

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. М. Т. КАЛАШНИКОВА

Информационные технологии и письменное наследие

El'Manuscript-2012

Материалы IV международной научной конференции
Петрозаводск, 3–8 сентября 2012 года

Петрозаводск, Ижевск
2012

УДК 004.9
ББК 81.11+81.2-0
И741

Изданы при поддержке гранта РФФИ (проект № 12-06-06061-г),
гранта РГНФ (проект № 12-04-14154-г) и в рамках реализации комплекса
мероприятий Программы стратегического развития ПетрГУ на 2012-2016 г.

Ответственные редакторы:

В. А. Баранов, д-р филол. наук, проф.

А. Г. Варфоломеев, канд. физ.-мат. наук, доц.

Информационные технологии и письменное наследие [Текст] :
И741 материалы IV междунар. науч. конф. (Петрозаводск, 3–8 сентября
2012 г.) / отв. ред. В. А. Баранов, А. Г. Варфоломеев. — Петрозаводск ;
Ижевск, 2012. — 328 с.

ISBN 978-5-8021-1402-5

Сборник содержит материалы конференции, посвященной современ-
ным электронным средствам хранения, описания, обработки, исследования
и публикации памятников письменности и исторических источников.

УДК 004.9
ББК 81.11+81.2-0

ISBN 978-5-8021-1402-5

© Петрозаводский государственный
университет, 2012
© Ижевский государственный технический
университет им. М. Т. Калашникова, 2012

ТЕХНОЛОГИЯ СТРУКТУРНО-ПАРАМЕТРИЧЕСКОГО СИНТЕЗА СМЫСЛОБРАЗУЮЩИХ КОМПОНЕНТ В СИСТЕМЕ РУССКОГО ЯЗЫКА

М. А. Шаронов, С. В. Моченов

Ижевский Государственный Технический университет, Ижевск

В данной статье авторами сделана попытка свести накопленный материал в единую концепцию для получения теоретического обоснования разработки технологии синтеза смыслообразующих компонент. На базе теоретических положений уточняется многоаспектность моделирования «интеллектуального человеко-машинного языкового интерфейса» и сопутствующие такому моделированию процессы, составляющие первый этап проектирования языкового процессора на основе открытой нейросетевой структуры.

Введение и актуальность

Язык, как семиотическое явление, имеет две стороны: материальную и смысловую. Через материальную сторону языка реализуется его коммуникативная функция. Через смысловую сторону — его когнитивная функция.

Потребность в коммуникации (общении) вынуждает человека использовать некоторый инструментарий: с одной стороны, необходимый для хранения знаний, с другой стороны, необходимый как форма для их передачи.

Основными инструментами формализации знаний человеку служат звуки речи, буквы алфавита языка и невербальные компоненты речевого сообщения. И хотя их относят к материальной составляющей языка, их функция — обслуживать его смысловую составляющую. Более того, на фундаментальном уровне компоненты материальной стороны языка связаны в единую систему, основной целью которой является транспортировка смысловой составляющей сообщения.

Как уже отмечалось, язык — явление системное. А результатом работы системы языка является «знаковый продукт» (т.е. сообщение), элементы которого подчинены законам функционирования системы языка. Именно в системности употребления знаковых элементов и обнаруживается их смысл. Таким образом, нельзя рассматривать речь или текст отдельно от пересекающихся с ними во время воспроизведения или восприятия невербальных компонент.

По этому поводу в работе [Горелов, 2009: 3] И.Н.Горелова неоднократно показано, что «...мы прибегаем к вербальному тексту тогда, когда невербального общения недостаточно». Таким образом, текст сообщения приобретает свой конечный смысл в пересечении с его невербальной составляющей.

Для наглядности (см. рис.1) приведена гипотетическая модель взаимодействия языковых компонент.

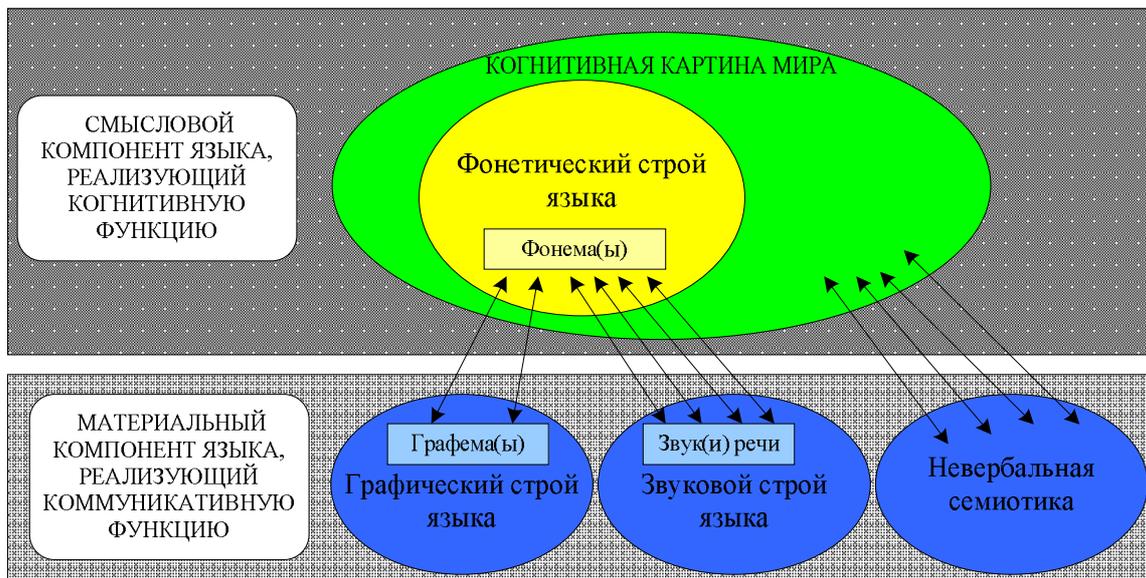


Рис. 1. Гипотетическая модель взаимодействия компонент макросистемы языка

Сама по себе языковая «материя» работать не способна. Равно также и звуки неизвестного вам языка не оживут в вашем сознании образами из привычной предметной реальности.

Современной лингвистике известно, что языковые «механизмы» приводятся в движение действием некоторой фундаментальной «силы», определяемой термином сема. Можно предположить, что сема и есть тот материал, который функционирует в когнитивных структурах мозга и изнутри приводит в действие все знаковые элементы и механизмы.

Очевидно, что материальная оболочка является транспортным уровнем в макросистеме языка. Она обеспечивает транспортировку смысловой (когнитивной) компоненты — семы.

Современный компьютер способен хранить информационные сообщения в файлах различных форматов: в виде текстовых документов, видео и аудио файлов, обеспечивая дополнительными средствами только коммуникативный (транспортный) компонент языка.

Проблема построения интеллектуальных языковых интерфейсов для вычислительных средств остаётся не разрешённой по двум причинам:

С одной стороны мы имеем тенденцию к чрезмерной детализации элементов коммуникативной компоненты языка и её функций.

С другой стороны мы имеем недостаточную проработку вопроса моделирования когнитивных функций, управляющих коммуникативными (транспортными) элементами.

Смысл сообщения по-прежнему извлекается пользователем в процессе восприятия информации, а адекватная модельная реализация процедуры, из-

влекающей смысловую компоненту из языкового сообщения, до сих пор остаётся предметом обширных дискуссий.

Постановка задачи

Поставленная в исследовании научная задача требует разработки аналога орудия коммуникации для «интеллектуального человеко-машинного интерфейса», цель применения которого — обслуживание семы — смысловой (когнитивной) компоненты сообщения.

Интерфейс необходимо адаптировать к работе с семьей как с «силой», приводящей механизмы смоделированной макросистемы языка в движение по аналогии с деятельностью человеческого сознания. Процесс извлечения семы в настоящем исследовании назван технологией синтеза смыслообразующих компонент (далее ТСС).

Отправным пунктом являются представления о структуре формальных текстов на русском языке, а так же представления о статистических закономерностях, наблюдаемых в их знаковой структуре.

Для интерпретации результатов, полученных статистическими методами, а также для понимания того, как происходит сама интерпретация текстов, необходимо выяснить, как такой материальный компонент, как графический алфавит, обслуживает смысловую сторону языка с точки зрения существующих теорий?

Тексты, написанные с применением стандартных алфавитов, оказывают определённое специфическое воздействие на сознание читающего. Проанализируем, в чём же состоит это воздействие? Что происходит в сознании человека во время прочтения текста?

Очевидно, что при прочтении текста происходит сразу несколько когнитивных процессов, так как между собой взаимодействуют сразу несколько уровней языковых объектов: звуковой, фонетический, слоговый, алфавитный, морфемный, словесный и др. От понимания, в каком соотношении находятся между собой объекты этих уровней, зависит адекватность разрабатываемой языковой модели, которую предстоит использовать в качестве орудия коммуникации.

Так же при моделировании необходимо учитывать тот факт, что управление элементами коммуникативной составляющей всегда происходит лишь с целью обслуживания смысловой (когнитивной) составляющей сообщения — семы. Соответствующий выбор коммуникативных средств реализации языкового сообщения предопределяется его когнитивным (смысловым) содержанием. При этом необходимо понимать, что письменность обеспечивает особую реализацию когнитивной компоненты.

Теоретическое обоснование

И.А. Бодуэном де Куртенэ впервые было установлено, «...что для каждого данного языка в сознании говорящих на нём существует не всё бесчисленное множество наблюдаемых исследователями звуковых оттенков, но лишь строго ограниченное их число». Эти звуковые оттенки были названы И.А. Бодуэном де Куртенэ в ранних его работах психологическими эквивалентами звуков, а позже Л.В. Щербой фонемами». [Реформатский, 1970:

128]. Вслед за этим исследователь горских языков Кавказа П.К. Услар в работе [Услар, 1887: 61] «...отмечает, что в каждом данном языке, при всём кажущемся звуковым его многообразии, существует лишь строго ограниченное количество таких звуков, которые служат для различения слов языка». Позже московский исследователь Н.П. Яковлев в своей работе [Яковлев, 1928] констатирует, «...что фонемы — это те звуки, с помощью которых происходит различение слов в языке».

Утверждения касательно роли фонем в структуре языков до сих пор являются дискуссионными в виду неясности природы фонемы. Нельзя точно констатировать: фонема — это звуковая единица или единица сознания и, соответственно, имеет она физические свойства как материальный объект или не имеет.

Приблизиться к ответу на вопрос — «что из себя представляет фонема?» — помогает теория «фонологических квантов» предложенная В.Я. Плоткиным [Плоткин, 1993]. Данная работа возвращает нас к идеям И.А. Бодуэна де Куртенэ.

В своих работах «большое внимание он уделил тем компонентам языка, которые начинают и завершают фонационную деятельность человека». [Плоткин, 1993: 11]. Словами самого Бодуэна первоочередной задачей фонетической науки является изучение «не столько звуков как акустических продуктов известных физиологических работ, сколько скорее самих этих физиологических работ, или артикуляций. Иначе: при решении нашей задачи стоят на первом плане не звуки, а только их производство, фонация» [Бодуэн де Куртенэ, 1963: 256]

Рассмотрим процесс фонации на начальном этапе. Для порождения определённого набора артикуляций человеческому мозгу необходимо выработать в нервной системе сигналы к управлению движением речепроизводящих органов. В результате таких движений фонема приобретает материальную оболочку и становится звуком речи.

В современной фонетике данный аспект речепроизводства изучается артикуляционным методом [Осипов, 1992: 5]. Данный метод представляет фонему как единицу системы языка с точки зрения порождающих её артикуляций. Исчерпывающее описание метода артикуляции дано в [Осипов, 1992: 14–20].

Сложность фонетической системы конкретного языка может порождать определённые трудности при создании для него графического алфавита, а в сопоставлении с другими языками — обуславливает наличие алфавитов неодинакового размера.

Фонемы языка опознаются и различаются коммуникантами по наборам дифференциальных и интегральных признаков (далее ДП и ИП) в различных их противопоставлениях.

Такие признаки отличают фонему на фоне всего фонетического строя системы языка и являются характерными её конфигурациями, которые сознание коммуниканта способно фиксировать.

Также, благодаря ДП и ИП фонема является наименьшим элементом, способным обслуживать передачу смысловой компоненты. Следовательно, передача смысловой компоненты основана на способности коммуниканта точно опознавать и различать индивидуальные особенности каждой фонемы в системе языка.

Далее следует этап транспортировки языкового сообщения. В итоге, сообщение принимается коммуникантом с помощью органов слуха. Если процесс фонации рассматривать на последнем его этапе, то мы увидим, что из полученного сообщения слуховые органы снова извлекают сигналы управления — конфигурации звуков, которые могут быть вновь использованы нервной системой.

Принципы формирования и извлечения последовательностей таких сигналов, обеспечивающих артикуляцию и слуховое (акустическое) восприятие, представляют особый научный интерес при использовании данного аспекта в производстве языковых технологий, так как содержат в себе множество ответов на вопросы касательно организации деятельности моделей макросистемы языка в целом.

Согласно приведённому теоретическому обоснованию, любой языковой знак имеет особую конфигурацию реализующих его нервных импульсов, которая задаётся нервной системой и определяет знак в системе языка.

Фонема в данном аспекте выступает как управляющий элемент, функционирующий как вспомогательный центр регулирования нервной деятельности.

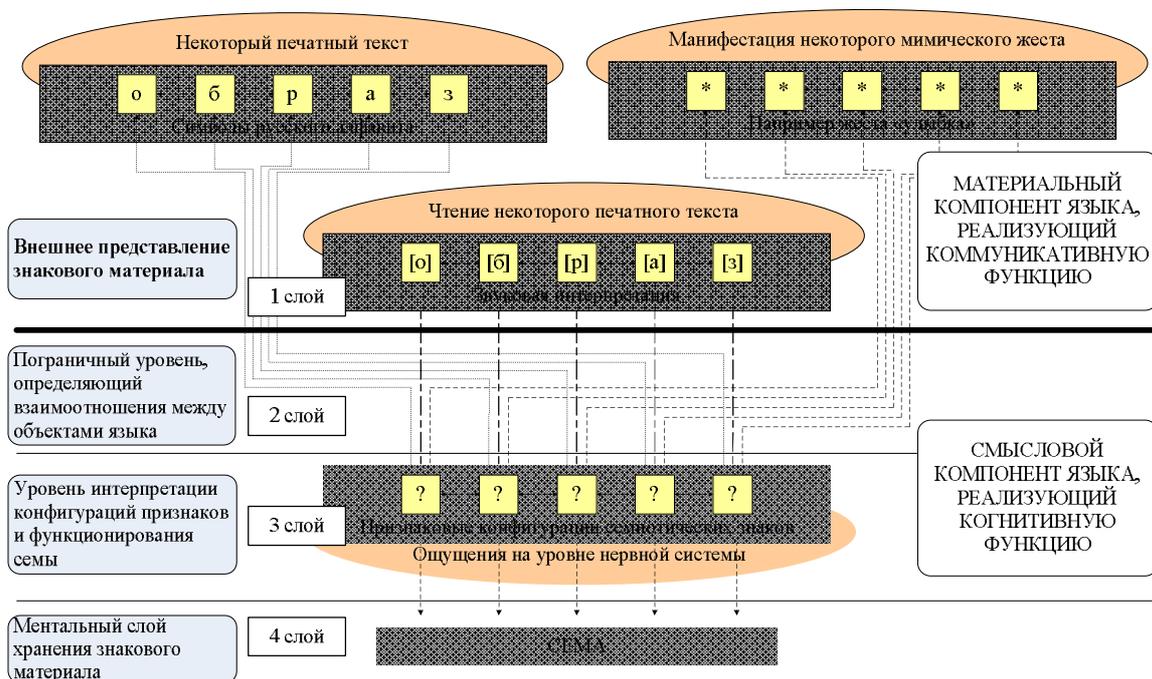


Рис. 2. Гипотетическая модель интерпретации семиотических знаков

Построение гипотетической модели

Основными функциями фонемы как управляющего элемента в макросистеме языка являются производство определённого звука речи или его

идентификация. Такими же функциями фонема обладает и по отношению к буквам алфавита.

Фонетический строй современного русского языка конструктивно значительно богаче его графического строя. Фонетика русского языка, по различным сведениям, насчитывает от 39 до 43 элементов. Графические элементы опосредованно воздействуют на элементы фонетического уровня через пограничные звуковые и слоговые конструкции. Из-за такого посредничества графический символ языка оказывается лишь грубой абстракцией фонемы, а не её прямым начертательным образом. Поэтому в текстах не наблюдается полной корреляции между последовательностями букв алфавита и их звуковыми эквивалентами.

Графика русского языка состоит из 33 элементов. Обычно графический алфавит через знаки-буквы стремится обозначить фонемы языка, и хотя такое однозначное соответствие наблюдается редко и имеет обыкновение утрачиваться в процессе развития устного языка — средств алфавита на практике оказывается достаточно для кодирования языковых сообщений. «... Звуки, ... выделяемые к значению, ... варианты фонем, ... звуковые оттенки, появляющиеся под влиянием смены соседних звуков» [Яковлев, 1926] являются скрытым за графическими символами алфавита материалом.

Последнее утверждение характеризует динамичность фонетического строя по сравнению со статичным характером графики языка и оправдывает необходимость разработки ТСС, обеспечивающей извлечение необходимого смыслообразующего материала и приводящей в соответствие структуру взаимоотношений между элементами языка разных уровней.

Фонема в нашем понимании — это некоторый синтетический объект (аналогия нервному центру), реагирующий на определённую конфигурацию признаков. Если фонема (нервный центр) обнаруживает допустимую конфигурацию признаков, то раздражающий звук распознаётся коммуникантом как данная фонема и, следовательно, функционирует далее как знаковый элемент языка.

С другой стороны, при чтении, раздражителем, действующим на фонему (нервный центр) выступают группы признаков, соответствующие звуковой интерпретации элементов текста, состоящих из букв алфавита. Таким образом, графические символы через признаки звуков ассоциируются с соответствующими им фонемами.

Графическое начертание языка в этом случае хоть и является грубой абстракцией его «фонетического» строя, для коммуниканта этой абстракции оказывается достаточно, чтобы оперировать графическими символами (имея в виду вполне конкретные психологические представления звуков) как фонемами и тем самым уменьшить количество графических элементов русского языка с 43 до 33.

Исходя из сказанного видно, что у человека фонетическое (шире — когнитивное) представление о языке первично, а графическое (символьное) — вторично. В этой связи одной из приоритетных целей при проектировании языкового интерфейса становится установление соответствий между

графическим, звуковым, слоговым и фонетическим строем языка через при-
знаковые характеристики фонем.

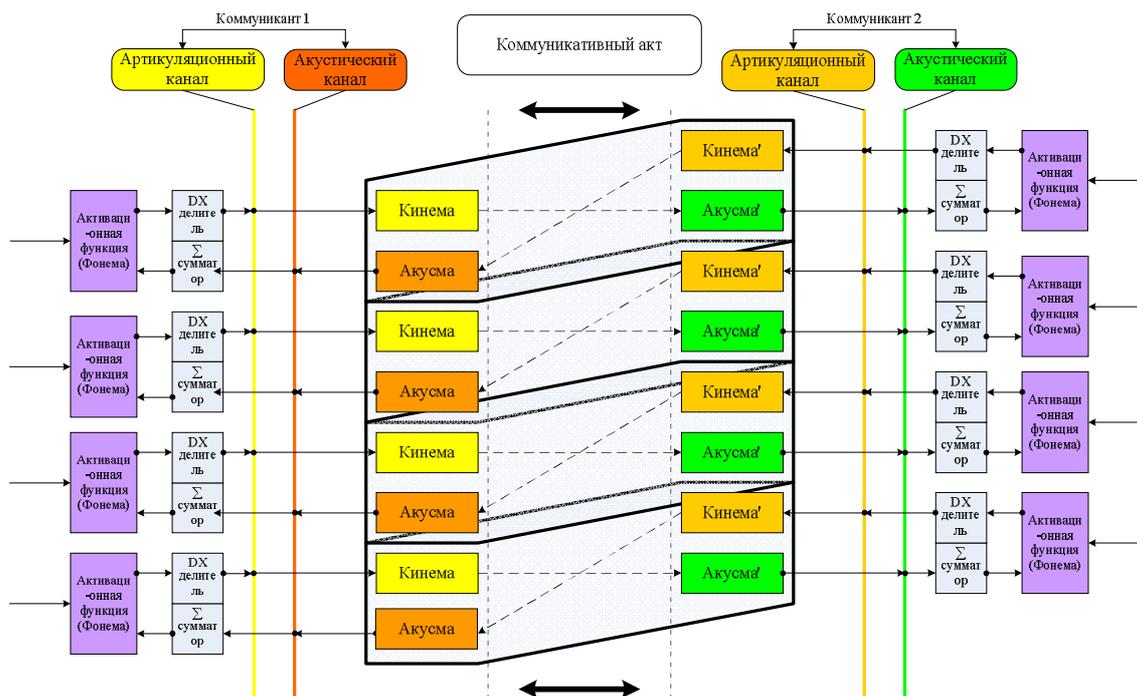


Рис. 3. Гипотетическая модель процесса фонации

В предлагаемом подходе к моделированию деятельности «интеллектуального человеко-машинного интерфейса» предлагается использовать вышеописанные конфигурации знаков как аналоговые характеристики объекта языковой реальности. С объектом языковой реальности таким образом можно отождествлять любую информацию, поступающую на вход моделируемого интерфейса. Наличие или отсутствие внутри анализируемого объекта этих признаков в определённой конфигурации будет строго указывать на то, с каким элементом языковой действительности (с какой фонемой) мы имеем дело.

Возвращаясь к моделированию когнитивной функции, осуществляющей сопоставление графического и фонетического строя языка, следует сказать, что суть такого моделирования состоит в определении фонетического строя графического письменного сообщения по вышеописанным признаковым конфигурациям фонем.

Искомые конфигурации признаков фонетического строя языка на практике оказываются завуалированными внутри графической конструкции сообщения, но за счет функционирования именно этих признаков осуществляется транспортировка смысла сообщения.

В аспектах анализа текстовых документов, синтез и поиск различительных (конфигурирующих) признаков фонем — чрезвычайно важные задачи фундаментального характера, обнажающие представление о передаче и извлечении смысловых компонент языка, так как выявляют скрытые графическими символами языка взаимоотношения между смыслообразующими элементами и элементами плана выражения.

Список литературы

- Бодуэн де Куртенэ, 1963 — Бодуэн де Куртенэ И.А. Избранные труды по общему языкознанию. Т. I. М., 1963.
- Горелов, 2009 — Горелов И.Н. Невербальные компоненты коммуникации. М., 2009.
- Зиндер, 1979 — Зиндер Л.Р. Общая фонетика. М.: Изд-во «Высшая школа», 1979.
- Комарцова и Максимов, 2002 — Комарцова Л.Г., Максимов А.В. Нейрокомпьютеры. М.: Изд-во МГТУ им. Н.Э.Баумана, 2002.
- Корочков, 2009 — Корочков А.В. Компьютерное моделирование графемно-фонемного преобразования в английском языке (на основе традиционных правил чтения). Саранск, 2009.
- Львовский, 1982 — Львовский Е.Н. Статистические методы построения эмпирических формул. М., 1982.
- Матусевич — Матусевич М.И. Современный русский язык. М., 1976.
- Осипов, 1992 — Осипов Б.И. Краткий курс русской фонетики из лекций по современному русскому языку. Омск, 1992.
- Плоткин, 1993 — Плоткин В.Я. Фонологические кванты. Новосибирск, 1993.
- Реформатский, 1970 — Реформатский А.А. Из истории отечественной фонологии. М., 1970.
- Трубецкой, 2000 — Трубецкой Н.С. Основы фонологии. М., 2000.
- Услар, 1887 — Услар П.К. Абхазский язык. Тифлис, 1887.
- Холл и Боденхамер, 2007 — Холл М., Боденхамер Б. Полный курс НЛП. Спб., 2007.
- Щерба, 1983 — Щерба Л.В. Русские гласные в качественном и количественном отношении. Л., 1983.
- Яковлев, 1928 — Яковлев Н.Ф. Математическая формула построения алфавита // Культура и письменность Востока. Книга I. М., 1928. С. 41–64.