

Первый
междисциплинарный семинар

«Анализ разговорной русской речи»

АР³-2007



29 августа 2007 года,
Санкт-Петербург, СПИИРАН

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Санкт-Петербургский институт информатики и автоматизации
Санкт-Петербургский государственный университет
аэрокосмического приборостроения

Первый
междисциплинарный семинар
«Анализ разговорной русской речи»

АР³ - 2007

29 августа 2007 года,
Санкт-Петербург, СПИИРАН

Санкт-Петербург
2007

УДК 004.522

РОССИЙСКАЯ АКАДЕМИЯ НАУК

Санкт-Петербургский институт информатики и автоматизации

Санкт-Петербургский государственный университет аэрокосмического приборостроения

Санкт-Петербург, 199178, 14 линия, 39.

<http://www.spiras.nw.ru/speech>

Труды первого междисциплинарного семинара

«Анализ разговорной русской речи» (АР³ - 2007). – СПб.: ГУАП, 2007. – 87 с.

ISBN 978-5-8088-0270-4

Издание представляет собой сборник докладов, сделанных на заседаниях первого междисциплинарного семинара «Анализ разговорной русской речи» (АР³ - 2007), проходившего 29 августа 2007 года в Санкт-Петербургском институте информатики и автоматизации Российской академии наук. Семинар посвящен обсуждению особенностей разговорной речи и возможных подходов к автоматическому анализу русской речи. Междисциплинарный подход к изучению речи позволит скорее продвинуться в моделировании речевой деятельности и решить фундаментальную проблему человеко-машинного диалога.

УДК 004.522

Статьи печатаются в авторской редакции.

Издание осуществлено за счет средств европейского гранта SIMILAR IST-2002-507609.

ISBN 978-5-8088-0270-4

© СПИИРАН, 2007

© Коллектив авторов, 2007

© ГУАП, 2007

Оглавление

<i>А.Л. Ронжин.</i> Особенности автоматического распознавания разговорной русской речи	4
<i>Н.А. Коротяев, А.А. Кибрик, В.И. Подлеская.</i> Полипредикативность и интеграция в устном нарративном дискурсе	14
<i>А.В. Венцов, Е.И. Риехакайнен, Н.А. Слепокурова.</i> Ментальный лексикон и восприятие редуцированных форм	21
<i>П.М. Эйсмонт.</i> Роль визуального и звукового восприятия при порождении спонтанного нарратива	25
<i>Е.М. Малов, Е.В. Горбова.</i> Дискурсивные слова в русской разговорной речи (на материале анализа спонтанной разговорной речи)	31
<i>С.В. Андреева.</i> Разговорная русская речь и ее единицы	37
<i>Е.П. Комовкина, Н.А. Слепокурова.</i> Об опыте составления частотного словаря устного спонтанного текста	43
<i>В.В. Евдокимова.</i> Вариативность формантной картины гласных в разных видах речи	49
<i>И.В. Жарков, С.Н. Грамницкий.</i> Обобщённый фонетический алфавит	55
<i>Б.М. Лобанов, Л.И. Цирульник.</i> Моделирование внутрисловных и межсловных фонетико-акустических явлений полного и разговорного стилей речи в системе синтеза речи по тексту «Мультифон»	57
<i>Л.А. Васильева, С.О. Тананайко.</i> Статистические характеристики русских согласных и динамика современных произносительных тенденций в русской разговорной речи	72
<i>Ал.Б. Леонтьева, И.С. Кипяткова.</i> Моделирование нефонемных речевых элементов и создание альтернативных транскрипций для распознавания спонтанной речи	77

Особенности автоматического распознавания разговорной русской речи

*Санкт-Петербургский институт информатики и автоматизации,
Санкт-Петербург, Россия,
ronzhin@iias.spb.su*

1. Введение

В течение нескольких лет работы над созданием системы автоматического распознавания русской слитной речи был накоплен некоторый опыт и выявлен ряд проблем, с которыми сталкиваются разработчики речевых технологий [1,2]. Существенное различие между обучающими речевыми данными и теми, что приходится обрабатывать в реальных условиях – является основной причиной ошибок систем распознавания. Вариативность таких факторов как произношение, темп, стиль речи, а также окружающие шумы зачастую невозможно учесть заранее на этапе обучения системы. Кроме того, для обеспечения естественности взаимодействия какой бы ограниченной не была предметная область система должна быть обучена на распознавание достаточно большого словаря, так как в диалоге пользователи могут употреблять различные наборы слов. В связи с чем возникают проблемы низкой скорости обработки и появления близких по звучанию слов. Наконец, в зависимости от задачи меняется допустимая точность распознавания – при стенографировании требуется распознавание всех значимых слов, а в диалоговых системах для формирования запроса к информационным ресурсам иногда достаточно распознать несколько ключевых слов во фразе. Эти и другие проблемы рассмотрим более подробно.

2. Основные проблемы автоматической обработки речи

Основная методологическая проблема, которая преследовала нас и остается актуальной до сих пор – это проблема выбора предметной области (задачи). К сожалению системы распознавания речи пока еще не достигли того уровня качества, чтобы в ближайшее время поступить в массовое использование. Коммерческие структуры пока не готовы инвестировать в речевые технологии, поскольку велик риск, что фундаментальные работы в этой области не окупятся достаточно быстро. Поэтому в ходе исследований самостоятельно ставились задачи и разрабатывались демонстрационные версии диалоговых систем.

Любая система распознавания речи требует предварительного этапа обучения или настройки. Для этого необходимы речевые и языковые материалы, близкие к условиям реальной эксплуатации. Однако в распоряжении разработчиков, в лучшем случае, могут быть только аудиозаписи начитанных фраз и тексты из предметной области. Поэтому система, обученная на таком материале, естественно, дает серьезные сбои и пользователь, неудовлетворенный качеством распознавания речи, преждевременно заканчивает сеанс и обычно больше не работает с такой системой.

Единственным выходом из этой ситуации видится постепенный ввод в эксплуатацию, основанный на методике Гудвин, по которой осуществляется скрытая экспертная поддержка на первых этапах работы системы [3]. Недостатком такого подхода является высокая себестоимость на начальном этапе внедрения, поскольку требуется содержать и штат операторов, и сервер системы распознавания речи, снабженный возможностью оперативной экспертной корректировки результатов.

Наиболее сложный случай составляют те задачи, при которых предъявляются жесткие требования ко времени обработки речи. Так, например, при использовании системы распознавания речи в информационно-справочных службах пользователь ожидает естественного диалога, и система должна обрабатывать речевой запрос, производить поиск информации и генерировать текст ответа (приемлемый для синтеза речи) в режиме реального времени. Такие диалоги или сессии обычно состоят из нескольких вопросов и ответов, и их

длительность в среднем не превышает нескольких минут. Обычно системы ориентированы на предоставление информации по заданной предметной области и ожидается, что пользователи будут использовать относительно небольшой лексикон. Поэтому систему распознавания часто настраивают на поиск и обработку ограниченного числа ключевых слов в потоке речи [4]. При этом результатом работы системы распознавания является массив ключевых слов, с большой вероятностью содержащихся во входном речевом сигнале. Этого набора слов достаточно для того, чтобы произвести поиск справочной информации.

С другой стороны, система массового обслуживания должна учитывать вариативность речи, которая проявляется на всех уровнях обработки от фонетики до семантики.

Предполагается, что при речеобразовании каждый человек использует примерно одинаковый для всех набор артикуляционных движений [6]. Тем не менее, речь каждого индивида в отдельности неповторима из-за врожденных и приобретенных особенностей как голосового тракта, так и стиля речи. Кроме того, на качество речи существенное влияние оказывают свойства канала приема-передачи речевого сигнала, также важную роль играют условия окружающей обстановки. К последним можно отнести уровень окружающего шума, количество и цели участников диалога и другие [5]. Вариативность в речи возникает вследствие изменений состояния и особенностей самого диктора и внешних условий записи речевого сигнала. Некоторая классификация факторов, влияющих на вариативность речевого сигнала, представлена в таблице 1.

Таблица 1. Основные факторы, влияющие на вариативность речевого сигнала.

Долговременные				Кратко-временные			Свойства канала приема и передачи сигнала		Свойства окружающей обстановки							
Анатомические особенности речевого тракта				Приобретенные в процессе общения в окружающем обществе		Нарушения в речи		Состояние диктора								
Объем грудной клетки	Длина голосового тракта	Объем и вес губ, языка	Другие параметры, характеризующие форму речевого тракта	Диалект	Стиль речи	Заикание	Афазия	Другие болезни	Физиологическое (простуда, анестезия речевой полости, интоксикация и др.)	Эмоциональное (радость, скука, подавленность, волнение)	Тип микрофона	Удаленность от микрофона	Помехи в канале передачи сигнала	Уровень внеязычного шума	Число участников диалога	Число модальностей, доступных для общения

Указанные в таблице факторы в конечном итоге сказываются на качестве речевого сигнала, который поступает в систему распознавания речи. Основными параметрами, которые подвержены изменению и могут быть количественно оценены, являются следующие:

- громкость;
- темп;
- частота основного тона;

- спектральный состав речи;
- уровень микроколебаний в спектральной и амплитудной области;
- разборчивость, как интегральный параметр качества речи.

Для учета акустической вариативности речевого сигнала наиболее популярными сейчас считаются методы нормализации частоты основного тона [7], методы спектрального вычитания, позволяющие частично устранить шумовую составляющую в сигнале, а также скрытые марковские модели (СММ), состояния которых описываются посредством смесей гауссовских распределений плотностей вероятностей, обеспечивающих достаточно полное покрытие возможных вариантов произношения фонем с учетом фонетических контекстов и междикторских различий [8].

Кроме того, следует учитывать, что в диалоговых системах мы имеем дело со спонтанной речью и, как правило, первые два-три предложения абонента являются невнятными и малоинформативными. Это может быть приветствие или пауза колебания, в ходе которой пользователь пытается сформулировать свой вопрос и найти контакт с оператором [9]. Наличие в речевом сигнале различных «помех», возникших в ходе речеобразования (хезитации, незаконченные слова, непреднамеренные повторы слов, и т.д.), существенно ухудшает точность распознавания ключевых слов, так как вероятность возникновения ложных гипотез значительно возрастает.

Особым случаем является дистанционное распознавание речи, при котором использование микрофона-гарнитуры невозможно и приходится записывать речь диктора на расстоянии. Например, при использовании голосовых функций в информационных киосках применение гарнитуры является не эргономичным по причине кратковременности сеансов, и пользователь не станет тратить время на то, чтобы одевать и настраивать микрофон. При дистанционной записи качество речи значительно ухудшается. Во-первых, следует учитывать, что информационные киоски обычно устанавливаются в общественных местах, на улице, где условия далеки от идеальных. Для того чтобы устранить шумовую составляющую, применяют узконаправленные микрофоны или массивы микрофонов. Микрофон с узкой диаграммой направленности будет хорошо записывать только сигнал, исходящий из определенной рабочей области рядом с информационным киоском. Недостатком такого микрофона является неустойчивость к вариативности дикторов по росту.

Более перспективным является использование массива микрофонов. Одновременная запись окружающей акустической обстановки по нескольким каналам позволяет не только локализовать источники речи и шумов, но и программно настроить диаграмму направленности таким образом, чтобы отфильтровывать речь пользователя от посторонних шумов. Массив микрофонов также эффективен при записи совещаний, семинаров. Однако стенографирование таких записей пока затруднительно, поскольку возможен случай одновременных реплик нескольких дикторов, и точность распознавания будет не велика. Поэтому дистанционное стенографирование пока рассматриваться не будет.

Кроме справочных служб, к диалоговым системам как классу задач, требующих обработки речи в реальном режиме времени, можно отнести голосовое управление техническими и бытовыми объектами, где в ответ на команду пользователя система должна выполнить некоторое действие, адекватное заданию. В такой задаче обычно круг пользователей ограничен и заранее известен, поэтому такие системы получили наибольшее развитие и даже существуют коммерческие системы. Например, некоторые мобильные телефоны и смартфоны содержат опцию распознавания телефонных номеров по ключевому имени. Кроме того, дикторонезависимая модель управления MP3-плеером была представлена на международной выставке IST'2006 компанией BMW.

Другим классом прикладных задач распознавания речи является стенографирование. Чаще всего при такой задаче производится обработка некоторого монолога, записанного в достаточно хороших акустических условиях при помощи микрофона-гарнитуры. Поэтому в отличие от систем массового обслуживания, где речь поступает через телефонные каналы и/или записывается на улице, системы автоматического стенографирования получают речевой сигнал с гораздо лучшим качеством записи. Так как здесь предъявляются более мягкие

требования по скорости распознавания, то система может обработать речевой сигнал за несколько проходов, используя методы адаптации к голосу диктора и прикладной задаче [10]. Путем постепенного переобучения моделей фонем и дополняя словарь неизвестными до этого системе словами достигается максимальная точность распознавания. Сравнительная таблица требований предъявляемых к диалоговым и стенографическим системам представлена в таблице 2.

Таблица 2. Требования к системам распознавания речи

Требование	Классы прикладных задач распознавания речи	
	Диалоговые системы	Стенографирование
Размер словаря	Средний (до 5000 слов)	Сверхбольшой (свыше 100000 слов)
Дикторнезависимость	Обязательно	Желательно. Возможна адаптация
Предметная область	Узкая (ограничена стратегией диалога)	Неограниченная (каждый монолог из своей предметной области).
Условия записи	Телефон, уличный шум	Микрофон, тихий кабинет
Время обработки	В режиме реального времени	Относительно неограниченное

Еще одной особенностью систем стенографирования является необходимость работы со сверхбольшим словарем, где возникают ошибки, связанные с появлением близких по звучанию слов, а также комбинаций слов. В результате одновременно могут появляться гипотезы с одинаково высокой вероятностью, но содержащие разные слова. В большинстве случаев такая неоднозначность снимается за счет применения стохастической модели языка и онтологии предметной области. Однако следует отметить, что формализация семантических знаний до сих пор остается камнем преткновения как для лингвистов, так и для специалистов в области искусственного интеллекта [11,12]. В данной статье и в наших исследованиях семантический уровень языка остается пока не задействованным. Данная область требует отдельного исследования и ей занимаются многочисленные специалисты по автоматической обработке текста и машинному переводу. Далее рассмотрим методику подготовки баз данных, необходимых для системы распознавания, а также сам декодер слитной русской речи, разработанный в ходе последних лет исследований.

3. Методика подготовки баз данных для системы распознавания речи

В этом разделе приведем методику подготовки баз данных и основные процедуры автоматической обработки текста, которые были разработаны в ходе исследований. Это позволит указать основные типы преобразований, которые используются в системе и осветить ряд проблемных мест, которые могут являться причинами возникновения ошибок уже на этапе распознавания речи.

Методика создания баз данных представлена на рисунке 1. Текущая версия программного комплекса использует базовый список основ и концовок, а также правила анализа и синтеза словоформ, полученные путем автоматизации печатного варианта грамматического словаря русского языка А.А.Зализняка [13]. Для более компактного представления словаря используется не окончание, а так называемая концовка, правила выделения ее границы представлены в [14]. Имеющиеся в распоряжении тексты или фразы из предметной области (ПО) служат для построения словаря и модели языка ПО. При обработке текстов собираются все встретившиеся основы. При этом анализ каждого слова производится следующим образом. От слова последовательно отрезаются все возможные концовки (содержащиеся в базовом списке), а оставшаяся часть словоформы (предполагаемая основа) ищется в списке основ. Время поиска пропорционально числу букв в основе, а не числу основ благодаря представлению списка основ в виде лексического дерева [15]. При успешном поиске

запоминается индекс найденной основы. В результате после отрезания всех возможных концовок для одного слова может быть найдено несколько вариантов основ. Если же не оказалось ни одной основы, которая бы синтезировала данное слово, то оно обрабатывается вручную. Для каждого такого слова формируется новый набор основ (при чередовании в основе будут созданы все варианты основ) и к ним приписываются соответствующие грамматические маркеры, необходимые для синтеза парадигмы [14]. Полученный список основ ПО далее используется для синтеза парадигм слов. По сути, этот список содержит все словоформы, которые могут быть распознаны в этой задаче.

Следующим этапом является транскрибирование полученного списка словоформ. За счет учета мягкости согласных и ударности гласных число вариантов основ и концовок несколько увеличивается после транскрибирования. Например, словарь Зализняка [13], который использовался в качестве базового, содержит в сумме 97473 различных основ и концовок (имена собственные были исключены из рассмотрения на данном этапе). После учета чередования в основе и преобразования в транскрипцию их число увеличилось почти в 2 раза. В таблице 3 приведены сведения о том, как увеличивалось число основ и концовок на каждом этапе преобразования. Каждая транскрибированная словоформа также содержит информацию о положении границы между основой и концовкой и индекс основы, от которой она синтезирована.

Таблица 3. Увеличение числа различных основ и концовок с учетом чередования в основе и фонетического представления.

Вид представления	Изменяемые				Неизменяемые	Всего элементов в базе данных
	Имена		Глаголы			
	Число основ	Число концовок	Число основ	Число концовок		
Базовые формы	66224	60	27799	199	3191	97473
С учетом чередования в основе	86337	60	35263	199	3191	125050
После транскрибирования	133123	118	51219	629	3191	188280

Русский язык относится к числу синтетических языков и имеет относительно высокий уровень флективности [16]. По этой причине обработка словаря, представленного в виде простого списка словоформ или их транскрипций, практически невозможна по причине низкой скорости. Наиболее популярным и эффективным способом представления словаря на сегодня считается использование префиксного лексического дерева, в котором одинаковые начальные участки слов объединяются [17,18,19]. При описании словоформ объединяются буквы, а при создании дерева транскрипций объединяются фонемы. В данной работе было впервые предложено использовать лексическое префиксное дерево с двухуровневой структурой, где первый уровень представляет собой граф основ, а второй – список концовок. В результате был разработан двухуровневый морфо-фонетический префиксный граф (ДМПГ), который наиболее компактно описывает все используемые словоформы и их транскрипции [14].

Генерация ДМПГ производится по списку транскрибированных словоформ ПО, и поэтому полученный граф способен генерировать только грамматически правильные слова. В предшествующих исследованиях, стремясь к более компактному представлению словаря, в качестве базовой единицы мы использовали морфы, а связи между ними задавались статистически [20,21]. Поэтому в ходе распознавания могли создаваться последовательности морфов, которые не являлись словами или содержали ряд грамматических ошибок. Более детально топология ДМПГ, процесс его создания и сравнение с базовыми моделями можно найти в [14].

Последним этапом подготовки баз данных для системы распознавания речи является создание модели языка по тексту предметной области, которое производится следующим образом. Каждое предложение преобразовывается в последовательность индексов основ и далее производится подсчет основ, пар основ, троек основ и т.д. (длина окна анализа зависит от типа N -граммной модели языка). Число встретившихся основ и их комбинаций сохраняется в разреженной матрице размерностью W^N , где W – число основ в словаре, а N – тип модели языка. После обработки всех текстов предметной области разреженная матрица преобразовывается в структуру, где хранятся только встретившиеся комбинации основ, а их значения нормализуются и приводятся к логарифмической шкале.



Рис. 1. Методика подготовки баз данных для системы распознавания речи.

Таким образом, по текстам из предметной области подготавливаются две основные базы данных: (1) модель языка; (2) двухуровневый морфо-фонетический префиксный граф. Кроме того, система распознавания речи также использует заранее подготовленные и независящие от предметной области следующие базы данных: (1) массив СММ моделей фонем, обученных по представительному корпусу русской речи (2) базовый список основ и концовок, а также правила синтеза словоформ.

В заключение этого раздела обсудим некоторые отрицательные моменты, связанные с применением разложения словоформы на основу и концовку. Компактность представления словаря и сокращение размера обучающего текстового материала в разработанной модели, к сожалению, может стать причиной возникновения некоторой неоднозначности в ходе анализа словоформы. Например, одна и та же словоформа может быть сгенерирована от нескольких

типов основ, которые имеют одинаковое или разное написание. Так при анализе словоформы «ясно» будет получено 3 варианта основы: (1) «ясно» – наречие с нулевой концовкой; (2) «ясно» – предикат с нулевой концовкой; (3) «ясн» – краткая форма прилагательного среднего рода с концовкой "о". Соответственно при синтезе для этой словоформы будет сгенерированы парадигмы от всех типов основ. Хотя возможно, что на самом деле в предметной области реально используется только один из типов основ.

Для обучения модели языка используется статистика частотности комбинаций основ. Использование статистики по комбинации основ, а не словоформ позволяет значительно снизить сложность модели и соответственно повысить скорость обработки. Инвариантность модели обучения к окончаниям также позволяет сократить текстовый материал, необходимый для обучения модели. Для описания вероятности пары смежных лексем теперь достаточно одного значения, а не списка вероятностей всех возможных комбинаций словоформ этих двух лексем. Недостатком такого подхода является потеря морфологической информации. В результате при обработке речевого сигнала окончания в распознанной последовательности слов могут оказаться не согласованными по числу, роду, падежу, времени и т.д. Кроме того, здесь также возможно возникновение неоднозначности по типу основы при анализе словоформы. Поэтому есть риск, что комбинации не совместимых типов основ некоторых словоформ будут иметь высокую вероятность.

Проблема неоднозначности при выборе основы и согласовании концовок словоформ может быть решена путем использования синтаксического анализатора, который позволит определить роль словоформы в предложении, ее часть речи и таким образом сократить допустимые варианты основ. На этапе распознавания в свою очередь использование синтаксической и морфологической информации в комбинации с акустической вероятностью распознанных концовок позволит подобрать ряд грамматически правильных гипотез фраз. В следующем разделе кратко рассмотрим основные уровни обработки речи в разработанной системе на базе ДМПГ декодера.

4. Система автоматического распознавания русской слитной речи

Применение строгих грамматических правил при формировании двухуровневого морфофонемного графа обеспечивает построение всех возможных словоформ и исключает возникновение грамматически некорректных комбинаций основы и окончания при декодировании речевого сигнала.

Для использования данного графа в задаче распознавания слитной речи вводится обратная связь, обеспечивающую генерацию последовательности словоформ с неограниченной длиной. Строго говоря, число слов в последовательности будет зависеть от длины записанного речевого сигнала и при поступлении последней фонемы, гипотеза распознанной фразы (путь по графу) заканчивается последним начатым словом. На рисунке 2 показана структурная схема ДМПГ, пригодная для декодирования русской слитной речи.

Полученный граф способен генерировать только грамматически правильные слова, а именно, все возможные последовательности фонем (произвольные пути по графу) образуют только существующие в русском языке словоформы или последовательности словоформ. Однако устная речь существенно отличается от письменной [5]. Для разговорной речи характерна редукция фонем и целых слов; чаще всего используется некоторый ограниченный словарь, обусловленный конкретной ситуацией; строятся лаконичные, логичные высказывания. Также в речи часто употребляются слова, недопустимые в литературном языке, такие как жаргонизмы, вульгаризмы, просторечные слова и т.д. На стиль речи также влияет число участников диалога, условия и цель общения. В целом, можно сказать, что форма и содержание устной речи выбирается по принципу коммуникативной целесообразности, стремясь минимизировать время, затрачиваемое на донесение информации до собеседника.

В результате на вход декодера поступает цепочка фонем, обозначающая слово, но содержащая ряд ошибок. Для того чтобы декодер, основанный на ДМПГ, мог обрабатывать любые последовательности фонем и формировать из них грамматически правильные слова, разработанная топология графа закладывается в основу скрытой Марковской модели, и для

всех соседних узлов вводятся вероятности переходов. Полная схема работы декодера построена с учетом хорошо зарекомендовавших себя в области распознавания речи методов: синхронизированного по времени поиска Витерби [10], метода передачи маркеров, методики отсеечения маловероятных гипотез, стохастической модели языка, учитывающей комбинации основ [22].

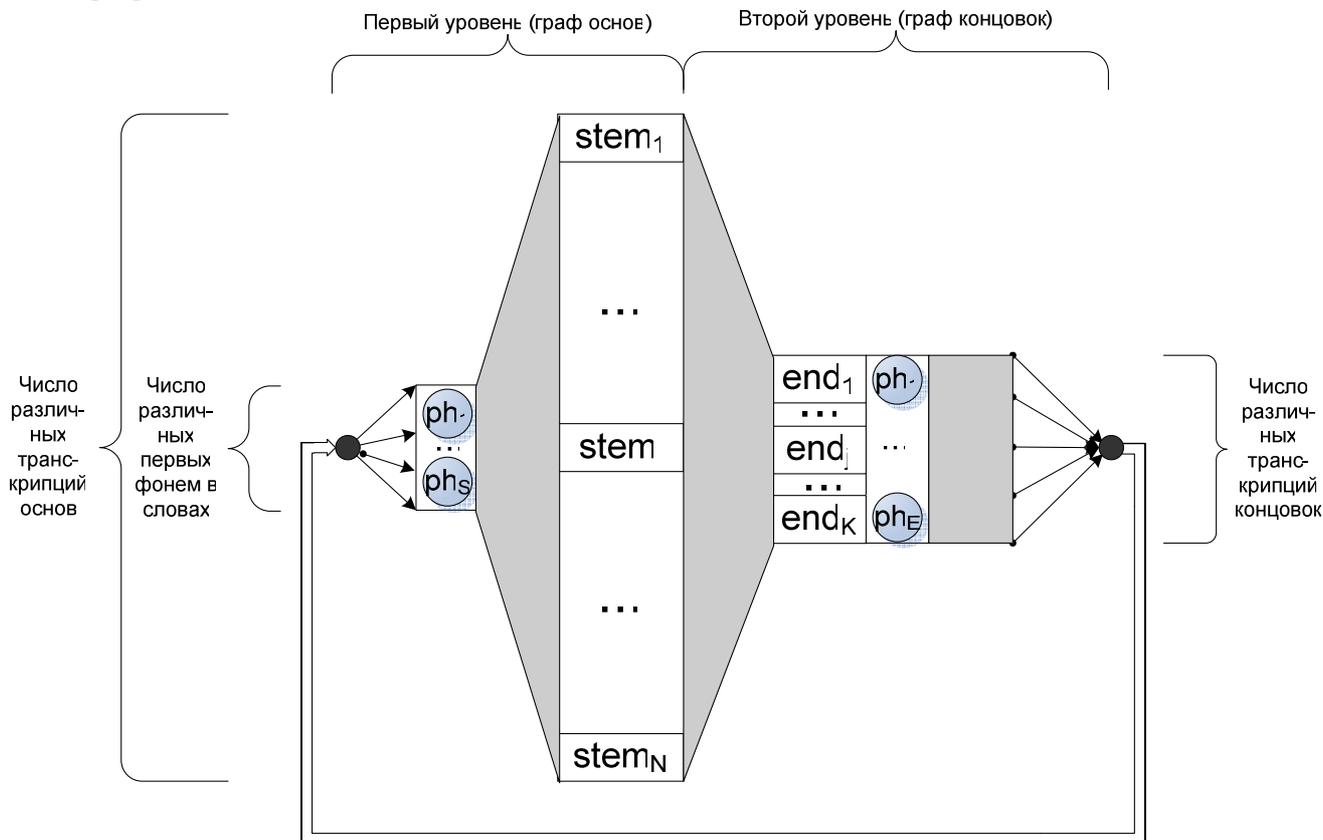


Рис. 2. ДМФГ для декодирования слитной речи.

Во время декодирования для каждого вектора наблюдений предварительно вычисляется набор вероятностей состояний для всех фонем, содержащихся в графе. Учитывая, что разные маркеры одновременно проходят по графу большое число одинаковых фонем, за счет предварительного расчета банка вероятностей всех состояний фонем достигается существенное ускорение. Гипотезы фразы содержат индексы распознанных основ S , концовок E , их акустические вероятности P_A , а также оценку по модели языка основ P_{SLM} . Этот поток данных далее поступает на вход синтаксического анализатора (Рис. 3.), где проверяется синтаксическая и грамматическая согласованность гипотезы фразы. Концовки, вызывающие сомнения, исправляются посредством синтаксического анализа и морфологического синтезатора, который генерирует все подходящие концовки для основ во фразе [14]. Если концовка была распознана неверно, то после ее замены на синтаксически правильный вариант акустическая вероятность концовки $P_A(E_i)$ будет перемножена с некоторым понижающим коэффициентом k_G . Если возможно несколько вариантов правильных гипотез фраз с разными концовками, то гипотеза фразы размножается.

При выставлении окончательной оценки гипотезы фразы учитывается: (1) акустическая вероятность последовательности основ и концовок; (2) вероятность комбинации основ с учетом модели языка; (3) соответствие гипотезы правилам синтаксиса; (4) число грамматически верных окончаний в последовательности словоформ. Последние два показателя нормализуются таким образом, чтобы их значения варьировали от нуля до единицы. Линейная комбинация всех четырех показателей используется для выбора гипотезы фразы с максимальной вероятностью, которая и является результатом распознавания.

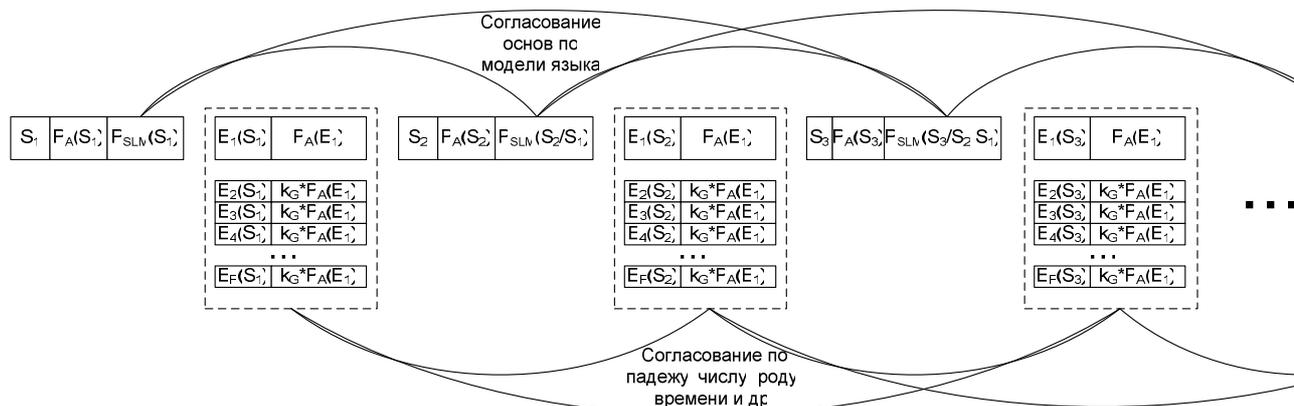


Рис. 3. Согласование гипотезы фразы на уровне синтаксической обработки

В данный момент разработанный декодер содержит в словаре свыше 2 миллионов неповторяющихся словоформ, тем не менее, экспериментальная проверка была проведена на достаточно тривиальной задаче – распознавании отдельно произнесенных фраз, тем не менее, полученные результаты по точности оказались не хуже, чем в общепринятой модели распознавания речи [10], а скорость обработки оказалась гораздо выше, особенно при получении списка лучших гипотез.

5. Заключение

Ключевыми проблемами автоматической обработки речи является ее высокая вариативность, возникающая из-за междикторских различий и изменчивости окружающей акустической обстановки. Для учета вариативности и обучения моделей фонем и слов требуются гигантские текстовые и речевые материалы, подготовка которых требует скрупулезной экспертной работы. Поэтому современные системы распознавания, обученные на ограниченных речевых корпусах, записанных в лабораторных условиях, пока еще не показывают качества, удовлетворительного для массовой эксплуатации. В ходе исследований были предложены новые подходы к автоматической обработке русской речи и языка.

Разработка компактного способа представления словаря особенно актуальна для флективных языков с относительно богатой морфологией. Декомпозиция словоформы на основу и окончание по грамматическим правилам позволяет хранить словарь в виде префиксного дерева основ и автоматически генерировать произвольную словоформу. При создании модели языка в обучающем тексте выделяются только основы, что позволяет учесть все возможные комбинации словоформ, даже при отсутствии таковых при обучении. В ходе исследований предложен декодер слитной русской речи на основе двухуровневого морфофонемного префиксного графа. Процедура его построения сводится к транскрибированию всех словоформ словаря и последующему объединению начальных участков основ и окончаний в двухуровневый ориентированный граф. Прохождение по графу обеспечивает генерацию только грамматически верных словоформ, а использование аппарата скрытых Марковских моделей позволяет оценить и выбрать наиболее вероятные гипотезы слов по входной последовательности фонем.

Последующая работа будет посвящена исследованию особенностей спонтанной речи и их учета в процессе подготовки баз данных и декодирования речевого сигнала. Изучение ассимилятивных процессов, происходящих как внутри слова, так и на стыке слов и построение альтернативных транскрипций, позволит учесть вариативность произношения разговорной речи. Кроме того, разрабатываются специальные модели мусора, обученные на распознавание артефактов речи, заполненных пауз и других элементов, характерных для спонтанной речи и серьезно ухудшающих качество работы системы. Своевременная локализация таких помех и удаление их из последующей обработки позволит существенно повысить производительность работы системы.

Данное исследование проводится при поддержке гранта Президента Российской Федерации № МК-9351.2006.9 и гранта РФФИ № 07-07-00073-а.

6. Литература

1. Косарев Ю.А. Естественная форма диалога с ЭВМ. Л.:Машиностроение, 1989, 143 с.
2. Ронжин А.Л., Карпов А.А., Ли И.В. Речевой и многомодальный интерфейсы. - М.: Наука, 2006 - (Информатика: неограниченные возможности и возможные ограничения), 173 с.
3. Ронжин А.Л., Леонтьева А.Б. Применение техники «Гудвин» для моделирования человеко-машинного взаимодействия // Известия вузов. Приборостроение. 2006. Т. 49, № 11. С. 70-75.
4. Foote, J. T., Jones, G. J. F., Sparck Jones, K., & Young, S. J. Talker-independent keyword spotting for information retrieval. In Proc. Eurospeech 95, volume 3, 1995, pp. 2145-2148.
5. Гойхман О.Я., Надеина Т.М. Речевая коммуникация // Инфра-М, 2006, 272 с.
6. Сорокин В.Н., Цыплихин А.И. Сегментация и распознавание гласных. Информационные процессы, Том 4 , № 2, С. 202-220.
7. Akira Kurematsu, Mariko Nakano-Miyatake, Hector Perez-Meana, Eric Simancas-Acevedo Performance analysis of Gaussian Mixture Model speaker recognition system with different speaker features. Electronic Journal «Technical Acoustics» 2005, vol. 14.
8. Mohamed Benzeghiba, Renato De Mori, Olivier Deroo, Stephane Dupont, Denis Jouvet, Luciano Fissore, Pietro Laface, Alfred Mertins, Christophe Ris, Richard Rose, Vivek Tyagi, Christian Wellekens. Impact of variabilities on speech recognition. In Proc. of 11-th International Conference SPECOM'2006, St. Petersburg: "Anatoliya", 2006, pp. 493-498.
9. Александрова О.А., Иваницкий В.В. Пауза колебания – комплексный феномен современной коммуникации // Вестник Новгородского государственного университета. Т. 25, С. 95- 101.
10. S.Young, G.Evermann, M.Gales, T.Hain, D.Kershaw, G.Moore, J.O.D.Ollason, D.Povey, V.Valtchev, P.Woodland, "The HTK book: version 3.3. Technical Report", Cambridge University, UK. <http://htk.eng.cam.ac.uk>, 2005.
11. Пинкер С. Язык как инстинкт. Пер. с англ. / Общ. Ред. В.Д. Мазо. – М.: УРСС, 2004. – 456 с.
12. Попов. Э.В. Общение с ЭВМ на естественном языке. – М.:УРСС, 2004 – 360 с.
13. Зализняк А.А. Грамматический словарь русского языка. М.: Русские словари, 2003. 800 с.
14. Ронжин А.Л., Леонтьева А.Б., Кагиров И.А., Леонтьева Ал.Б. Двухуровневый морфофонемный префиксный граф для декодирования русской слитной речи. Труды СПИИРАН. Вып. 4, т.1, Санкт-Петербург, 2007.
15. Ortmanns, S., Eiden, A., Ney, H. "Improved Lexical Tree Search for Large Vocabulary Recognition". IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Seattle, WA: 817-820, 1998.
16. Гринберг Дж. Квантитативный подход к морфологической типологии языков // НЛ. Вып. III. М., 1963.
17. Gelbukh A, Sidorov G. Approach to construction of automatic morphological analysis systems for inflective languages with little effort // Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science 2588: Springer-Verlag, 2003. P. 215–220.
18. Сокирко А.В. Морфологические модули на сайте www.aot.ru // Труды межд. конф. Диалог-2004. М., 2004.
19. <http://www.rco.ru/>
20. Карпов А.А., Ронжин А.Л., Ли И.В. SIRIUS - система дикторнезависимого распознавания слитной русской речи // Известия ТРТУ, № 10, 2005, С. 44-53.
21. A.L. Ronzhin, A.A. Karpov, "Russian Voice Interface". Pattern Recognition and Image Analysis, 2007, Vol. 17, No. 2, pp. 321–336.
22. Pražák A., Psutka J., Hoidekr J., Kanis J., Müller L., Psutka, J. Adaptive Language Model in Automatic Online Subtitling // Proc. 2nd IASTED International Conference on Computational Intelligence CI 2006. San Francisco, California, USA, 2006. P. 479–483.

Н.А. Коротаев, А.А. Кибрик, В.И. Подлеская

Полипредикативность и интеграция в устном нарративном дискурсе*

*Российский государственный гуманитарный университет,
Московский государственный университет,
г. Москва, Россия,
n_korotaev@hotmail.com*

1. Введение

Всякая попытка изучения устного дискурса неизбежно сопряжена с достаточно очевидной проблемой: речь, зафиксированная только в форме звука, слишком многоаспектна, чтобы человеческий взгляд и сознание могли одновременно удержать какой-то ее значимый фрагмент. Поэтому едва ли не первой задачей при анализе устной речи становится создание дискурсивной транскрипции – принципов переложения звукового потока в письменную форму. Эта, казалось бы, техническая задача в действительности ставит перед исследователем целый ряд нетривиальных теоретических вопросов. Какие компоненты звука должны быть отражены в транскрипции, а какими можно пренебречь? Каков инвентарь единиц, которые необходимо отражать в транскрипции? Как перейти от континуальности звука к дискретной транскрипции? Разработка систематической и последовательной системы записи устного дискурса требует введения десятков и даже сотен различных правил, которые должны отражать содержательные свойства анализируемых явлений.

В настоящей статье мы рассмотрим лишь некоторые аспекты этой работы, которую мы проводим на материале корпуса устных рассказов детей и подростков о сновидениях.

2. «Рассказы о сновидениях»

Предлагаемая нами система транскрипции устного русского дискурса разработана на материале корпуса аудиозаписей «Рассказов о сновидениях» (название призвано установить преемственность со знаменитыми «Рассказами о грушах» У. Чейфа, см. Chafe 1980, во многом определившими направление корпусных дискурсивных исследований). Запись рассказов была произведена сотрудницей (ныне – доцентом) кафедры нервных болезней Московской медицинской академии им. И.М.Сеченова Е.А.Корабельниковой (см. Корабельникова 1997). Корпус включает 129 рассказов детей и подростков в возрасте от 7 до 17 лет, при этом 69 из них взяты от детей с невротическими расстройствами, 60 – от здоровых детей из контрольной группы. Суммарная продолжительность всех аудиозаписей составляет около двух часов, а общее число словоупотреблений – примерно 15 тысяч единиц.

Несмотря на довольно небольшой объем корпуса, мы полагаем, что с количественной точки зрения он все же является репрезентативным по отношению к исследуемому типу устного русского дискурса – личным рассказам. Подробнее о корпусе см. работу Kibrik et al. 2002.

3. Элементарная дискурсивная единица (ЭДЕ) как центральное понятие дискурсивной транскрипции

Никакая дискурсивная транскрипция заведомо не может отразить всю информацию, содержащуюся в потоке устной речи. Поэтому при разработке принципов письменной фиксации устного дискурса необходимо заранее знать, какие исследовательские задачи предстоит решать на базе создаваемой транскрипции. Наша транскрипционная система призвана в первую очередь отражать устройство дискурса – его локальную и (в меньшей степени) глобальную структуру. Играет роль не только линейный, но и иерархический аспект структуры; последний мы понимаем в духе теории риторической структуры (TRC) У. Манна и С. Томпсон (см. Mann, Thompson 1988; Mann, Matthiessen, Thompson 1992), согласно которой

* Работа выполнена при поддержке гранта РФФИ 06-06-80470.

дискурс можно представить в виде иерархически организованной сети единиц, связанных так называемыми риторическими отношениями¹.

Ключевая задача при таком подходе к описанию дискурса – выделение минимальных узлов этой иерархической структуры. Подобные узлы, соответствующие минимальным квантам в порождении дискурса, мы называем **элементарными дискурсивными единицами (ЭДЕ)**². Каждая ЭДЕ записывается в транскрипции отдельной строкой, на конце которой ставится пунктуационный знак, указывающий на то, как данная ЭДЕ вписывается в локальную структуру рассказа. Подчеркнем, что правила расстановки пунктуационных знаков существенно отличаются от правил стандартной русской пунктуации и не выводимы из них. Кроме того, в транскриптах отмечаются заполненные и незаполненные паузы, движения тона на акцентированных словоформах и некоторая другая информация.

Более подробное описание используемой нами системы дискурсивной транскрипции можно найти в работе Кибрик, Подлесская 2003. В настоящей статье мы остановимся исключительно на вопросах выделения ЭДЕ.

4. Канонические ЭДЕ. Малые ЭДЕ

Сущность ЭДЕ как минимальных шагов в продвижении дискурса проявляется в различных аспектах процесса речепорождения: физиологическом (ЭДЕ произносится на одном выдохе), когнитивном (ЭДЕ вербализует один «фокус сознания», см. Chafe 1994), семантическом (ЭДЕ описывает одно событие или состояние), синтаксическом (ЭДЕ является простой клаузой) и просодическом (ЭДЕ организована как один произносительный контур с точки зрения движения тона, наличия основного акцентного центра, темпа и громкости – подробнее см. Кибрик, Подлесская 2006). В нашей работе мы сосредоточиваемся на координации просодического и семантико-синтаксического аспектов и говорим соответственно о просодическом и семантико-синтаксическом **критериях** выделения ЭДЕ.

В прототипическом случае ЭДЕ отвечает обоим этим критериям, то есть заключается в один интонационный контур и представляет собой простую клаузу, описывающую одну ситуацию. Таковы, например, все 9 ЭДЕ в приведенном ниже фрагменте:

- (1) Z14³
2. был \сон значит,
3. ..(0.2) ээ(0.4) ..(0.4) {ЧМОКАНЬЕ} ..(0.2) корочев ..(0.1)
на /бал меня пригласили,
4. ...(0.5) а у меня нет –\пла-атъя.
5. ...(0.8) Значит ...(0.6) Вдумаю:
6. «–/Ба-ал ну всё-таки,
7. \как я \пойду?»
8. ..(0.4) \Ну-у /одела там /юбку короткую,
9. /пришла,
10. а они там все в таких вот /юбках там,

Такие ЭДЕ, совпадающие по объему с простой клаузой, мы называем **каноническими**. Согласно нашим данным (см. Кибрик, Подлесская 2006), доля канонических ЭДЕ в «Рассказах

¹ Отметим, однако, что применение ТРС к устному дискурсу потребовало определенной адаптации первоначальной модели, – см. статью Литвиненко 2001.

² В литературе об устном дискурсе чаще используются другие термины – фраза, интонационная фраза, интонационная группа, интонационная единица (ИЕ) (см. Chafe 1994: 57), синтагма. Мы предпочитаем термин «элементарная дискурсивная единица», поскольку он наиболее точно отражает функциональную природу данных единиц, а также позволяет единым образом говорить об устном и письменном дискурсе.

³ Буква N перед номером обозначает принадлежность рассказов подкорпусу, взятому у детей с невротическими заболеваниями; буква Z означает, что рассказ взят у здорового ребенка из контрольной группы. В левой колонке указывается номер строки в рассказе.

о сновидениях» составляет около 70%. Однако существуют и отклонения от указанного выше стандарта.

Неканонические ЭДЕ делятся на **малые** и **большие**. Малые ЭДЕ по своему объему меньше клаузы, а большие – превосходят ее. Типология малых ЭДЕ включает в себя **сопутствующие, разрывные и регуляторные** ЭДЕ. Сопутствующие ЭДЕ возникают в том случае, когда говорящий оформляет клаузу не за один шаг, а за два или более. Приведем пример сопутствующей малой ЭДЕ, следующей за клаузой:

(2) Z19

12. И /вдруг я увидела какую-то ..(0.4) \к-коробку.

13. ..(0.3) С /бантиком \сверху.

В (2) рассказчица сначала произнесла полноценную клаузу, но затем добавила к ней новый компонент, как бы «сделав вид», что клауза еще не была завершена, а завершается только теперь. При этом, однако, интонационное оформление этих двух речевых отрезков не позволяет считать, что перед нами одна ЭДЕ, а не две.

Подробнее о типах малых ЭДЕ можно прочитать в статье Кибрик, Подлеская 2006. Общим для всех малых ЭДЕ является то, что, не являясь клаузами, они вместе с тем не могут быть включены в состав ни одной другой, канонической, ЭДЕ. Можно сказать, что при дискурсивном анализе подобные речевые фрагменты выделяются в ЭДЕ по «остаточному принципу».

5. Большие ЭДЕ

Природа возникновения больших ЭДЕ – совершенно иная. Если малые ЭДЕ являются проявлением свойственной устному дискурсу фрагментации, то большие ЭДЕ – это продукт противоположного явления, феномена интеграции (о понятиях фрагментации и интеграции в дискурсе см. Chafe 1982). Чаще всего образование больших ЭДЕ связано с утратой некоторым предикатом существенной части предикативных свойств.

Рассмотрим несколько примеров.

В примере (3) последняя ЭДЕ представляет собой **сериализованную** конструкцию:

(3) N40

25. \Хорошо чувствовать

26. то что ты сидел ..(0.3) с \папой как будто_

27. с \мамой_

28. с \Димкой_

29.(1.8) сидишь \телевизор смотришь_

30. ..(0.1) а /оказывается,

31. это всё \неправда.

Несмотря на наличие двух глаголов – сидишь и смотришь, – данная конструкция, говоря неформальным языком, описывает одну ситуацию, не отличаясь в этом отношении от простых клауз с одним предикатом. При этом первый по порядку глагол частично делексикализуется и фактически превращается в маркер способа действия. Важный лексико-грамматический сигнал слитности этой конструкции – отсутствие сочинительного союза между предикатами; основной просодический симптом – наличие лишь одного акцента, расположенного на зависимом второго по порядку глагола.

В примере (4) появление большой ЭДЕ связано с **грамматикализацией матричного предиката** говорят:

(4) Z14

15. «Это говорят новая /мо-ода там.»,

16. в восемнадцатом /веке новая такая –\мо-ода пошла.

В случаях, подобных (4), налицо семантическая, просодическая и когнитивная редукция цитационных предикатов: они теряют ряд фонем, произносятся без акцента, как клитики, и понижаются в статусе до уровня цитационных маркеров – таких как мол, типа и др.

Еще один тип больших ЭДЕ – **бессоюзные инфинитивные конструкции**. С семантико-синтаксической точки зрения они близки к моноклаузальным конструкциям, а тот факт, что такие ЭДЕ произносятся в рамках единого интонационного контура, обеспечивает их просодическую целостность – см. пример (5):

(5) Z28

2.(2.6) /потом /мы вышли \отдохнуть,

В концентрированном виде рассмотренные выше явления иллюстрирует пример (6), в котором на две ЭДЕ приходится пять глаголов. В первой строке имеется комбинация финитного глагола с инфинитивом, а во второй есть цитационный маркер и случай сериализации:

(6) N67

21. ..(0.3) /подойти <по-моему> <\меня | \мне> она /попросила,

22. говорит ..(0.3) «/Пойди \помоги ей!».

При всей ослабленности одного из глаголов в приведенных выше примерах, большие ЭДЕ все же во многом близки к стандартным полипредикативным конструкциям (а зачастую – и выводимы из них). Можно сказать, что они находятся на стыке моно- и полипредикативности. Нередко граница между «обычными» полипредикативными конструкциями и подобными частично грамматикализованными комплексами оказывается размытой: наряду с достаточно очевидными случаями типа (2)-(6) для каждого явления, ведущего к образованию больших ЭДЕ, находится и целый ряд неочевидных, промежуточных. Задача транскрибирования дискурса, тем не менее, требует принятия четкого, бинарного решения – «одна строка» или «две строки». Поэтому в арсенале исследователя должен иметься набор критериев, с помощью которых можно отличить стандартную полипредикацию от слитных комплексов из нескольких предикатов и провести границу в определенной точке континуума переходных случаев.

Рассмотрим критерии такого рода на примере одного из классов интересующих нас конструкций – конструкций с эпистемическими предикатами.

6. Эпистемические конструкции и большие ЭДЕ⁴

Под **эпистемическими** в настоящем исследовании понимаются конструкции, выражающие довольно широкий спектр языковых значений, среди которых можно выделить три основных семантических центра. Это значения эпистемической гипотезы и эпистемического ожидания, а также важный для анализируемого корпуса подкласс эвиденциальных значений, связанный с указанием на принадлежность рассказываемого к миру сна⁵.

Носителем указанных выше значений в каждой отдельной конструкции является **эпистемическое выражение** (ЭВ). Эпистемические выражения различаются по своему синтаксическому и дискурсивному статусу. С одной стороны, это могут быть заведомо непредикативные единицы: вводные слова (по-моему), частицы (как будто) и проч. Такие слова, разумеется, не образуют отдельных ЭДЕ, а примыкают к тем клаузам, значения которых модифицируют.

С другой стороны, роль эпистемических выражений могут играть предикаты, формирующие полноценные клаузы, то есть канонические ЭДЕ. С синтаксической точки зрения они подчиняют клаузы, в которых выражается объект эпистемической оценки, см. пример (7):

⁴ Более подробный анализ эпистемических конструкций в «Рассказах о сновидениях» содержится в работе Коротаяв 2007.

⁵ Определения используемых терминов см. в книге Плулунян 2000: 309-325; там же рассматриваются пути сближения эвиденциальных значений со значениями эпистемической и ирреальной модальностей.

(7) N28

3. ..(0.3) мне /кажется,

4. что –кто-то-о ко мне \подбира-ается.

Наконец, в определенных контекстах эпистемические предикаты утрачивают часть своих предикативных свойств и грамматикализуются в маркеры соответствующих отношений. Этот процесс и приводит к появлению одного из видов больших ЭДЕ. Так, в примере (8) глагол кажется уже не образует отдельной клаузы, а служит эпистемическим маркером, выражающим неполную уверенность говорящего в том, что у него появилась новая машина:

(8) Z36

20. ..(0.3) потом я' ..(0.2) \’это ... (0.6) стал' || ... (0.9) у меня

уже' ..(0.1) другая кажется /машина стала,

Каковы же формальные корреляты содержательного различия между примерами типов (7) и (8)? В каких случаях предикатные эпистемические выражения образуют отдельные ЭДЕ и где проходит граница между полноценным предикатом и грамматикализированным эпистемическим маркером?

Для ответа на эти вопросы мы предлагаем использовать два набора критериев, применение которых позволяет определить статус эпистемических выражений и последовательно провести разбиение эпистемических конструкций на ЭДЕ.

Основные критерии

Факторы сохранения предикативности эпистемического выражения:

1) Наличие в эпистемической конструкции подчинительного союза или же наличие у ЭВ зависимых слов, за исключением подлежащего, выраженного личным местоимением (**синтаксический критерий**).

2) Глагольная форма прошедшего или будущего времени (**морфологический критерий**).

Если **оба** приведенных выше условия выполнены, эпистемическое выражение признается несомненным предикатом и выносится в отдельную ЭДЕ, см. пример (9):

(9) N66

14. ... (1.2) {ЦОКАНЬЕ} ..(0.2) эээ(0.5) ..(0.1) мне показалось

15. что это ..(0.2) почему-то был \мужчина.

Факторы утраты предикативности эпистемического выражения:

1) Отсутствие в эпистемической конструкции подчинительного союза, а также отсутствие у ЭВ собственных зависимых (**синтаксический критерий**).

2) Невозможность образования формы прошедшего или будущего времени (**морфологический критерий**).

При выполнении **обоих** этих условий эпистемическое выражение признается очевидно непредикативным и, соответственно, не выделяется в отдельную ЭДЕ, см. последнюю строку примера (10):

(10) N26

95.(1.3) Говорю «А Лоткуда ты знаешь,

96. что мне её решила \мама?

97. Может я её \–сама-а сделала?»

Как нетрудно заметить, основные критерии охватывают только самые явные случаи сохранения/утраты предикативности. Для менее очевидных примеров наряду с основными критериями предлагается использовать дополнительные. Итоговое решение о делении на ЭДЕ принимается по простому большинству факторов в пользу утраты или сохранения предикативности. При равном числе противодействующих факторов признается, что предикативность эпистемического выражения сохранена.

Дополнительные критерии

1) **Позиционный критерий.** Расположение ЭВ перед клаузой, содержащей объект эпистемической оценки, говорит в пользу сохранения предикативного статуса ЭВ. Расположение ЭВ после клаузы, содержащей объект оценки, или внутри нее, говорит в пользу утраты предикативного статуса ЭВ.

Именно позиционный критерий – наряду с синтаксическим – оказывается решающим при анализе приведенных выше примеров (7) и (8). На предикативный статус ЭВ кажется в (7) указывает не только наличие в конструкции союза что, но и препозиция ЭВ. В (8), напротив, грамматикализация кажется подчеркивается как отсутствием маркеров синтаксической связи, так и положением ЭВ внутри другой клаузы.

2) **Критерий ненулевой глагольной формы.** Отсутствие эксплицитной формы глагола свидетельствует о принадлежности ЭВ к маркерам, а не к предикатам. Этот критерий позволяет выявить меньшую предикативность ЭВ типа видно, похоже по сравнению с было похоже или кажется.

3) **Критерий предикативности языкового выражения, содержащего объект оценки.** Еще один фактор, говорящий против выделения эпистемического выражения в отдельную ЭДЕ, – непредикативный характер языкового выражения, содержащего объект эпистемической оценки.

Так, в примере (11) кажется объединяется в одну ЭДЕ с на даче именно вследствие непредикативного характера последнего выражения, тогда как значения синтаксического и позиционного критериев нейтрализуют друг друга:

(11) Z55

1. Значит мне /снится,

2. что я где-то на –приро-оде вот_

3. ..(0.6) кажется на /даче,

7. Заключение

Анализ устной речи требует выработки последовательных и систематических принципов дискурсивной транскрипции. Центральная задача транскрипции, ориентированной на отражение локальной структуры текста, – выделение элементарных дискурсивных единиц (ЭДЕ). Прототипическая ЭДЕ представляет собой простую клаузу. Однако в определенных обстоятельствах могут встречаться ЭДЕ как меньшего, так и большего объема. Один из видов ЭДЕ, превышающих по объему клаузу, возникает при грамматикализации эпистемических предикатов. Для опознания случаев грамматикализации предлагается использовать два набора критериев, учитывающих синтаксические, морфологические и позиционные свойства эпистемических конструкций.

Литература

1. Кибрик А.А., Подлеская В.И. 2003. К созданию корпусов устной русской речи: принципы транскрибирования // *Научно-техническая информация (серия 2)*, 10, 5-12.
2. Кибрик А.А., Подлеская В.И. 2006. Проблема сегментации устного дискурса и когнитивная система говорящего // Соловьев В.Д. (ред.) *Когнитивные исследования*. Вып. 1. М.: Институт психологии РАН. – 138-158.
3. Корабельникова 1997. Сновидения при невротических расстройствах у детей и подростков. Диссертация на соискание ученой степени кандидата медицинских наук. Москва: ММА им. Сеченова.
4. Коротаев Н.А. 2007. Сегментация полипредикативных конструкций в корпусах устных текстов: конструкции с эпистемическими предикатами в аудиокорпусе «Рассказов о сновидениях» // *Научно-техническая информация (серия 2)*, 2, 30-36.
5. Литвиненко А.О. 2001. Описание структуры дискурса в рамках теории риторической структуры: применение на русском материале // *Труды Международного семинара*

- «Диалог-2001» по компьютерной лингвистике и ее приложениям. Т. 1. Аксаково. – 159-168.
6. Плунгян В.А. 2000. *Общая морфология*. Введение в проблематику. М.: Эдиториал УРСС.
 7. Chafe W. (ed.) 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
 8. Chafe, W. 1982. Integration and involvement in speaking, writing, and oral literature // D. Tannen(ed.) *Spoken and written language: Exploring orality and literacy*. Norwood: Ablex, 35-54.
 9. Chafe, Wallace. 1994. *Discourse, consciousness, and time*. Chicago: University of Chicago Press.
 10. Kibrik A.A., Podlesskaya V.I., Kal'kova T.M., Litvinenko A.O. 2002. Cognitive structure of narrative discourse: the analysis of children's night dream stories // Нариньяни А.С. (ред.) *Компьютерная лингвистика и интеллектуальные технологии. Труды международного семинара «Диалог-2002»*. Т. 1: Теоретические проблемы. М.: Наука. – 635-647.
 11. Mann W.C., Thompson S.A. 1988. Rhetorical structure theory: toward a functional theory of text organization. *Text*, 8, 243-281.
 12. Mann W.C., Matthiessen C., Thompson S.A. 1992. "Rhetorical structure theory and text analysis". In: Mann W.C., Thompson S.A. (eds.) *Discourse description: Diverse linguistic analyses of a fund-raising text*. Amsterdam: Benjamins, pp. 39-78.

А.В. Венцов, Е.И. Риехакайнен, Н.А. Слепокурова

Ментальный лексикон и восприятие редуцированных форм

*Санкт-Петербургский государственный университет,
г. Санкт-Петербург, Россия,
av.ventsov@mail.ru, nataliars@inbox.ru,*

В настоящее время большинство исследователей признают, что важной (или важнейшей) частью системы восприятия речи человеком является ментальный лексикон. И хотя о структуре этого словаря и образующих его элементах мало что известно, множество работ посвящено проблеме поиска способов вхождения в словарь и возможностей его использования при описании процессов восприятия речи.

Говоря далее о ментальном лексиконе, мы будем иметь в виду только ту его гипотетическую составляющую, которая предположительно используется на ранних этапах обработки речевого сигнала: при идентификации слов и, возможно, при сегментации через идентификацию. Обе эти операции достаточно успешно реализуются в так называемой «теории когорты» (Венцов, Касевич, 2003) при условии, что в словаре имеется область, где все элементы заданы в терминах (символах) первичного (фонетико-психоакустического) описания исходного речевого сигнала. Назовем эту область ментального лексикона перцептивным словарем.

Показано, что, используя механизмы «когорты», удается успешно справиться с задачей идентификации изолированно предъявляемых слов и с сегментацией непрерывного речевого потока через идентификацию (Kassevich, Ventsov, Yagounova, 2000). Отметим, что указанные процедуры успешно работают лишь при условии полного совпадения способов описания входного речевого сигнала и элементов, образующих перцептивный словарь. При этом в качестве последних выступают словоформы.

Эти исследования были выполнены на искусственном «речевом» сигнале, представленном полным орфографическим описанием, и предпринимались, главным образом, для проверки эффективности подобных процедур в качестве элементов предполагаемой модели восприятия речи. Естественный же речевой сигнал характеризуется высокой степенью вариативности в диапазоне от кодифицированной литературной речи до спонтанной. Последней свойственны различные степени редукиции сегментного состава, вплоть до утраты значительной части отдельных словоформ (Русская разговорная речь, 1983).

Однако повседневный опыт любого носителя языка убедительно свидетельствует о том, что даже в условиях предельной редукиции, свойственной спонтанной речи, процесс естественного речевого общения не становится невозможным. Из этого можно заключить, что либо в перцептивном словаре каждого носителя языка содержатся все редуцированные варианты соответствующих словоформ, либо система восприятия речи использует некие процедуры, обеспечивающие надежный поиск в словаре по сохраняющейся части сегментной информации даже в спонтанной речи.

На материале русского языка был проведен эксперимент по восприятию частотных словоформ вне контекста (Бондарко и др., 1988). Целью этого эксперимента было определение возможности правильного распознавания редуцированных частотных словоформ, а также выделение тех лингвистических признаков, сохранность которых важна для правильного восприятия редуцированной словоформы в этих условиях. Испытуемым предъявлялись для прослушивания изолированные редуцированные словоформы, выделенные из записей спонтанных диалогов. Им сообщалось, что это наиболее часто употребляемые в современной речи словоформы, и предлагалось в случае опознания записать их в орфографии. Допускалась запись в транскрипции или отказ. Каждый стимул предъявлялся 10 испытуемым. Правильным ответом считалось только полное соответствие ответа нормативному варианту предъявленного стимула. Опознание словоформы считалось состоявшимся при наличии 8 и более правильных ответов из 10. Число правильно опознанных словоформ составило 54% от общей суммы

ответов. В качестве объяснения хорошей опознаваемости некоторых форм слов было высказано предположение о наличии в сознании говорящих (в перцептивном словаре) редуцированных форм слов как возможных вариантов реализации соответствующих полных форм.

Здесь надо заметить, что 10 процентов использованных в этом эксперименте стимулов соответствовали нормативному произнесению, а большая часть правильно опознанных сигналов отличалась, главным образом, качеством реализованных сегментов при сохранении их нормативного количества.

На голландском материале был проведен эксперимент, основной целью которого также было определение того, как влияет степень редукции и наличие/отсутствие контекста на восприятие редуцированных словоформ (Ernestus, Baayen, Schreuder, 2002). На основе полученных данных голландские исследователи сделали вывод, что в ментальном лексиконе не могут храниться все возможные фонетические варианты каждой словоформы, поскольку редуцированные словоформы плохо распознаются при изолированном предъявлении. По их мнению, в перцептивном словаре хранятся некие канонические формы, из которых в процессе восприятия речи выбираются варианты, сопоставимые с редуцированными словоформами.

Первоначально нашей задачей было повторение эксперимента голландских исследователей на материале русского языка. Однако в дальнейшем оказалось, что принятая нами методика работы со спонтанной речью позволяет подобраться к исследованию возможных принципов поиска в ментальном лексиконе и процедур обработки речевого сигнала в системе восприятия речи. Некоторые результаты этих экспериментов излагаются ниже.

Эксперимент I

Задачей данного эксперимента было непосредственное сравнение особенностей восприятия редуцированных словоформ, предъявляемых изолированно и в контексте. В качестве материала использовались фрагменты нескольких спонтанных диалогов, в каждом из которых участвовало два человека. Общая продолжительность записи около 16 минут.

Тестовая последовательность состояла из двух частей. В первой части испытуемым предъявлялись расположенные в произвольном порядке изолированные словоформы, выделенные из диалогов, а во второй – они же в коротком контексте. Размер контекста выбирался так, чтобы, по нашему мнению, редуцированная словоформа могла быть однозначно распознана. Большинство контекстов представляли собой синтагмы. Порядок предъявления словоформ в первой и второй частях не совпадал. Каждый сигнал в стимуле повторялся три раза: в первой части с интервалом в 0,5 секунды, а во второй – в 1 секунду. Каждый стимул был представлен в наборе единственной реализацией. Межстимульный интервал составил 3,5 и 10 секунд для первой и второй частей, соответственно, что давало испытуемым возможность спокойно фиксировать ответы в протоколах.

Всего в эксперименте было использовано 48 стимулов (по 24 в каждой части), общая продолжительность звучания составила 9 минут 20 секунд. Стимулы, записанные на магнитную ленту, предъявлялись испытуемым через громкоговорители.

Участникам эксперимента предлагалось прослушать предъявляемые стимулы и записать то, что они услышали, буквами русского алфавита. Разрешалось пропускать сигналы (ставить прочерк в анкете), если испытуемый не успевал или не мог идентифицировать стимул. Испытуемыми были студенты и преподаватели Филологического и Восточного факультетов Санкт-Петербургского государственного университета: всего 82 человека. Каждый испытуемый слушал экспериментальный набор только один раз.

В целом эксперимент показал, что вне контекста редуцированные словоформы не могут быть правильно опознаны носителями языка: относительная частота правильного опознавания словоформ не превышала 12%. При этом из 24 предъявленных словоформ 12 не были опознаны ни один раз, для 8 частота правильных ответов не превышала 5%, для трех – не превышала 10% и только одна была правильно опознана 12-ю процентами испытуемых.

В контексте все словоформы были правильно опознаны не менее чем тремя испытуемыми. При этом для 14 словоформ частота правильных ответов превышала 50% (из них 5 словоформ

при изолированном предъявлении не были правильно опознаны ни одним испытуемым).

Такой результат дает основание предположить, что перцептивный словарь действительно не содержит редуцированных словоформ и что существует какой-то механизм выбора из словаря подходящей «кодифицированной» словоформы. Можно предположить следующую последовательность осуществляемых при этом операций. По сохранившемуся сегментному составу редуцированной словоформы из перцептивного словаря выбираются все подходящие словоформы (вместе и указанием их частотности). Затем из полученного списка отбирается один вариант, соответствующий контексту. Перебор начинается с наиболее частотных словоформ. Очевидно, что для конкретной реализации редуцированной словоформы могут быть выбраны различающиеся «кодифицированные» аналоги в зависимости от контекста.

Эксперимент II

Цель данного эксперимента – подтвердить гипотезу об определяющей роли контекста при восприятии редуцированных словоформ и убедиться, что изменение контекста действительно приводит к изменению интерпретации редуцированной словоформы.

В качестве материала использовались стимулы из предыдущего эксперимента, а также записи диалогов, служившие исходным материалом. При составлении тестового набора сигналов были применены следующие операции: сокращение контекста, расширение контекста и изменение контекста.

Процедуре сокращения контекста были подвергнуты стимулы, в которых редуцированные словоформы были хорошо распознаны в предыдущем эксперименте. Из контекста удалялись компоненты, которые представлялись необходимым материалом для определения не только семантики, но и грамматических характеристик редуцированной словоформы и сводили таким образом к минимуму число возможных интерпретаций редуцированной словоформы. Например, стимул «а он гэ[овори]т: – Да, но эт[о] не в первы[й] раз!» трансформировался в «гэ[овори]т не в первы[й] раз».

Контекст увеличивался в стимулах, где в предыдущем эксперименте исследуемые словоформы не имели левого контекста и воспринимались плохо. К примеру, «[бук]пугвальн[о] за час» трансформировалось в «это был[о] [бук]пугвальн[о] за час». В обоих приведенных примерах в квадратные скобки заключены отсутствующие в звучащем тексте сегменты.

Наибольший интерес представляли стимулы, в которых редуцированные словоформы помещались в измененный контекст. Необходимо было так сконструировать его для каждой редуцированной словоформы, чтобы получить в ответах испытуемых ее интерпретацию, отличающуюся от оригинальной, но допустимую с точки зрения контекста и сохранившейся сегментной информации редуцированной словоформы. Так, из стимула «на сл[едую]щ[ий] день» был создан стимул «на сл[едую]щ[ий] картин[у]». Предполагалось, что созданный таким образом сигнал будет восприниматься как «на следующую(ей,) картину(е)».

Тестовые последовательности были созданы так же, как в предыдущем эксперименте, и была полностью сохранена методика самого эксперимента. Испытуемыми по-прежнему были студенты и преподаватели Филологического и Восточного факультетов Санкт-Петербургского государственного университета: всего 78 человек. Ни один из испытуемых не принимал участия в предыдущем эксперименте.

В целом результаты эксперимента подтвердили правильность наших предположений относительно роли контекста при восприятии редуцированных словоформ (спонтанной речи): сужение контекста снижает надежность правильного опознавания словоформы, расширение контекста повышает опознаваемость, а в измененном контексте словоформа воспринимается в соответствии с семантико-грамматической структурой контекста.

«а он гэ[овори]т: – Да, но эт[о] не в первы[й] раз!	– 82,9% правильных ответов
гэ[овори]т не в первы[й] раз	– 37,2% правильных ответов
[бук]пугвальн[о] за час	– 32,9%
это был[о] [бук]пугвальн[о] за час	– 94,9%
на сл[едую]щ[ий] картин[у]	– 70% ответов «ей, ую».

Таким образом, полученные нами данные позволяют с достаточной долей уверенности говорить о том, что в ментальном лексиконе носителей русского языка не содержатся все возможные варианты редуцированных словоформ, встречающиеся в спонтанной речи.

Можно предположить, что для поиска и первичного отбора подходящих вариантов в перцептивном словаре используется вся сохранившаяся информация о сегментном составе редуцированной словоформы, а окончательный выбор производится с учетом частотности и ограничений, налагаемых семантико-грамматической структурой контекста.

Литература

1. Бондарко Л.В., Вербицкая Л.А., Гейльман Н.И. и др. Фонетика спонтанной речи / Под ред. Н.Д. Светозаровой.- Л.: Изд-во Ленингр. ун-та, 1988.- 248 с.
2. Венцов А.В., Касевич В.Б. Проблемы восприятия речи.– Изд. 2-е.– М.: Едиториал УРСС, 2003.– 240 с.
3. Русская разговорная речь: Фонетика. Морфология. Лексика. Жест / Отв. ред. Е.А. Земская. М.: Наука, 1983.
4. Ernestus M., Baayen H., Schreuder R. The Recognition of Reduced Word Forms // Brain and Language.- 2002.- V. 81.- P. 162-173.
5. Kassevich V.B., Ventsov A.V., Yagounova E.V. The simulation of continuous text perceptual segmentation: A model for automatic segmentation of written text // Language and Language Behavior.- 2000.- V 3, P II.- P 48-59

Роль визуального и звукового восприятия при порождении спонтанного нарратива

*Институт филологических исследований
Санкт-Петербургского государственного университета,
г. Санкт-Петербург, Россия,
polina272@hotmail.com*

В основе исследования, некоторые результаты которого отражены в настоящей работе, лежит серия экспериментов по изучению и анализу семантических и синтаксических характеристик спонтанного нарратива. Под термином «нарратив» понимается связный устный рассказ-монолог о некоторой последовательности действий.

Одной из основных задач этого исследования стало выяснение взаимодействия механизмов восприятия испытуемыми визуальной и звуковой информации и механизмов порождения связного текста. Кроме того, проведение серии экспериментов – эксперимента с взрослыми носителями языка и эксперимента с детьми младшего школьного возраста – позволило сделать выводы о возрастных особенностях порождения спонтанного нарратива и о влиянии на них уровня языковой компетенции.

Для решения этой задачи была разработана специальная методика проведения эксперимента. В первом эксперименте в роли испытуемых выступили 20 студентов гуманитарных и технических вузов Санкт-Петербурга в возрасте от 19 до 25 лет, а во втором – дети в возрасте 7-8 лет. Выбор испытуемых именно этих возрастных категорий был обусловлен уровнем развитости языковой личности к моменту поступления в школу, когда ребенок, по свидетельству Седова, «становится обладателем языкового механизма, своего рода персонального компьютера, который открывает перед ним новые когнитивно-коммуникативные возможности» (Седов 2004: 20). Кроме того, именно в возрасте 7-8 лет, по Пиаже, ребенок переживает первый критический период, когда эгоцентризм вступает в противоречие с появляющимся социальным инстинктом.

Материалом для экспериментов послужил 4-минутный фрагмент мультфильма, к которому прилагалась инструкция, идентичная для всех этапов эксперимента, - комментировать происходящее на экране (без каких-либо ограничений на форму и содержание рассказа). Каждый эксперимент состоял из трех этапов, между которыми был перерыв в 2 месяца, чтобы избежать запоминания испытуемыми содержания мультфильма и того, что они говорили на каждом этапе⁶. Различия между демонстрируемыми испытуемым мультфильмами заключались в их звуковом сопровождении: на первом этапе мультфильм демонстрировался без звука, на втором – в сопровождении модифицированной звуковой дорожки и на третьем – с оригинальным звуком мультфильма.

Второй этап – в сопровождении модифицированной звуковой дорожки – был проведен только с взрослыми испытуемыми⁷. При модификации оригинальной звуковой дорожки были сохранены интенсивность и основной тон исходного звукового сигнала, но отсутствовали все прочие фонетические характеристики, т.е. сохранилась вся просодика оригинального речевого сигнала, но вся информация, относящаяся к гласным и согласным, была удалена. Предъявление

⁶ Первый этап эксперимента с детьми проводился в январе, т.е. после полугодия обучения детей в первом классе, второй этап – в апреле-мае, т.е. в конце первого класса.

⁷ Эксперимент со студентами проводился в 2002-2003 годах, а с детьми – в 2004-2005. По результатам эксперимента с взрослыми было решено этап в сопровождении модифицированной звуковой дорожки с детьми не проводить, так как оказалось, что такой звуковой режим не оказывает значительного влияния на семантические и синтаксические характеристики порождаемого нарратива (см. далее).

испытуемым исходного видеоряда в сопровождении таким образом модифицированной звуковой дорожки дало возможность проверить гипотезу о том, что воспринимаемая говорящим в момент порождения речи просодическая информация влияет на просодику, а соответственно, и тема-рематическое членение порождаемых им высказываний.

Полученные от испытуемых на всех этапах эксперимента тексты были записаны на аудиопленку. Всего было получено более 6 часов записей 94 спонтанных связных рассказов.

К числу анализируемых синтактико-семантических параметров спонтанного нарратива были отнесены следующие: выбор испытуемым общей коммуникативной стратегии при построении текста, используемые тема-рематические структуры высказывания и текста, заполнение глагольных валентностей, использование эллипсиса и сходных с ним явлений, вторичная номинация. Проведенный анализ показал, что определяющим для тема-рематической структуры текста и количества тех или иных тема-рематических структур высказывания является совершаемый испытуемыми на первичном этапе выбор коммуникативной стратегии.

Под коммуникативными стратегиями нами понимаются те общие коммуникативные установки, которые были избраны говорящим для реализации полученного в ходе эксперимента коммуникативного задания – комментирования мультфильма. По материалам экспериментов можно говорить о 3-х основных стратегиях.

Стратегия 1 отражает нарратив от 3-го лица, и тексты, созданные с использованием такой стратегии, являются описанием происходящего с точки зрения стороннего наблюдателя. Стратегия 2 приводит к построению текстов, имеющих в своей основе перволичную повествовательную форму, но в более расширенном варианте, так как в таких текстах испытуемый говорит не от собственного лица, а от лица каждого появляющегося на экране персонажа. Эту стратегию можно также назвать «озвучиванием» текстов.

В качестве стратегии 3 был выделен такой подход к комментированию мультфильма, который можно назвать «критикой». Эта стратегия привносит в тексты частые ссылки на личное мнение испытуемого по поводу происходящего (*по-моему, я так думаю, не понимаю*), критический анализ изображения (*какое небо голубое / ой оно и вовсе не голубое, ой это даже не сосна это наверно елка только почему-то палка с какими-то / лепешками / а не ветками*), поступков героев (*медведь какой-то странный / какие-то у него идеи интересные очень // залезть на елку // развлекись / называется // раз заняться нечем / надо по елкам ползать*) и даже иногда самого испытуемого (*я тупая наверно очень но я не понимаю зачем он залез на дерево*)⁸.

Эти стратегии часто пересекались в рамках одного текста. Чаще всего сочетались стратегии 1 и 2, когда испытуемый и озвучивал персонажей, и просто описывал происходящее на экране. Кроме того, могли в рамках одного текста пересекаться и стратегии 1 и 3, т.е. нарратив от третьего лица и критические комментарии. Невозможным оказалось только сочетание стратегий 2 и 3 – озвучивание персонажей и привнесение личной критики испытуемого, что, видимо, вполне логично, так как говорящему трудно находиться одновременно и во внутренней ситуации мультфильма (при озвучивании персонажей), и во внешней ситуации эксперимента, оставаясь самим собой, что необходимо для критического взгляда на мультфильм⁹.

Выбор коммуникативной стратегии для построения текста сохраняется, независимо от различных условий эксперимента, практически у всех взрослых испытуемых. Иная картина наблюдалась в «детском» эксперименте. В отличие от студентов, практически все дети избранную на первом этапе стратегию во втором этапе, когда слушали оригинальную звуковую

⁸ С определенной долей обобщения эту стратегию можно соотнести с одним из выделенных У. Лабовом элементов нарратива – evaluation (Labov 1997).

⁹ В 3 нарративах, полученных в ходе проведения эксперимента с взрослыми, была использована комбинация всех трех основных коммуникативных стратегий – «нарратив от третьего лица», «озвучивание» и «критика», однако базовой стратегией в них являлся «нарратив от третьего лица», к которому присоединялись лишь некоторые элементы двух других стратегий.

дорожку мультфильма, изменяли на другую. Основные изменения произошли в связи с включением в нарратив элементов прямой речи, которая чаще всего повторяла оригинальный текст мультфильма.

Интересно также отметить, что дети старались воспроизводить оригинальный текст мультфильма полностью, особо старательно копируя интонационные контуры. Один испытуемый настолько увлекся текстом мультфильма, что проговаривал его как бы вместе с героями, в результате чего произносимые вслух высказывания оказывались усеченными до отдельных слов или даже слогов. Кроме того, в текстах большинства испытуемых-детей присутствуют паузы во время речи персонажей, что также свидетельствует о повышенном внимании к оригинальному тексту мультфильма. В эксперименте со студентами подобной картины не наблюдалось.

Одной из причин такого поведения детей можно предположить то, что интонация рассматривается ими как неотъемлемая внутризная характеристика, т.е. стратегия испытуемых-детей – с данной точки зрения – заключается в том, что они «копируют» текст со всеми его характеристиками, включая интонационные.

Для анализа тема-рематической структуры текста использовалась психолингвистическая модель тема-рематического структурирования текста Л.В. Сахарного, которая «позволяет представить текст в виде иерархической организации тема-рематических структур разной степени сложности, переводя, таким образом, разговор об анализе цельности, ее структурировании и характере отражения в тексте на операционный уровень» (Сахарный 1998: 7).

По этой модели, в тексте выделяются простые немаркированные монокомплексы, состоящие только из наименования предмета (*Мальчик*), и простые маркированные монокомплексы, состоящие из наименования предмета и какого-то одного его признака (*Мальчик¹⁰ ест яблоко*). Простые монокомплексы могут комбинироваться в сложные монокомплексы различной структуры: структурой «куста» называется комбинация из двух или более простых маркированных монокомплексов, имеющих общий предмет (*Шалтай-Болтай сидел на стене, Шалтай-Болтай свалился во сне*); структурой «цепочки» называется комбинация из двух или более простых маркированных монокомплексов, у которых признак первого выступает как предмет второго монокомплекса (*Это синица, что ворует пшеницу, что в дальнем чулане хранится в доме, который построил Джек*); комбинированной (смешанной) структурой называется сочетание структуры куста со структурой цепочки.

При анализе текстов, полученных в ходе эксперимента с детьми, было обнаружено преобладание структур куста над всеми другими возможными типами. Возможной причиной этого может быть то, что в данной структуре не требуется повторения темы, создания новой синтаксической и семантической структуры, сохраняется закономерность текстообразования. Иными словами, «известная» информация (тема) сохраняется в памяти говорящего, а вербализуется только «новая» информация (рема).

Анализ тема-рематической структуры полученных нарративов и их статистические изменения на разных этапах эксперимента выявил зависимость используемых структур от выбранной испытуемым коммуникативной стратегии. В случае выбора стратегии «нарратива от третьего лица» в тексте преобладали сложные тема-рематические структуры – структуры «куста» и смешанные структуры в «детских» нарративах и структуры всех типов в текстах, полученных при эксперименте со взрослыми. Стратегия же «озвучивания» увеличивала число простых маркированных и немаркированных монокомплексов в текстах как детей, так и взрослых, так как именно этот тип структур является более типичным для разговорной речи.

При выделении и классификации тема-рематических структур высказывания учитывались интонационные характеристики высказывания и его морфо-синтаксическая структура. В

¹⁰ Подчеркиванием отмечен предмет монокомплекса.

результате было выделено 9 типов простой тема-рематической структуры (одна тема – одна рема) и 4 типа сложной тема-рематической структуры высказывания

Тип А: самый частотный, классический тип: тема в нем обозначает субъект действия и совпадает с подлежащим, а рема – либо действие, которое оно совершает, либо признак, который характеризует тему. Такие высказывания демонстрируют «совпадение актуального и грамматического членений предложения» (Кронгауз 2001: 247).

Тип В: тема обозначает место действия и выражена существительным в косвенном падеже с предлогом либо наречием места, а рема сообщает, что происходит в обозначенном темой месте. О предложениях такого типа упоминал В. Матезиус: «Иногда в предложении, своеобразно предвосхищая ещё не раскрытую предметную ситуацию высказывания, отбираются обстоятельства места или времени, которые ставятся в начало бытийного предложения в качестве исходной точки высказывания» (Матезиус 1967: 241). Падучева называет такие предложения «локальными» и выделяет в них три компонента: локализатор, являющийся в то же время темой (или «началом» по Падучевой), подлежащее и бытийный глагол, являющиеся ремой (Падучева 2002: 128).

Тип С: тема выражена указательным местоимением *это*, а способы выражения ремы ограничений не имеют.

Тип D: рема выражает признак предмета, обозначенного в теме, а тема выражена либо существительным, либо личным местоимением в косвенном падеже.

Тип E: тема обозначает определенный момент времени и выражена чаще всего наречием, а рема сообщает то, что в указанное время происходит.

Тип F: высказывания с «невыраженной темой» (*звонит телефон; пришла бабушка*). Борботько считает, что эти высказывания выполняют выделительную функцию и основной их целью является привилегированная презентация объекта (ср. *il reste trois places* и *trois places restent*), и называет их «презентирующими» [Борботько 1981]. Кроме того, высказывания с такой структурой характерны для театральных ремарок: *vint un personnage etrange* (пример Реферовской), *появляются лакеи* (Е. Шварц). Хотя в подобных предложениях часто выделяют «обратный порядок» тема-рематического членения или говорят об отсутствии в нем темы как таковой, вероятно, более справедливым оказывается предположение В.Б. Касевича о наличии в подобных предложениях эксплицитно не выраженной темы – некоего локализатора описываемой высказыванием ситуации.

Тип G: тема только называется, но ничего о ней не сообщается.

Тип H: в отличие от высказываний предыдущего типа, в этом типе в высказываниях обозначена только рема, определяющая всё предыдущее высказывание. Сюда же относятся восклицательные высказывания, которые включают в свою структуру особые восклицательные лексемы и конструкции, например, *какой!, ну и*.

Тип I: вопросительные высказывания с обратным тема-рематическим порядком, где ремой является вопросительное слово, занимающее первую позицию в синтагме (см. Янко 2001).

Сравнение данных, полученных от испытуемых-студентов, с данными эксперимента с детьми показало, что некоторые типы простой тема-рематической структуры оказываются одинаково частотными в обоих случаях – например, тип А. Увеличение в текстах, полученных от испытуемых-детей, числа высказываний с простой тема-рематической структурой типа I (вопросительная конструкция) можно объяснить повышенным вниманием детей к воспроизводству чрезвычайно эмоциональной оригинальной речи персонажей мультфильма. Одной из причин такого поведения детей можно предположить то, что интонация рассматривается ими как неотъемлемая внутризная характеристика, т.е. стратегия испытуемых-детей – с данной точки зрения – заключается в том, что они «копируют» текст со всеми его характеристиками, включая интонационные.

Интересно отметить, что в материале «детского» эксперимента реже встречаются высказывания с простой тема-рематической структурой типа D (где тема выражена существительным или местоимением в косвенном падеже), а высказывания с простой тема-рематической структурой типов В и E (где тема выражена наречием места или времени соответственно) – чаще, чем у взрослых. Таким образом, дети этого возраста, вероятно, уже

умеют отделить синтаксический способ выражения темы от ее коммуникативной нагрузки и овладели навыками построения высказываний, где в качестве «известной» информации выступает обстоятельство места или времени, однако еще испытывают некие затруднения с выражением темы посредством использования существительного или местоимения в «неименительном» падеже.

Особое внимание привлекает практически полное отсутствие в материале «детского» эксперимента высказываний с простой тема-рематической структурой типа С, где тема выражена указательным местоимением *это*. Тема может указывать на какое-то событие или предмет на экране (*это сосна*), на описанную в предыдущих высказываниях ситуацию (*ведь это гораздо привлекательнее нежели сидеть сейчас*) или часть предыдущего высказывания (*[очень похоже / на звуки которые издаёт кошечка // всё-таки да /] это пожалуй кошечка*)¹¹. Как отмечает Кормилицына, «выполняя анафорическую функцию, местоимение *это* отсылает к ситуации, названной в предтексте, причем иногда оно способно обобщать содержание даже нескольких предложений» (Кормилицына 2003: 43). Именно эта способность к столь широкой абстракции, где содержится указание на тему – либо картинку, либо предтекст, вероятно, еще не полностью выработана у детей 7-8 лет.

Высказывания со сложной тема-рематической структурой могут быть 4 типов: первый вариант структуры «куста» состоит из одной темы, к которой привязывается несколько рем; вторым вариантом «структуры куста» является такой тип сложных тема-рематических конструкций, который представляет собой структуру с вторичной предикацией, выраженной деепричастием или деепричастным оборотом. В таких случаях в русском языке тема совпадает всегда, можно даже сказать, «в принудительном порядке», а второй предикат имеет подчиненный характер. К оставшимся двум типам относится «структура цепочки», которая состоит из последовательности тема-рематических структур, где рема 1 равна теме 2, и «структура матрешки», включающая в себя конструкцию с предикатным актантом, от которого зависит отдельная тема-рематическая структура, являющаяся неотъемлемой частью высказывания¹².

Сравнение данных, полученных от испытуемых-студентов, и «детского» эксперимента показало, что из сложных тема-рематических структур самой редкой является второй вариант структуры «куста», содержащий в себе деепричастный оборот (у детей он встретился только один раз на оба этапа эксперимента¹³). Кроме того, ни разу детьми не была употреблена структура «цепочки», хотя у студентов она встречалась неоднократно. Что же касается двух других сложных тема-рематических структур, то и у детей, и у студентов были отмечены одинаковые тенденции к уменьшению количества первого варианта структуры «куста» и увеличению употребления структуры «матрешки».

Сопоставление тема-рематической структуры высказывания с выбранной испытуемыми коммуникативной стратегией показало следующую зависимость, которая подтверждается как материалами эксперимента со студентами, так и материалами детского эксперимента. Для стратегии «нарратив от третьего лица» оказались более характерны сложные типы тема-рематической структуры высказывания («куст 1», «куст 2», «цепочка» и «матрешка») и следующие простые типы тема-рематической структуры – тип А, В и Е («ответственные» за установку пространственно-временных параметров текста), тип С и тип F. Выбор для построения нарратива стратегии озвучивания приводил к увеличению в тексте числа простых тема-рематических структур: типы А, D и Е, тип G (только тема), тип H (только рема) и тип I,

¹¹ Примеры из «студенческих» текстов.

¹² Между выделяемыми по модели Сахарного видами тема-рематической структуры текста и классификацией тема-рематической структуры высказывания можно провести некоторые параллели. Структура куста у Сахарного близка по своему составу структуре «куст 1», а структура цепочки у Сахарного – одноименной структуре сложного тема-рематического высказывания. Их близость подтверждается также и результатами анализа их употребления студентами и детьми.

¹³ Хотя, как отмечает С.Н. Цейтлин, появление деепричастных оборотов в детской речи происходит в возрасте 5-6 лет (Цейтлин 2000: 220), вероятно, к возрасту 7 лет ребенок все же усваивает эти структуры еще недостаточно прочно.

соответствующий вопросу; из сложных тема-рематических структур при стратегии «озвучивания» оказывались возможными только небольшие структуры типа «куст 1».

Полученную зависимость тема-рематической структуры текста и тема-рематической структуры высказывания от выбора коммуникативной стратегии можно представить на схеме¹⁴:



Проведенный анализ семантических и синтаксических особенностей спонтанного нарратива у взрослых носителей русского языка (студентов 19-25 лет) и детей в возрасте 7-8 лет показал, что дети с большим вниманием относятся к звуковой информации, чем студенты, вероятно, причиной этого является необходимость для ребенка чем-то дополнить визуальную информацию, чтобы лучше объяснить, понять «картинку» и сделать правильный вывод. Кроме того, в этом возрасте для детей характерна стратегия «копирования», что проявляется как в попытках воспроизведения интонации воспринимаемых высказываний, так и в отсутствии конструкций с анафорическим *это*, где последнее заменяло бы название ситуации.

Результаты сравнения нарративов, полученных при трех акустических режимах, подтвердили влияние воспринимаемого звукового сигнала на выбор коммуникативной стратегии, которая, в свою очередь, определяет последующий выбор тема-рематической структуры текста и тема-рематической структуры высказывания.

Литература

1. Борботько В. Г. Элементы теории дискурса, Грозный, 1981.
2. Кормилицына М.А. Семантически осложненное (полипропозитивное) простое предложение в устной речи. Изд. 2-е. М., УРСС, 2003.
3. Кронгауз М.А. Семантика. М., 2001.
4. Крушельница К.Г. К вопросу о смысловом членении предложения // Вопросы языкознания, 1955, №5.
5. Матезиус В. О так называемом актуальном членении предложения // Пражский лингвистический кружок, М., 1967.
6. Падучева Е.В. Высказывание и его соотношенность с действительностью. М., 2002.
7. Сахарный Л.В. Тема-рематическая структура текста: основные понятия // Язык и речевая деятельность, 1998, №1.
8. Седов К.Ф. Дискурс и личность. М., Лабиринт, 2004.
9. Цейтлин С.Н. Язык и ребенок. Лингвистика детской речи. СПб., 2000.
10. Labov W. Some Further Steps in Narrative Analysis / Journal of Narrative and Life history, 1997.

¹⁴ Для двух других рассматривавшихся параметров нарратива – глагольной структуры и вторичной номинации – зависимости от выбора коммуникативной стратегии не обнаружилось. Однако есть основания полагать, что на их выбор зависит употребление той или иной тема-рематической структуры высказывания.

Дискурсивные слова в русской разговорной речи (на материале анализа спонтанной разговорной речи)¹⁵

*Санкт-Петербургский государственный университет,
г. Санкт-Петербург, Россия,
verde2002@yandex.ru*

1. ВВЕДЕНИЕ

1.1. Материал и основная цель исследования.

Целью данного фрагмента исследования спонтанной русской разговорной речи явилось описание *дискурсивных слов*¹⁶ (далее – ДС) русского языка с применением теоретической парадигмы, предложенной авторами коллективной монографии «Дискурсивные слова русского языка. Опыт контекстно-семантического описания» под редакцией К.Л. Киселевой и Д. Пайара. Материалом для работы послужили тексты ток- и реалити-шоу «К барьеру» (91 стандартная страница текста) и «Семь под солнцем» (114 страниц), записанные летом и осенью 2005 года¹⁷.

На первом этапе работы были изучены текстовые расшифровки телепередач и выделен ряд дискурсивных слов, куда вошли как уже описанные в литературе единицы, так и новые, ранее не рассматривавшиеся. Далее был проведен сопоставительный анализ обнаруженных ДС.

1.2. Предмет исследования

Предметом исследования является класс слов, важной особенностью которых является «непосредственная связь с функционированием дискурса» [Киселева, Пайар 1998: 7]. Благодаря этим словам речь принимает логическое и эмоциональное оформление, а говорящий получает гибкий инструмент для выражения целого спектра значений, в том числе и оценочных.

В русской лингвистической традиции слова, явившиеся предметом нашего анализа, принято именовать в соответствии с одной из двух категорий, в рамках которых их описывают: служебные (неполнозначные) слова и, по определению В.В. Виноградова (см. [Виноградов 1947/1986]), модальные слова. Подход «со стороны» служебных (неполнозначных) слов обладает тем недостатком, что «включает и собственно служебные слова (такие, как предлоги, союзы, связки, различные аналитические форманты), и разнообразные частицы, функции которых не укладываются в понятие ‘служебные слова’, и ‘полузнаменательные’, ‘строевые’ слова» [Леденев 1973]. Виноградовская же классификация модальных слов основана, как отмечается в [Киселева, Пайар 1998: 8] на функционально-семантических критериях. В.В. Виноградов выделяет слова, выражающие субъективную передачу чужой речи, мысли и ее оценку со стороны говорящего (*мол, де, дескать, будто, будто бы, слышно*), эмоциональное освещение действительности (*чего доброго, право, как-никак, авось, небось, пожалуй*), логическую оценку сообщения (*вероятно, несомненно, видимо*), отношение содержания речи к общей последовательности мыслей (*кстати, кроме того, например*), оценку способа выражения (*буквально, вообще говоря, словом*) и др. [Виноградов 1947/1986: 603-607].

Поскольку соответствующий фрагмент лексической системы русского языка еще не получил подробного освещения в литературе, используемая терминология и методы описания в настоящее время не обладают должной концептуальной стройностью и варьируют в

¹⁵ Работа выполнена в рамках проекта «Русский язык и современная Россия» при поддержке федеральной целевой программы «Русский язык».

¹⁶ Отметим, впрочем, что в русистике термин *дискурсивное слово* уступает по употребительности другим терминам – логическая частица, модальная частица, коннектор, дискурсивный маркер и др.

¹⁷ Более подробное описание материала см. в [Горбова et al. 2006].

зависимости от выбранного подхода. В рамках данного исследования принимается терминология и методология контекстно-семантического подхода, послужившая основой для анализа ряда ДС, описанных в коллективной монографии [Киселева, Пайар 1998].

2. ДИСКУРСИВНЫЕ СЛОВА В ТОК-ШОУ «К БАРЬЕРУ» И РЕАЛИТИ-ШОУ «СЕМЬ ПОД СОЛНЦЕМ»

2.1. Критерии выделения ДС

В данном исследовании, вслед за авторами [Киселева, Пайар 1998], были приняты следующие критерии для выделения ДС.

Дискурсивными признавались единицы, обладающие признаками:

- отсутствия какого-либо денотата в общепринятом смысле;
- установления отношений между двумя (или более) составляющими дискурса (в частности, отражение того, как высказываемое соотносится с истиной; отражение процесса взаимодействия говорящего и слушающего; выражение этической оценки, мнения; выражение пресуппозиций; выражение эмоциональных отношений; выражение логических взаимосвязей; выражение отношения фрагмента речи к общей последовательности мыслей; выражение оценки способа выражения; сопоставление и противопоставление разных утверждений говорящего или говорящих друг с другом и т.п.).

Указанные признаки ДС в то же время могут служить критериями и при разделении дискурсивных и недискурсивных употреблений слов, реализующих оба типа употребления.

2.2. Инвентарь дискурсивных слов, выявленных в исследованном материале

На основании этих функционально-семантических критериев из текстов исследованных ток- и реалити-шоу был выделены дискурсивные единицы, список которых представлен в Таблице 1.

Таблица 1

Инвентарь ДС, выявленных в исследованном материале

№ n/n	«К барьеру»	Кол-во вхождений	«Семь под солнцем»	Кол-во вхождений
1	буквально	3	буквально	6
2	ведь	20	ведь	2
3	видимо	2	видимо	4
4	вообще(-то)	21	вообще/ваше(-то)	164
5	в общем(-то)	8	в общем(то)	14
7	вовсе	3	-	0
8	во всяком случае	1	-	0
9	в первую очередь	4	в первую очередь	3
10	в (этом) смысле	2	-	0
11	вновь	3	-	0
12	всё-таки	11	всё-т(а)ки	11
13	вот	220	вот	333
14	-	0	вот именно	4
15	всё равно	10	всё равно	12
16	в том числе	10	-	0
17	-	0	в остальном	1
18	-	0	в целом	1
19	всего лишь	2	всего лишь	1
20	в принципе	2	в принципе	19
21	даже	34	даже	45

22	действительно	15	действительно	25
23	естественно	2	естественно	4
24	дело в том, что	5	дело в том, что	1
25	же	97	же	68
26	именно	13	именно	25
27	итак	2	итак	27
28	как раз	1	как раз	1
29	-	0	как ни странно	1
30	конечно	54	конечно	46
31	к сожалению	9	к сожалению	3
32	кстати	16	кстати	18
33	кстати говоря	4	-	0
34	короче	1	короче	77
35	(...) ладно	5	(...) ладно	47
36	лишь (бы)	2	-	0
37	лично	10	лично	4
38	между прочим	4	между прочим	4
39	(...)-то	58	(...)-то	54
40	наверно(е)	8	наверно(е)	39
41	наоборот	5	-	0
42	наверняка	3	наверняка	1
43	на самом деле	8	на самом деле	55
44	неужели	1	неужели	1
45	ну	196	ну	350
46	наконец(-то)	5	наконец(-то)	9
47	несомненно	1	несомненно	1
48	однако	8	-	0
49	особенно	2	особенно	2
50	-	0	по-любому	7
51	по крайней мере	3	по крайней мере	6
52	почти	2	почти	3
53	потом	23	потом	33
54	по всей видимости	1	-	0
55	-	0	по-настоящему	5
56	практически	3	практически	4
57	просто	56	просто	287
58	прямо	6	прямо	19
59	-	0	пускай	1
60	пусть	14	пусть	8
61	разве	3	разве	1
62	скорее/й (всего)	1	скорее/й (всего)	2
63	снова	6	снова	2
64	собственно	1	-	0
65	совершенно	11	совершенно	7
66	совсем	6	совсем	5
67	-	0	соответственно	5
68	тем более	2	тем более	2
69	тем не менее	3	-	0

70	только	35	только	64
71	так вот	9	так вот	8
72	у точно (не) z	10	у точно (не) z	7
73	уж	5	уж	4
74	уже	59	уже	77
75	фактически	5	-	0
76	хорошо	25	хорошо	46
77	хотя/ь	18	хотя/ь	24
78	чё/чево	19/0	чё/чево	91/1
79	что ли	5	что ли	14
80	явно	2	явно	1

2.3. Сопоставительный анализ употребления дискурсивных слов в текстах ток-шоу «К барьеру» и реалити-шоу «Семь под солнцем»

Анализ приведенных в Таблице 1 списков ДС можно обобщить в виде следующих наблюдений.

1. Частотность использования ДС в спонтанной разговорной речи обнаруживает значительное варьирование: от количества вхождений, равного 1, до нескольких сотен. Наибольшей частотностью характеризуются следующие ДС: *ну, вот, вообще/ваще(-то), просто* (последнее ДС высокочастотно лишь в речи участников реалити-шоу «Семь под солнцем»).
2. ДС разбиваются на несколько групп по представленности / непредставленности того или иного ДС в речи участников обеих телепередач и по частотности их использования:

I группа – соотношения типа X-0 или 0-Y: Ia (X-0, т.е. ДС, использованные в «К барьеру» и не представленные в «Семь под солнцем») - *вовсе, во всяком случае, в (этом) смысле, вновь, в том числе, кстати говоря, лишь (бы), наоборот, однако, по всей видимости, собственно, тем не менее, фактически*; Ib (0-Y, т.е. ДС, использованные в «Семь под солнцем» и не представленные в «К барьеру»); - *вот именно, в остальном, в целом, как ни странно, по-любому, по-настоящему, пускай, соответственно*;

II группа – соотношение типа X-Y, причем X и Y не равны нулю, и количественное превышение одного над другим менее или равно 3: *буквально, видимо, в общем(-то), в первую очередь, всё-таки, вот, всё равно, всего лишь, даже, действительно, естественно, же, именно, как раз, конечно, к сожалению, кстати, лично, между прочим, (...) -то, наверняка, неужели, ну, наконец(-то), несомненно, особенно, по крайней мере, почти, потом, практически, пусть, разве, скорее/й (всего), снова, совершенно, совсем, тем более, только, так вот, у точно (не) z, уж, уже, хорошо, хотя/ь, что ли, явно*;

III группа – соотношение типа X-Y, причем X и Y не равны нулю, и количественное превышение одного над другим более 3: IIIa (количественное преобладание в «К барьеру»): *ведь, дело в том, что*; IIIb (количественное преобладание в «Семь под солнцем»): *вообще(-то), в принципе, в целом, итак, короче, (...) ладно, наверно(е), на самом деле, просто, прямо, чё/чево*.

По-видимому, наибольший интерес в ходе дальнейшего анализа должны представлять группы I и III как отражающие различные (социолингвистически значимые) предпочтения говорящих в выборе ДС русского языка. С другой стороны, инвентари ДС в обоих списках обнаруживают существенно больше общего, чем расхождений, и наиболее многочисленной группой является II группа. Это свидетельствует о том, что в языке существует некая центральная зона дискурсивных слов, куда входят единицы, используемые большинством говорящих.

Поскольку ограничения в объеме не позволяют привести здесь всех наблюдений, сформулированных нами в ходе сопоставительного анализа ДС (а тем более сколько-нибудь полно проиллюстрировать употребления ДС участниками обеих телепередач), ограничимся лишь немногими замечаниями.

Во-первых, на использовании тех или иных ДС зачастую сказываются индивидуальные предпочтения говорящего относительно определенных слов, что часто приводит к девальвированию этих единиц и обретению ими статуса паразитического словесного материала; ДС становятся, по выражению Л.В. Щербы, “упаковочным материалом” [Щерба 1957: 44, в примечании]. См., например:

(1) В: А был ли такой момент, когда ты удивилась сама себе? Не тому, что ты ошиблась, а тому, что ты сделала какой-то поступок, которого сама от себя не ожидала?

Ю: Нет, *в принципе*, я тут делала то, что я могу сделать, *в принципе*... (Семь под солнцем)

Ср. также ниже (3-6) отрывки из высказываний одной из участниц реалити-шоу, по которым можно отметить, что использование ДС *ну* вошло в привычку, стало частью индивидуального стиля речи, и не приносит какой-либо модификации значения высказывания, выступая, вероятно, как проявление неких особенностей модели мышления (когнитивного стиля).

Во-вторых, в первую очередь в речи участников реалити-шоу «Семь под солнцем», для которой характерна высокая степень “раскованности”, можно отметить употребление неблизких по значению ДС, см. (2). Однако, несмотря на кажущуюся избыточность и “логические скачки”, интерпретация таких высказываний не вызывает у слушающего затруднений, хотя и производит впечатление “нагроможденности”.

(2) А: Не, Микки Маус, *вообще, на самом деле* лучше всего, *наверно*, будет.

С: Сделаю Микки Мауса только ради вас. (Семь под солнцем)

Далее, нередко в качестве ДС используются маркеры хезитации или, по выражению В.И. Подлесской (см., напр., [Подлесская 2006]), маркеры препаративной подстановки, – когда говорящий, не находя в момент речи наиболее подходящей экспликации своей мысли, вынужден воспользоваться подстановкой, неким средством приблизительной номинации, чтобы “выиграть время”, обдумать выбор нужного слова. Одним из примеров таких маркеров может служить слово *ну* (чем, по-видимому, и обусловлено его безусловное количественное лидерство в нашем материале: 196 вхождений в «К барьеру» и 350 – в «Семь под солнцем»). Ср.:

(3) Аня: *Ну-у*, меня все здесь раздражает. Я не могу здесь находиться, я брезгую просто. (Семь под солнцем)

(4) Аня: С Сашей изначально мы подружками не были. Я ее никогда не называла подружкой, она то есть знакомая моя девочка, да, человек, с которым, *ну*, я нашла общий язык на проекте. Я думаю, что дальше как бы у нас будут отношения, *ну*, продолжаться, вот, общение, но она как бы очень, не знаю, возможно, глупая еще. (Семь под солнцем)

(5) В: *Ну*, а в жизни на проекте тебе все нравится?

Аня: Нет, не все.

В: *Ну*, у тебя есть какие-нибудь жалобы или пожелания? (Семь под солнцем)

(6) Аня: *Ну*, мне не нравится то, что просто многие люди, когда включаются камеры, ведут себя совершенно, совершенно по-другому.

В: Кто?

Аня: *Ну*, это-о-о Кирилл, Юлианна, Саша. И вот они просто меняются, просто сразу же. (Семь под солнцем)

(7) *Ну* надо же показать тем людям, которые стоят вокруг нас, да? нашей молодежи там и так далее, *ну* подлинную историю надо рассказать. (К барьеру)

Отметим, что иногда в устной спонтанной речи могут возникать сбои как следствие несоответствия некоего фрагмента высказывания изначальному плану, и говорящие для их корректировки применяют дискурсивные слова – этой теме посвящена, к примеру, совместная работа В.И. Подлесской и А.А. Кибрика “Коррекция сбоев в устной спонтанной речи: опыт корпусного исследования” [Кибрик, Подлесская 2005]. Среди возможных причин таких сбоев приводятся следующие: говорящий сделал выбор в пользу неправильной или неточной

номинации; порожденный фрагмент высказывания не может быть адекватно встроен в дискурс; фрагмент был завершен прежде, чем говорящий успел спланировать дальнейшее развертывание текста. Во время корректировки таких сбоев говорящий вынужден прибегать к помощи элементарных дискурсивных единиц, наиболее подходящей из которых, как мы ранее уже отметили, является ДС *ну*.

Вообще в речи молодежи, с характерной для нее тенденцией к частой смене темы разговора, со слабой контекстной и ситуационной связностью, большую роль играют ДС, помогающие отчасти сгладить эти недостатки, ввести “объективно несвязанное как субъективно связанное” [Киселева, Пайар 1998: 249]: *кстати, вообще, ну, итак, между прочим, вообще, лично (...)*. Отметим, что в речи образованных людей эти же лексические средства могут выражать объективную связность. И тогда для одних и тех же ДС в случае молодежной речи будут действительны грани ‘ни с того, ни с сего’, ‘отступление от ассоциации’, ‘поворот разговора’, а для “взрослой речи” - грани ‘неожиданно, но неслучайно’, ‘существенная деталь’, ‘заодно’, ‘удачно’, ‘хочу подчеркнуть’.

В заключении отметим, что изложенное выше небольшое частное исследование является собой самый общий обзор употребленных в имеющемся в нашем распоряжении корпусе русской разговорной речи дискурсивных слов с краткими пояснениями относительно особенностей их функционирования.

Данное исследование носит пилотный характер и в некоторой степени предполагает небесспорность принятых решений. Тем не менее, авторы надеются, что эта работа внесет свой скромный вклад в изучение многообразия дискурсивной лексики русского языка и послужит основой для дальнейших исследований в этой области.

Список литературы

1. Баранов et al. 1993 – *А.Н. Баранов, В.А. Плунгян, Е.В. Рахилина*. Путеводитель по дискурсивным словам русского языка. М., 1993.
2. Виноградов 1947/1986 – *В.В. Виноградов*. Русский язык (Грамматическое учение о слове). Изд.3-е. М., 1986.
3. Горбова et al. 2006 – *Е.В. Горбова, Н.А. Слепокурова, Т.В. Черниговская, Е.П. Комовкина, Т.В. Матвеева, Е.И. Риехакайнен, А.С. Романова*. Предварительные результаты мониторинга современной русской устной спонтанной речи // Современная русская речь: состояние и функционирование. / Под ред. С.И. Богданова, Н.О. Рогожиной, Е.Е. Юркова. СПб., 2006. С.3-31.
4. Кибрик, Подлеская 2005 – *А.А. Кибрик, В.И. Подлеская*. Коррекция сбоев в устной спонтанной речи: опыт корпусного исследования // Компьютерная лингвистика и информационные технологии: Труды международной конференции «Диалог’05». М., 2005
5. Киселева, Пайар 1998 - Дискурсивные слова русского языка. Опыт контекстно-семантического описания / Под ред. Д. Пайара, К.Л. Киселевой, М., 1998
6. Котов 2003 – *А.Е. Котов*. Социокультурные особенности функционирования дискурсивных маркеров. Доступно на: <http://science.ncstu.ru/articles/hs/10>
7. Леденев 1973 – *Ю.И. Леденев*. Состав и функциональные особенности класса неполнозначных слов в современном русском литературном языке. М., 1973.
8. Подлеская 2006 – *В.И. Подлеская*. О грамматикализации и «прагматизации» маркеров речевого затруднения: феномен препаративной подстановки // Третья Конференция по типологии и грамматике для молодых исследователей. Материалы. СПб., 2006. С.189-210.
9. Феденко 2002 – *П.В. Феденко*. Опыт контекстно-семантического описания дискурсивного маркера *Ну*. <http://www.dialog21.ru/materials/archive.asp?id=7399&y=2002&vol=6077>
10. Щерба 1957 - *Л.В. Щерба* Опыты лингвистического толкования стихотворений: «Воспоминание» Пушкина / Избранные работы по русскому языку. М., 1957. С. 26–44.

Разговорная русская речь и ее единицы

*Саратовский государственный университет,
г. Саратов, Россия,
svandreeva@rambler.ru*

В статье предлагается классификация синтаксических единиц на основе их строения и функций применительно к устной речи (развиваются мысли автора, опубликованные в журнале «Вопросы языкознания» и монографии «Речевые единицы русской речи: Система, зоны употребления, функции»)¹

Поскольку в понимании саратовской школы изучения РР разговорная речь – это только устная речь, встает вопрос об особых единицах, свойственных устной речи в отличие от речи письменной. Ядром устной речи является РР, анализ которой невозможен без описания всех ее структурных единиц. В связи с этим одной из основных считаем проблему выработки новых универсальных оснований выделения структурных единиц устной речи.

Членение потока устной речи на предложения оказывается довольно затруднительным. Выделенные из этого потока единицы далеко не всегда обладают признаками грамматического предложения. Вместе с тем, несмотря на большое количество работ по устной речи (разговорной, научной, телевизионной и т.д.)² сама система синтаксических единиц еще не разработана, а разработанная для письменной речи плохо подходит к речи устной.

Материалом нашего исследования послужила устно продуцируемая речь (магнитофонные записи производились автором в период с 1997 по 2006 г.), а также опубликованные тексты устной речи. Общий объем фактического материала приближается к ста пятидесяти тысячам словоупотреблений.

На основе проведенного анализа конкретных устных дискурсов и литературных данных нами установлены четыре вида информации, передаваемой единицами речевой коммуникации. Под **фактуальной информацией** понимается всё то, что пополняет интеллектуальный запас знаний человека или содержит сведения бытового характера, необходимые в данный момент.

Как установили исследователи, со стороны адресата коммуникация предполагает не только декодирование семантики языковых знаков, но и раскрытие использованных глубинных смыслов текста, пресуппозиций, авторских стратегий, учет фоновых знаний, ситуации общения и структуры дискурса. В рамках нашего исследования – это распознавание разнообразной коммуникативной и дискурсивной информации. К этим выводам близки теоретические положения О.Б. Йокоямы о семи видах знания, в том числе метаинформационном, совершенно необходимом для успешного осуществления информационного дискурса³. Для нас важно утверждение исследовательницы о том, что средства передачи метаинформационного знания (знания кода и дискурсивной ситуации) вырабатываются в каждом языке, следовательно, само это знание универсально для разных языковых систем и без него невозможен прием фактуальной информации.

Коммуникативная информация отражает взаимодействие между автором речи и ее адресатом и, следовательно, направлена на организацию общения: информативность в контакто-регулирующем плане (поддержание контакта, оформление этапов интеракции); информативность в оценочно-интерпретационном плане речи/ситуации; информативность в плане интерперсональных отношений (интимизация общения, смягчение категоричности суждений и т.п.).

Дискурсивная информация направлена на организацию дискурса и ориентацию адресата в нем для оптимального восприятия как фактуальной, так и коммуникативной информации: информативность в плане структурирования дискурса, обозначения роли фрагмента в тексте, отношения к нему автора и т.д.

Сигнальная «информация» – это информативность произвольных речевых проявлений (в том числе табуированных восклицаний и их заменителей) – выражение психофизиологического и эмоционально-чувственного состояния говорящего.

В результате исследования «многоликой» речевой реальности выделены два типа единиц: основные, составляющие «костяк» синтаксиса речи, и вспомогательные, используемые говорящим факультативно, т.к. их функционирование ограничено конкретными речеорганизующими задачами.

Основные единицы речевой коммуникации условно назовем «конструктивно-синтаксическими единицами» (КСЕ). Данный термин свидетельствует о том, что в фокусе внимания находятся не общепринятые синтаксические единицы (синтаксема, словосочетание, простое предложение, сложное предложение и т.д.), а строевые элементы дискурса, функционирующие самостоятельно или в составе более сложных единиц. Компонент *конструктивно-* указывает не на соединение неких элементов в составе единицы, а на их роль в организации, конструировании дискурса. Этот термин позволяет охватить как предложенческие, так и непредложенческие, как предикативные, так и непредикативные речевые факты, в то время как термин «предикативная единица» исключает исследование непредложенческих единиц, термин «предложение» исключает исследование единиц не только непредложенческих, но и в составе единиц сложного предложения, а термин «высказывание» связан только с коммуникативным уровнем синтаксиса.

Основные (конструктивно-синтаксические) единицы речевой коммуникации можно определить следующим образом: это наименьшие структурные единицы как предложенческого типа на основе свободного конструирования, характеризующиеся реальной, формальной или актуально-ситуативной предикативностью, так и клишированные единицы релятивного типа, не обладающие признаками грамматической моделируемости и предикативности, но отличающиеся семантической и интонационной завершенностью и вместе с предложенческими единицами образующие «костяк» («тело») дискурса, особенно в устной диалогической речи.

Непредложенческие единицы релятивного типа, выполняющие речерегулирующие функции, большинство ученых называют коммуникативами (А. *Ну будешь/хочешь в команду?* – Б. *А мо!*). Конструктивно-синтаксическими единицами в строгом смысле этого слова могут называться только предложенческие структуры. Коммуникативы релятивного типа (*Еще чего! Да ну!* и др.) должны быть противопоставлены предложенческим КСЕ на основании отсутствия общей для всех них формы и предикативного признака. Коммуникативы обязательны для диалога, но могут отсутствовать в монологе (особенно письменном). Однако, используя термин «конструктивно-синтаксические единицы», мы стремились подчеркнуть, что указанные предложенческие и непредложенческие структуры являются основными единицами речевой коммуникации в отличие от единиц вспомогательных, факультативных. Проведенное исследование позволяет не только предположить, но и доказать, что, являясь функциональным эквивалентом предложения, принадлежащий сфере дискурсивной модальности непредикативный нечленимый коммуникатив представляет собой самостоятельную речевую единицу наряду с инвариантной единицей языка – предложением (обладающим семантической и формальной структурой, предназначенной для формирования и передачи диктумного, номинативного содержания).

Наш анализ позволил выделить пять типов конструктивно-синтаксических единиц. В основу классификации положены следующие ведущие дифференциальные признаки: наличие/отсутствие коммуникативной самодостаточности единицы, особенности реализации в ней категории предикативности, соотношение диктумной (пропозициональной) и модусной составляющих, соответствие грамматической модели, характер передаваемой информации, функциональная направленность.

Первый тип КСЕ – **типичная предикативная единица** – это автосемантическая/синсемантическая КСЕ с реальной эксплицитной или имплицитной предикативностью; сочетающая диктумные и модусные смыслы; реализующая инвариантную грамматическую модель (возможна системная или контекстуальная ее неполнота); несущая в

основном фактуальную информацию, т.е. выполняющая информативную функцию, но иногда используемая для передачи коммуникативной информации и в этом случае имеющая речерегулирующую функциональную направленность: *Я слышал/ (1) (что) завтра обещают дождь// (2); А. Когда мы встречаемся? (3) – Б. Только когда она придет// (4); Как концерт? (5) = Какое впечатление произвел концерт?; Аня подарила сестренке/ сотовый/ (6) (а) сестренка ей/ ручку// (7); Оставайтесь с нами! (8)*. Приведенные примеры иллюстрируют: автосемантическую (3,5,6,8), синсемантическую (1,2,4,7) типичных предикативных единиц; соответствие грамматическим моделям предложения: системно-языковая неполнота модели наблюдается в (5) и контекстуальная неполнота – в (7). Типичные предикативные единицы реализуют имплицитную (7) или эксплицитную предикативность (в остальных случаях). Приведенные единицы характеризуются сочетанием диктумных и модусных смыслов; представлением фактуальной информации и информативной функциональной направленностью в (1–7), или коммуникативной информации и речерегулирующей направленностью в (8).

Второй тип – **структурно модифицированная предикативная единица**. Это автосемантическая/синсемантическая КСЕ с реальной эксплицитной/имплицитной или актуально-ситуативной предикативностью, сочетающая диктумные и модусные смыслы, базирующаяся на структурно модифицированной модели предложения (типизированный разговорный вариант модели, слабооформленное построение, ситуативно редуцированное и т.п.), несущая в основном фактуальную информацию, т.е. имеющая информативную функциональную направленность: *Вот этот вот термин/ он введен мною// (1); Ну-у/ твоя тушь/ конечно... (2) = очень высокого качества/дорогая; (В аптеке) Панангин// (3)*. Приведены структурно модифицированные предикативные единицы: типизированные разговорные варианты моделей – конструкция с именительным темы (1), намеренно структурно незавершенное высказывание с обрывом синтагматической цепи (2); предельно редуцированное построение, в котором вербализован только предмет востребования, остальная часть конструкции семантизируется ситуационно (3). Все КСЕ предикативны, но характеризуются разными способами оформления предикативного значения: (1) – эксплицитной предикативностью, выраженной формально-грамматически; (2) – эксплицитной предикативностью, реализованной посредством суперсегментных средств; (3) – актуально-ситуативной предикативностью, поскольку ситуация приобретения лекарственного препарата выступает неязыковым средством выражения предикативности. Все единицы характеризуются информативной направленностью, так как передают фактуальную информацию.

Третий тип – **функционально-семантически модифицированная предикативная единица (коммуникатив с формальной предикативностью)** – это автосемантическая/синсемантическая КСЕ с эксплицитной, но всегда формальной, а не реальной предикативностью, с функционально активизированной модусной и угнетенной диктумной составляющей, реализующая грамматическую модель предложения, однако несущая коммуникативную информацию и соответственно имеющая только речерегулирующую направленность: (По телефону) А. *До свидания дорогая!* – Б. *Целую тебя!* (1); *Так что вы думаете?* (2) *Этот альбом у нее сохранился!*; (Обращаясь к входящему в комнату) *Пришел?* (3); *Вообрази/ (4) идет она вчера по проспекту...; Она мне ничего не отдала/ представляешь?* (5). КСЕ (1,3) функционируют как автосемантические, (2,4,5) – синсемантические (предваряют информативные высказывания или находятся внутри них в качестве сегментов, играющих роль фатических маркеров). Коммуникатив с формальной предикативностью реализует грамматическую модель предложения, однако это и не предложение, и не коммуникатив в их обычном понимании. У КСЕ этого типа, несмотря на их эксплицитную предикативность, информативный компонент и диктумная составляющая функционально угнетены: (1) приближается к статусу клишированных этикетных формул; (2,4) подготавливают собеседника к восприятию последующей информативной части высказывания, обеспечивая «активного слушателя»; (3) используется для установления контакта; (5) позволяет говорящему апеллировать не только к вниманию собеседника, но и к его чувствам.

Четвертый тип – **предикативно-релятивная единица (гибридный коммуникатив)** – это автосемантическая/синсемантическая КСЕ с реальной эксплицитной или имплицитной предикативностью, с доминирующей модусной и ослабленной диктумной составляющей, реализующая фразеологизированную грамматическую модель предложения, сочетающая фактуальную и коммуникативную информацию, поэтому характеризующаяся информативно-речерегулирующей функциональной направленностью: А. *Как твои?* (1) *Привет им!* (2) – Б. *Спасибо//*. В эллиптических вариантах предложений *Как твои?* (в полной форме: Как твои близкие поживают/чувствуют себя?) и *Привет им!* (Передавай им привет!) фатическая составляющая доминирует над информативной. Вопросительное высказывание (1) – это фактически не запрос информации о состоянии дел и здоровья, а этикетная формула выражения внимания, участия, о чем свидетельствуют как последующая реплика, так и реакция собеседника (*Спасибо*). А. *Пока!* – Б. *До завтра!* (3) Информативное содержание гибридного информатива (3) сохраняется (*Расстаемся до завтра* – до определенного срока). Однако на самом деле это формула прощания, поэтому можно говорить о том, что фатическая составляющая усилена, а информативная ослаблена. А. *Тебе эта музыка не нравится?* – Б. *Почему не нравится?* (4) – А. *Ну ты как я включаю/ всё время на кухне торчишь//*. Гибридный коммуникатив *Почему не нравится?* (4) осуществляет запрос информации (информативная составляющая), вместе с тем в позиции ответной реплики, так же как в частных вопросах, в нем ощутимо модальное значение. Информативная составляющая осложнена коммуникативными параметрами: это не просто запрос информации, а еще и «реакция на модус полагания» (Н.Д. Арутюнова) высказывания собеседника (= Непонятно, почему ты так считаешь; ты не права).

Из приведенных примеров видно, что гибридный коммуникатив может характеризоваться как автосемантичностью (1,2,3), так и синсемантичностью (4); как эксплицитной предикативностью (1,2,3), так и имплицитной (4); реализовывать фразеологизированную эллиптическую модель предложения в ее системно-неполном (1,2,3) или контекстуально неполном (4) варианте; сочетать информативную и речерегулирующую функциональную направленность.

Пятый тип – **релятивная единица (собственно коммуникатив)** – это синсемантическая непредикативная КСЕ, выражающая модусные смыслы, несущая прежде всего коммуникативную информацию, имеющая речерегулирующую направленность. Вследствие денотативной опустошенности коммуникатив не является членом предложения (элементом субъектно-предикатной структуры): А. *Пока!* (1) – Б. *Ну давай!* (2). В реактивных единицах (1,2), выполняющих функцию размыкания контакта, фатический компонент покрывает всё высказывание.

Вспомогательные единицы речевой коммуникации, не составляющие «костяк» синтаксиса речи, функционируют в устной речи наряду с основными единицами (выделенными пятью типами КСЕ). Это дискурсив, диалогический повтор и звуковой жест.

В последнее время **дискурсивные слова** и близкие к ним единицы стали объектом пристального внимания исследователей-русистов. В роли дискурсивно-структурирующих единиц выступают прежде всего предложно-падежные формы, служебные и полуслужебные лексемы, частицы, модальные слова, некоторые знаменательные слова, организующие дискурс (*всего-навсего, по всей видимости, только, лишь, разумеется, скажем* и др.), а также типичные предикативные единицы дискурсивной направленности (*Вот это я говорила о детстве// Потом...* – предикативные единицы ориентируют адресата в потоке речи: информируют об исчерпанности темы и переходе к новому тематическому фрагменту сообщения).

Наши наблюдения над дискурсивами (включая так называемые метатекстовые конструкции, средства речевой рефлексии) показали, что они имеют самую разнообразную структурную организацию и предикативное оформление. Это может быть 1) предикативно оформленная единица: *так теперь его называют*; 2) полупредикативная единица: *попросту / условно говоря*; 3) инфинитивная единица непредикативного характера: *проще / точнее сказать*; 4) частично десемантизированные глаголы: *скажем, глядишь*; 5) сочетания словоформ: *по словам учителя; дело в том* и т.п.

Второй вспомогательной единицей речевой коммуникации является **диалогический повтор**, выступающий как вторая реплика диалога, т.е. речевая реакция одного коммуниканта на высказывание другого. Такой повтор – широко распространенное речевое явление, обусловленное диалогическим видом общения. Как считают исследователи, использование в речи слушающего слов говорящего («чужого материала») является процессом естественным и практически неизбежным.

Попытка типологии разнообразных коммуникативных функций диалогического повтора предпринята в одной из наших работ⁴. Так, диалогический повтор в качестве «маркера несоответствия» указывает на несоответствие сообщения представлениям слушающего о предмете речи: А. *Между вторым и третьим этажом/ лежит переноска черная//* – Б. *Между вторым и третьим?* (В доме только два этажа) – А. *Ой! Между первым и вторым//*. «Эхо-повторам», напротив, свойственны автоматизм и информативная опустошенность, что не способствует реализации определенных стратегий и тактик: А. *Сколько тебе сарделек/ одну/ две?* – Б. *Две//* – А. *Две //*.

В отличие от коммуникатива, который как конструктивно-синтаксическая единица характеризуется совокупностью структурно-семантических и функциональных признаков, дискурсивы и диалогические повторы не образуют единого функционального класса единиц, поскольку не имеют четких границ и собственных дифференциальных структурно-семантических признаков. Данные единицы не приобретают самостоятельного значения в дискурсе, а только организуют его (могут находиться в составе других единиц); они факультативны, хотя и очень важны. На этих основаниях делаем вывод о том, что они являются не основными, а вспомогательными единицами дискурса и речевой коммуникации в целом.

К вспомогательным единицам речевой коммуникации мы относим и **звуковые жесты** – произвольные восклицания, выполняющие только сигнальную функцию: сигналы психофизиологического и эмоционально-чувственного состояния говорящего. **Безусловно-рефлекторные** речевые проявления – это первичная система произвольных восклицаний, связанная с подсознанием и сближающая человека с животным миром. Речевые проявления такого рода, когда звучание опережает осмысление эмоций, мы условно называем «докоммуникативными» знаками: (Сморщившись) **О-ох!** (1) *Что-то опять колет!;* (Автомобиль «подрезала» другая машина. Водитель, вспыхнув) **У-ух! Блин!** (2). Безусловно-рефлекторные речевые реакции (1,2), произвольно возникающие в ответ на неожиданное явление, ощущение, нельзя считать «подлинной речью». Как показывают наблюдения, «докоммуникативные» знаки отличаются спаянностью с конкретной ситуацией. Эти непредикативные речения *Фу!* (ощущение отталкивающего запаха) *Ай! Ой! Фу ты! Ха-ха!* и т.п. характеризуются отсутствием диктумной части (пропозиционального содержания). «Докоммуникативный» знак в отличие от коммуникативного не направлен на речевое взаимодействие собеседников. Следовательно, звуковой жест – это самостоятельная непредикативная нечленимая междометная речевая единица, имеющая сигнальную функцию (часто это сигнал о неблагоприятном психо-физиологическом или эмоционально-чувственном состоянии говорящего).

В основе проведенного анализа – функциональная направленность единицы (информативная, фатическая, регулятивная, дискурсивно-структурирующая или их сочетание), разграничивающая составные части высказывания и не позволяющая объединять их в единое элементарное синтаксическое целое. Поэтому члены предложения в качестве самостоятельных речевых единиц нами не вычленились, парцелляты считались неполными КСЕ, а речевые жесты, разные виды дискурсивов вычленились в качестве особых вспомогательных речевых единиц, поскольку они не входят в диктум предложения, а следовательно, не видоизменяют фактуальную информацию, привнося информацию другого вида. При вычленении вспомогательных единиц речевой коммуникации учитывались также следующие моменты: 1) коммуникативное, интонационное их выделение и образование ими особой синтагмы, 2) возможность при определенных условиях быть самостоятельной единицей.

Таким образом, в континууме коммуникации выделяются три зоны речевых единиц: коммуникативно-информативная зона (передают фактуальную информацию),

коммуникативно-регулятивная (передают коммуникативную информацию), дискурсивно-структурирующая (передают дискурсивную информацию). Кроме того, в континууме речевой коммуникации существует, хотя и стоит особняком и реализуется далеко не во всех типах устного дискурса, зона безусловно-рефлекторных проявлений («докоммуникативных» знаков). В результате взаимодействия выделенных зон образуются пространства их наложения. Комплекс зон имеет признаки полевой системности.

Наряду с многофункциональной типичной предикативной единицей, используемой в большинстве зон речевой коммуникации, в языке–речи определенные функции формируют специализированные единицы зон. Особый интерес представляют синкретичные единицы, которые еще не были в сфере внимания исследователей: гибридный коммуникатив, типичные предикативные единицы информативно-регулятивной или регулятивно-дискурсивной направленности, дискурсив и диалогический повтор регулятивно-дискурсивной направленности. Например, типичные предикативные единицы могут синкретично контаминировать фактуальную и коммуникативную информацию. В речи телеведущих они используются в качестве конвенционализованной формулы прощания (коммуникативный регулятор) и информирования о следующем выходе в эфир: *Завтра в это же время дня обязательно увидимся//*.

Основные антиномии «говорящий – слушающий», «экономия – избыточность» разрешаются в разговорной речи благодаря клишированности, конвенционализованности большого числа единиц. Поскольку в стандарте заинтересованы как говорящий, так и слушающий, естественно расширение общественного диапазона функционирования клишированных и конвенционализованных единиц: универсальная многофункциональная типичная предикативная единица всё чаще «уступает место» специализированной маркированной речевой единице – коммуникативу.

Процесс формирования коммуникативов продолжается. На пути «коммуникативизации» и «дискурсивизации» замечены единицы с разной степенью утраты номинативного содержания и свернутости структуры. В связи с ускорением темпа жизни наблюдается активизация коммуникативных процессов: 1) функциональной конвенциализации типичных предикативных единиц, приводящей к угасанию в них информативной функции, 2) формирования коммуникативов как собственно речевого, так и языкового уровней. Названные процессы так и хочется обозначить емкой формулой В.Г. Костомарова – «*Наш язык в действии*».

Таким образом, разработанная типология единиц речевой коммуникации (детализация существующей классификации) позволила описать наименьшие структурные единицы устной речи, определить их системное взаимодействие и коммуникативный потенциал, тем самым способствуя решению актуальной для русистики проблемы. Эта типология речевых единиц служит опознаванию «неопознанных синтаксических объектов» реальной устной речи. Совокупность полученных результатов создает фундамент нового направления синтаксиса – изучения не только отдельных синтаксических единиц, а всей реальной коммуникации как динамично меняющегося полевого устройства.

Примечания

1. См.: Андреева С.В. Типология конструктивно-синтаксических единиц в русской речи // ВЯ. 2004. №5; Речевые единицы устной русской речи: Система, зоны употребления, функции. Изд. 2-е, испр. М., 2006.
2. См., напр.: Русская разговорная речь / Под ред. Земской Е.А. М., 1973; Русская разговорная речь. Фонетика. Морфология. Лексика. Жест. М., 1983; Лаптева О.А. Русский разговорный синтаксис. М., 1976; Разговорная речь в системе функциональных стилей современного русского литературного языка. Грамматика. Саратов, 1992; М., 2003 и др.
3. Йокояма О.Б. Когнитивная модель дискурса и русский порядок слов. М., 2005.
4. См.: Андреева С.В., Плотникова А.В. Дискурсивный повтор в системе средств устной коммуникации // Русский язык и славистика в наши дни. М., 2004.

Е.П. Комовкина, Н.А. Слепокурова

Об опыте составления частотного словаря устного спонтанного текста*

Санкт-Петербургский государственный университет,
г. Санкт-Петербург, Россия,
nataliars@inbox.ru, ekomovkina@yahoo.com

В ходе выполнения проекта по исследованию русской спонтанной речи авторами настоящего доклада была выполнена письменная расшифровка видеозаписей телевизионных спонтанных диалогов, прозвучавших в эфире летом-осенью 2005 г. Эти записи составляют достаточно надежную репрезентативную выборку (время звучания каждой – не менее 4 час., расшифровка обеих занимает в сумме около 200 стр. текста), в которой зафиксирован определённого рода срез реальной звуковой «материи» русского языка на сегодняшнем этапе его функционирования.

Указанные записи представляют, по нашему предположению, разные пласты живой устной речи – молодежный, «раскованный» (передача *Большой Брат* (канал ТНТ) и относительно более «респектабельный» (*Культурная революция* (канал Культура)). Об этом свидетельствует как разница в возрасте, социальном статусе, образовании, культуре участников телепередач, так и характер ведения в них диалогов. Поскольку в передаче *Культурная революция* обсуждаются некоторые интересные для собеседников темы и вопросы мировоззренческого характера, реплики отдельных участников этой программы часто представляют собой спонтанные мини-монологи, в то время как в реалити-шоу *Большой Брат*, не обременявшем своих участников никакими серьезными интеллектуальными проблемами, ведется как будто бы обычный бытовой диалог с частой сменой собеседников. Об указанной разнице наглядно говорят цифры: по данным первых 50 стр. письменных расшифровок, средняя длина реплики одного участника диалога в *Большом Брате* в 5 раз короче, чем в *Культурной революции*.

Первый этап исследования полученного речевого материала включал в себя общий качественный обзор и сплошную фиксацию замеченных отклонений от соответствующих фонетических, грамматических и лексических норм кодифицированного литературного языка. Состав этих отклонений в обоих текстах вполне соответствовали той картине, которая нарисована в классических работах московских и саратовских исследователей русской разговорной речи [Горбова, Слепокурова и др. 2006].

Далее, естественно, возник вопрос о выявлении степени «спонтанности», «разговорности» этих текстов и о сравнении их по этому параметру друг с другом. Представляется совершенно нереалистичным осуществлять это сравнение в виде некоего подсчета по текстам предварительно отмеченных «разговорных явлений» и их последующего сопоставления по количеству и составу. Это обусловлено тем, что указанные «разговорные явления», представляющие собой отклонения от норм кодифицированного языка, имеют разную степень очевидности и достоверности как для носителей языка, так и для лингвистов. Критерии их выделения предельно прозрачны в сфере фонетики и морфологии (понятно, например, что зафиксированные нами в текстах выражения *бóшку ломит* или *в московском списке* «Родины» содержат в себе абсолютно недопустимые речевые ошибки), значительно менее очевидны в сфере лексики и словообразования (в самом деле, *к проблеме нечитания книг* или *ты не выпендриваешься перед всеми* – это ошибки или нормативные для разговорной речи явления?) и совершенно размыты и неопределенны в сфере синтаксиса, где разного рода эллипсис, употребление именительного темы, более свободный, чем при письме, порядок слов и т.п. уже давно признаются многими авторитетными исследователями устной речи

* Работа выполнялась в рамках проекта «Русский язык и современная Россия» при поддержке федеральной целевой программы «Русский язык» и Благотворительного фонда В.Потанина.

«сущностными» и потому, очевидно, нормативными атрибутами русской спонтанной речи. Эти соображения, основанные на предварительном, но в то же время далеко не поверхностном знакомстве с имеющимся в нашем распоряжении речевым материалом, заставляют предполагать, что единственным по-настоящему продуктивным подходом к сформулированной задаче является **сплошной** просмотр и анализ текстов с целью выявления в них некоторых строго определенных лексико-грамматических явлений. Последнее было осуществлено нами в виде попытки создания **частотного словаря** устного спонтанного текста.

Основная трудность, с которой пришлось столкнуться в процессе выполнения работы, – отсутствие прецедентов. Хотя в литературе и были обнаружены некоторые данные о попытках составления словарей устной русской речи, относящиеся к 60-ым гг. прошлого века, не удалось найти сведений о конкретных процедурах. Нельзя было ориентироваться и на опыт создания немногих имеющихся частотных словарей русского языка, основанных на письменных текстах, в которых единицей текста считается любое слово «от пробела до пробела» и каждая входящая в грамматическую парадигму словоформа возводится к основной словарной форме, – в результате же получается точная количественная информация о представленности в исследованных текстах неких «инвентарных» лексико-грамматических единиц непонятной психолингвистической природы.

Поэтому в данной работе мы опирались, прежде всего, на те принципы, которыми руководствуются разработчики и создатели существующих в настоящее время независимых отечественных корпусов русского языка – Национального корпуса русского литературного языка (НКРЛЯ, Санкт-Петербург, руководитель – проф. В.Б. Касевич) и Национального корпуса русского языка (НКРЯ, Москва, руководитель – проф. В.А. Плунгян). В обоих этих корпусах, создающихся на основе огромных массивов письменных текстов, не предусмотрена процедура лемматизации и отдельной словарной единицей считается **словоформа**. Это решение соответствует последним экспериментальным психолингвистическим данным, согласно которым единицей перцептивного лексикона (а равным образом и единицей текста) у носителей языков с развитым словоизменением является не лексема, а словоформа, и, возможно, – фонетическая словоформа, представляющая собой, в общем виде, просодически единый комплекс из знаменательного слова и клитик [Венцов, Касевич 2003]. Совершенно очевидно, что и любые современные попытки создания словарей **устных** разговорных текстов должны учитывать реальную психолингвистическую природу входящих в словник единиц. Поэтому в настоящей работе в качестве базовой словарной единицы рассматривалась отдельная словоформа.

Две других весьма серьезных трудности, с которыми пришлось столкнуться при подходе к практическому решению поставленной задачи, связаны с проблемами лексико-грамматической омонимии и так называемых «составных слов». Эти трудности, являющиеся главным камнем преткновения и при разработке корпусов русского языка, на наш взгляд, вполне адекватно решаются создателями НКРЛЯ. Проблема омонимии преодолевается в нем путем приписывания **каждой** словарной единице так называемого дескриптора, в котором указываются все лексико-грамматические признаки этой единицы; что же касается «составных слов», под которыми понимаются единицы, представляющие собой сочетания формальных слов, но не образующиеся по правилам, а воспроизводимые в тексте в «готовом» виде (например, *друг друга, в целом, в самом деле, может быть, не по себе* и т.д.), то они учитываются как отдельные текстовые и словарные единицы и их список по мере вхождения в корпус новых текстов постоянно пополняется.

Ориентируясь на этот подход и полностью с ним солидаризируясь, мы, тем не менее, не могли его копировать в данной работе – во-первых, из-за ограниченности времени и ресурсов и, во-вторых, из-за того, что, в отличие от составителей НКРЛЯ и НКРЯ, мы имели дело с записями **просодически размеченных устных** разговорных текстов. И хотя эта разметка была достаточно примитивной (отмечались паузы между синтагмами и высказываниями; коммуникативный тип высказывания – утверждение, побуждение, вопрос; ненормативные лексические ударения), опираясь на нее, а также на некоторые литературные данные и собственную интуицию, мы сочли возможным и необходимым использовать в своей работе

более широкую интерпретацию понятия «составное слово», чем это допускается при составлении словарей на основе письменных текстов.

Что касается омонимии, то ее фактический масштаб стал очевидным лишь при нацеленной на выполнение поставленной задачи дальнейшей обработке текста: выяснилось, что количество случаев употребления говорящими в устной речи лексически и грамматически омонимичных словоформ огромно: ср., например, семантику и частеречную принадлежность таких частотных и общеупотребительных лексем, как *это* и *хорошо*, в следующих примерах: *Что это ты сегодня такая грустная?*, *Что это значит?*, *Это дело всем вместе решать*, *Кингстон это такой город на Ямайке*, *Она хорошо поет*, *Здесь хорошо*, *Хорошо / я ему скажу*. См. также примеры грамматической омонимии, т.е. случаи совпадения одинаковых словоформ в разных грамматических значениях: *сигареты на столе*, *возьми там сигареты*; *большой перерыв*, *с большой надеждой*, *к большой радости*, *на большой тарелке* и т.д.). Естественно, каждая из отмеченных выше словоформ должна учитываться в словаре в качестве самостоятельной единицы. Для «разведения» омонимичных словоформ было решено ввести систему цифровых маркеров. Различные маркеры использовались для разных падежей существительных, прилагательных и местоимений, некоторых глагольных форм (например, инфинитива в функции императива – в отличие от инфинитива после модальных и фазовых глаголов). Особыми пометами в случаях необходимости отмечались также род прилагательного, субстантивированные прилагательные, наречия в предикативной функции и т.д.

Возвращаясь к лексической омонимии, следует оговориться, что, поскольку целью работы была попытка составления **частотного словаря**, мы не ставили перед собой задачу проводить тонкую семантическую дифференциацию на основе, скажем, имеющихся толковых словарей русского языка. Кроме того, характер текста позволял предполагать, что в словарях, составленных в основном по письменным источникам и отражающих функционирование письменного русского языка в лучшем случае трех – четырехдесятилетней давности, едва ли могут фиксироваться все лексико-семантические особенности современной, «раскованной», иногда шокирующей речевой «стихии» (весьма показательно, например, что встретившиеся в тексте личные глагольные словоформы *свалит* и *тормозит* употреблены лишь в сленговых значениях, зафиксированных в «Большом словаре русской разговорной речи» В.В. Химика и отсутствующих в академических толковых словарях). К тому же, объем рассматриваемого материала и ограниченные сроки исполнения работы делали такую дифференциацию совершенно нереалистичной. Поэтому сплошное (по тексту) разграничение омонимичных словоформ проводилось в основном путем тщательных повторных просмотров текста и в опоре на лингвистический опыт и интуицию исполнителей работы.

Наиболее трудным, серьезным и, разумеется, дискуссионным стал для данной работы вопрос о выделении и учете в рассматриваемом материале текстовых и словарных единиц, представляющих собой сочетания двух и более отдельных лексических элементов. И собственный речевой опыт лингвистов – авторов данной работы, и давно ставшие классикой наблюдения над русской разговорной речью ее московских и саратовских первооткрывателей, и тонкая интуиция известных исследователей-филологов (см., в частности, понятие коммуникативного фрагмента в [Гаспаров 1996]), и, наконец, вынужденно глубокое погружение исполнителей работы в анализируемый текст – все говорило о высокой степени клишированности разговорной речи, из чего с неизбежностью вытекала необходимость ее отражения в задуманном частотном словаре.

Поэтому в рамках выполнения поставленной задачи встал вопрос о получении некоего списка лексически объединенных форм – аналогов того, что в начале данного раздела было названо «составными словами», а далее будет именоваться рабочим термином «идиоматизмы». Мы сочли необходимым прибегнуть к такому переименованию потому, что, судя по описаниям подходов к созданию НКРЛЯ (см., например, [Баскулина 2006]), понятие «составного слова», возникшее из потребности адекватной лексической интерпретации прежде всего письменных текстов, является, несмотря на отсутствие полного списка таких слов, гораздо более строгим и отрефлексированным, чем то условное и недостаточно определенное с точки зрения наличного

лингвистического инструментария содержание, которое вкладывается нами в термин «идиоматизм». И «составные слова», и «идиоматизмы» – это предположительно единицы ментального лексикона, функционально равные слову и всякий раз воспроизводимые (и воспринимаемые) в речи в виде неких цельных комплексов. Разумеется, в состав «идиоматизмов» входят и «составные слова», но круг первых, на наш взгляд, может и должен быть гораздо более широким – хотя бы потому, что они принадлежат разговорной речи с ее высокой долей автоматизма, следствием чего является стремление говорящих к использованию «готовых» единиц. Ср.: «Говорящий, находясь в условиях непринужденного неподготовленного общения, стремится упростить и облегчить свое «речевое поведение», поэтому он легко и часто прибегает к готовым языковым формулам, в том числе всякого рода клише, шаблонам и стереотипам» [Земская, Китайгородская, Ширяев 1981: 6].

Априорно можно было предполагать, что выделенные нами «идиоматизмы» будут обладать разным языковым статусом и разной функциональной нагруженностью и, скорее всего, не будут жестко отграничиваться друг от друга, образуя своеобразный континуум. На одном полюсе этого континуума – грамматически не изменяемые единицы с внутренними связями такой силы, которая не позволяет вносить в эти единицы какие-либо изменения, на другом – единицы с возможностью грамматического изменения и с ослабленными внутренними связями, что выражается в спонтанной речи как возможностью перестановок образующих их элементов, так и инкорпорированием «посторонних» элементов.

Приступив к задаче непосредственного выделения «идиоматизмов» в анализируемом тексте, мы пытались использовать для формального обоснования этой процедуры все известные и доступные нам правила, критерии и средства. В их число вошли: 1) критерии разделения слова и словосочетания, предложенные В.Б. Касевичем (см. [Касевич 1988] и [Касевич 2006]), критерии разделения слова (=словоформы) и морфемы, предложенные В.А. Плунгяном [Плунгян 2003], критерии И.А. Мельчука, «противопоставляющие словоформу, с одной стороны, частям словоформы (морфам и цепочкам морф), а с другой стороны – группам словоформ (=словосочетаниям)» [Мельчук 1997], и 2) критерий просодической слитности. Из первой группы наиболее полезным и имеющим почти универсальную силу действия оказался для наших целей лишь критерий идиоматичности (неаддитивности); соответствие вычлененных «идиоматизмов» второму критерию (отчасти параллельному выделенному И.А. Мельчуком такому свойству словоформы, как звуковая связанность) было полным. Кроме того, при выполнении данной работы практически неизбежной оказалась сильная опора на субъективный фактор – лингвистическую интуицию и семантическую интроспекцию исполнителей.

В общей сложности в анализируемом тексте было выделено более 900 «идиоматизмов». По критерию идиоматичности полученные нами объединенные формы можно разделить на две группы.

Первая – аддитивные **аналитические единицы**, образующиеся в речи с большой степенью регулярности. В эту группу вошли следующие аналитические формы глагола (следует оговориться, что приводимый далее перечень далеко не полностью соответствует принятым в русистике взглядам, отраженным, в частности, в Грамматике-80): отрицательная форма; темпоральные формы (глагол *быть* + инфинитив; в основном так образуется будущее время); сослагательное наклонение (*б(ы)* + глагольная форма на *-л*); повелительное наклонение (*давай(те)* + инфинитив); желательное наклонение (*пусть/пускай* + личная форма глагола). В эту же группу вошли формы превосходной степени прилагательного (*самый/ая/ое* + прилагательное), формы превосходной степени наречия (типа *больше всего, быстрее всех, меньше всех*) и формы сравнительной степени прилагательного и наречия (*более* + прилагательное или наречие).

Единицы этой группы характеризуются большой регулярностью образования и силой внутренних связей. Хорошим примером этого может служить отрицательная форма глагола, которая образуется регулярно и при этом не допускает ни вставок, ни перестановок (в сочетаниях со вставкой типа *не очень хотел, не он хотел* и т.д. отрицание относится не к глаголу, а к «вставке»).

Во вторую группу вошла большая группа неаддитивных «идиоматизмов» (*грубо говоря, голову морочить, дело за вами, чё-то не так, вроде как, какого чёрта, всё что угодно, по этой причине, таким макаром, ни фига себе, задавай вопрос, вся в меня, ничего/ничё страшного, водишь машину, без проблем, в первых рядах, внештатные ситуации, ничего не поделаешь, чувство юмора, горячие точки, ничё/ничего себе* и т.д.). В отличие от единиц первой группы, они не образуются регулярно и каждый из них уникален по своей структуре. С точки зрения как действия критериев переставимости и вставимости, так характеризующих возможности изменения линейного порядка группы единиц, как и возможности грамматического словоизменения компонентов, эта группа представляет собой пеструю неструктурированную картину. Мы отдаем себе отчет в том, что выделение по меньшей мере некоторых «идиоматизмов», отнесенных к этой группе, может стать объектом дискуссии, в рамках которой главным аргументом авторов в пользу данного решения может служить лишь интуиция, а оправданием в случае доказанной неудачи – абсолютно пионерский характер работы и необходимость «с чего-то начинать».

Основным результатом проделанной работы является пробная версия частотного словаря, полученная в результате компьютерной обработки размеченного текста по программе, созданной старшим научным сотрудником лаборатории моделирования речевой деятельности СПбГУ (руководитель – проф. В.Б. Касевич) канд. биол. наук А.В. Венцовым. Этот словарь включает в себя всего 4778 расположенных в алфавитном порядке и снабженных индексом частотности единиц.

Полученная версия частотного словаря, разумеется, не может претендовать на какую-либо даже весьма условную полноту охвата реального разговорного словника современной русской речи – как из-за ограниченного объема анализируемого материала, так и из-за достаточно сильной тематической привязки рассмотренного текста к событиям, происходившим в реалити-шоу *Большой Брат*. Однако мы надеемся, что в качестве некоего первого опыта, словарь может оказаться полезным прежде всего с методической точки зрения, а принципы, положенные в основу его составления, могут стать подготовительным материалом для дальнейших исследований в этой области. Ближайшей задачей для нас является составление подобного же словаря (а затем и получение «объединенного» частотного словаря) и для записей телепрограммы *Культурная революция*.

Ниже в качестве образца приведен небольшой фрагмент полученного частотного словаря: в столбцах цифр направо от словарных единиц указаны индексы частотности, межсловное нижнее подчеркивание соединяет в единое целое «идиоматизмы», цифра, соединенная со словарной единицей нижним подчеркиванием, соответствует грамматическому / лексическому маркеру, маркирующему, падеж (в случае омонимии падежных форм) а также некоторые особенности глагольных форм. Знаком + обозначен ударный гласный.

вошли	1	всё_ещё	1
вперёд	4	все_за_одного	1
впереди	1	всё_необходимое_3	3
впрочем	1	всё_нормально	2
вращаться	1	всё_равно	11
вре+менном_5	1	всё_хорошо	1
времени_1	7	всё_что_угодно	1
время	13	всё-всё-всё	1
время_3	8	всегда	12
врешь	5	всего_1	1
вроде_как	1	всего_2	1
вру	1	всего_лишь	1
вряд_ли	1	всего_хорошего	1
все	35	всём	2
всё	78	всем_02	17
все_03	5	всем_4	1
всё_1	20	всеми	4
всё_2	27	всё-таки	5
всё_3	1	всё-то	1
всё_3	1	всех	9
всё_время	1	всех_01	6

Литература

1. Баскулина Ю.Н. «Составные слова» в современном русском языке (на материале Национального корпуса русского литературного языка // Материалы XXXV Международной филологической конференции. 13-18 марта 2006 г. Вып. 22. Секция общего языкознания. Ч. 2. С.17-23.
2. Венцов А.В., Касевич В.Б. Проблемы восприятия речи. М., 2003.
3. Гаспаров Б.М. Язык. Память. Образ: Лингвистика языкового существования. М., 1996.
4. Горбова Е.В., Слепокурова Н.А. и др. Предварительные результаты мониторинга современной русской устной спонтанной речи // Современная русская речь: состояние и перспективы. Часть II. СПб, 2006.
5. Земская Е.А., Китайгородская М.В., Ширяев Е.Н. Русская разговорная речь. Общие вопросы. Словообразование. Синтаксис. М., 1981.
6. Касевич В.Б. Семантика. Синтаксис. Морфология. М., 1988.
7. Мельчук И.А. Курс общей морфологии. Т. 1. М.; Вена, 1997.
8. Плунгян В.А. Общая морфология. Введение в проблематику. М., 2003.

Вариативность формантной картины гласных в разных видах речи

Санкт-Петербургский государственный университет,
г. Санкт-Петербург, Россия,
postmaster@phonetics.pu.ru

Спонтанная речь – это основное речевое поле, которое доминирует в человеческом общении. Она – многовариантна, сильно зависит от экстралингвистических факторов и, вследствие этого, обладает рядом фонетических особенностей [1]. Знание закономерностей спонтанной речи необходимо для решения прикладных задач, поэтому изучение ее особенностей стало целью фонетических исследований [2]. Изучались сама манера произнесения (весомость произнесения отдельных фонем), частотный состав речевого сигнала, длительности реализаций аллофонов фонем и пауз, изменения темпа, уровни фонации. Использовался сравнительный метод – те или иные параметры спонтанной речи сравнивались с кодифицированным литературным языком, обычно с чтением текста [1].

Целью проведенного исследования является поиск различий в разных видах речи на уровне вариативности формантных картин гласных. Задача анализа особенностей работы артикуляторного аппарата требовала разработки процедуры построения формантных картин гласных путем автоматизированной обработки аудиосигнала. Анализ существующих методов обработки речевого сигнала показал, что ни один из них не может в полной мере решить такую задачу [3, 5, 6]. Чтобы получить результат, был предложен базирующийся на частотных методах теории автоматического управления способ построения передаточных функций и частотных характеристик основных отделов речевого тракта [4]. В его основу была заложена двухзвенная динамическая модель, состоящая из голосового источника и артикуляторного аппарата, представляющих собой два последовательно соединенных динамических звена, каждое из которых обладает своими частотными свойствами [5, 6, 7, 8].

Процесс фонации рассматривается как вынужденные акустические колебания, возникающие на резонансных частотах голосового источника под действием флуктуаций иницирующего воздушного потока, а процесс артикуляции – как фильтрация выходного сигнала голосового источника (глоттальной волны) артикуляторным аппаратом [3].

Отнесение частотной избирательности, проявляющейся в процессе фонации, к свойствам голосового источника позволяет рассматривать речевой тракт как обобщенную динамическую систему, в которой выделяются два последовательно соединенных звена, каждое из которых обладает своими динамическими свойствами (Рис.1).

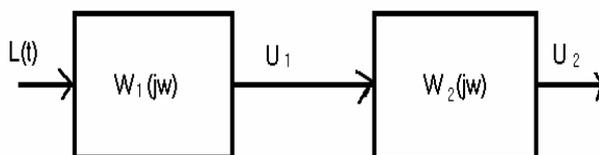


Рис.1. Динамическая система речевого тракта человека, состоящая из двух динамических звеньев

Введены обозначения:

$W_1(j\omega)$ – частотная передаточная функция голосового источника,

$W_2(j\omega)$ – частотная передаточная функция артикуляторного аппарата,

$L(t)$ – иницирующее внешнее воздействие,

$U_1(t)$ – выходной сигнал голосового источника,

$U_2(t)$ – выходной речевой сигнал.

Выработанный подход позволил обратиться к детально проработанным частотным методам теории управления как на этапе определения структуры и параметров эквивалентных передаточных функций $W_1(p)$ и $W_2(p)$, так и при синтезе с их помощью речевого сигнала [4, 7, 8].

Предложенный способ построения АЧХ артикуляторного аппарата базируется на совместной обработке статистических характеристик двух речевых файлов, в которых степень присутствия голосовой и артикуляторной составляющих заведомо различна.

Передаточные функции голосового и артикуляторного аппаратов могут быть получены путем обработки экспериментального аудиоматериала с помощью соотношений

$$|W_1(j\omega)|^2 = \frac{S_{U_{20}}(\omega)}{k_0^2 |W_{2П}(j\omega)|^2 S_L(\omega)}, \quad (1)$$

$$|W_2(j\omega)|^2 = \frac{k_0^2 S_{U_{2a}}(\omega) |W_{2П}(j\omega)|^2}{k_a^2 S_{U_{20}}(\omega)}, \quad (2)$$

где $W_1(j\omega)$ и $W_2(j\omega)$ – передаточные функции голосового и артикуляторного аппаратов соответственно,

$S_{U_{20}}(\omega)$ – спектральная плотность выходного сигнала речевого тракта, полученная путем применения процедуры определения расчета спектральной плотности ко всему достаточно длинному речевому сообщению в целом (около 100 с),

$S_{U_{2a}}(\omega)$ – спектральная плотность выходного сигнала речевого тракта, полученная в результате обработки отдельного сегмента сообщения – гласной либо ее части (около 30 мс),

k_0 и k_a – масштабные коэффициенты усиления при обработке всего речевого сообщения и его отдельного сектора соответственно,

$S_L(\omega)$ – спектральная плотность входного воздействия на речевой тракт, заданного в форме «белого шума»,

$W_{2П}(j\omega)$ – сглаженная при статистической обработке всего сообщения в целом передаточная функция артикуляторного аппарата.

Эти соотношения позволяют, используя стандартные процедуры определения спектральных плотностей с помощью авторегрессионной модели на основе минимизации ошибки предсказания сигнала, построить амплитудно-частотные характеристики (АЧХ) голосового и артикуляторного аппаратов [9]. Такие АЧХ описывают динамику системы и, в свою очередь, могут быть использованы как исходный материал при решении задачи создания аналитических моделей этих отделов речевого тракта [8].

В результате процедуры подготовки был выделен набор частот реальных резонансных пиков АЧХ артикуляторного аппарата. Далее проводилась процедура преобразования набора резонансных частот АЧХ артикуляторного аппарата в присущий фонетическим исследованиям набор формантных частот. В процедуре оценивается соответствие каждой из резонансных частот априорно заданным частотным диапазонам формант. На выходе процедуры получен для каждого изучаемого сегмента набор идентифицированных частот формант. Таким образом, методика дает возможность получить формантную картину, в которой для каждой формантной составляющей найдены ее частота и относительная мощность.

Высокая степень соответствия этой информации каждому временному срезу достигается выбором ширины окна обработки и уровнем дискретизации первичного аудиоматериала.

Достоинством описанного метода является возможность последовательного построения АЧХ артикуляторного аппарата в любой временной момент на протяжении реализации аллофона гласной. Получаемый информационный массив, представляющий собой описание трансформации эквивалентной АЧХ артикуляторного аппарата во времени, демонстрирует вариативность формантной картины гласного как по спектральному составу, так и по сравнительной интенсивности отдельных составляющих. Для восприятия речи слушателем

равно важны оба эти фактора. Частотный набор формант служит кодом, по которому слушатель распознает гласную [2], но для его правильного восприятия необходимо, чтобы каждая формантная составляющая имела в речевой посылке определенную интенсивность, поэтому также важны энергетические соотношения между формантами [10].

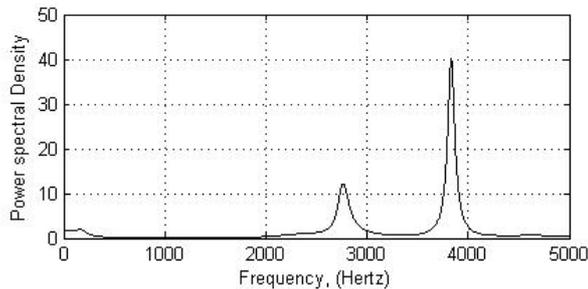


Рис. 2. Пример реализации АЧХ артикуляторного аппарата. Доминирующая F3. Реализация аллофона [e].

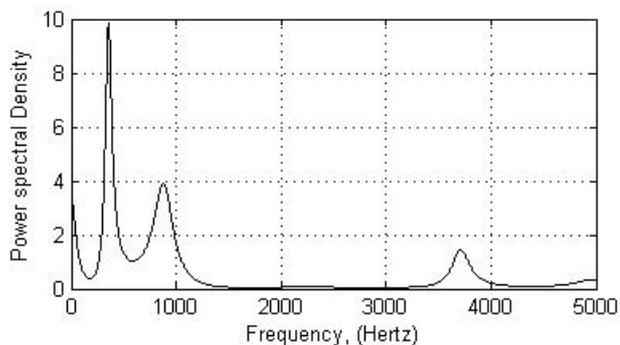


Рис. 3. Пример реализации логарифмической АЧХ артикуляторного аппарата. Доминирующая F1. Реализация аллофона [u].

Накапливаемый при поточечной обработке сегмента реализации гласной массив АЧХ содержит информацию, требуемую для изучения сложных фонетических явлений, что открывает возможность установления соответствия между перцептивно выявленными особенностями речи и объективными изменениями формантной картины гласных.

Был проведен сравнительный анализ реализаций гласных в разных видах речи, в частности, сравнение спонтанной речи и чтения текста.

Материал для исследования отбирался из фонетической базы данных, разработанной в рамках проекта РФФИ 04-06-80111 «Спонтанная речь как источник изменения произносительной нормы», проведенного на кафедре фонетики и методики преподавания иностранных языков филологического факультета СПбГУ под руководством проф. Л.В. Бондарко и проф. П.А. Скредина [11]. В ней приведены аудиозаписи спонтанной речи и чтения текста группой дикторов. В разных видах речи текст речевых посылок идентичен.

Методика проведения сравнительного анализа разных видов речи базируется на применении процедуры определения формантных картин гласных и предусматривает следующие этапы:

1. Отбор фонетического материала. Из аудио и текстовых материалов отобраны и сгруппированы в массивы реализации аллофонов с соответствующим указанием перцептивной транскрипции. Иерархически группировка проводилась по дикторам, по типам аллофонов.
2. Из аудиоматериалов выделены соответствующие отобраным аллофонам сегменты.
3. С использованием описанной процедуры для каждого сегмента, для каждого момента времени (для каждого шага окна в 512 отсчетов) построена АЧХ артикуляторного аппарата, а по ней определялась формантная картина.

4. На каждой формантной картине выделена наиболее интенсивная доминирующая в сигнале форманта, которой в основном определяется в этот момент речевой сигнал. На АЧХ ей соответствует самый интенсивный резонансный пик. Критерием самой интенсивной форманты выбрана площадь резонансного пика.

5. На протяжении сегмента проведена относительная весомость каждой доминирующей форманты, а также характер возможного чередования доминирующих формант. При оценке весомости каждой из формант за событие принималась регистрация форманты в качестве доминирующей в данном цикле расчета. По каждому аллофону определялась весомость I -ой (I – номер форманты) доминирующей форманты по формуле

$$FI_k = \frac{N_{ik}}{N_k} 100\%, \quad (3)$$

где N_{ik} - число циклов расчета в K -ом опыте (в K -ом сегменте аудиоматериала), в котором зарегистрирована доминанта I -ой форманты,

N_k – общее число циклов расчетов в K -ом опыте;

Весомость I -ой доминирующей форманты для группы, соответствующей конкретному аллофону, определяем по формуле (4), где m – общее число опытов в группе данного аллофона.

$$F_i = \frac{\sum_{k=1}^{k=m} N_{ik}}{\sum_{k=1}^{k=m} N_k}, \quad (4)$$

6. Проведено сравнение наборов и весомостей доминирующих формант для текстуально сопряженных реализаций гласных из разных видов речи, причем за базовую принималась группа реализаций определенного аллофона предварительно отобранная в режиме чтения текста.

По описанной методике для каждого опыта (для каждого отобранного сегмента реализации аллофона), для каждого момента времени с дискретностью ширины окна обработки материала проведен цикл расчетов по определению доминирующей форманты. Каждому выбранному из режима чтения текста аллофону в таблицах сопоставлен текстуально сопряженный аллофон из режима СР. Обработка проведена для вариантов мужского и женского голоса.

По результатам сравнения сделаны выводы о характере воспроизведения аллофона диктором в разных видах речи:

1. Рассмотрение массива амплитудно-частотных характеристик артикуляторного аппарата подтверждает высокую вариативность формантной картины, ее специфичность для каждого аллофона, для каждого вида речи и дает возможность проследить изменение каждой форманты по частоте и интенсивности.

2. Полученная стабильная объективная информация о структуре и трансформации формантной картины подтверждает положение о множественности способов реализации диктором каждого аллофона гласных.

3. Поаллофонное сравнение реализаций гласных позволяет определить в формантных картинах особенности, характерные для разных видов речи.

4. Разработанная методика выделения доминирующей – наиболее интенсивной форманты, как признака, характеризующего особенности работы артикуляторного аппарата, дает возможность описать характер вариации формантной картины. В качестве доминирующей форманты, которой в данный момент в основном определяется энергия акустической речевой посылки, может выступать каждая из трех основных формант F1, F2, F3.

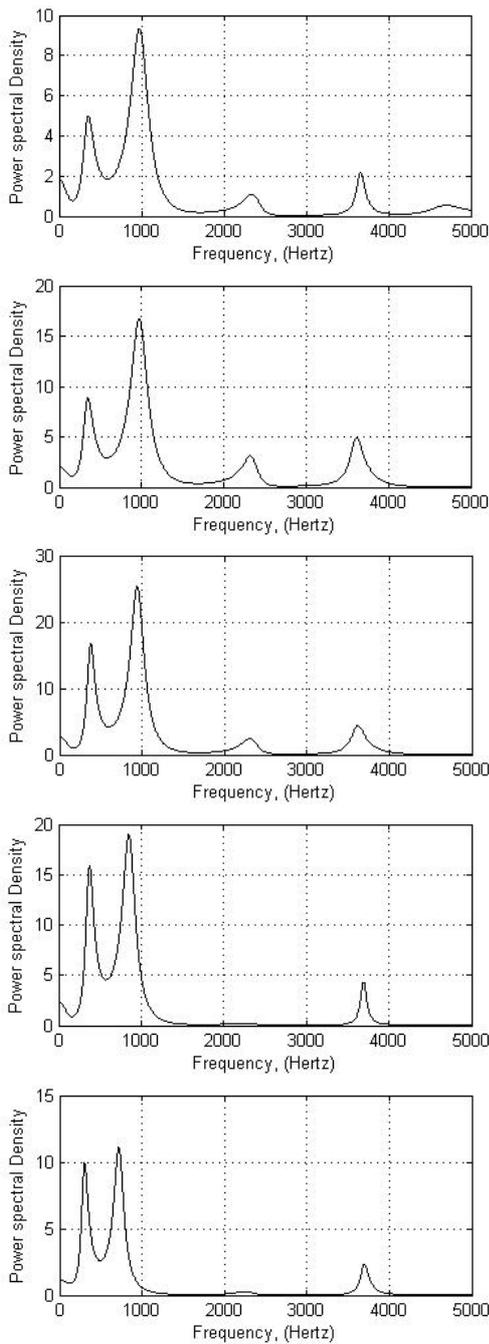


Рис. 4. Последовательная обработка окном гласного. Реализация аллофона [o]. Спонтанная речь.

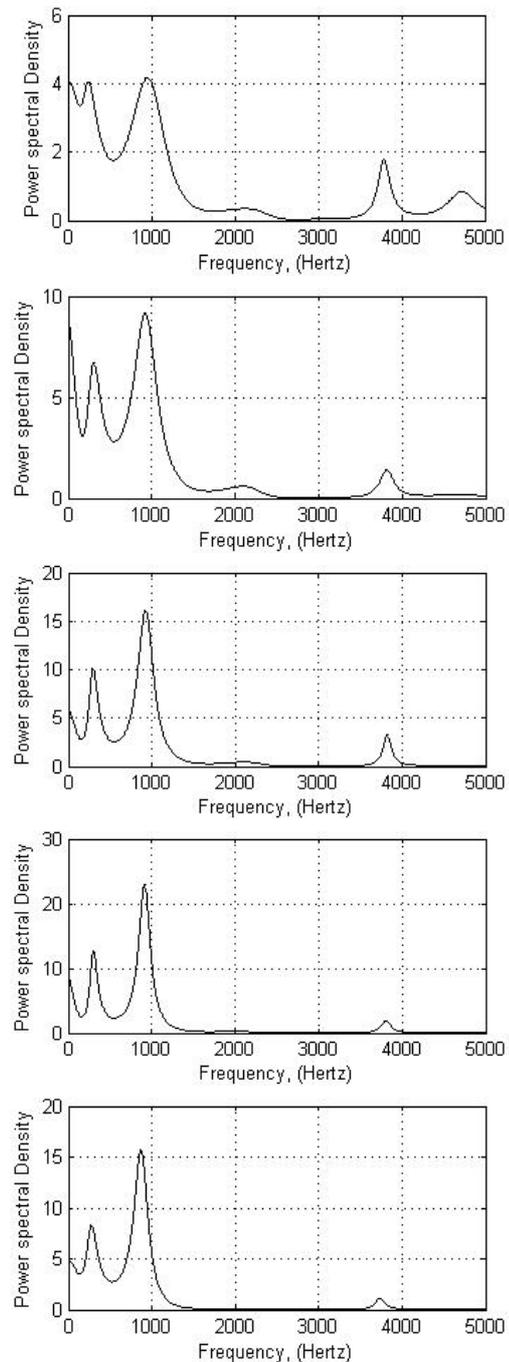


Рис. 5. Последовательная обработка окном гласного. Реализация аллофона [o]. Диктор-мужчина. Чтение текста.

5. Сравнительная оценка показывает, что весомость каждой из основных формант как доминирующей различна для разных видов речи. Обобщение результатов позволяет считать:

5.1. Чаще всего в роли доминирующей в большинстве аллофонов выступает F2. Ее весомость составляет от 50% до 100%.

5.2. Весомость F1 в качестве доминирующей форманты в спонтанной речи почти во всех группах реализаций аллофонов в 2-5 раз выше, чем при чтении текста. Исключение составляет группа реализаций аллофона [ɣ], в которой отмечена обратная картина, что возможно определяется предельной краткостью и высокой интенсивностью сравниваемых реализаций аллофонов в спонтанной речи, а также группа реализаций аллофона [o] (диктор – женщина), в которой практически нет отличий между видами речи в картине выбора доминирующей форманты. Это может быть объяснено индивидуальной манерой речи, а также

тем, что аллофон [o] встречается как правило в ударной позиции в русском языке и подвергается меньшей количественной и качественной редукции, по сравнению с безударными аллофонами.

5.3. Форманты F2 и F3 значительно чаще доминируют в режиме чтения текста, тогда как в режиме спонтанной речи в ряде случаев доминирующая F3 вообще не регистрируется. При этом нужно особо отметить, что такой вывод сделан после того, как было выявлено и исключено из рассмотрения другое явление, близкое по эффекту, но имеющее иную фонетическую природу, – разовое проявление доминирующей F3 в самом начале и в конце сегмента гласной, на границе с согласной, обусловленное коартикуляцией.

5.4. Помимо трех основных формант в структуре формантной картины обычно присутствует F4, составляющая которой иногда сравнима с первыми тремя формантами. Зарегистрирован случай доминирования F4 в реализациях аллофона [u] (мужской голос, режим CP).

5.5. Частотные диапазоны существования доминирующих формант в разных видах речи практически не отличаются.

6. Полученные результаты хорошо согласуются с положением «принципа экономии усилий» [1]. Чтению текста диктором, как режиму, требующему большего контроля и, соответственно, более равномерного распределения энергии, чаще соответствует формантная картина с доминированием F2 и F3, в то время как спонтанной речи, режиму более свободному и рациональному, сравнительно более низкочастотная формантная картина с высокой весомостью доминирования F1.

7. Сравнением частот доминирующих формант гласных в разных видах речи установлено, что ударные гласные в спонтанной речи более энергетически насыщены, чем при чтении – доминирующие форманты в них по частоте превосходят доминирующие форманты соответствующих гласных при чтении. Соответственно, безударные гласные в спонтанной речи менее энергетически насыщены, чем при чтении, т.е. их доминирующие форманты лежат ниже, чем доминирующие форманты соответствующих гласных при чтении.

Полученные результаты могут найти применение при решении прикладных фонетических задач распознавания и высококачественного синтеза речи.

Литература

1. Фонетика спонтанной речи. Под ред. Светозаровой Н.Д. СПб., 1988, 245 с.
2. Бондарко Л.В. Фонетика современного русского языка. С-Петербург, 1998.
3. Евдокимова В.В. Возможности методов моделирования голосового источника // Вестник Санкт-Петербургского университета. Выпуск. 2, часть. 1, серия 9, СПб., 2007, с. 58-63.
4. Бесекерский В.А., Попов Е.П. Теория систем автоматического регулирования. М., Наука, 1972.
5. Фант Г. Акустическая теория речеобразования. Москва, 1964.
6. Fant G. The voice source in the connected speech // Speech Communication, № 22, 1997, p. 125-139.
7. Евдокимова В.В. Способ построения динамической модели речевого тракта. Фонетический лицей, выпуск 2, СПбГУ, 2006, с. 33-40.
8. Евдокимова В.В. Динамические модели составных частей речевого тракта. Фонетический лицей, выпуск 2, СПбГУ, 2006, с. 22-32.
9. Сергиенко А.Б. Цифровая обработка сигналов. М., 2003. с.
10. Галунов В.А., Станкевич С.А. О модели речеобразования // Сборник докладов VIII Всесоюзной акустической конференции. – Л. 1973, с. 42-44.
11. Bondarko L.V., Volskaya N.B., Tananaiko S.O., Vasilieva L.A. Phonetic properties of Russian Spontaneous Speech // Proceedings of the 15th ICPHS. Barcelona, 3-9 August 2003, p. 2973-2976.

Обобщённый фонетический алфавит

ООО «Одитек»,
Г. Санкт-Петербург, Россия,
work@auditech.ru

В процессе работы над европейскими проектами создания баз транскрибированных данных “SpeechDat”, “Speecon”, “LC-STAR” и разработки систем распознавания речи компания «Одитек» столкнулась с недостатками существующих фонетических алфавитов, таких как алфавит МФА [1] или SAMPA [2].

Выбор фонетического алфавита является базовой задачей при создании речевых технологий как для систем распознавания речи, так и для систем синтеза. Фонетический алфавит должен обладать множеством свойств, а именно:

- быть полным, то есть подходящим для описания любого языка;
- адекватно отражать явления акустического уровня;
- быть удобным для ввода или редактирования в стандартных редакторах;
- быть гибким, то есть обладать набором правил для создания специализированного алфавита конкретного языка, трифонного алфавита и т.п.

Кроме того, алфавит в современных технологиях должен обеспечивать возможность:

- суммировать/разделять различные базы транскрибированных аудио данных;
- интегрировать различные системы акустической сегментации;
- сравнивать характеристики автоматических транскрипторов;
- сравнивать характеристики различных систем акустической сегментации.

Основные недостатки алфавита МФА:

- Ориентирован на обозначения фонем только одного языка, то есть, в частности, не различает взрывные переднеязычные согласные по месту образования. Например, русский зубной и английский альвеолярный (*t*) обозначаются одним значком. В то же время фрикативные переднеязычные (*s*) представлены разными знаками, видимо, на том основании, что есть языки, где они представляют разные фонемы. Отсюда вывод: МФА предназначен для описания фонетической системы одного языка — произвольного, но одного.
- Само описание звуков в МФА при этом является артикуляторным (а следовательно, и акустическим, поскольку акустические фонетические особенности звука — следствие его артикуляции), но не описанием с точки зрения системы языка. Например, для русских фонем, противопоставленных по твердости/мягкости, отдельных знаков в МФА нет.

Таким образом, имеется внутреннее системное противоречие: набор символов МФА представляет собой фонемно ориентированное описание звуков человеческой речи, в то время как классификация звуков в МФА базируется на артикуляторных основаниях, безотносительно к фонетическим системам языков. Алфавит SAMPA является машинно-читаемой формой алфавита МФА, а следовательно, обладает теми же недостатками.

В предлагаемом алфавите, ориентированном на звуки речи, и прежде всего на их акустические и артикуляторные особенности, целесообразно оказалось ввести принцип представления **качества** звука особым базовым алфавитным знаком, допускающим комбинации с диакритическими и иными вспомогательными элементами, ортогонально

представляющими некие **дополнительные артикуляторные признаки**, безотносительно к тому, могут ли эти признаки быть дифференциальными признаками фонем в каком-либо языке:

- палатализацию, двухфокусность, лабиализацию, инспираторность («щелчковость»), долготу, слабошумность и пр. для согласных;
- огубленность, степень редукции, влияние фонетического контекста, назализацию, инспираторность, шепотность, продвинутость и пр. для гласных.

В предлагаемом обобщенном фонетическом алфавите вводятся отдельные базовые знаки, представляющие качественно разные звуки с учетом следующих артикуляторных признаков:

- для согласных:
 - способ артикуляции (смычные, щелевые, аппроксиманты, аффрикаты);
 - активный действующий орган и место образования (губно-губные, губно-зубные, переднеязычные зубные, переднеязычные альвеолярные, переднеязычные постальвеолярные, ретрофлексные, среднеязычные, заднеязычные, увулярные, фарингальные, эпиглоттальные, глоттальные);
 - работа голосовых связок (звонкие, глухие);
- для гласных:
 - ряд (передний, средний, задний);
 - подъем (4 градации).

Прочие артикуляторно-акустические признаки рассматриваются как дополнительные и обозначаются вспомогательными элементами.

Представление набора фонем конкретного языка или диалекта в обозначениях предлагаемого обобщенного алфавита может быть построено как подмножество знаков алфавита, снабженных вспомогательными элементами, соответствующими дифференциальным признакам данной фонемы в данном конкретном языке.

Набор аллофонов фонемы будет строиться как упорядоченное множество аллофонов по числу фонетических позиций, специфических для данной фонемы, определяемых с учетом влияния фонетического контекста и степени редукции.

[1] IPA, <http://www.arts.gla.ac.uk/IPA/ipa.html>

[2] SAMPA, <http://www.phon.ucl.ac.uk/home/sampa/index.html>

Б.М. Лобанов, Л.И. Цирульник

Моделирование внутрисловных и межсловных фонетико-акустических явлений полного и разговорного стилей речи в системе синтеза речи по тексту «Мультифон»

*Объединённый институт проблем информатики НАН Беларуси,
г. Минск, Беларусь,
lobanov@newman.bas-net.by , l.tsirulnik@newman.bas-net.by*

В статье приводятся результаты исследований внутрисловных и межсловных стыковых явлений в естественной русской речи на фонетическом и акустическом уровнях. Предлагаются модернизированные алгоритмы преобразования «буква-фонема» для случаев полного и разговорного стилей слитой речи. Исследуются акустические характеристики фонемных сочетаний на стыках слов в слитной речи, и предлагается расширенный набор аллофонов и мультифонов, используемый для моделирования акустических явлений в синтезаторе речи по тексту «Мультифон».

Введение

Современные требования, предъявляемые к синтезаторам речи по тексту, делают актуальной задачу высококачественного воспроизведения слитной речи не только для полного стиля речи, но и для стиля, близкого к разговорному. Одной из особенностей слитной речи является наличие специфических явлений на стыках слов внутри синтагмы, иногда существенно отличающихся от подобных внутрисловных явлений. Эти явления проявляются на фонетическом уровне (особенности преобразования «буква-фонема») и на акустическом (особенности преобразования «фонема-аллофон», или «фонема-мультифон»).

Внутрисловные закономерности изменения звуков русской речи к настоящему времени изучены достаточно подробно как на фонетическом, так и на акустическом уровне [1-4]. Построенные на их основе правила преобразования «буква-фонема» и «фонема-аллофон» успешно используются в существующих системах синтеза речи по тексту.

Фундаментальным исследованиям фонетики слитной разговорной речи, в том числе стыковых явлений, посвящена монография [5]. Однако приведенные в ней правила часто являются противоречивыми и неоднозначными. В данной работе делается попытка систематизации некоторых из этих правил, показаны особенности фонетических и акустических явлений как на стыках слов, так и внутри слов для разговорного стиля речи. Разработанные правила реализованы в системе синтеза речи по тексту «Мультифон» [6], соответственно, в блоках преобразования «буква-фонема» и «фонема-аллофон».

1. Моделирование внутрисловного преобразования «буква-фонема»

Предварительно разбитый на синтагмы и нормализованный орфографический текст (расшифрованы аббревиатуры, преобразованы к текстовому виду числа и т.д.) поступает на вход алгоритма преобразования «буква-фонема» (ПБФ). Результатом работы ПБФ являются цепочки фонем, отображающих каждую синтагму входного текста. Используется общепринятый набор 42-х фонем русской речи, для обозначения которых приняты символы, показанные в таблице 1.

Таблица 1. Обозначение фонем русской речи

Обозначение	Значение	Обозначение	Значение	Обозначение	Значение	Обозначение	Значение
B	[б]	J'	[й]	R	[р]	CH'	[ч]
B'	[бь]	K	[к]	R'	[рь]	SH	[ш]
V	[в]	K'	[кь]	S	[с]	SH'	[щ]
V'	[вь]	L	[л]	S'	[сь]	A	[а]
G	[г]	L'	[ль]	T	[т]	O	[о]
G'	[гь]	M	[м]	T'	[ть]	U	[у]
D	[д]	M'	[мь]	F	[ф]	E	[э]
D'	[дь]	N	[н]	F'	[фь]	Y	[ы]
ZH	[ж]	N'	[нь]	X	[х]	I	[и]
Z	[з]	P	[п]	X'	[хь]		
Z'	[зь]	P'	[пь]	C	[ц]		

Алгоритм ПБФ опирается на известные правила внутрисловного преобразования букв-фонем [2], в которых для каждой буквы учитывается её левое и правое окружение. Слова, не подчиняющиеся регулярным правилам преобразования, и их фонемная транскрипция содержатся в списке исключений.

Алгоритм ПБФ вначале проверяет каждое слово поступившей синтагмы на наличие в списке исключений и заменяет, в случае нахождения, на соответствующую цепочку фонем. Если же слово в списке не найдено, к нему применяются регулярные правила ПБФ. При этом алгоритм учитывает категории фонем по месту и способу образования, показанные в таблице 2.

Таблица 2. Категории фонем русской речи по месту и способу образования

Согласные											Гласные				
Способ образования		Глухие			Звонкие						Огубленные	Неогубленные	Способ		Место
		Взрывные	Аффрикаты	Щелевые	Взрывные	Щелевые	Дрожащие	Носовые	Боковые	Плавающие					
Место образования	Мягкие	K'	~	X'	G'	~	~	~	~	~	J'	~	I	Высокое	Переднее
	Твёрдые	K	~	X	G	~	~	~	~	~	~				
Альвеолярные	Мягкие	~	Ch'	Sh'	~	~	R'	~	~	~	~	E	Низкое	Переднее	
	Твёрдые	~	~	Sh	~	Zh	R	~	~	~					
Зубные	Мягкие	T'	~	S'	D'	Z'	~	N'	L'	~	U	Y	Высокое	Заднее	
	Твёрдые	T	C	S	D	Z	~	N	L	~					
Губные	Мягкие	P'	~	F'	B'	V'	~	M'	~	~	O	A	Низкое	Заднее	
	Твёрдые	P	~	F	B	V	~	M	~	~					

2. Правила межсловного преобразования «буква-фонема» для полного стиля речи

2.1. Правила ПБФ на стыке служебного и знаменательного слов

Объединение 2-х или 3-х орфографических слов в одно фонетическое слово осуществляется путём присоединения служебных (предлоги и частицы) к знаменательным словам. При этом предлоги {без, безо, близ, в, во, вне, для, до, за, из, изо, к, ко, меж, на, над, надо, о, об, обо, от, ото, по, под, подо, пред, предо, при, про, с, со, у, чрез, ... } и частица {не} присоединяются к последующему слову, а частицы: {бы, де, ли, же, -то, -ка, -либо, -нибудь, -таки, } - к предшествующему слову.

Ниже рассматриваются правила межсловного преобразования для букв, разделённых пробелом, на стыке служебного и знаменательного слов внутри фонетического слова.

2.1.1. Последняя буква первого слова

а) гласная:

– после объединения полностью сохраняются внутрисловные правила.

Примеры: *по дороге* [PADAROG'Е], *на столе* [NASTAL'Е].

б) согласная:

– внутрисловные правила ассимиляции по глухости-звонкости сохраняются полностью.

Примеры: *в саду* [FSADU], *в лесу* [VL'ESU], *с дороги* [ZDAROG'І], *с Васей* [SVAS'EJ'].

– в отличие от внутрисловных правил, твёрдая согласная не переходит в мягкую фонему перед гласными Я, Ё, Ю, Е, И.

Примеры: *под ёлкой* [PADJ'OLKAJ'], но: *потёмки* [PAT'OMK'І].

– в отличие от внутрисловных правил, отсутствует ассимиляция с последующей мягкой согласной, одинаковой по месту образования.

Примеры: *под сеном* [PAT S'ENAM], *без деда* [B'EZD'EDA], но: *везде* [V'EZ'D'E].

2.1.2. Первая буква второго слова

а) гласная:

– внутрисловные правила в сохраняются, но не полностью.

– в отличие от внутрисловных правил, О и И после Ъ не переходят в сочетания фонем [J'О] и [J'І] соответственно. Гласная О переходит в [О] в ударной позиции и в [А] в безударной, гласная И переходит в [І].

Примеры: *иль осень наступила* [IL' OS'EN' ...], *иль иволга поёт* [IL' IVOLGA ...], но *чьи* [CH'J'І], *бульон* [BUL'J'ON].

– гласная И после всех согласных, кроме Ё, переходит в фонему [Y].

Примеры: *под ивой* [PADYVAJ'], *с Иваном* [S YVANOM], но: *ведь Иван* [VED'IVAN].

б) согласная:

– внутрисловные правила сохраняются полностью.

2.2. Букво-фонемные правила на стыке фонетических слов

2.2.1. Последняя буква первого слова

а) гласная:

– внутрисловные правила сохраняются полностью.

б) согласная:

– внутрисловные правила ассимиляции по глухости-звонкости сохраняются, но не полностью.

– озвончение глухих согласных П(ь), Т(ь), К(ь), Ф(ь), С(ь), Ш(ь), Щ, Ц, Ч(ь) происходит, когда первой буквой второго слова являются Б, Д, Г, З или Ж (как и для внутрисловного преобразования). Особенностью реализации межсловной ассимиляции в этом случае является

появление «неканонических» фонем русской речи, возможных только на стыках слов при озвончении согласных Ц, Ц, Ч, а именно: Ц > [ZH'], Ц > [DZ], Ч > [DZH']

Примеры: борц горячий [BORZH' GAR'ACH'U'], заяц беляк [ZAJADZ B'EL'AK], дочь гуляет [DODZH' GUL'AJ'ET].

– в отличие от внутрисловных правил ассимиляции, звонкие согласные Б(ь), В(ь), Г, Д(ь), Ж(ь), З(ь) не оглушаются только в тех случаях, когда первой буквой второго слова являются Б, Д, Г, З или Ж.

Примеры: дуб зелёный [DUBZ'EL'ONYJ'], бровь Зои [BROV'ZOI]. Но: бровь Нины [BROF'N'INY].

– в отличие от внутрисловных правил твёрдая согласная не переходит в мягкую фонему перед гласными Я, Ё, Ю, Е, И.

Примеры: сад ёлок [SAT J'OLAK], город Ереван [GORAT J'ER'EVAN], но: хотели [XAT'EL'I].

– в отличие от внутрисловных правил отсутствует ассимиляция с последующей мягкой согласной, одинаковой по месту образования.

Примеры: город Симбирск [GORAT S'IM'B'IRSK], но: вместе [VM'ES'T'E].

2.2.2. Первая буква второго слова

а) гласная:

– внутрисловные правила сохраняются, но не полностью.

– в отличие от внутрисловных правил, О и И после Ъ не переходят в сочетания фонем [J'O] и [J'I] соответственно. Гласная О переходит в [O] в ударной позиции и в [A] в безударной, гласная И переходит в [I].

Примеры: щадить оленя [SH'AD'IT' AL'EN'A], любить игру [L'UB'IT' IGRU].

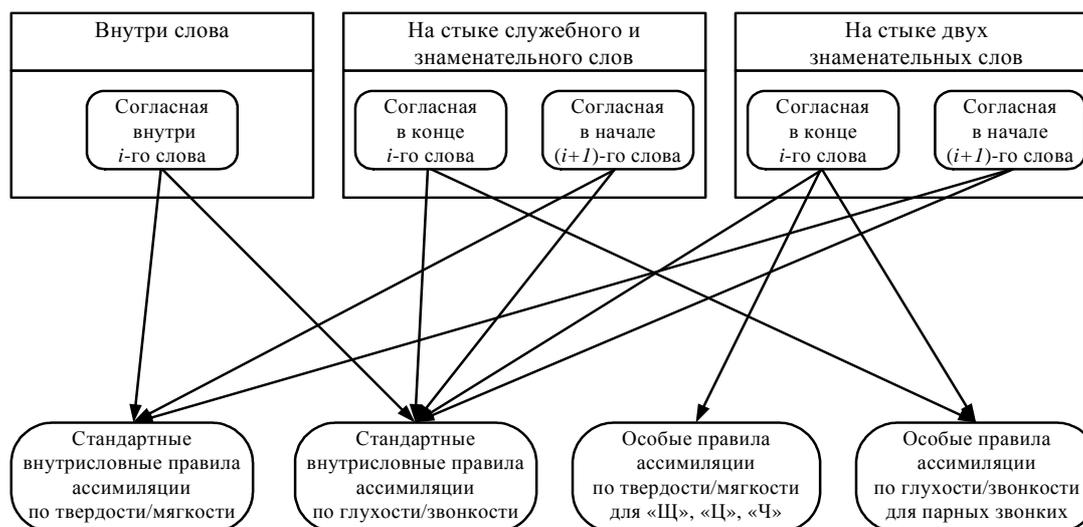
– гласная И после всех согласных, кроме Ё, переходит в фонему [Y].

Примеры: лист ивы [L'IST YVY], нос ивана [NOS YVANA]. Но: мать ивана [MAT'IVANA].

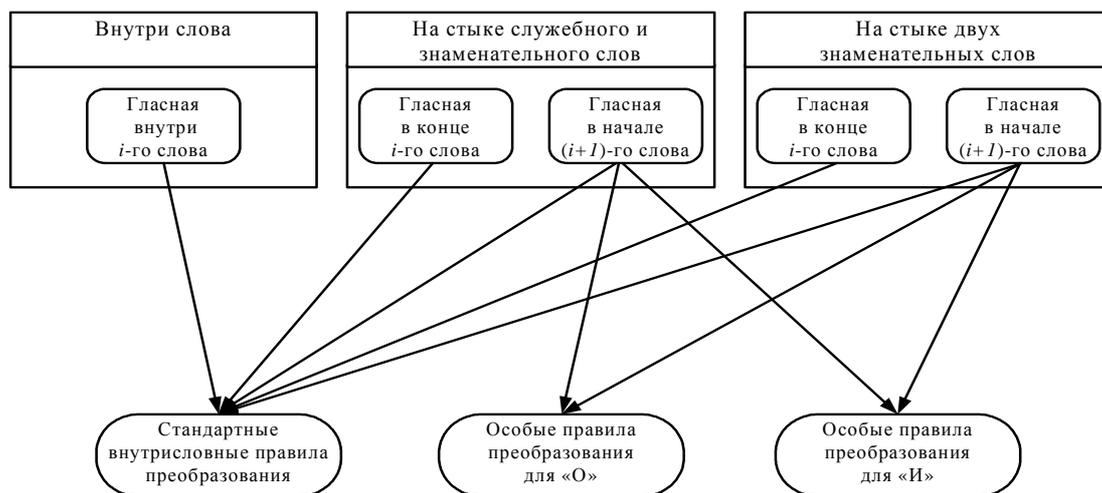
б) согласная:

– внутрисловные правила сохраняются полностью.

Описанные в разделах 2.1 и 2.2 особенности ПБФ на стыках орфографических слов внутри фонетического слова и на стыках фонетических слов внутри синтагмы показаны на рис.1(а) для согласных и на рис.1(б) для гласных.



а)



б)

Рис. 1. Особенности внутрисловных и межсловных правил преобразования «буква-фонема»: а) для согласных, б) для гласных

3. Внутрисловные правила обработки фонемного текста для разговорного стиля речи

При моделировании ПБФ для разговорного стиля речи необходимо учитывать, что некоторые фонемы, находящиеся, в основном, на стыках слов, редуцируются до полного исчезновения, некоторые ассимилируются. Закономерности таких преобразований целесообразно рассматривать не для орфографического текста, а уже для фонемного текста, полученного с помощью правил ПБФ для полного стиля речи. В данном разделе приводятся внутрисловные правила обработки фонемного текста, в следующем разделе – межсловные правила обработки фонемного текста, отражающие эти закономерности.

3.1. Полная редукция безударных гласных

Безударные гласные редуцируются до полного исчезновения, если они находятся:

а) между одинаковыми согласными

Примеры: *нанизал* [NAN'IZAL] → [NN'IZAL], *филологический* [F'ILALAG'ICH'ESK'IY] → [F'ILLAG'ICH'ESK'IY], *каким бы ни был* [KAK'IM BY N'I BYL] → [KK'IM BY N'I BYL].

б) после одной из парных по глухости-звонкости согласных и перед соответствующей парной согласной.

Примеры: *когда* [KAGDA] → [KGDA], *создать* [SAZDAT'] → [SZDAT'], *шажок* [SHAZHOK] → [SHZHOK].

в) после мягкой согласной и перед согласной.

Примеры: *лцстать* [L'ISTAT'] → [L'STAT'], *деревянный* [D'ER'EV'ANNYJ'] → [D'R'V'ANNYJ'], *полем* [POL'EM] → [POL'M], *девять* [D'EV'AT'] → [D'EV'T].

г) после сонорной согласной и перед согласной.

Примеры: *ракета* [RAK'ETA] → [RK'ETA], *кузница* [KUZ'N'ICA] → [KUZ'N'CA].

д) после фрикативной согласной и перед согласной.

Примеры: *собака* [SABAKA] → [SBAKA], *пуфик* [PUF'IK] → [PUF'K], *отеческий* [AT'ECH'ESK'IJ'] → [AT'ECH'SK'IJ']

3.2. Полная редукция звонких взрывных согласных

Звонкие согласные редуцируются до полного исчезновения, если они находятся после гласной и перед безударной гласной.

Примеры: *надал* [PADAL] → [PAAL], *погоди* [PAGAD'I] → [PAAD'I].

3.3. Полная редукция сонорных согласных

Сонорные согласные, стоящие перед \N\, редуцируются до полного исчезновения.

Примеры: *четырнадцать* [CH'ETYRNACCAT'] → [CH'ETYNACCAT'], *восемнадцать* [VAS'EMNACCAT'] → [VAS'ENACCAT']

3.4. Полная редукция согласных |V|, |V'|

Фонемы |V|, |V'|, стоящие между гласными, редуцируются до полного исчезновения.

Примеры: *совесть* [SOV'ES'T'] → [SOES'T'], *травы* [TRAVY] → [TRAY].

3.5. Полная редукция согласных |F|, |F'|

Фонемы |F|, |F'| редуцируются до полного исчезновения, если они находятся после глухой согласной.

Пример: *сфабриковать* [SFABR'IKAVAT'] → [SABR'IKAVAT'].

3.6. Полная редукция согласной |T'|

Фонема |T'|, стоящая после |S'|, редуцируется до полного исчезновения.

Примеры: *в новостях* [V NAVAS'T'AX] → [V NAVAS'AX], *жимолость* [ZHIMALAS'T'] → [ZHIMALAS'].

3.7. Полная редукция согласной |D'|

Фонема |D'|, стоящая после |Z'|, редуцируется до полного исчезновения.

Примеры: *надзиратель* [NAD'Z'IRAT'EL'] → [NAZ'IRAT'EL'], *подземелье* [PAD'Z'EM'EL'J'E] → [PAZ'EM'EL'J'E].

Описанные правила обработки фонемного текста показаны на рис.2.

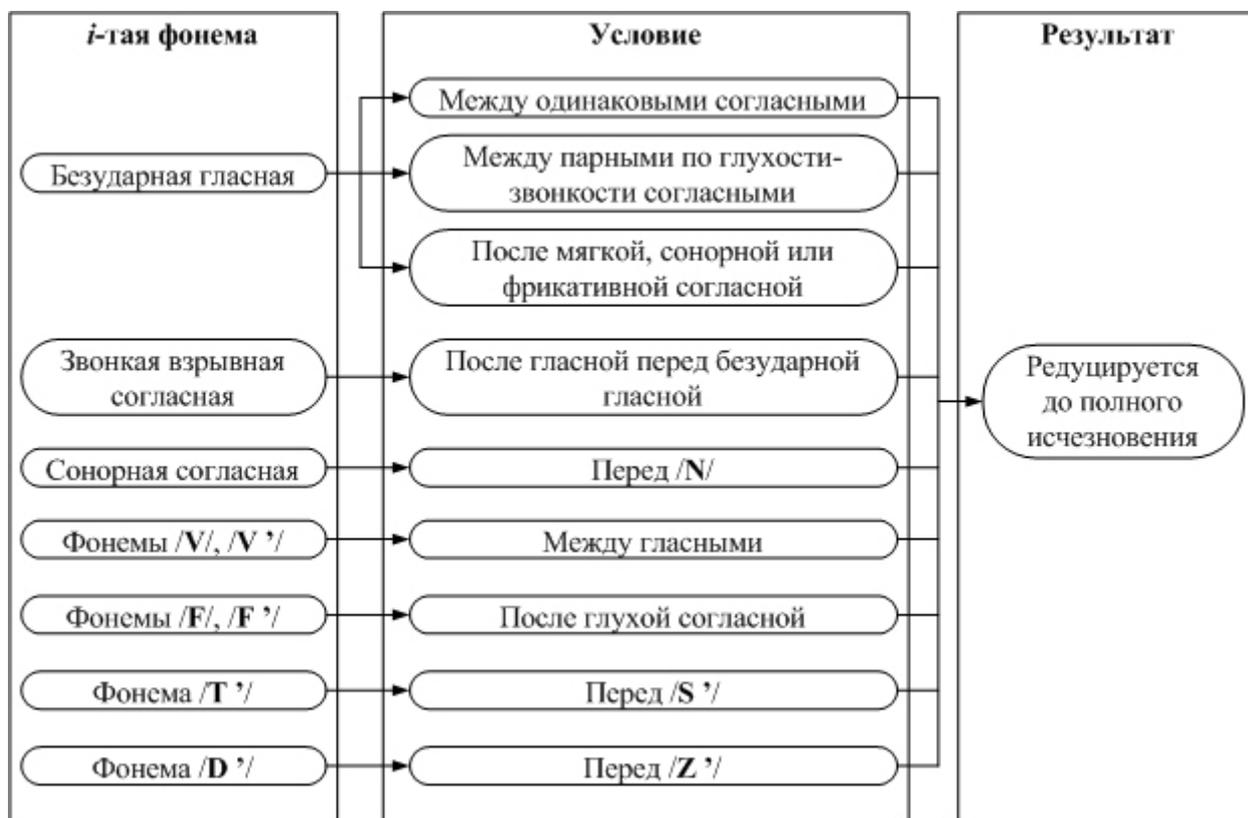


Рис. 2. Внутрисловные правила обработки текста для разговорного стиля речи

4. Межсловные правила обработки фонемного текста для разговорного стиля речи

4.1. Правила обработки фонемного текста на стыке фонетических слов

4.1.1. Последняя фонема первого слова гласная

Внутрисловные правила обработки фонемного текста выполняются полностью.

Примеры: *надо тащить* [NADA TASH'IT'] → [NAD TASH'IT'], *мясо сырое* [M'ASA SYROJ'E] → [M'AS SYROJ'E], *сначала получи* [SNACH'ALA PALUCH'I] → [SNACH'AL PALUCH'I], *после полутора часов* [POS'L'E PALUTARA CH'ASOF] → [POS'L'E PALUTAR CH'ASOF], *плохо спишь* [PLOXA SP'ISH] → [PLOX SP'ISH], *афишу видел* [AF'ISHU V'ID'EL] → [AF'ISH V'ID'EL], *Колю видел* [KOL'U V'ID'EL] → [KOL' V'ID'EL], *в поле стоит* [F POL'E STAIT] → [F POL' STAIT], *стены большие* [S'T'EP'I BAL'SHYJ'E] → [S'T'EP' BAL'SHYJ'E], *басня Крылова* [BAS'N'A KRYLOVA] → [BAS'N' KRYLOVA], *убрали дорогу* [UBRAL'I DAROGU] → [UBRAL' DAROGU].

4.1.2. Последняя фонема первого слова согласная

Внутрисловные правила обработки сохраняются лишь частично, а именно:

– выполняется правило 3.5 для фонем \F, F'\.

Примеры: *царств много* [CARSTV MNOGA] → [CARST MNOGA], *доказательств уже нет* [DAKAZAT'EL'STF UZHE N'ET] → [DAKAZAT'EL'ST UZHE N'ET].

– выполняется правило 3.6. для фонемы \T'\.

Пример: *есть порох* [J'ES'T' PORAX] → [J'ES' PORAX];

– выполняется правило 3.7. для фонемы \D'\.

Пример: *гвоздь был* [GVOZ'D' BYL] → [GVOZ' BYL].

Кроме того, добавляются следующие правила:

– фонемы \T, D\, стоящие после \S,Z\ соответственно, редуцируются до полного исчезновения.

Примеры: *хвост коровы* [XVOST KAROVY] → [XVOS KAROVY], *поезд готов* [POJ'EZD GATOF] → [POJ'EZ GATOF]

– фонемы \V', F'\, стоящие после глухой согласной, редуцируются до полного исчезновения.

Примеры: *ветвь дерева* [V'ETV' D'ER'EVA] → [V'ET D'ER'EVA], *ветвь тополя* [V'ETF' TOPOL'A] → [V'ET TOPOL'A].

– сочетание фонем \S'T'\ в конце слова переходит в фонему \SH'\, если следующее слово начинается с \CH'\.

Примеры: *есть чему* [J'ES'T' CH'EMU] → [J'ESH' CH'EMU]

– согласная \J'\ в конце слова редуцируется до полного исчезновения, если ей предшествует безударная гласная, а следующее слово начинается с любой фонемы, кроме ударной гласной.

Примеры: *красный шар* [KRASNYJ' SHAR] → [KRASNY SHAR], *хороший аргумент* [XAROSHYJ' ARGUM'ENT] → [XAROSHY ARGUM'ENT].

4.1.3. Первая фонема второго слова гласная.

Внутрисловные правила обработки выполняются полностью.

4.1.4. Первая фонема второго слова согласная.

В этом случае выполняется следующее правило:

– если в начале слова стоит сочетание фонем \J'E\ или \J'A\, причём гласная безударная, оно переходит в фонему \Y\ в случае, если первое слово заканчивается на согласную.

Примеры: *вниз его опусти* [VN'IS J'EVO APUST'I] → [VN'IS YVO APUST'I], *северней Якутии* [S'EV'ERN'EJ' J'AKUT'II] → [S'EV'ERN'EJ' YKUT'II].

Описанные правила обработки фонемного текста показаны на рис.3.

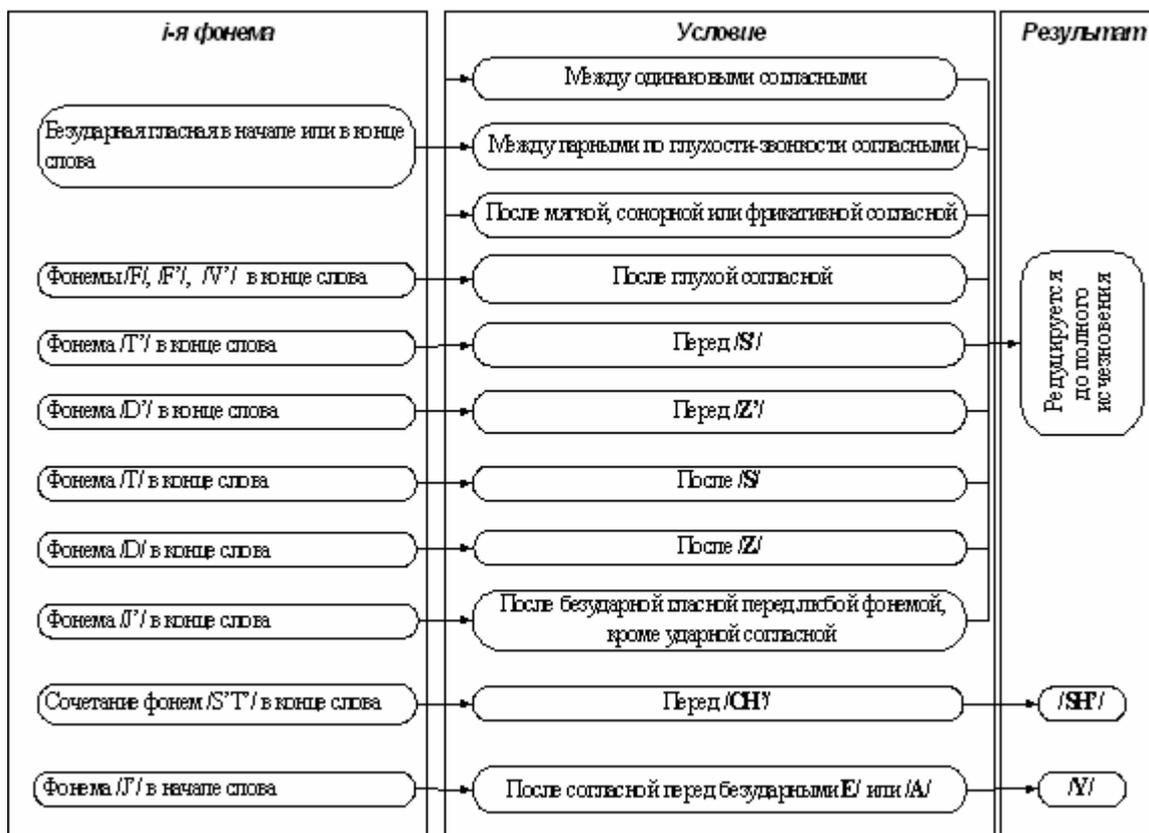


Рис. 3. Межсловные правила обработки фонемного текста для разговорного стиля речи

5. Внутрисловные и межсловные правила преобразования «фонема-аллофон»

Внутрисловные правила преобразования «фонема-аллофон» (ПФА), успешно используемые в разработанной системе синтеза речи [6], описываются приведенными ниже схемами.

Для гласных фонем в слове выделяются следующие типы позиционных аллофонов:

- V0 – полноударный, или V1 – частично-ударный;
- V2 – первый предударный;
- V3 – не первый предударный;
- V4 – первый заударный;
- V5 – не первый заударный.

Здесь V обозначает аллофон любой гласной (A, O, U, Y, E, I), V0 – полноударный аллофон гласной, употребляется в большинстве знаменательных слов, а V1 – частично-ударный, употребляется, как правило, в многосложных служебных словах, местоимениях, а также в сложных двухакцентных знаменательных словах. Употребление V1 вместо V0 может быть обусловлено особенностями акцентной структуры синтагмы и фразы.

На вход алгоритма ПФА подаётся фонемный текст синтагмы, на выходе получается соответствующий аллофонный текст. При этом для обозначения аллофонов согласных используются следующие символы: обозначение фонемы, за которым следует два индекса: первый индекс показывает левый контекст, второй – правый контекст согласной.

Правила ПФА для комбинаторных аллофонов согласных показаны на рис. 4.

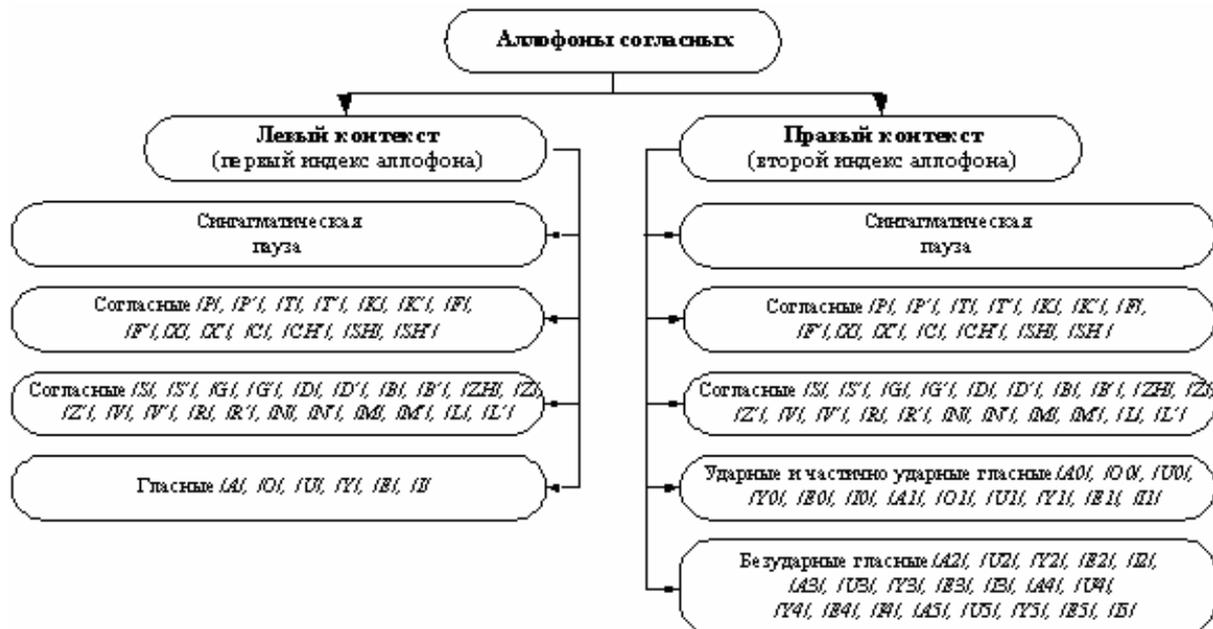


Рис. 4. Правила преобразования «фонема-аллофон» для согласных фонем

Для обозначения аллофонов гласных используется четыре символа: первый — буква — обозначает фонему, следующие три — цифры — обозначают индексы фонемы, причём первый индекс показывает позицию гласной по отношению к ударению, второй индекс — левый контекст, третий индекс — правый контекст гласной.

Правила ПФА для позиционных и комбинаторных аллофонов гласных показаны на рис. 5.

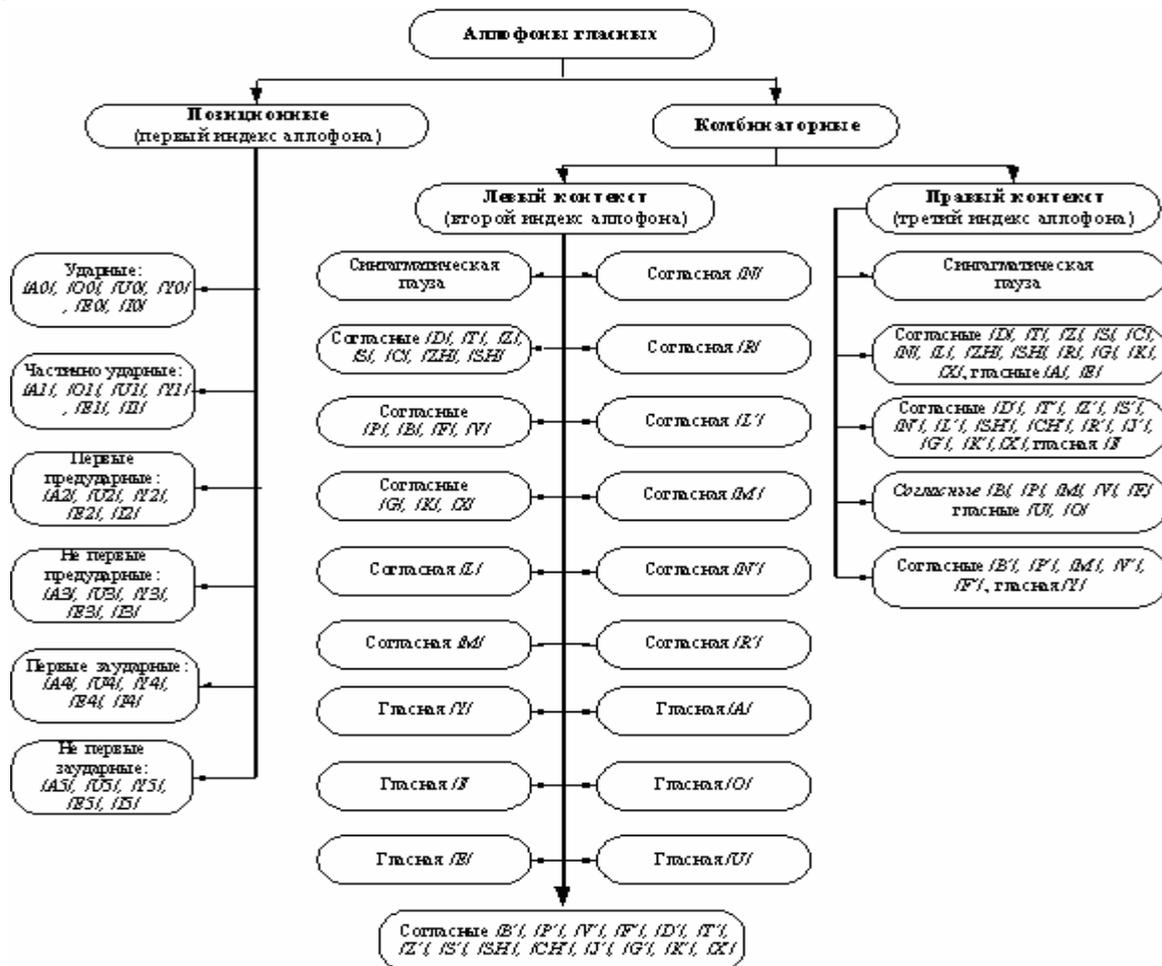


Рис. 5. Правила преобразования «фонема-аллофон» для гласных фонем

6. Сравнительная оценка акустических характеристик фонем внутри слова и на стыке слов

6.1. Гласные в сочетаниях «гласная-гласная»

В отличие от внутрисловной позиции, где сочетания «гласная-гласная» встречаются достаточно редко, в позиции на стыках слов это явление проявляется весьма часто (например, на стыках с часто употребляемыми служебными словами: *на, о, и, не* и др.).

В соответствие с вышеприведенной классификацией позиционных аллофонов гласных внутри знаменательных слов возможны следующие сочетания:

- | | |
|---------|---------|
| • V0,V4 | • V2,V1 |
| • V1,V4 | • V3,V2 |
| • V2,V0 | • V3,V3 |
| • V4,V5 | • V5,V5 |
- (1)

На стыке служебного и знаменательного слов (внутри фонетического слова) возможны следующие комбинации гласных (где символом «←» обозначается стык служебного и знаменательного слов):

- | | |
|-------------------|------------------|
| – для проклитиков | – для энклитиков |
| • V2-V0 | • V4-V5 |
| • V2-V1 | • V5-V5 |
| • V3-V2 | |
| • V3-V3 | |
- (2)

Сравнивая списки (1) и (2), можно отметить, что каждая из комбинаций гласных на стыке слов внутри фонетического слова (2) присутствуют также и в списке (1). На рис. 6(а,б) для сравнения приведено два примера реализации акустических явлений внутри и на стыке слов для комбинаций V2,V0, V2-V0 и V3,V2, V3-V2. Как видно из приведенных примеров, динамика спектральных переходов внутри слова и на стыке служебного и знаменательного слов практически совпадает, что даёт основание к использованию в таких случаях последовательности одних и тех же аллофонов.

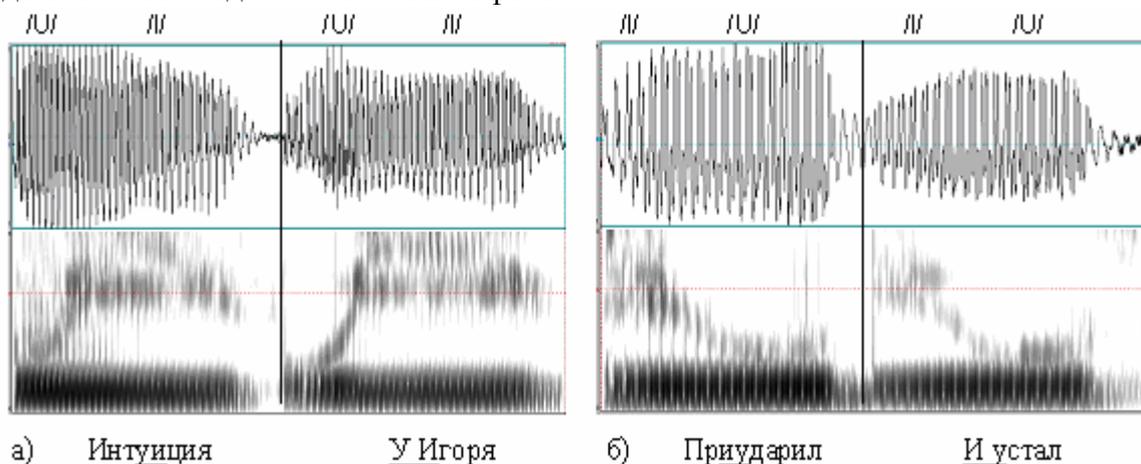


Рис. 6. Примеры реализации акустических явлений на стыке служебного и знаменательного слов а) для сочетаний гласных V2,V0; V2-V0; б) для сочетаний гласных V3,V2; V3-V2

На стыке двух знаменательных слов возможны следующие сочетания комбинаторных аллофонов гласных (где символом «_» обозначается стык двух знаменательных слов):

- | | |
|---------|---------|
| • V0_V0 | • V1_V1 |
| • V0_V1 | • V1_V2 |
| • V0_V2 | • V1_V3 |
| • V0_V3 | • V4_V0 |
| • V1_V0 | • V4_V1 |
- (3)

- V4_V2
- V4_V3
- V5_V0
- V5_V1
- V5_V2
- V5_V3

Сравнивая списки (3) и (1), необходимо отметить, что ни одна из комбинаций гласных на стыке слов из списка (3) не присутствуют в списке (1), что выдвигает требование использования при синтезе речи специфических межсловных аллофонов.

На рис. 7 (а-г) приведены примеры реализации акустических явлений на стыке слов для комбинаций V0_V0, V4_V0, V0_V2, V4_V2. При рассмотрении приведенных примеров становится очевидным, что на стыке двух знаменательных слов отсутствует даже какой-нибудь намёк на межсловную паузу, а динамика спектральных переходов вполне непрерывна. Это ещё раз подтверждает, что на стыках слов необходимо использовать специфические межсловные аллофоны гласных.

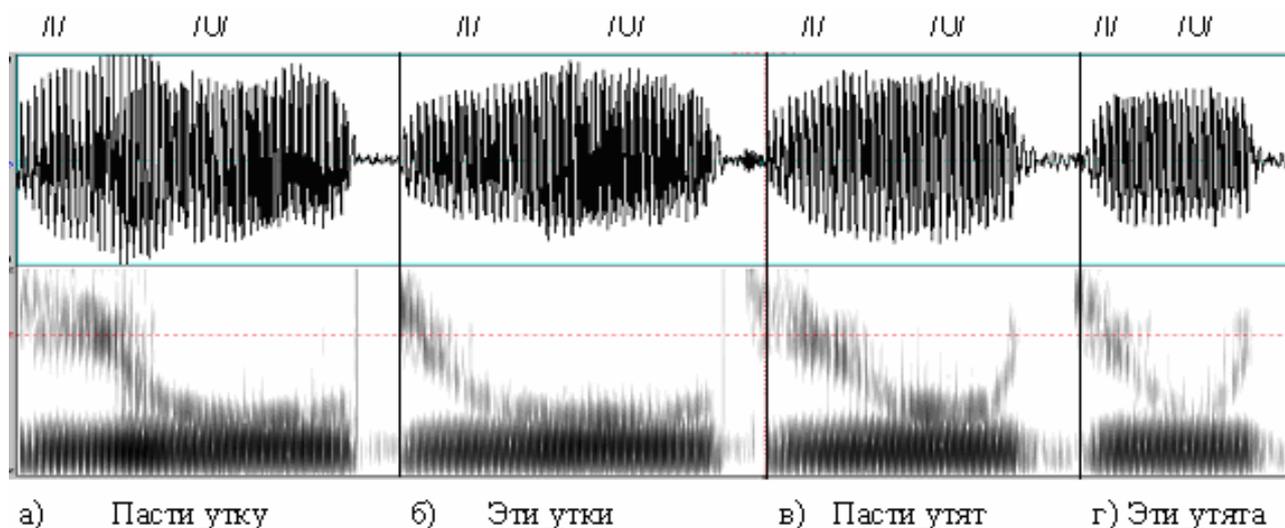


Рис. 7. Примеры реализации акустических явлений на стыке двух знаменательных слов а) для сочетаний гласных V0_V0; б) для сочетаний гласных V4_V0; в) для сочетаний гласных V0_V2; г) для сочетаний гласных V4_V2

6.2. Гласные в сочетаниях «согласная-гласная», «гласная-согласная»

На рис. 8. приведены примеры реализации акустических характеристик гласной, следующей после согласной, внутри слова (а, в) и на стыке двух знаменательных слов (б, г). Как видно из приведенных примеров, влияние предшествующей согласной на гласную внутри слова существенно сильнее, чем на стыке слов. Однако это влияние на стыках слов всё-таки присутствует, что проявляется не только на уровне сравнения спектральных характеристик, но и на уровне слухового восприятия синтезированной речи.

Подобно рассмотренному выше случаю, акустические характеристики гласной изменяются также и под воздействием последующей согласной, как внутри, так и на стыке слов. На рис. 9 приведены примеры реализации акустических явлений на стыке слов (а, в) и внутри слова (б, г). Как видно из приведенных примеров, влияние последующей согласной на гласную внутри слова существенно сильнее, чем на стыке слов.

Как показали эксперименты, при добавлении в акустическую базу синтезатора речи специфических межсловных аллофонов для сочетаний «согласная-гласная», «гласная-согласная» синтезированная речь становится более естественной, «гладкой», менее «отрывистой».

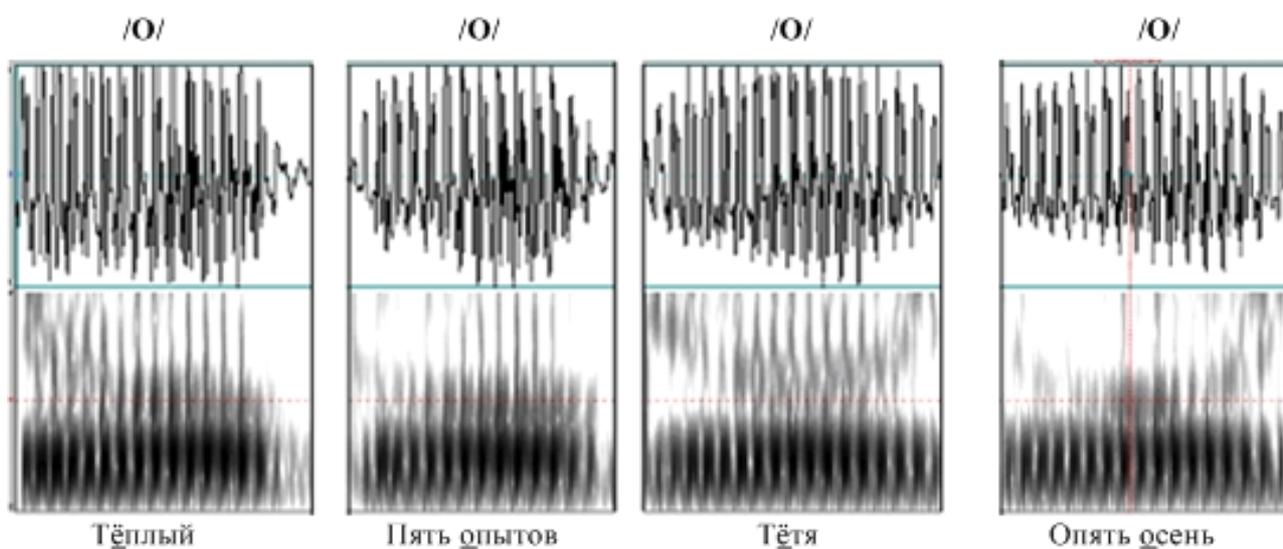


Рис. 8. Примеры акустических характеристик гласной внутри слова (а, в) и на стыке двух знаменательных слов (б, г) при реализации а) гласной V041; б) гласной V04_1; в) гласной V043; г) гласной V04_3

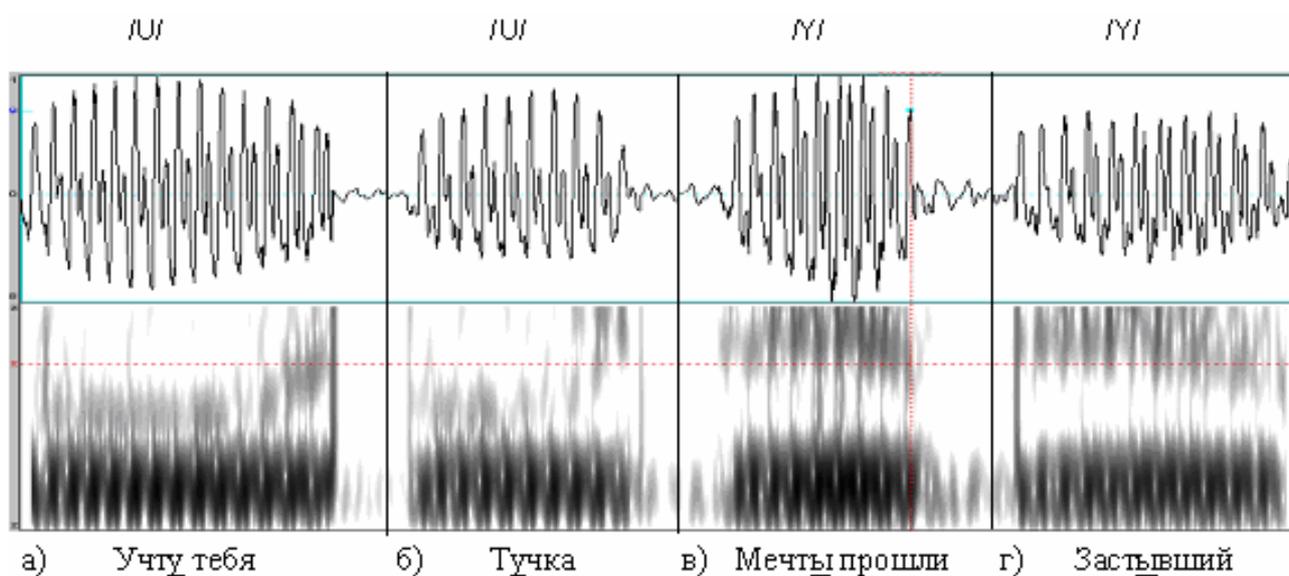


Рис. 9. Примеры акустических характеристик гласной на стыке двух знаменательных слов (а, в) и внутри слова (б, г) при реализации а) гласной V01_3; б) гласной V013; в) гласной V01_2; г) гласной V012

6.3. Акустические характеристики согласных

Как показали исследования, акустические характеристики согласных внутри и на стыках слов различаются, в отличие от гласных, в значительно меньшей степени, так что межсловный переход не приводит к появлению сколько-нибудь значимых для восприятия специфических аллофонов. Это утверждение иллюстрируется рис. 10–12 на примерах различных межсловных и внутрисловных сочетаний согласных.

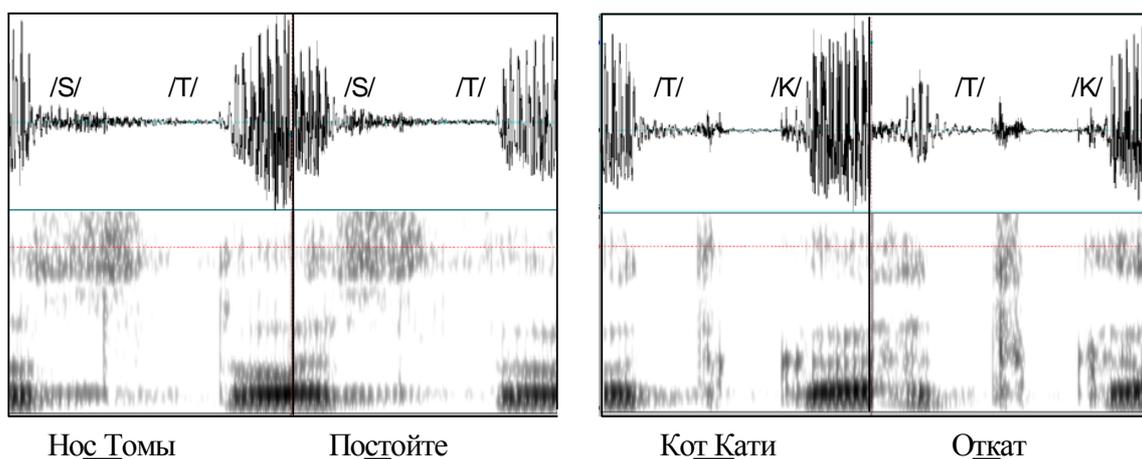


Рис. 10. Примеры реализации акустических явлений на стыке двух знаменательных слов и внутри слова для сочетаний согласных фонем «глухая - глухая»

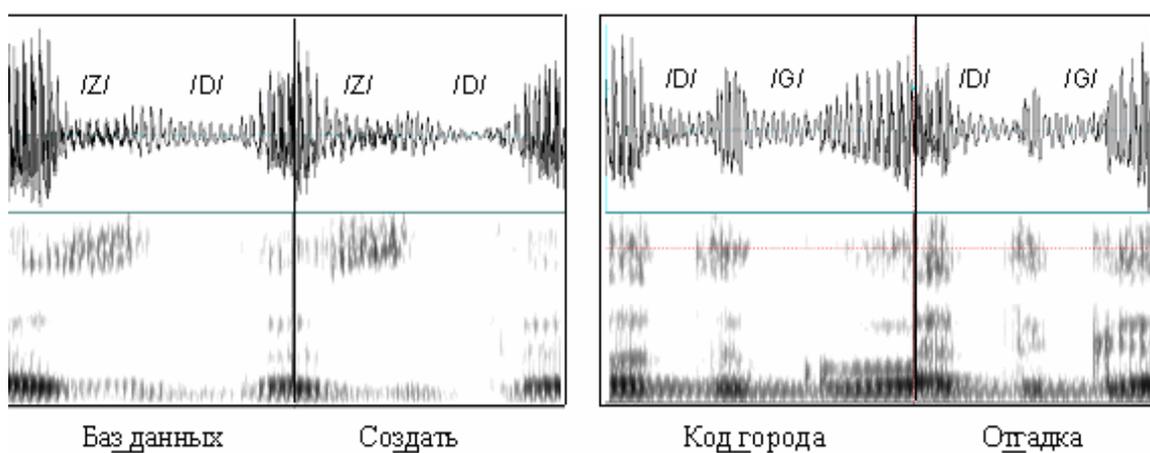


Рис. 11. Примеры реализации акустических явлений на стыке двух знаменательных слов и внутри слова для сочетаний согласных фонем «звонкая – звонкая»

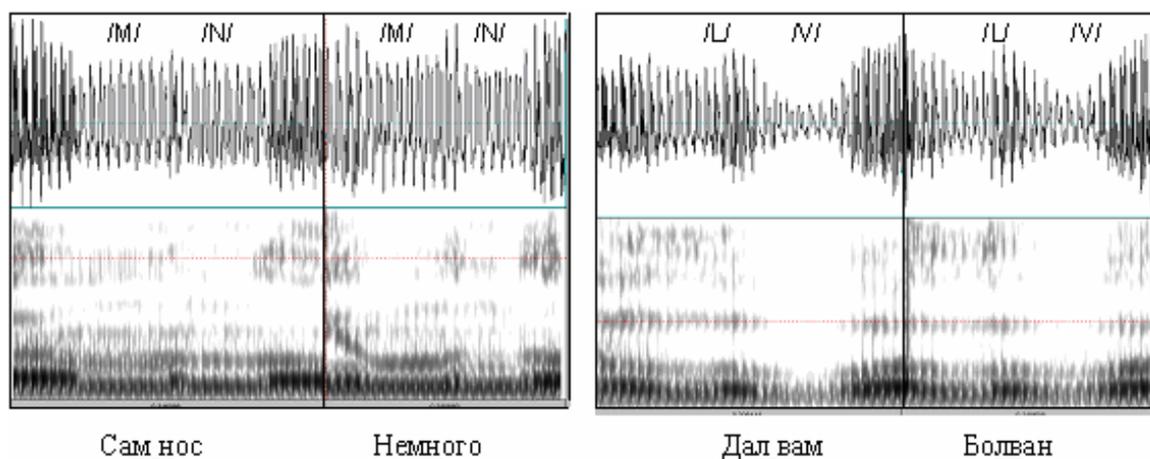


Рис. 12. Примеры реализации акустических явлений на стыке двух знаменательных слов и внутри слова для сочетаний согласных фонем «сонорная-сонорная»

Заключение

Современные системы синтеза достигли такого уровня развития, когда разборчивость синтезированной речи уже не является определяющим показателем (хотя и остаётся одним из важнейших). На рис. 13 приведены результаты оценки звуковой разборчивости синтезированной речи для различных систем синтеза русской речи по тексту (см.: www.sintezator.narod.ru), полученные в соответствии с методикой ГОСТ 16600-72 [7] на основе слухового восприятия фонетически сбалансированного набора псевдослов.

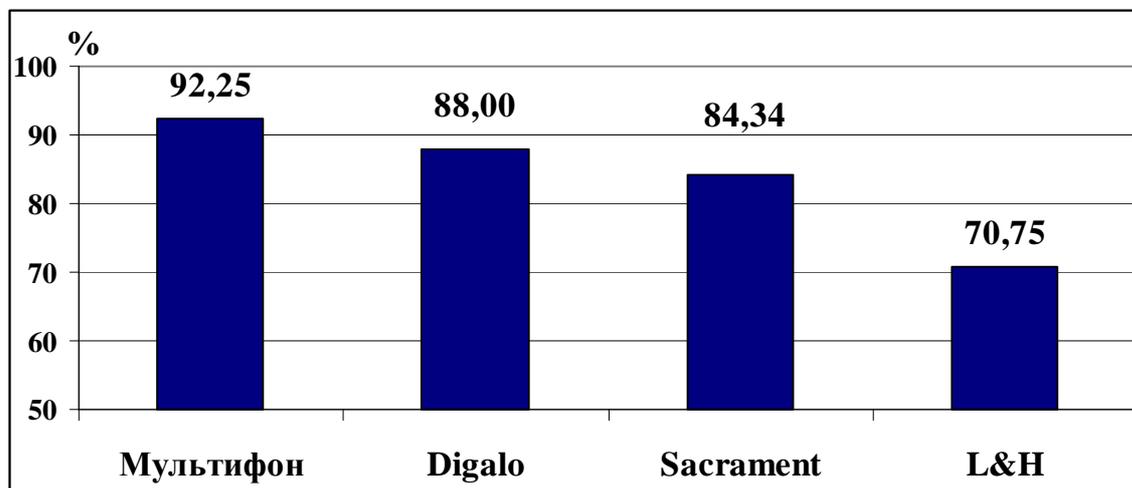


Рис. 13. Сравнительная характеристика разборчивости синтезированной речи

Достигнутое в современной модели синтезатора «Мультифон» значение звуковой разборчивости, превышающее 95%, обеспечивает, в соответствии с ГОСТ 16600-72, «понимание передаваемой речи без малейшего напряжения внимания». Однако такая оценка синтезированной речи будет справедлива лишь для случая синтеза изолированно произносимых слов. В реальности при машинном чтении произвольных текстов мы имеем дело с синтезом слитной речи, точность и комфортность восприятия которой зависит от полноты моделирования в синтезаторе фонетико-акустических явлений естественной слитной речи. В настоящей работе мы попытались обобщить исследования, направленные на выявление особенностей букво-фонемного и фонемно-аллофонного преобразований и их алгоритмизацию для синтеза полного и разговорного стилей слитной речи, которые реализованы в системе синтеза речи по тексту «Мультифон».

Дальнейшее совершенствование системы синтеза речи по тексту «Мультифон» связано не только с решением задач достижения высокого качества и естественности речи, но также и с приданием ей индивидуальности звучания. В последнее время развивается новое направление речевых исследований и разработок, позволившие представить синтезатор речи как средство компьютерного клонирования персонального голоса и дикции личности [8, 9].

Литература

1. Златоустова, Л.В. и др. Алгоритм преобразования русских орфографических текстов в фонемную запись. – М.: МГУ, 1970.
2. Лобанов, Б.М., Панченко, Б.В. Преобразователь графема-фонема для синтеза речи по орфографическому тексту // Тр. Междунар. конф. «Автоматическое распознавание слуховых образов АРСО-8». – Ч. 4. – Львов, 1974. – С. 15–18.
3. Зиновьева, Н.В., Кривнова, О.Ф. Прикладные системы с использованием фонетических знаний // Проблемы фонетики: сб. статей. – М.: Прометей, 1993. – С. 288-300.
4. Захаров, Л.В. Акустическая вариативность звуковых единиц в русской речи // Язык и речь: проблемы и решения: сб. науч. тр. – М.: Макс-Пресс, 2004. – С. 240–269.

5. Русская разговорная речь // Под редакцией Е.А.Земской.- М.: Наука, 1973.- 485 с.

6. Lobanov, B., Tsirolnik, L. Development of multi-voice and multi-language TTS synthesizer (languages: Belarussian, Polish, Russian) // Proceedings of the 11-th International conference SPECOM'2006, St. Petersburg, Russia.– P. 274-283.

7. ГОСТ 16600-72. Передача речи по трактам радиотелефонной связи.- М.: Государственный комитет стандартов Совета министров СССР, 1973.- 90 с.

8. Лобанов Б.М. Компьютерное клонирование персонального голоса и речи. // Новости искусственного интеллекта. №5(55).– Москва, 2002. – С. 35-39.

9. Lobanov, B., Tsirolnik, L. Phonetic-Acoustical Problems of Personal Voice Cloning by TTS // Proceedings of the 9-th International conference SPECOM'2004, St. Petersburg, Russia. – P. 17-21.

Л.А. Васильева, С.О. Тананайко

Статистические характеристики русских согласных и динамика современных произносительных тенденций в русской разговорной речи

*Санкт-Петербургский государственный университет,
г. Санкт-Петербург, Россия,
lavasil@mail.ru*

Фонетическая природа согласных

С общefonетической точки зрения природа согласных принципиально отличается от природы гласных. В артикуляторном аспекте гласные характеризуются разлитой артикуляцией и отсутствием преграды в надгортанных полостях для выходящей из легких струи воздуха, согласные же всегда имеют определенный фокус образования, и наличие той или иной преграды на пути выдыхаемого воздуха при их артикуляции является обязательным. Наличие определенного фокуса артикуляции и необходимость противопоставлять в пределах консонантных систем большое количество согласных (в русском языке, например, их 36) обеспечивает относительную устойчивость артикуляции согласных по сравнению с гласными. В акустическом аспекте существенно то, что согласные являются преимущественно шумами, а гласные – тонами и обладают формантной структурой, при этом значения частот формант могут варьировать в зависимости от фонетического контекста и нести информацию о характеристиках предшествующих согласных. В теоретико-информационном отношении согласные нагружены больше, чем гласные. Общеизвестно, что слова как лексические единицы во многих случаях могут быть опознаны только по согласным, а гласные играют в передаче информации, и, соответственно, в опознании слов, несколько меньшую роль [2]. Об этом же свидетельствует и консонантный коэффициент, равный для русской звучащей речи 1.3. [2]. Но гласные часто играют роль носителей той информации, которая позволяет опознавать разные словоформы.

Этим объясняется тот общеизвестный факт, что согласные в потоке речи менее подвержены изменениям, чем гласные [1].

Тем не менее, при изучении различных видов русской устной речи в экспериментальном материале – а именно в спонтанной речи и в чтении – не могли не появиться некоторые вариации количественных параметров употребления согласных, очевидно, связанные с различием двух изучаемых типов речи.

В настоящее время на филологическом факультете СПбГУ осуществляется изучение русской спонтанной речи с использованием пакета программ для редактирования и обработки речевых сигналов PRAAT, а также акустических баз данных. Целью этого проекта является теоретическое осмысление закономерностей образования и изменения произносительной нормы русского языка, разработка проблем нормализации произношения, подготовка материалов для заполнения банка данных русского языка (Фонетический Фонд русского языка). Данное исследование проводилось в рамках этого проекта на материале записей спонтанной речи и чтения 10 дикторов (мужчин и женщин различных возрастных групп). Наличие записей двух типов речи позволило сопоставлять их фонетические характеристики, так как в чтении были употреблены те же синтаксические конструкции и словоформы, что и в спонтанной речи.

В ходе данного этапа исследования была изучена общая частотность употребления различных согласных, их типов в спонтанной речи и чтении, а также частотность употребления согласных в различных фонетических контекстах.

Частотность употребления русских согласных в разных типах речи

На первом этапе было осуществлено общее ранжирование (по частоте встречаемости) согласных в двух типах транскрипции: акустической и идеальной (использовались

обозначения Международного фонетического алфавита IPA). Необходимо отметить, что в акустической транскрипции ранги звуков в спонтанной речи и чтении совпадали как для отдельных дикторов, так и по всему материалу. Идеальная же транскрипция в двух типах речи абсолютно идентична. При сравнении рангов звуков в акустической и идеальной транскрипциях видно, что расхождения между ними невелики (см. табл. 1).

Ранги	Акустическая транскр.	Идеал. транскрипция	Ранги	Акустическая транскр.	Идеал. транскрипция	Ранги	Акустическая транскр.	Идеал. транскрипция
1	t	t	19	g	ɣ'	37	h	
2	p	p	20	j	b	38	g'	
3	k	m	21	tʃ	g	39	ʏ	
4	s	k	22	d'	d'	40	ø	
5	ɾ	v	23	ž	f'	41	ʃ	
6	m	s	24	x	ž	42	θ	
7	p'	n'	25	f'	x	43	x'	
8	n'	ɾ	26	m'	v'	44	β	
9	v	p	27	ts	m'	45	ž'	
10	l	j	28	ʂ	ts	46	tʂ	
11	l'	l	29	~	k'	47	ʏ'	
12	d	l'	30	v'	p'	48	β'	
13	t'	d	31	k'	ʂ	49	θ'	
14	ʂ	t'	32	p'	z'			
15	s'	s'	33	z'	b'			
16	z	ʂ	34	b'	g'			
17	ɾ'	tʃ	35	w	f'			
18	b	z	36	f'	x'			

Таблица 1. Результаты ранжирования по частоте встречаемости (1-й ранг присвоен самому частотному звуку) звуков в акустической транскрипции и фонем в идеальной транскрипции.

Затем устанавливалась корреляционная зависимость между частотными рангами каждого согласного в разных типах речи и разных видах транскрипции. Сопряженными считались ранги, присвоенные одному и тому же согласному в сравниваемых видах транскрипции (см. табл. 2).

Дикторы	I	II	III	IV	V	VI
Д1	0,968	0,94	0,935	0,948	0,908	0,932
Д2	0,958	0,933	0,94	0,937	0,898	0,938
Д3	0,944	0,974	0,942	0,987	0,834	0,976
Д4	0,922	0,935	0,964	0,955	0,955	0,946
Д5	0,959	0,962	0,968	0,968	0,869	0,869
Д6	0,9	0,935	0,958	0,968	0,875	0,889
Д7	0,92	0,965	0,974	0,976	0,906	0,952
Д8	0,974	0,93	0,962	0,977	0,923	0,917
Д9	0,962	0,945	0,945	0,975	0,89	0,924
Д10	0,952	0,928	0,967	0,954	0,93	0,934

Таблица 2. Коэффициент корреляции Спирмена для сопряженных выборок согласных в разных типах речи и в разных транскрипциях. I – отношение данных для каждого диктора в спонтанной речи и чтении по акустической транскрипции; II – отношение усредненных данных к данным для каждого диктора в спонтанной речи по акустической транскрипции; III – отношение усредненных данных к данным для каждого диктора в чтении по акустической

транскрипции; IV - отношение усредненных данных к данным для каждого диктора по идеальной транскрипции; V - отношение данных для каждого диктора по акустической транскрипции к сопряженным данным по идеальной транскрипции в спонтанной речи; VI - отношение данных для каждого диктора по акустической транскрипции к сопряженным данным по идеальной транскрипции в чтении.

Надо отметить, что во всех случаях корреляция оказывается высокой или очень высокой. Тогда, когда сопоставлялись акустическая транскрипция в спонтанной речи и чтении для каждого диктора, усредненные значения идеальной транскрипции, а затем и акустической транскрипции в обоих типах речи и соответствующие значения для каждого диктора, то коэффициент корреляции /r/ превышает 0,9. Тогда же, когда сопоставлялись акустическая и идеальная транскрипция в спонтанной речи и чтении, то коэффициент корреляции в спонтанной речи оказывался несколько меньше, чем в чтении (0,899 и 0,928 соответственно). Можно сделать вывод, что акустическая транскрипция согласных несколько больше отклоняется от идеальной, чем в чтении. Что же касается сопоставления с данными Скрелина П.А., полученными при изучении текста, включающего более 1000000 фонемупотреблений, то коэффициент корреляции очень высок (0,92), что еще раз подтверждает фонетическую представительность нашего материала. Полученные данные целиком согласуются с выявленными ранее тенденциями [3].

Что касается частотности употребления звуков различных типов, то наиболее частотными являются глухие смычные, а наименее частотна твердая аффриката. В целом, звонкие шумные согласные менее частотны, чем глухие, а мягкие в целом менее частотны, чем твердые. Второе место по частотности занимают твердые носовые сонанты, а четвертое – твердые плавные. Кроме того, любопытно отметить, что частотность звонких мягких щелевых оказалась меньше, чем частотность мягкой аффрикаты. Полученные данные вполне согласуются с описанными ранее тенденциями.

Гораздо более информативным оказалось исследование частотности употребления согласных в различных фонетических контекстах.

Во-первых, нужно сказать, что все типы согласных употреблялись во всех возможных контекстах, за исключением мягких звонких смычных, которые ни разу не встретились между согласными. Другим очевидным выводом стало то, что процент употребления твердых согласных был почти всегда значительно больше, чем мягких. Исключения составляют мягкие плавные между согласными, которые и в спонтанной речи, и в чтении встречались несколько чаще, чем твердые, а также мягкая аффриката, которая в позиции между согласными, после гласного перед согласным и между гласными в обоих типах речи встречалась чаще, чем твердая. Что касается контекста после согласного перед гласным, то в нем твердая аффриката встречалась чаще мягкой в обоих типах речи. Говоря об аффрикатах, однако, надо помнить, что в русском языке твердая и мягкая аффриката не составляют пары, аналогичной абсолютному большинству пар твердых и мягких согласных; кроме того, частота встречаемости аффрикат по сравнению с другими типами согласных вообще очень мала.

Зависимость частотности употребления согласных от фонетического контекста

На втором этапе рассматривалась частотность употребления других типов согласных во всех возможных контекстах (см. табл. 3). Наиболее частотны в обоих типах речи твердые глухие смычные между гласными. На втором месте в спонтанной речи твердые глухие смычные после согласного перед гласным, а в чтении – твердые звонкие смычные между гласными. На третьем месте по частотности и в спонтанной речи, и в чтении стоят твердые глухие щелевые после гласного перед согласным. На четвертом месте в спонтанной речи оказались твердые звонкие смычные между гласными, а в чтении – твердые глухие смычные после согласного перед гласным.

Спонтан. речь		Чтение	
1	VPV	1	VPV
2	CPV	2	VP+V
3	VFC	3	VFC
4	VP+V	4	CPV
5	VPC	5	VAppV
6	VAppV	6	VNV
7	VNV	7	VPC
8	CNV	8	CNV
9	CAppV	9	VN'V
10	VApp'V	10	VNC
11	VN'V	11	VApp'V
12	VFV	12	CAppV
13	VNC	13	VFV
14	CP+V	14	CP+V
15	VF+V	15	VF+V
16	VP+C	16	VP+C
17	CFC	17	VP'+V
18	CPC	18	VAppC
19	VAppC	19	CApp'V
20	VP'+V	20	VF+C
21	CApp'V	21	CP'V
22	CN'V	22	VP'V
23	VF+C	23	VFC
24	VP'V	24	CFC
25	CP'V	25	VAff'V
26	CFV	26	CFV
27	VFC	27	CN'V
28	VF'V	28	CPC

Спонтан. речь		Чтение	
29	VAff'V	29	CP'+V
30	VApp'C	30	VF'V
31	CF'V	31	CF'V
32	CP'+V	32	VApp'C
33	VAff'C	33	VAffV
34	VN'C	34	VAff'C
35	VAffV	35	CF+V
36	VP'C	36	VP'C
37	CP+C	37	CAffV
38	CNC	38	VF'+V
39	CF+V	39	VN'C
40	CApp'C	40	CP+C
41	CP'C	41	VF'+C
42	VF'+V	42	CAff'V
43	CAffV	43	CF+C
44	VF'+C	44	VAffC
45	CAff'V	45	CApp'C
46	CN'C	46	CNC
47	CF+C	47	CN'C
48	VAffC	48	CAppC
49	CAppC	49	CP'C
50	VP'+C	50	VP'+C
51	CFC	51	CF'C
52	CAff'C	52	CAff'C
53	CAffC	53	CF'+V
54	CF'+V	54	CAffC
55	CP'+C	55	CF'+C
56	CF'+C	56	CP'+C

Таблица 3. Частотность употребления типов согласных в разных фонетических контекстах (ранжирование сделано по убыванию частотности). В таблице использованы следующие обозначения: С – согласный; V – гласный; P – твердый глухой смычный; P+ – твердый звонкий смычный; P' – мягкий глухой смычный; P'+ – мягкий звонкий смычный; F – щелевой; App – носовой сонант; N – носовой сонант; Aff – аффриката.

Надо отметить, что частотность большинства фонем в одних и тех же контекстах практически совпадает или близка в обоих типах речи. Наибольшие различия отмечаются для сочетаний трех согласных (контекст между согласными) и для твердой аффрикаты. Общеизвестно, что частотность трехкомпонентных сочетаний согласных, а также аффрикат в произношении вообще невелика. Интересно, однако, что даже эти сильные отличия не разделяют перечисленные фонемы больше, чем на 8 рангов. Во всяком случае, можно сказать, что эти отличия никак не влияют на общее сходство полученных распределений. Для большинства же наиболее частотных согласных ранги или вообще совпадают, или различаются незначительно.

Таким образом, можно сделать вывод, что максимальные различия в частотности употребления тех или иных согласных в разных контекстах связаны со сложностью артикуляции и редкостью употребления как групп согласных, так и сложных звуков – аффрикат, что и вызывает различные способы их упрощения в разных типах речи.

Литература

1. Bondarko L.V., Volskaya N.B., Tananaiko S.O., Vasilieva L.A. Phonetic Properties of Russian Spontaneous Speech //Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, 3-9 August 2003. P. 2973-2976
2. Бондарко Л.В. Фонетика современного русского языка. СПб, 1998. С.99.
3. Бондарко Л.В., Зиндер Л.Р., Штерн А.С. Некоторые статистические характеристики русской речи. // Слух и речь в норме и патологии. Л., 1977

Моделирование нефонемных речевых элементов и создание альтернативных транскрипций для распознавания спонтанной речи

*Санкт-Петербургский институт информатики и автоматизации,
Санкт-Петербург, Россия,
 {leonty, kipyatкова}@ias.spb.su*

1. Введение

Автоматическое распознавание спонтанной речи осложнено тем, что в ней содержатся неинформативные элементы, такие как невербальные и вокализованные паузы, артефакты, фоновый шум. Междометия и вводные слова, особенно характерные для разговорной речи и выполняющие определенную дискурсивную роль, для диалоговой системы, настроенной на решение узкой задачи также неинформативны. Другая проблема при распознавании разговорной речи заключается в вольности формирования как фраз, так и слов. Часто для подтверждения или отрицания используются нефонемные аналоги междометий «да» и «нет». Произношение же слов отличается от диктора к диктору, и транскрипция произнесенных слов зачастую не совпадают с транскрипциями, образованными по правилам русского языка. Все это нужно учитывать при создании моделей различных таксономических единиц речи, участвующих в процессе распознавания для более точного описания разговорной речи. В данной статье описываются способы моделирования нефонемных элементов спонтанной речи, а также методы учета вариативности произношения с помощью создания альтернативных транскрипций слов.

2. Подход к учету структуры спонтанной речи

При обработке спонтанной речи на вход распознавателя поступают звуки фонемной и нефонемной природы. Кроме того, это могут быть звуки, как производимые диктором, так и посторонние шумы и речь тех, кто не контактирует непосредственно с системой. Возможные компоненты входного сигнала показаны на рис. 1. Сигнал будет содержать шумы окружающей обстановки, в которой производится запись. Также в ходе речеобразования возникают «шумы» органов голосового аппарата, так называемые артефакты речи. Другими элементами, производимыми диктором, являются: озвученные и невербальные паузы, нефонемные подтверждения и отрицания, ключевые слова, различные междометия и вводные слова. Помимо этого возможно присутствие незнакомых системе слов. Какой бы большой не был взят словарь – он не в состоянии покрыть все слова, которые пользователь может употребить в диалоге с системой для получения желаемого результата.

Рассмотрим каждый элемент подробнее. Акустический фон записанного речевого сигнала может содержать шумы улицы, города или даже речь людей, находящихся неподалеку от диалоговой системы, однако не общающихся с ней напрямую, а также шумы канала передачи данных. Все эти звуки накладываются на речь пользователя, затрудняя распознавание.

Вокализованные (озвученные) паузы могут быть вызваны различными причинами: колебания, размышления и др. [1]. Чтобы не допустить разрыва во фразе и диалоге с собеседником, образовавшаяся пауза заполняется разного рода звуками. Это могут быть как растянутые звуки, напоминающие фонемы («а-а», «э-э», «м-м»), так звуки явно нефонемной природы (кряхтение, хриплые «а», «о», «м») или даже комбинации звуков («хм», «гм», «ма»). При диалоге между людьми озвученные паузы помимо того, что не позволяют разорвать разговор, давая собеседнику понять, что оратор не закончил свое высказывание, или могут служить неким сигналом помощи, обращенным к собеседнику. Для системы же

вокализованные паузы не несут информативной нагрузки и должны быть устранены на ранних уровнях обработки сигнала.

Невербальные паузы, как правило, не являются элементами дискурса, и могут возникнуть в любой момент диалога [1]. Они могут быть вызваны смехом, покашливанием, прочищение горла. Длительность подобных явлений может сильно варьироваться. В связи с этим от невербальных пауз отделяем артефакты, преимущественно короткие, неречевые элементы, например, причмокивание, цоканье языком, звуки, связанные с громким дыханием, а также различные кинестезические явления.

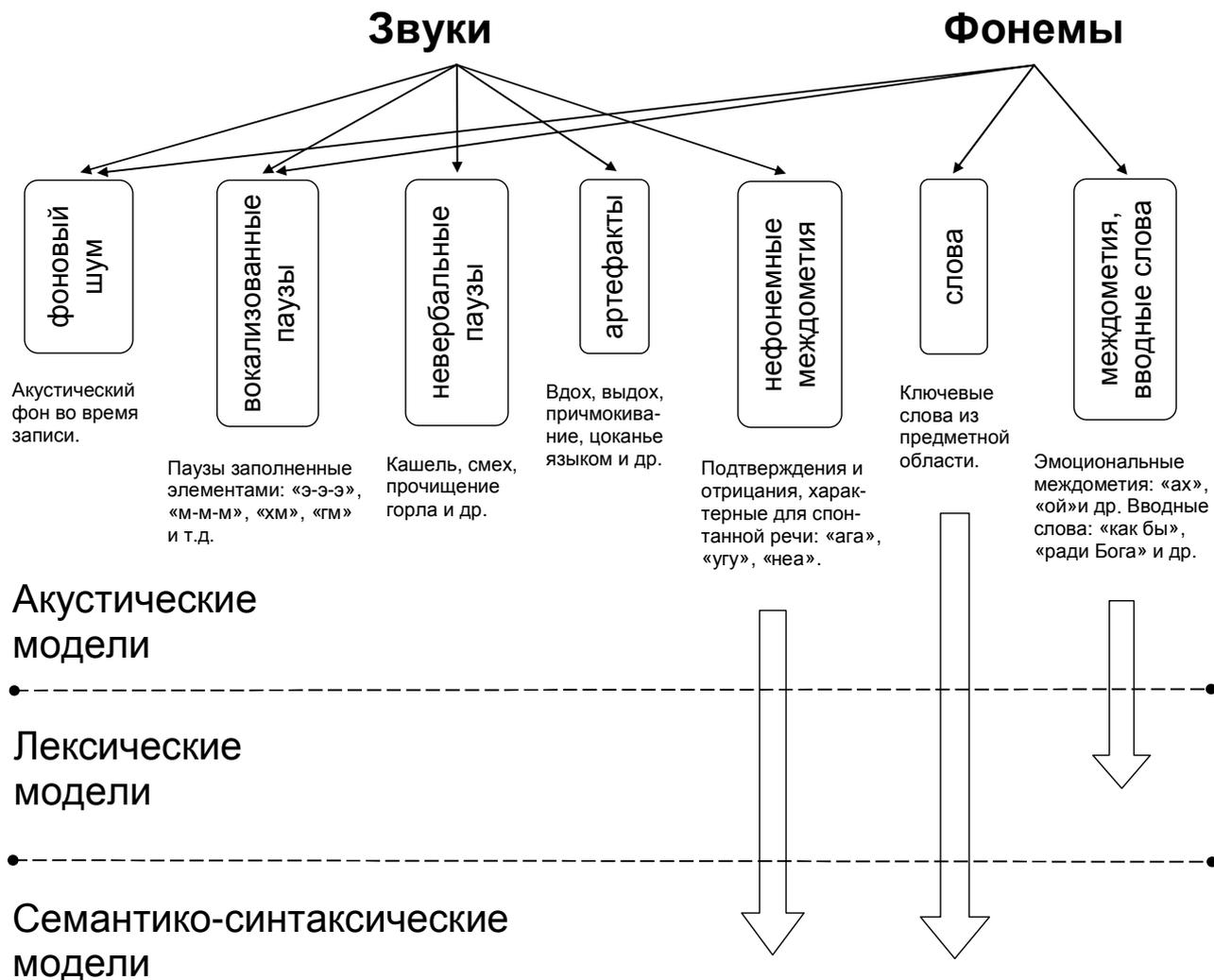


Рис. 1. Модели для работы со спонтанной речью

Для подтверждения и отрицания в разговорной речи вместо междометий «да» и «нет» часто используются их нефонемные аналоги «угу», «ага» и «неа». Сложность выделения этих элементов заключается в их схожести с вариативными вокализованными паузами «у-у», «а-а», «м-м».

Словарь составляется исходя из области применения диалоговой системы. Ключевые слова должны в достаточной степени описывать предметную область и учитывать возможные формулировки пользователя, направленные на решение конкретной задачи. Появление неучтенных слов (out-of-vocabulary words) во входном речевом сигнале требуют отдельного решения.

Даже полностью распознанная фраза диктора может быть трудна для понимания смысла, так как может содержать избыточную информацию, поэтому необходима предварительная обработка для исключения неинформативных, с точки зрения системы,

элементов. Речь идет, в первую очередь об эмоциональных междометиях и некоторых вводных словах. Словарь исключаемых междометий, так же как и словарь ключевых слов следует подстраивать под конкретную задачу.

Большинство нефонемных элементов могут быть обработаны на акустическом уровне. Для исключаемых междометий и вводных слов требуется создать словарь, аналогичный словарю ключевых слов, с той лишь разницей, что они обрабатываются внутри декодера, но не передаются на последующие уровни обработки. Нефонемные подтверждения и отрицания используются для понимания смысла ответа наравне с «да» и «нет».

Все эти модели могут быть реализованы в модифицированном гибридном декодере [2], который состоит из: словаря, фонемного декодера, модели мусора и модели тишины. Словарь распознавателя строится на базе двухуровневого морфофонемного префиксного графа (ДМПГ) [3]. Исключаемые междометия, вводные слова, а также ключевые слова обрабатываются в ДМПГ. Фонемный декодер позволяет распознавать слова, неизвестные системе (OOV). Модель мусора покрывает фоновый шум, вокализованные и невербальные паузы, артефакты. Модель тишины введена для учета беззвучных пауз между словами. Сигнал, поступающий на вход системы, обрабатывается одновременно всеми четырьмя компонентами декодера. Значимость каждой из компонент определяется заранее на этапе настройки системы, а в процессе распознавания вероятность гипотезы, поступающей с некоторой компоненты декодера, перемножается с соответствующим коэффициентом значимости.

3. Методы создания альтернативных транскрипций

Помимо зашумленности речи, связанной с наличием посторонних слов, множества акустических эффектом и шумов, в спонтанной речи сильно варьируется произношение слов, что осложняет процесс распознавания. Большое несоответствие между наблюдаемым произношением и принятыми фонетическими транскрипциями является одной из главных причин низкой производительности систем распознавания спонтанной речи.

Стандартные подходы на базе скрытых марковских моделей, содержащих смеси гауссовских плотностей вероятностей, связанные состояния и контекстно-зависимые модели фонем (например, трифоны), позволяют до некоторой степени учесть вариативность речи [4], [5]. В том числе учитываются эффекты контекстной зависимости фонем, их замещение и изменчивость по длительности. В [6] было показано, что возможно достичь точности распознавания свыше 90% для задач с большим словарем и с ограниченной вариативностью речи, например для обработки аудиозаписи прочитанного текста. Однако в спонтанной речи вариативность произношения и темпа речи намного больше, и поэтому стандартная схема распознавания не справляется с отклонениями такого уровня. Описанный ранее подход к моделированию элементов спонтанной речи позволяет частично решить проблему отсутствия беглости в спонтанной речи, но никак не учитывает вариативность произношений ключевых слов.

Некоторые исследователи пытались учитывать вариативность произношения на уровне сублексических единиц путем введения сложных структур связанных состояний [7]. Однако эти методики позволяют улучшить обработку при замещении слов, но не подходят для работы с другими важными явлениями спонтанной речи как пропущенные и вставленные слова. Поэтому для моделирования вариативности произношения часто используются методы, основанные на сложных лексических моделях, в которых альтернативные транскрипции заданы в явном виде.

В последние годы были предложены некоторые методы генерации альтернативных транскрипций из канонической формы. Эти методы базируются на фонетических или лингвистических знаниях [8], а также используются статистические методы [9], при которых варианты произношения получаются автоматически при анализе речевой базы данных. Рассмотрим эти группы методов более подробно.

3.1 Генерация альтернативных транскрипций по правилам русского языка и речи

Транскрипции слов создаются по базовым правилам транскрибирования [10], однако в разговорной речи некоторые звуки могут редуцироваться, кроме того, на стыке слов также часто возникают явления ассимиляции и редукции звуков [11]. Правила, позволяющие учесть явления редукции и ассимиляции в процессе синтеза русской речи, сформулированы в статье [12]. Используя эти правила, можно получить альтернативные транскрипции для расширения словаря в задачи распознавания речи. На рис. 2 показан укрупненный алгоритм транскрибирования словоформ с учетом ассимиляции и редукции.

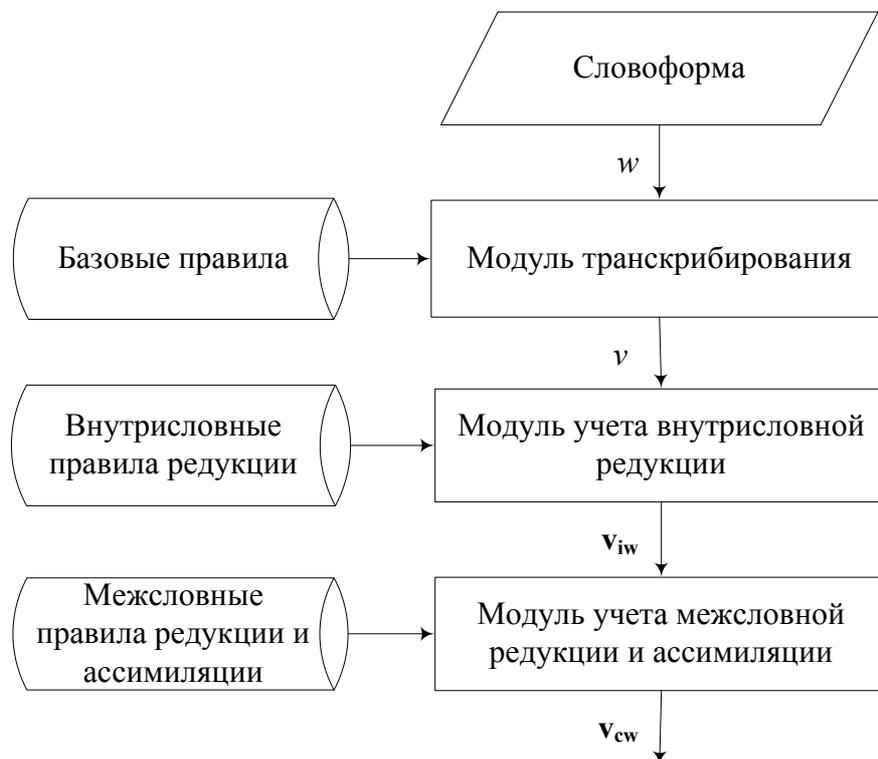


Рис. 2. Алгоритм транскрибирования словоформы с учетом ассимиляции и редукции

Модуль транскрибирования, используя базовые правила, преобразует поступающую на его вход словоформу w в последовательность фонем. Полученная таким образом транскрипция v поступает в модуль учета внутрисловной редукции, где с помощью правил внутрисловной редукции мы получаем набор альтернативных транскрипций V_{iw} данной словоформы. Поскольку для одной словоформы может быть применено несколько правил, то на выходе модуля получаются транскрипции, учитывающие все возможные варианты произношения. Далее эти альтернативные транскрипции в модуле учета межсловной редукции и ассимиляции преобразуются в альтернативные транскрипции словоформы V_{cw} . Полученный таким образом набор транскрипций теоретически должен содержать все варианты произношений, которые могут возникать в разговорной речи различных людей.

3.2 Статистические методы генерации альтернативных транскрипций

Альтернативные транскрипции в явном виде также могут быть созданы с помощью фонемного распознавателя. Процесс создания альтернативных транскрипций представлен на рис. 3, где $w_1 \dots w_n$ – слова, содержащиеся в фразе, $v_1 \dots v_n$ – транскрипции слов, $ph_{11} \dots ph_{ns}$ – распознанные фонемы, $v_{a1} \dots v_{an}$ – альтернативные транскрипции для каждого слова.

Фраза, которая в данный момент используется для получения дополнительных транскрипций, состоит из n слов. Для каждого слова этой фразы создается транскрипция. Последовательность фонем, которая получается в результате распознавания этой фразы,

совмещается с ее транскрипцией помощью динамического программирования или скрытых марковских моделей. Таким образом, получается альтернативная транскрипция для каждого слова. Из-за того, что одно и тоже слово может встречаться в нескольких фразах из обучающего корпуса речи, и вследствие вариативности произношения различных дикторов для каждого слова может быть создано несколько альтернативных транскрипций:

$$v_{ai} = \{v_{ai_1}, \dots, v_{ai_m}\}$$

где m – число альтернативных транскрипций для i -го слова.

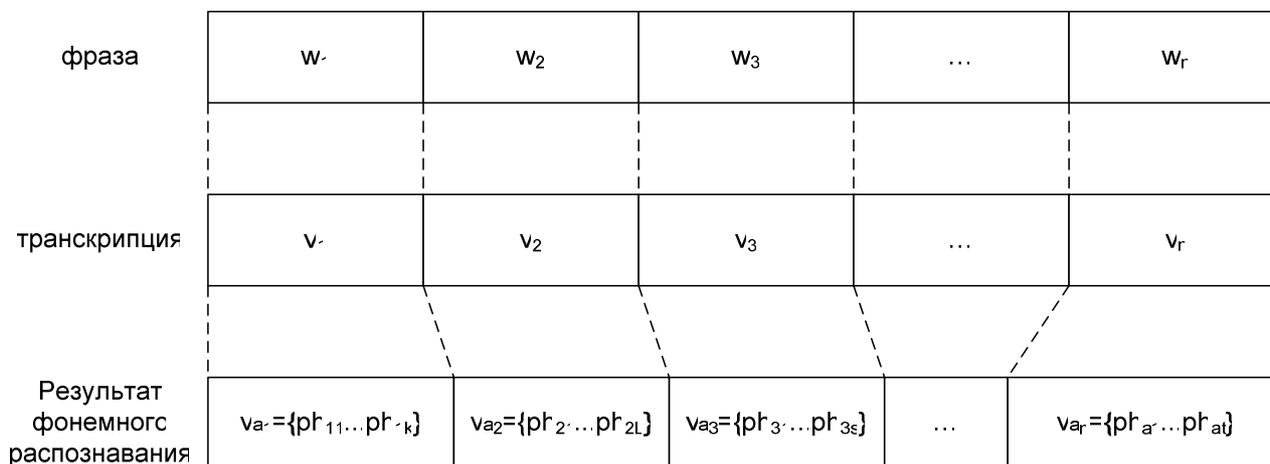


Рис. 3 – Процесс создания альтернативных транскрипций с помощью фонемного распознавателя

Фонемный распознаватель может использовать только акустическую информацию или накладывать некоторые ограничения на последовательность распознаваемых фонем, например, путем применения биграммной фонетической грамматики. Однако, фонетические транскрипции, полученные с помощью этой техники, содержат много зашумленных участков. Для решения этой проблемы в [9,13], было предложено дерево решений, основанное на правилах генерации сглаженных альтернативных транскрипций из основных форм транскрипций.

Возможно также автоматическое обучение робастным фонетическим транскрипциям по речевым данным, используя стандартный метод поиска Витерби со свободным выбором между альтернативными произношениями. При этом увеличение числа вариантов произношений достигается за счет ослабления ограничений при совмещении входного акустического сигнала с существующими в словаре транскрипциями слов, например, делая присутствие каждой фонемы в транскрипции необязательным. Это позволяет получить дополнительные транскрипции в ходе распознавания. Этот подход был использован в [14] для определения правил редукции фонем в спонтанной речи. При распознавании голландского языка в задаче со средним размером словаря (около 1000 слов) это дало улучшение на 8% по сравнению с системой без альтернативных транскрипций.

Проверка описанных методов создания альтернативных транскрипций будет проведена в дальнейшей работе при обучении системы распознавания русской разговорной речи.

4. Моделирование нефонемных элементов спонтанной речи

На базе собранного корпуса было проведено исследование нефонемных элементов спонтанной речи. Собранный корпус содержит 103 диалога пользователей с операторами справочных служб заказа авиабилетов. Было записано 39 диалогов пользователей с женскими голосами и 64 - с мужскими. Возраст дикторов-женщин варьировался от 17 до 44 лет, и у мужчин - от 17 до 59 лет. Общая продолжительность записей в корпусе составляет 2 часа 50 минут, включая обращения пользователей, ответы оператора и паузы, вызванные ожиданием ответа. Корпус содержит 970 высказываний пользователей и примерно столько же

высказываний операторов. Большинство диалогов содержат от 5 до 24 обращений и длятся от 30 секунд до 4 минут. Также в корпусе имеется 11 диалогов длиной менее 30 секунд, связанных с переадресацией запроса, и несколько диалогов длиной до 7 минут в связи с длительной обработкой запроса оператором.

На основании данных из корпуса был выделен ряд наиболее часто встречающихся нефонемных элементов в записанных диалогах (табл.1). Нами были рассмотрены только элементы, которые встречались в корпусе более двух раз. В связи с этим не рассматривается, например, нефонемное отрицание «неа», содержащееся в корпусе только 1 раз. Среди артефактов были выделены чмоканье и вдох. Невербальные паузы (НП) представлены только кашлем. Больше всего было выделено различных вариантов вокализованных пауз. Отчасти это вызвано тем, что ситуация заказа авиабилетов была смоделирована, и людям общающимся с оператором не требовалось куда лететь. Поэтому выбор рейса и даты представлял некоторые затруднения, так как у дикторов не было четких установок по этому поводу. К шумовым паузам были отнесены паузы, заполненные звуками неясного характера, производимые диктором в процессе речеобразования и записи сигнала.

Таблица 1. Элементы спонтанной речи в собранном корпусе.

Сравнительная характеристика	Артефакты		НП	Вокализованные паузы							Подтверждения		Шумовые паузы	Слова	Всего
	Чмоканье	Вдох		Кашель	АААА	ААММ	ММММ	ММАА	Шва	ЕЕЕЕ	ЭЭУУ	Ага			
Количество элементов	5	62	4	67	12	30	2	11	29	3	45	145	56	5199	5670
Суммарная длительность, с	1.36	26.88	1.65	22.81	6.83	12.43	0.92	2.95	13.14	0.90	16.81	58.43	24.99	1781.10	1971.20
Средняя длительность, с	0.27	0.41	0.41	0.34	0.53	0.41	0.46	0.27	0.45	0.30	0.37	0.40	0.45	0.34	0.35

На эффекты спонтанной речи в корпусе приходится 8.6% от общего числа файлов. Из них 4.96% приходится на артефакты, вокализованные, невербальные и шумовые паузы. По продолжительности эффекты спонтанной речи составляют 7.9%, из них 5.8% принадлежат неинформативным элементам. Средняя длительность файлов составила 0.39 с как для слов, так и для элементов спонтанной речи.

Данные по количеству и длительности элементов спонтанной речи приведены для пользователя, анализ речи оператора не проводился. Тишина была исключена из корпуса в данном исследовании, используя разработанный алгоритм определения границ речи [15].

Для описания выделенных элементов спонтанной речи были введены 11 дополнительных фонем. В таблице 2 представлен ряд моделей, описывающих эти

элементы. На данном этапе с учетом небольшого количества нововведенных "фонем" был составлен собственный алфавит для транскрибирования нефонемных элементов, без учета существующих фонетических алфавитов [16], [17]. Варианты альтернативных транскрипций целиком основаны на данных из корпуса.

Таблица 2. Описание моделируемых элементов спонтанной речи.

Название элемента	Альтернативные транскрипции	Описание
АААА	^a ^o ^a ^a ^e ^a ^o	Растянутый звук «а», с некоторыми вариациями «оа», «аэ», «ао», иногда хриплое «а».
ААММ	^a ^mm ^o ^mm	Комбинация звуков «ам», либо короткое замыкающее «м» при закрытии губ.
АГА	^a ^h ^a ^o ^h ^a ^a ^g ^a	Нефонемное подтверждение «аха», иногда редуцированное до «а-а», с вариациями первого гласного, либо хриплое «ага».
КАШЕЛЬ	^k ^u ^k ^k ^u	Одинарное или двойное покашливание.
ММММ	^m	Мычание, длинное «м».
ММАА	^m ^a	Комбинация звуков «ма» с длинным «м».
УГУ	^u ^h ^u ^u ^g ^u	Нефонемное подтверждение «уху», иногда редуцированное до «у-у», либо хриплое «угу».
ЧМОКАНЬЕ	^ch	Чмокание по звуку напоминает щелчок.
ВДОХ	^w ^ch ^w	Звук от набирания воздуха через рот или нос, а также комбинация чмокания и вдоха
ШВА	^o	Модель гласного среднего ряда «schwa», по звучанию похожего на нечто среднее между «а», «э» и «о».
ЭЭЭЭ	^e	Растянутый звук «э», иногда хриплое «э».
ЭЭУУ	^e ^uu ^e ^o ^o ^uu	Звуки «эу», «эо», «оу».

Каждая модель нефонемного звука строится на основе лево-правой скрытой марковской модели с тремя состояниями. Любой элемент составляется из последовательности некоторых звуков, в дальнейшем планируется введение моделей с более сложными топологиями. Пока было создано пять моделей фонем для «гласных» звуков: ^a, ^e, ^o, ^u, ^uu, шесть моделей для «согласных» звуков: ^g, ^k, ^kl, ^m, ^mm, ^h, и одна фонема для вдоха ^w.

Был проведен закрытый эксперимент, в котором обучающие данные выступали одновременно и проверочными. В этом случае общий процент правильного распознавания элементов спонтанной речи составил 80.48%. Применение гауссовских смесей для описания состояний марковской модели позволило повысить точность до 92.77%. Несмотря на полученный достаточно высокий процент распознавания, при обработке спонтанной речи важнее будет отделить «мусорные» элементы от ключевых слов, при этом на сколько путаются эти элементы между собой не так важно. В проведенном эксперименте артефакты и невербальные паузы были распознаны правильно во всех случаях. Среди вокализованных пауз 2.6% были приняты за нефонемные подтверждения. В основном из-за редуцированных форм «а-а» и «у-у» 5.3% нефонемных подтверждений были распознаны как вокализованные паузы, 0.5% как артефакты и 0.5% как невербальные паузы.

5. Заключение

Представленные результаты предварительного исследования подтвердили целесообразность использования моделей помех и нефонемных подтверждений при обработке разговорной речи. Разработанные модели способны выделять в высказывании такие элементы и распознавать их подобно ключевым словам. В будущем будут исследованы модели с более сложной топологией и проведены эксперименты по распознаванию спонтанной русской речи, где словарь будет состоять из ключевых слов, а также нефонемных элементов и помех, характерных для разговорной речи. Также будет проведена апробация методов создания альтернативных транскрипции на основе экспертных знаний и статистических моделей распознавания речи. Описанные подходы направлены на расширение словаря системы распознавания русской речи с целью повышения робастности декодирования входного сигнала, записанного в реальных условиях эксплуатации.

6. Литература

1. О. А. Александрова, В. В. Иваницкий. Пауза колебания – комплексный феномен современной коммуникации. Вестник Новгородского государственного университета. Т. 25, С. 95- 101.
2. I. Bazzi, J. Glass. Modeling out-of-vocabulary words for robust speech recognition. Proc. ICSLP, Beijing, 2000.
3. А. Л. Ронжин, Ан. Б. Леонтьева, И. А. Кагиров, Ал. Б. Леонтьева. Двухуровневый морфофонемный префиксный граф для декодирования русской слитной речи. Труды СПИИРАН. Вып. 4, т.1, Санкт-Петербург, 2007.
4. D. Jurafsky, W. Ward, Z. Jianping, K. Herold, Y. Xiuyang, Z. Sen. What kind of pronunciation variation is hard for triphones to model? Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Vol. 1, P. 577–580. Salt Lake City, UT, USA, May 2001.
5. T. Hain. Implicit pronunciation modeling in ASR. Proc. ISCA Tutorial and Research Workshop on Pronunciation Modeling and Lexicon Adaptation for Spoken Language. P. 129–134, Estes Park, Colorado, USA, September 2002.
6. R. Schwartz, T. Colthurst, N. Duta, H. Gish, R. Iyer, C. L. Kao, D. Liu, O. Kimball, J. Ma, J. Makhoul, S. Matsoukas, L. Nguyen, M. Noamany, R. Prasad, B. Xiang, D. Xu, J. L. Gauvain, L. Lamel, H. Schwenk, G. Adda, L. Chen. Speech recognition in multiple languages and domains: The 2003 BBN/LIMSI EARS system. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing. Vol. 3, P. 17–21, Montreal, Canada, May 2004.
7. H. Yu, T. Schultz. Enhanced tree clustering with single pronunciation dictionary for conversational speech recognition. Proc. European Conf. on Speech Communication and Technology. P. 1869–1872. Geneva, Switzerland, 2003.
8. J. M. Kessens, M. Wester, H. Strik. Improving the performance of a Dutch CSR by modeling within-word and cross-word pronunciation variation. In Speech Communication, Vol. 29, P. 193–207, 1999.
9. J. E. Fosler-Lussier. Multi-level decision trees for static and dynamic pronunciation models. Proc. European Conf. on Speech Communication and Technology, P. 463–466. Budapest, Hungary, 1999.
10. Академия наук СССР Институт русского языка «Русская грамматика». М.: Наука, 1980.
11. Русская разговорная речь. Под редакцией Е.А.Земской. М.: Наука, 1973. 485 с.
12. Б. М. Лобанов, Л. И. Цирульник. Моделирование внутрисловных и межсловных фонетико-акустических явлений полного и разговорного стилей в системе синтеза речи по тексту. Труды первого междисциплинарного семинара «Анализ разговорной русской речи». Санкт-Петербург, 2007.
13. M. Riley, W. Byrne, M. Finke, S. Khudanpur, A. Ljolje, J. McDonough, H. Nock, M. Saraclar, C. Wooters, G. Zavaliagkos. Stochastic pronunciation modeling from hand-labelled phonetic corpora. In Strik (eds.), Speech Communication, Vol. 29, P. 209–224, 1999.

14. J. Kessens. Making a difference on automatic transcription and modeling of Dutch pronunciation variation for automatic speech recognition. Ph.D. dissertation, University of Nijmegen. Netherlands, 2002.
15. А. А. Карпов. Робастный метод определения границ речи на основе спектральной энтропии. Научно-теоретический журнал «Искусственный интеллект», №4. Донецк, Украина, 2004, С. 607-613.
16. В. Л. Цуканова. Опыт применения методов дискурсивно ориентированной транскрипции к материалу неиндоевропейского языка. Труды международной конференции «Диалог 2006». Бекасово, Россия, 2006.
17. V. Podlesskaya, A. Kibrik. Methods of oral speech corpora research: discourse transcription development experience. Proc. of Cognitive Modeling in Linguistics. Varna, Bulgaria, 2003.

Научное издание

Первый междисциплинарный семинар
«Анализ разговорной русской речи» АР³ - 2007

Труды семинара

Составители: А.Л. Ронжин, И.А. Кагиров

Издание осуществлено за счет средств
европейского гранта SIMILAR IST-2002-507609.

Отпечатано в типографии «Микроматикс»
Санкт-Петербург, Большой пр. В.О., д. 55, тел../факс 328-52-63
Подписано в печать 19.07.2007 Печ. л. 5,5 тираж 200 экз., формат 291x207