

ББК 81
УДК 81.23
П 78

Печатается по решению научно-технического
совета Череповецкого государственного университе-
та, протокол № 1 от 20.09.2017.

Ответственный редактор: *Е.В. Грудева*, доктор филологических наук, профессор, зав. кафедрой отечественной филологии и прикладных коммуникаций Гуманитарного института Череповецкого государственного университета

П 78 Проблемы порождения и восприятия речи: Материалы XIV выездной школы-семинара (г. Череповец, 1–3 декабря 2016 г.) / Отв. ред. Е.В. Грудева. Череповец: Череп. гос. ун-т, 2017. 144 с.
ISBN 978-5-85341-763-2

Сборник включает статьи как профессиональных, так и начинающих лингвистов, принимавших участие в работе XIV выездной школы-семинара «Порождение и восприятие речи». В сборник вошли работы, посвященные проблемам порождения и восприятия звучащей речи и письменных текстов; изучению дискурса и коммуникации, а также вопросам билингвизма и кодовых переключений.

Издание адресовано специалистам, а также аспирантам, магистрантам и студентам, специализирующимся в области психо- и онтолингвистики, билингвизма и кодовых переключений, а также всем тем, кто интересуется вопросами изучения дискурса и коммуникации.

ББК 81
УДК 81.23

ISBN 978-5-85341-763-2

© ФГБОУ ВО «Череповецкий государственный университет», 2017

МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ПОРОЖДЕНИЯ И ВОСПРИЯТИЯ РЕЧИ

А.В. Венцов

Санкт-Петербургский государственный университет

Моделирование процессов порождения и восприятия речи: проблемы

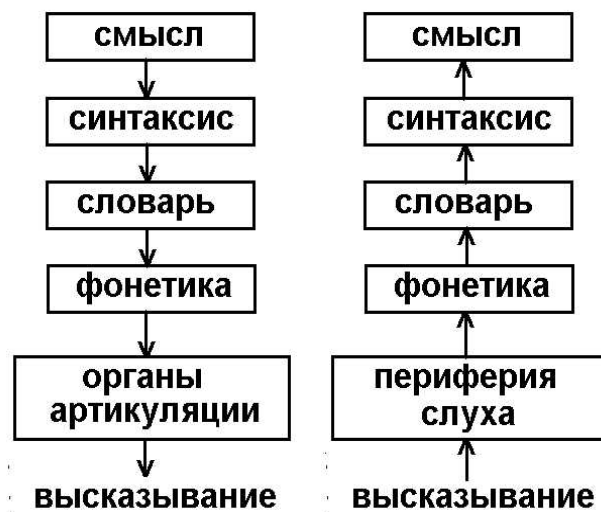
Когда говорят о моделировании процессов порождения и восприятия речи, следует иметь в виду, что из-за специфики объекта речь на самом деле идет о создании **функциональных моделей** этих речевых систем. Природа этих моделей требует детального описания операций, предположительно имеющих место при обработке информации в системах порождения и восприятия речи и обеспечивающих «преобразование» некоторого смысла в акустический речевой сигнал для говорящего и извлечение некоего смысла из речевого сигнала для слушающего.

Создание подобной модели, помимо получения новых научных знаний, может оказаться полезным для дефектологии, для обучения языку, при решении некоторых технических задач (автоматический синтез и распознавание речи).

Порождение и восприятие речи представляют собой многозвенные «закрытые» системы с практически исключенным прямым доступом ко всем протекающим в них внутренним процессам (рисунк).

Одним из способов исследования таких систем является «метод черного ящика». Конструкт «черный ящик» рассматривается как имеющий некий «вход» для ввода информации и «выход» для отображения результатов работы. При этом предполагается, что состояние выходов функционально зависит от состояния входов, но

происходящие в системе процессы наблюдателю принципиально неизвестны.



Условные схемы модели порождения (слева)
и восприятия (справа) речи

Если рассматривать когнитивные системы восприятия и порождения речи как «черные ящики», то приходится признать, что у каждой из них имеется не один «вход», а множество «внешних» (и речь — только один из них) и, в дополнение, неизвестное число «внутренних» (сумма всех знаний носителя языка, его способность выбирать оптимальную стратегию поведения применительно к конкретной ситуации, не контролируемые экспериментатором его подсознательные внутренние установки и т.д.). В таком случае для успешной интерпретации «выхода» с необходимостью следует представлять себе содержание «сигналов» на всех «входах».

При этом весьма распространенным является «представление о деятельности слушающего как о зеркальном отражении деятельности говорящего: говорящий идет от смысла к форме (звуковой или буквенной), а реципиент, наоборот, идет от звуков или букв к словам и складывает из тех целые предложения с искомым смыслом» (Норман 2011: 103).

Однако указанная «зеркальность» проявляется лишь в том банальном обстоятельстве, что, действительно, сначала некий смысл (известный только говорящему) преобразуется в речь, а затем речь

преобразуется в некий смысл (известный только слушателю). При более глубоком подходе ни о какой «зеркальности», разумеется, речи быть не может, поскольку последний модуль системы порождения речи (органы артикуляции на рис.) и первый модуль в системе восприятия (периферия слуха там же) и физиологически, и функционально не имеют между собой ничего общего.

Более того, человек обладает врожденной способностью анализировать звуки (в частности, речь). В онтогенезе сначала формируется система восприятия речи, только позднее – система речеобразования. Благодаря этому обеспечивается первичность слуха: субъект с моторной алалией может воспринимать речь, а от рождения глухой не может естественным путем научиться говорить.

Таким образом, мы сталкиваемся с практически неразрешимой проблемой: в системе порождения речи доступен только «выход» – акустический речевой сигнал, а в системе восприятия речи – только «вход», т.е. тот же акустический речевой сигнал. Какой смысл собирался выразить в речевом сигнале говорящий, т.е. «вход» системы порождения речи, строго говоря, неизвестно, но так же неизвестно и то, какой смысл («выход») извлек в конце концов слушатель из конкретного высказывания, если только этот смысл не проявляется в тех или иных внешних реакциях слушателя. Ведь порождает речевой сигнал один субъект, а воспринимает – другой, т.е. действуют две разные физиологические и когнитивные системы со своими собственными «знаниями» и «опытом».

В сущности, результаты любых фонетических инструментальных исследований соотносятся не с замыслами носителя языка, создавшего анализируемое высказывание, а с тем, как это высказывание воспринято самими исследователями. Поэтому субъективные оценки экспериментаторов, видимо, нельзя признавать единственно правильными, и вообще, вероятно, далеко не всегда есть возможность достоверно оценить, что же именно хотел сказать говорящий.

Помимо описанных «глобальных» проблем моделирования речевой деятельности человека, при описании гипотетических процессов в каждом отдельном «модуле» обсуждаемых систем возникают свои «локальные» проблемы.

Способность периферической слуховой системы (наружное, среднее и внутреннее ухо, преобразование сигнала на базилярной мембране) анализировать спектральную картину речевого сигнала (Чистович и др. 1976), а системы нейронов восходящего слухового пути – маркировать моменты резких изменений в сигнале (Чистович и др. 1986) обеспечивает возможность сегментации непрерывного речевого сигнала на интервалы «гласных» и «согласных» (не путать с соответствующими фонемами!) с одновременной фиксацией признаков, характеризующих свойства этих звуков.

Поскольку собственно акустический сигнал не может храниться в памяти долго, система преобразует акустические параметры каждого такого интервала в некоторое «символьное» описание, хранящее в сжатом виде всю информацию о звуке.

Предполагается, что, слушая звучащую речь, носитель русского языка «...должен представить ее в виде последовательности фонем (по Л.В. Щербе. – *А. В.*), т.е. преобразовать разнообразные акустические характеристики услышанного, свести их к возможностям фонемного инвентаря, имеющегося в его распоряжении» (Бондарко 1977: 115).

Каков же этот фонемный инвентарь? Могут ли это быть фонемы, принятые в теоретической лингвистике?

Практика акустико-фонетического транскрибирования русской спонтанной речи свидетельствует, что согласные фонемы не имеют аллофонов. Об этом же говорят и некоторые литературные данные. «Например, русские согласные [p], [t], [к] в позиции абсолютного конца слова – [dup], [vot], [vrak] – произносятся как сильно придыхательные» (Бондарко и др. 1991: 9), но придыхательность при этом представляет собой скорее дополнительный оттенок звучания и не влияет на идентификацию основного фонемного качества согласного.

«Сильное артикуляторно-акустическое различие между лабиализованными и нелабиализованными аллофонами согласных „бесполезно“ при восприятии, ибо огубленные гласные надежно отличаются от неогубленных по своим собственным характеристикам. Именно поэтому разные по огубленности аллофоны согласных в

перцептивном отношении как бы одинаковые» (Бондарко и др. 1991: 92).

Оглушенные звонкие согласные не являются аллофонами звонких согласных, а представляют собой соответствующие парные глухие (Венцов 2015).

Иначе обстоит дело с гласными. «Так, носители русского языка опознают и по-разному обозначают не 6 гласных (по числу гласных фонем), а 18 разных звуковых единиц (эталонов): комбинаторные аллофоны одной и той же фонемы в соседстве с твердыми или мягкими согласными опознаются как разные по качеству гласные» (Бондарко и др. 1991: 91). Прибавим к этому тот факт, что очень частотный аллофон «шва» ([ъ], [э]) в естественной речи может «представлять» сразу несколько фонем (/a/, /y/, /э/, /ы/) и до принятия лексического решения нет никакой возможности определить, какую именно.

Если исходить из определения фонемы, данного Бодуэном де Куртенэ («Фонема = единое представление, принадлежащее миру фонетики, которое возникает в душе (носителя языка. – А. В.) посредством слияния впечатлений, полученных от произношения одного и того же звука, – психический эквивалент звуков языка» (Бодуэн де Куртенэ 1963: 271)), аллофон «шва» следует включить в качестве самостоятельной единицы в набор «символов» (психологических фонем), которым оперирует слушатель при первичном описании речевого сигнала.

Таким образом, приходится признать, что при первичном «символьном» описании акустического речевого сигнала носитель языка может пользоваться набором «символов», существенно отличающимся от принятого в теоретической лингвистике (фонологии). Проблема в том, что точный объем и состав этого набора неизвестен, как неизвестны и процедуры его формирования в процессе освоения устной речи.

Далее. Если с помощью подобного набора символов описан некий отрезок речевого сигнала, возникает необходимость лексического членения такого отрезка. Все предложенные на сегодня модели сегментации непрерывного речевого потока на лексические

единицы предполагают использование внутреннего (ментального) словаря (McQueen et al. 1995; Weber, Scharenborg 2012). К сожалению, в публикациях отсутствуют какие-либо сведения об эффективности предлагаемых моделей при работе с естественным речевым материалом.

Чтобы лексический поиск и сегментация были успешными, следует допустить, что единицами этого словаря должны быть словоформы, описанные тем же набором «символов», которыми описывается речевой сигнал на выходе первичного слухового анализатора. Трудности появляются при определении внутренней структуры такого словаря.

Если единицами словаря являются словоформы в кодифицированной «фонетической» транскрипции (с учетом описанного выше разнообразия психологических фонем), то объем словаря оказывается предсказуемым, но очевидным образом возникает проблема идентификации редуцированных словоформ, составляющих значительную часть естественного речевого потока: так, например, едва ли окажется простой процедура идентификации редуцированной формы *ситна* [s'itna] с соответствующим кодифицированным вариантом *действительно* [d'ijstv'it'il'nə].

С другой стороны, еще сто лет назад Л.В. Щерба отмечал, что в словаре могут храниться все возникающие в речи варианты словоформ: «Выше я сказал, что для нашего сознания в большинстве случаев ясен идеальный фонетический состав слов – значит, не всегда. В самом деле, представим себе, что ребенок никогда не слышал отчетливого произношения слова *говорит*, а слышал лишь формы: gər'it и gr'it; он легко может себе представить на основании опыта со словом *мэчит* (*мычит*), что идеальная форма слова будет gyr'it, и если никто не поправит его соответственного отчетливого произношения, то он так и останется с *гырит* вместо *говорит*; но если у него будет смутное воспоминание о gəvAr'it, то сознание может колебаться, могут возникнуть две параллельные формы и т.п. Так, например, по-моему, у нас обе формы – и *здравствуйте* и *здрасьте* – существуют в сознании, тогда как того же нельзя сказать про *говорит* и *грит*, хотя это последнее зафиксировано даже и в литера-

туре: *grit*, по крайней мере мною, чувствуется как диалектизм» (Щерба 1957 (1915): 23).

Другими словами, Щерба пишет о том, что в ментальном лексиконе некоего воображаемого носителя языка может вообще не быть редуцированных форм, если в его речевом пространстве они не встречаются достаточно часто, но при этом хранятся какие-то признаки, характеризующие речь конкретного говорящего, что позволяет узнавать его по голосу и даже имитировать его речь.

Тогда возникает вопрос: а что же в действительности происходит с редуцированной формой в словаре, когда в нем появляется соответствующая полная форма? Удаляется ли редуцированный вариант из словаря? Если удаляется, то по каким алгоритмам он «восстанавливается» при лексическом поиске и сегментации, а если нет, то почему редуцированные формы не опознаются аудиторами при изолированном предъявлении?

Вместе с тем при наличии в словаре всех вариантов словоформ их идентификация, а с ней и сегментация непрерывного речевого потока, существенной проблемы не представляет, но при этом значительно увеличивается объем необходимой памяти для хранения словаря (правда, реальные возможности человеческой памяти все еще неизвестны, и есть основания считать, что они практически неисчерпаемы), а главное – остается непонятным механизм пополнения словаря новыми редуцированными вариантами каждой словоформы. Чтобы включить новый вариант в описание уже существующей в словаре конкретной словоформы, необходимо прежде всего выделить его как «новое» слово, а затем установить, вариантом какой именно словоформы данная звуковая последовательность является. Это, очевидно, окажется возможным лишь после «восстановления» значения этого «нового» слова в результате грамматического и семантического анализа всего контекста.

Таким образом, объединяющим фактором для всех редуцированных вариантов одной словоформы является ее лексико-грамматическое значение.

К сожалению, о функциональном моделировании процедур лексико-грамматического анализа на сегодняшний день литература молчит.

Итак, описанная выше «перцептивная» часть ментального лексикона обеспечивает сегментацию непрерывного речевого потока, идентификацию лексических единиц и в конечном счете – понимание речевого сигнала в условиях естественного речевого общения.

Как следует из рисунка, должна существовать и другая его часть, участвующая в порождении речевого сигнала, – «речеобразовательная». Иногда высказывается мнение, что обе системы обладают одним и тем же, общим, словарем.

Гипотетическую структуру перцептивного словаря оказывается возможным построить исходя из особенностей его функционального назначения, при этом такой структуры вполне достаточно и для решения задачи функционального моделирования процесса понимания речи. Возможно, такой подход поможет понять и принципы построения словаря, используемого при порождении речи.

В процессе овладения членораздельной речью ребенок имитирует речь окружающих взрослых, «подбирая» методом проб и ошибок такие движения и положения органов речеобразования, которые обеспечивают появление адекватного (либо с точки зрения акустических характеристик, либо с точки зрения способности вызывать требуемые реакции окружающих) акустического сигнала. В результате у него вырабатывается набор и структура команд для управления этими органами, а их реализация становится автоматической (подобно тому, как это происходит в процессе выработки автоматизмов у пианистов, фигуристов, у всех, умеющих ездить на велосипеде, наконец, у всех людей, умеющих ходить на двух ногах). “A speech motor program is defined herein as a set of neurons and their interconnections that, when activated, generate a learned set of articulator movements that carry out a speech sound” (Guenther 2016: 107).

Таким образом, можно полагать, что каждая словоформа порождаемого речевого сообщения описывается своим набором артикуляторных команд и некоторым лексико-грамматическим значением, вызывающим определенные реакции окружающих, а вместе все они образуют «речеобразовательный» словарь. Если в речевом пространстве ребенка встречаются редуцированные формы, они должны появляться в его речи и словаре наравне с полными формами.

При этом подходе модуль «фонетика» в структуре модели порождения речи (рис., слева) оказывается лишним.

Проверить справедливость такой гипотезы и, соответственно, работоспособность модели можно, сопоставив акустические параметры речи в окружающей ребенка среде с акустическими параметрами речи самого ребенка. Проблема состоит в том, что для подобного анализа нет материала: речевые корпуса, одновременно фиксирующие полное речевое окружение ребенка и его собственную речь, к сожалению, отсутствуют.

Это отсутствие не позволяет решить еще одну проблему: чью именно речь и каким образом имитирует ребенок. Известно, что частоты формант гласных в детской речи далеко не совпадают с таковыми в речи, скажем, мужчин. Значит, можно предполагать, например, что успешно имитировать речь окружающих будет невозможно, если среди них преобладают мужчины.

В моделях, предлагаемых психолингвистами, подобная структура словаря допускается лишь применительно к ранним стадиям овладения речью: "...word production emerges from a coupling of two initially independent systems, a conceptual system and an articulatory motor system" (Levelt et al. 1999: 1), а затем из-за ограниченного (!) объема памяти "...it is increasingly hard to keep all the relevant whole-word gestures apart. The child conquers this strain on the system by dismantling the word gestures through a process of phonemization; words become generatively represented as concatenations of phonological segments" (там же).

При этом единицами словаря оказываются лексемы (леммы), причем, насколько можно понять из структурных схем предлагаемых моделей, в графической форме, а затем обычно следует типичная цепочка преобразований: морфология – фонология – фонетика – слоговоеделение – словарь слоговых моторных команд.

О возникающих при этом критических вопросах речь обычно не ведется. Как и по каким правилам ребенок, еще не полностью овладевший речью, осуществляет лемматизацию? Откуда могут появиться графические формы лексем у носителей языка, не знающих грамоты, а у ребенка – до овладения письмом? Существуют ли универсальные правила слоговогоделения? И т.д.

Еще одна заслуживающая упоминания проблема заключается в том, что описание моделей (или даже теорий) в психолингвистической литературе обычно сводится к перечислению того, ЧТО будет происходить на каждом этапе, однако ничего не говорится о том, КАК это будет происходить.

Литература

1. Бодуэн де Куртенэ И.А. Опыт теории фонетических альтернатив. Глава из психофонетики // Избранные труды по общему языкознанию. Т. 1. М.: Изд-во АН СССР, 1963. С. 265–347.
2. Бондарко Л.В. Звуковой строй современного русского языка. М.: Просв., 1977. 175 с.
3. Бондарко Л.В., Вербицкая Л.А., Гордина М.В. Основы общей фонетики: Учеб. пособие. СПб.: СПбГУ, 1991. 152 с.
4. Венцов А.В. Важен ли признак «звонкости» согласных для восприятия речи? // Проблемы порождения и восприятия речи: Материалы XIII выездной школы-семинара (г. Череповец, 4–6 декабря 2014 г.) / Отв. ред. Е.В. Грудева. Череповец: ЧГУ, 2015. С. 6–11.
5. Норман Б.Ю. Основы психолингвистики: Курс лекций. Минск: БГУ, 2011. 131 с.
6. Чистович Л.А., Венцов А.В., Гранстрем М.П. и др. Физиология речи. Восприятие речи человеком. Л.: Наука, 1976. 388 с.
7. Чистович Л.А., Венцов А.В., Люблинская В.В., Столярова Э.И., Чистович И.А. Слуховые уровни восприятия речи. Функциональное моделирование // Акустика речи и слуха: Сб. науч. работ / Ред. Л.А. Чистович. Л.: Наука, 1986. С. 97–127.
8. Щерба Л.В. О разных стилях произношения и об идеальном фонетическом составе слов // Щерба Л.В. Избранные работы по русскому языку. М.: Учпедгиз, 1957. С. 21–25.
9. Guenther Frank H. Neural Control of Speech. MIT Press, 2016. 424 p.
10. Levelt Willem J. M., Roelofs A., Meyer Antje S. A theory of lexical access in speech production // Behavioral and Brain Sciences. 1999. Vol. 22. P. 1–75.
11. McQueen James M., Cutler A., Briscoe T., Norris D. Models of Continuous Speech Recognition and the Contents of the Vocabulary // Language and Cognitive Processes. 1995. Vol. 10. No 3/4. P. 309–331.
12. Weber A., Scharenborg O. Models of spoken-word recognition // WIREs Cogn. Sci. 2012. Vol. 3. P. 387–401 (doi: 10.1002/wcs.1178).