

БОНДАРКО Л. В.

ФОНЕТИЧЕСКИЙ ФОНД СОВРЕМЕННОГО РУССКОГО ЯЗЫКА

В контексте обсуждаемой проблематики фонетический фонд понимается как машинная реализация сведений, касающихся звукового уровня системы русского языка и необходимых для изучения этого уровня с самых разных точек зрения — от анализа вариативности акустических характеристик речевых сигналов и перцептивной оценки этой вариативности до количественной интерпретации функциональной нагрузки фонем при образовании и различении морфем и словоформ или создания модели интонационной организации высказывания или текста.

Машинная реализация фонетического фонда — не дань компьютерной моде, а единственный реальный способ осуществить целенаправленный и систематический поиск в интересующей нас области. Наличие фонетического фонда — одно из необходимых условий для создания Машинного фонда современного русского языка, поскольку именно фонетический фонд может обеспечить «озвучивание» любого из подфондов (словообразовательного, синтаксического и др.) в случае, когда возникнет необходимость выяснения основных тенденций в развитии того или иного уровня языковой системы в связи с реальными факторами, связанными с речевой деятельностью на русском языке.

Фонетический фонд, являясь одним из элементов Машинного фонда русского языка, представляет интерес и как самостоятельная структура, необходимая для решения не только теоретических, но и практических задач — таких, как обучение фонетике русского языка как неродного, социолингвистические исследования на звуковом уровне, проблемы автоматического распознавания и синтеза речи.

По своему содержанию¹ в собственно лингвистическом отношении фонетический фонд может быть охарактеризован следующим образом:

1. Набор фонем и их релевантные признаки известны, как и правила фонемных чередований, аллофонических изменений и основные виды вариативности в пределах нормы [1—3].

2. Фонетический фонд обеспечивает такое представление имеющихся данных, которое необходимо для проверки спорных вопросов на всех уровнях языковой организации — от собственно фонемной системы до реализаций спонтанной речи [4, 5].

3. Фактический материал, являющийся содержанием фонетического фонда, должен быть представительным и лингвистически корректным. Это обеспечивает возможность оценки теоретической интерпретации исходных предпосылок (п. 1) и «конкуренцию» при решении спорных вопросов: только то решение может считаться более адекватным истине, которое объясняет свойства материала, содержащегося в фонетическом фонде.

¹ Примеры и все дальнейшее изложение строится на материале сегментных единиц, однако общая структура предполагает и адекватное включение просодики, и учет взаимодействия сегментных и супraseгментных единиц.

Как любая машинная система, фонетический фонд содержит не только материал, но и программы, обеспечивающие его ввод, хранение, обработку и расширение как за счет увеличения количества звуковых реализаций, так и за счет введения новых сведений, полученных в результате обработки данного материала. В настоящей статье описываются только характеристики материала, составляющего основу фонетического фонда.

Предлагаемая здесь структура фонетического фонда русского языка не претендует на то, чтобы считаться окончательной, тем более, что она вплоть до настоящего времени больше ориентирована на проблемы сегментных единиц. Представляется, однако, что введение дополнительных блоков, необходимых для интонационного уровня, не изменит принципиальным образом общую структуру фонетического фонда.

Фонетический фонд состоит из четырех блоков, связанных между собой и в то же время представляющих самостоятельные «пучки» информации.

Первый блок — акустические характеристики звуковых единиц русского языка. В качестве материала здесь выступает цифровое представление речевых сигналов различной протяженности, различной структуры и функциональной нагруженности. В настоящее время этот материал содержит слоги типа CV, представляющие сочетания всех русских согласных со всеми гласными, около 150 слов, интересных в орфоэпическом плане (т. е. такие, где возможны различные звуковые реализации в пределах нормативного произношения, например, /sv'ičá/ — /s'v'ičá/ и т. д.), а также текст, составленный на основе 200 наиболее частых слогов, характерных для связной речи. Этот материал записан на магнитную ленту в произношении четырех дикторов — двух москвичей и двух ленинградцев (для каждого варианта произносительной нормы записывается произношение диктора-мужчины и диктора-женщины). Осуществляется процедура оцифровки записанного материала и его сегментации, которая необходима для того, чтобы из всего записанного материала можно было выбирать интересующие исследователя звуковые последовательности.

Текст оценен как с точки зрения представительности при анализе аллофонических изменений (т. е. определено число позиций и качество фонетических контекстов для гласных и согласных), так и с точки зрения лексико-морфологической представительности: слова, содержащие частотные слоги, подбирались на основе частотного словаря русского языка В. В. Морковкина с соавторами [6], а морфологические структуры словформ текста сопоставлены с данными типичных морфологических структур в словаре русского языка [7].

Второй блок фонетического фонда — машинная версия словаря морфем современного русского языка. В настоящее время основой являются два словаря — RDD [8] и Грамматический словарь А. А. Зализняка [9]. На материале этих словарей, где каждое слово представлено как последовательность морфем, произведен анализ некоторых существенных фонетических признаков морфем, в первую очередь — корневых. Определены наиболее типичные обобщенные фонетические структуры (в терминах C и V), получены данные об их распределении в таких частях речи, как существительное, прилагательное, глагол, наречие. Получены сведения о статистике корневых морфем с разным числом слогов и разными ударными гласными. Составлены списки корневых морфем с историческими чередованиями и произведена их классификация (с указанием количества корней с чередованием каждого типа). Созданные в ходе исследования программы позволяют получить подобные сведения для

любого вида морфем (частично имеются соответствующие сведения относительно префиксов). Поскольку в настоящее время количество машинных версий словарей морфем значительно увеличилось, окончательное формирование второго блока завершится в результате сопоставления и корректировки как упомянутых двух словарей, так и словаря А. И. Кузнецовой и Т. В. Ефремовой [10], частотного словаря корневых морфем Л. В. Кулешовой [11] и словаря служебных морфем, необходимость создания которого становится все более очевидной. Заполнение второго блока фонетического фонда позволяет получать практически исчерпывающие сведения о дистрибуции фонем внутри морфемы, об их функциональной активности, о законах оформления краевых позиций морфем и вокалических позиций по отношению к ударению. Такие сведения необходимы для того, чтобы выйти за рамки привычной аксиоматики, основанной на ограниченном круге примеров, кочующих из одного фонологического эссе в другое, и оценить количественно те фонологические процессы, которые характеризуют полный список морфем современного русского языка. Возможность оперировать с полным (или почти полным) перечнем морфем русского языка имеет решающее значение для понимания законов звуковой организации морфемы, с одной стороны, и для понимания собственно фонологических явлений — с другой. Это особенно актуально для отечественной фонологической традиции: представители одного из двух ведущих фонологических направлений — Московской фонологической школы (основной принцип, постулируемый сторонниками этого направления, — неизменность фонемного состава морфем) в последние десятилетия уделяют особое внимание слабым фонемам, хотя и само их определение, и вероятность появления слабых фонем вместо сильных не фундаментальным языковым материалом (ср. расхождения в понимании сильных и слабых фонем у Р. И. Аванесова и С. Н. Дмитренко, о чем пишет Л. Л. Буланин [12]). Было бы очень интересно на уровне описания фонемного состава морфем дать полное описание фонемных рядов гласных и согласных фонем (по Р. И. Аванесову), определить, сколько из них и каких именно являются наиболее представительными для системы русских морфем и сопоставить эти сведения с позднейшими интерпретациями.

Третий блок фонетического фонда — фонетические структуры русской словоформы. Эти данные могут быть получены в результате соединения наших знаний о правилах организации морфемы (второй блок) со сведениями о сочетаемости и позиционном распределении морфем в пределах словоформы. Указанные сведения, «хранящиеся» в морфемно-словообразовательном фонде, могут быть использованы в фонетическом фонде для получения разнообразных данных, необходимых исследователю: в настоящее время получены матрицы сочетаемости фонем на стыках морфем разных типов (префикс + префикс, префикс + + корень, корень + суффикс, суффикс + суффикс), что дает возможность сравнить сочетаемость фонем внутри морфемы с сочетаемостью на границах морфем. Нет нужды оценивать значение этих данных для описания закономерностей, действующих при формировании фонетической структуры словоформы. Возможность количественной оценки фонологических концепций на примерах, иллюстрирующих функции фонем при образовании звукового облика слова, не менее важна, чем изучение этого вопроса на материале морфем (второй блок).

Одним из интереснейших примеров анализа фонетических характеристик словоформы является анализ заударных морфемных комплексов на материале машинной версии RDD. Из 110 000 слов, содержащихся

в этом словаре, около 74 000 обнаруживают заударные гласные лексы. Известно, насколько сильна фонетическая редукция в заударной части слова, несущей в русском языке основную грамматическую информацию. Именно поэтому исчерпывающие сведения об организации заударных частей слов необходимы для понимания фонетических процессов. Обработка машинной версии словаря позволила установить как слоговую организацию этих комплексов (односложные встречаются в 42 752 словах, двухсложные — в 25 677 словах, трехсложные — в 5044 словах, четырехсложные — в 1226 словах), так и количество различных по сегментному составу типов (односложных — 392, двухсложных — 747, трехсложных — 283, четырехсложных — 70). При этом наиболее часто встречаются всего 311 морфемных комплексов, которые могут быть описаны 147 обобщенными фонетическими структурами.

Важнейшей задачей является анализ тех изменений фонетического облика слов, которые обусловлены влиянием формообразования. В настоящее время разрабатывается программа, позволяющая включать данные об акцентной парадигме слова и «синтезировать» возможные изменения в ритмической структуре словоформ, приводящие к фонемным чередованиям.

Четвертый блок фонетического фонда должен обеспечить возможность привлечения новых данных, представленных в виде текстов. Они необходимы как для пополнения предшествующих блоков, так и для экспериментальной проверки исследовательских гипотез. Этот блок включает в себя автоматический транскриптор, позволяющий преобразовывать орфографическую запись текста в последовательность фонетических символов — от С и V до детальных транскрипционных знаков комбинаторно-позиционных и «ситуативных» изменений. В настоящее время фонемный автоматический транскриптор практически действует, а фонетический находится в стадии активной разработки; как показал опыт автоматического транскрибирования текстов, возможны различные ситуации, и узкие задачи автоматического транскриптора могут быть существенно разными в зависимости от цели транскрибирования (пополнение первого блока, транскрипция для синтеза, транскрипция для обучения). Как следует из сказанного, все четыре блока тесно связаны между собой и в то же время допускают расширение структуры как за счет усложнения данных внутри каждого (или некоторых) блоков, так и за счет введения новых.

В заключение мне хотелось бы остановиться на некоторых практических аспектах использования теоретической основы и фактических данных фонетического фонда. Тот факт, что в основе фонда лежит информация о фонетически представительном тексте, с одной стороны, и о фонетических характеристиках морфемы и словоформы — с другой, позволяет использовать этот фонд при обучении на разных уровнях — как при обучении основам фонемного анализа русского языка, так и при обучении русской фонетике иноязычных учащихся. Например, введение в качестве самостоятельного фрагмента фонетического фонда частотного словаря В. В. Морковкина позволяет сформировать компактную программу обучения с элементами количественного описания фонологических свойств звуковых единиц (употребительность каждой из фонем в сильной или слабой позиции, реализация различительной функции в разных типах морфем, фонетические характеристики разных типов морфем и т. д.). Сопоставление такой «количественной» фонологии с данными, касающимися фонетически представительного текста, открывает хорошую перспективу для понимания

того, что именно находится в центре фонологической системы, а что — на ее периферии. Использование данных о частотности тех или иных фонетических цепочек при обучении русскому языку как иностранному позволяет облегчить речевую коммуникацию как на уровне речепроизводства, так и на уровне восприятия [13].

Создание фонетического фонда современного русского языка требует коллективных усилий. В настоящее время фонетический фонд должен рассматриваться как способ организации и упорядочивания актуальных фонетических исследований, требующих не только отчетливого исследовательского плана, но и соответствующего аппаратного и программного обеспечения.

СПИСОК ЛИТЕРАТУРЫ

1. Бондарко Л. В. Звуковой строй современного русского языка. М., 1977.
2. Вербицкая Л. А. Русская орфоэпия. Л., 1976.
3. Зиндер Л. Р. Общая фонетика. М., 1979.
4. Проблемы и методы экспериментально-фонетического анализа речи. Л., 1980.
5. Бондарко Л. В. Фонетическое описание языка и фонологическое описание речи. Л., 1981.
6. Морковкин В. В., Беме Н. О., Дорогонова И. А. и др. Лексическая основа русского языка. М., 1984.
7. Уровни языка в речевой деятельности. Л., 1986.
8. Worth D. S., Kozak A. S., Johnson D. B. Russian derivational dictionary. N. Y., 1970.
9. Зализняк А. А. Грамматический словарь русского языка. М., 1977.
10. Кузнецова А. Н., Ефремова Т. В. Словарь морфем русского языка. М., 1986.
11. Кулешова Л. В. Описание корней русского языка с помощью статистических методов: Автореф. дис. ... канд. филол. наук, М., 1976.
12. Буланин Л. Л. // ВЯ. 1987. № 4. Рец. на кн.: Дмитренко С. Н. Фонемы русского языка. Их сочетаемость и функциональная нагрузка. М., 1985.
13. Бондарко Л. В., Вербицкая Л. А. Два аспекта прикладных исследований фонетики. // Структурная и прикладная лингвистика. Вып. 2. Л. 1983.