СПб., 2009 (в печати).
  16. Щерба Л. В. О разных стилях произношения и об идеальном фонетическом составе слова // Л. В. Щерба. Языковая система и речевая деятельность. Л., 1974. С. 141–146.
  17. Asinovsky A., Bogdanova N., Rusakova M., Stepanova S., Ryk A., Sherstinova S. The ORD Speech Corpus of Russian Everyday Communication «One Speaker's Day»: Creation Principles and Annotation // LNCS/LNAI series. «Text, Speech and Dialogue» TSD-2009. Springer-Verlag, Berlin, Heidelberg. 2009. Pp. 250–257.
  18. Bally Ch. Le langage et la vie. Genève-Lille, 1952.
  19. Sherstinova T. The Structure of the ORD Speech Corpus of Russian Everyday Communication // «Text, Speech and Dialogue» TSD-2009. LNAI 5729. Springer-Verlag, Berlin, Heidelberg. 2009a. Pp. 258–265.
  20. Sherstinova T. Quantitative data processing in the ORD speech corpus of Russian everyday communication // Proc. of International Quantitative Linguistics Conference QUALICO 2009. Graz, Austria. 2009b. Pp. 56–57.