

## VII. К фонологической статистике

### 1. Два способа подсчета

Проблемы статистики и функциональной нагрузки фонологических элементов теснейшим образом связаны с учением о сочетаемости фонем. Статистикой звуков занимались и раньше, используя ее в различного рода практических и научных целях. Разумеется, при использовании ее в фонологических целях она должна быть соответствующим путем преобразована: подсчитывать следует не буквы и звуки, а фонемы и сочетания фонем. В специальной фонологической литературе значение статистики впервые было подчеркнуто В. Матезиусом<sup>1</sup>.

Определенный вклад в фонологическую статистику английского языка сделал Б. Тринка в уже упоминавшейся выше работе<sup>2</sup>. В.Ф. Твэдделл подверг статистической обработке согласные фонемы и их сочетания в немецком языке<sup>3</sup>. Фонологической статистикой в общем плане занимался Д.К. Ципф<sup>4</sup>. Таким образом, в настоящее время мы уже не испытываем недостатка в статистических работах по фонологии. И все же число их весьма невелико, а методы, применяемые в этих работах, весьма различны. Единого фонологически-статистического метода до сих пор все еще не выработано. По этой причине мы вынуждены ограничиться лишь отдельными соображениями по данному вопросу.

В области фонологии статистика имеет двоякое значение. Она должна выявить частотность определенных фонологических элементов в данном языке (фонем, сочетаний фонем, определенных типов слов и морфем), с одной стороны, и степень функциональной загруженности этих элементов и определенных фонологических оппозиций — с другой стороны. Для решения первой задачи статистическому обследованию должны подвергаться связные тексты, для решения второй задачи — словарный состав. И в том и в другом случае учету подлежит либо только абсолютное число фак-

<sup>1</sup> Ср. его работы «La structure phonologique du lexique du tchèque moderne», TCLP, I, стр. 67—85 и «Zum Problem der Belastungs- und Kombinationsfähigkeit der Phoneme», там же, IV, стр. 148 и сл.

<sup>2</sup> B. Tr n k a, A phonological analysis of present day standard English, «Práce z vědeckých ústavů», XXXVII, 1935, стр. 45—175.

<sup>3</sup> W.P. T w a d d e l l, A phonological analysis of intervocalic consonant clusters in modern German (Actes du IV Congrès international de linguistes à Copenhague, 1938).

<sup>4</sup> G.K. Z i p f, Selected studies of the principle of relative frequency in language, Cambridge, Mass., Harvard University Press, 1932; и его же. Psycho-biology of language, Boston—Cambridge, Mass., Riverside Press, 1935.

тической встречаемости, либо отношение этого числа к теоретически ожидаемому числу встречающихся элементов, выведенному на основании правил сочетаемости.

## 2. Стилистически обусловленные и структурно обусловленные числа

Любая разновидность фонологической статистики имеет свои специфические трудности. При определении частотности данного фонологического элемента в связных текстах первостепенное значение приобретает **выбор текста**.

Я открываю наудачу «Теорию языка» К. Бюлера<sup>1</sup> и беру на стр. 23 первый попавшийся отрывок в 200 слов (от «soll es also...» до «im Schoße der Sprachwissenschaften längst», 3–28 строки сверху). Этот отрывок содержит 248 ударных морфем: 204 односложных, 37 двухсложных и 7 трехсложных. Затем я беру другой текст, содержащий то же количество слов, а именно начало первой сказки из сборника Дирра<sup>2</sup>, и нахожу, что этот отрывок содержит всего лишь 220 ударных морфем, из них 210 односложных, 10 двухсложных и ни одной трехсложной. Точно такое же различие между двумя выбранными мною текстами обнаруживается и в отношении размера слов. У Бюлера я нахожу слова разной длины: от одного до девяти слогов, у Дирра — только слова в один, два и три слога; подавляющее большинство из них приходится на односложные:

	К. Бюлер		А. Дирр	
	Число слов		Число слов	
	абс.	в %	абс.	в %
Односложные	95	47,5	134	67,0
Двусложные	57	28,5	56	28,0
Трехсложные	27	13,5	10	5,0
Четырехсложные	7	3,5	—	—
Пятисложные	6	3,0	—	—
Шестисложные	6	3,0	—	—
Семисложные	1	0,5	—	—
Восьмисложные	—	0,0	—	—
Девятисложные	1	0,5	—	—
	200	100,0	200	100,0

Общее число слогов в данных отрывках составляет: у Бюлера 400, у Дирра 276; таким образом, средний размер слова у Бюлера — 2 слога, у Дирра — 1,4 слога. Так как носителями слога в немецком являются глас-

<sup>1</sup> К. B ü h l e r, Sprachtheorie, Jena, 1934.

<sup>2</sup> «Kaukasische Märchen», ausgewählt und übersetzt von A. Dirr («Die Märchen der Weltliteratur», herausg. von Friedrich von Leyen und Paul Zaunert, Jena, 1920).

ные (слоговые п, г, л в безударных слогах фонологически трактуются как эп, эг, эл), то число слогов, следовательно, указывает на число гласных (400 у Бюлера и 276 у Дирра). Что же касается согласных, то в отрывке из Бюлера их насчитывается 636, а в отрывке из Дирра — 429; иначе говоря, каждое слово у Бюлера содержит в среднем 3,2 согласных, у Дирра — 2,1 согласных. Отношение числа согласных к числу гласных в обоих текстах оказывается примерно одинаковым: согласные составляют 61%, а гласные — 39% общего числа фонем. Но общее число фонем в отрывке из Бюлера выражается числом 1036, а в отрывке из Дирра — числом 705 — отношением, примерно равным 3 : 2. Не следует думать, что эта разница сгладится, если взять отрывки большего размера. Она самым тесным образом связана с различиями в стиле. Ученая, рассчитанная на довольно высокий интеллектуальный уровень слушателя речь включает длинные слова, тогда как простой, рассчитанный на довольно примитивный уровень слушателя рассказ предпочитает короткие слова. Другой особенностью немецкой ученой речи является ее насыщенность сочетаниями согласных. В обследованном отрывке из Дирра всего лишь 55 сочетаний согласных, в которых участвуют 116 фонем, то есть 27% всех согласных фонем, тогда как в обследованном отрывке из Бюлера 127 сочетаний согласных, в которых участвует в общей сложности 281 фонема, то есть 44% всех согласных фонем. Что же касается распределения этих согласных в составе слова и морфемы, то большая часть сочетаний и в том и в другом тексте приходится на стык (Дирр — 40%, Бюлер — 42%) и на исход (Дирр — 33%, Бюлер — 32%) морфем. Но совершенно иные отношения наблюдаются в начале и середине морфем: у Дирра на начало приходится 22%, а на середину лишь 5% всех сочетаний согласных, тогда как у Бюлера на начало приходится 12%, а на середину 14% всех сочетаний согласных. При этом у Бюлера в середине морфем оказываются такие сочетания, как *cj* (*Situation*), *gm* (*Dogma*), *skr* (*deskriptiv*), которые не встречаются у Дирра не только в обследованном отрывке, но и во всем сборнике сказок. Это является следствием усиленного употребления заимствованных слов, чем отличается вообще ученый язык.

Выбранные нами в качестве примера речевые стили — интеллектуализованная ученая речь и намеренно упрощенная, воспроизводящая примитив повествовательная речь — представляют собой два полюса, между которыми располагаются прочие стили, обладающие своими специфическими признаками. Каждый текст имеет какой-нибудь определенный стиль. Поставив перед собой задачу исследовать на материале какого-либо текста частотность определенного фонологического элемента в данном языке, необходимо прежде всего выяснить, какой текст является наиболее подходящим для этой цели. Проблема допускает как будто два решения: нужно выбрать или «стилистически нейтральный» текст, или отдельные отрывки из многих текстов, характеризующихся разными стилями. Оба решения, однако, являются мало удовлет-

ворительными. В самом деле: какой текст считать «стилистически нейтральным» и в какой пропорции следует брать отрывки из разностилевых текстов?

Таким образом, полное высвобождение фонологической статистики из-под влияния стилистики, стилистических типов речи кажется неосуществимой задачей. При фонологически-статистическом исследовании всегда надо учитывать специфические особенности того или иного стиля речи. Поэтому в первую очередь следует установить, какие фонологические явления стилистически обусловлены, а какие независимы от стиля. Так, мы уже видели, что по крайней мере в немецком языке длина значимых единиц (слов или морфем) и частота сочетаний из согласных стилистически обусловлены. Наоборот, частотность отдельных фонем, видимо, мало зависит от стилевых особенностей текста.

Сравним частотность употребления гласных фонем в рассмотренных выше отрывках из Бюлера и Дирра (цифры даются в процентах):

	К. Бюлер		А. Дирр			
а	57 {	37 { 15	63 {	40 { 18		
э		22		22		
и		20 {		7	23 {	9
о				10		10
au		3		4		
і	43 {	39 { 17	37 {	35 { 16,5		
е, ä		18		11		
ei		4		7,5		
ü		4 {		3	2 {	1
ö	0,5		1			
öü	0,5		0			
		100		100		

Незначительные различия между текстами в отношении фонем а, е, ei едва ли можно приписать стилистическим особенностям названных текстов. Возможно, что при статистическом обследовании более значительных кусков текста это различие сгладилось бы.

Следовательно, частотность употребления фонем, по крайней мере в немецком языке, видимо, не обусловлена стилистически и при вычислении этой частотности пригоден любой текст (за исключением поэзии и особо изысканной прозы, где намеренно

искусственная деформация естественной частотности рассчитана на то, чтобы вызвать специфический эффект<sup>1</sup>. Ради осторожности, однако, следует искать для этих целей стилистически нейтральный материал. Наиболее пригодными в этом отношении оказываются записи различных разговоров<sup>2</sup> или газеты, где представлены самые разные стили (передовица, телеграмма, научно-популярная статья, официальные сообщения, спортивная хроника, экономический отдел, фельетон и т.д. и т.п.)<sup>3</sup>.

### 3. Предложенные интерпретации частотности фонем

Статистическому обследованию на частотность фонем подверглось пока очень небольшое число языков. Поэтому нам представляется преждевременной любая интерпретация статистических данных и обобщающие суждения по данному вопросу. Однако подобного рода попытки предпринимаются уже и теперь. Одна из таких теорий о причинах различий в частотности отдельных фонем в разных языках предложена Ван Гиннекеном<sup>4</sup>. Согласно этой теории, каждый человек якобы обладает унаследованной склонностью к определенным артикуляциям и инстинктивно выбирает такие слова, в которых имеются соответствующие этим артикуляциям звуки. Далее, Ван Гиннекен считает, что так как все народы возникли из смешения рас, то любой представитель того или другого народа несет в себе определенный комплекс унаследованных расовых признаков, причем этим признакам соответствуют и артикуляторные склонности. Поскольку расовые признаки одинаковы у всех представителей данного народа, то общей является и система фонем; индивидуальные же колебания в частотности фонем объясняются различиями в числовых соотношениях между расовыми признаками у отдельных представителей народа. Эта теория создана не индуктивным путем, она исходит не столько из конкретных фактов, сколько из априорных соображений. Привлекаемый фонематический материал служит здесь не для обоснования и проверки теории, а лишь для истолкования этой теории, причем истолкования эти во всех случаях остаются целиком гипотетическими: если какая-либо фонема в том или

<sup>1</sup> Ср. J. Mukařovský, La phonologie et la poétique, TCLP, IV, стр. 280 и сл.

<sup>2</sup> Звуковую статистику на материале записей случайных разговоров для русского языка дает Пешковский («Десять тысяч звуков русского языка», Сборник статей, Л., 1925, стр. 167—191). Аналогичная работа имеется и для шведского языка, она проделана на основании стенографической записи выступлений в шведском парламенте. К сожалению, в обоих случаях речь идет о статистике звуков, а не фонем.

<sup>3</sup> Ср. B. Eldridge, Thousand common English words, Buffalo, 1911.

<sup>4</sup> J. van Ginneken, Ras en Taal, «Verhandl. d. Kon. Akad. van Wetensch. te Amsterdam», Aft. Letterkunde, N. R. XXXVI, 1935, его же, De Ontwikkelingsgeschiedenis van de systemen der menschelijke Taalklanken, Amsterdam, 1932; его же, De Oorzaken der taalveranderingen, Amsterdam, 1930, его же, La biologie et la base d'articulation, «Journ. de Psychol.», XXX, стр. 266—320.

ином языке имеет особенно высокую или особенно низкую частотность, то строится предположение, что расовые признаки данного народа в одном случае благоприятствуют, в другом — препятствуют соответствующим артикуляциям. Но ведь это является логическим *petitio principii*, ибо сперва надо еще доказать, что та или иная частотность определенной фонемы в связной речи зависит от расовых признаков говорящего. Если частотность фонем в языках негров оказывается иной, чем в языках североамериканских индейцев, то это еще далеко не свидетельствует о зависимости фонемной частотности от расовых признаков, поскольку языки негров отличаются от языков североамериканских индейцев не только разной частотностью фонем, но и составом фонем, грамматической структурой и т.д. Объективное доказательство мог бы дать лишь эксперимент, в процессе которого удалось бы изолировать исследуемые факторы от всех остальных. Следовало бы, например, рассмотреть фонемную частотность у двух лиц, принадлежащих к разным расам, но говорящих на одном языке и имеющих один уровень образования (учитывая при этом стилистически равноценные высказывания). Результаты такого эксперимента, однако, имели бы научное значение лишь в том случае, если бы эксперимент был повторен многие сотни раз на представителях разных рас и различных языков. Только тогда имело бы смысл обсуждать этот вопрос.

Другая теория частотности фонем предложена Д.К. Ципфом<sup>1</sup>. Согласно его теории, частотность той или иной фонемы тем выше, чем проще реализация этой фонемы. Ципф, таким образом, целиком стоит на естественно-научных позициях. Это значит, что при проверке данной теории, чтобы быть последовательным, необходимо рассматривать факты с естественно-научной точки зрения. Но со строго естественно-научной точки зрения измерить степень сложности артикуляции невозможно. Звонкие смычные свидетельствуют о напряжении голосовых связок и одновременном расслаблении мускулатуры в полости рта; прямо противоположную картину дают глухие смычные: расслабление голосовых связок и напряжение в полости рта. Что здесь является более сложным? При произношении придыхательных согласных голосовая щель широко открыта, оставаясь в том же положении, какое она занимает при нормальном дыхании; при произношении же непридыхательных согласных голосовая щель в момент отступа должна принять другое положение, чтобы воспрепятствовать образованию придыхания. Но, с другой стороны, при более сильном потоке воздуха органы в полости рта оказываются, как правило, более напряженными. Таким образом, и в отношении оппозиций с участием придыхания трудно сказать, что здесь является «более сложным»: придыхательные или непридыхательные согласные. Все сказанное можно повторить в отношении любого противоположения по способу преодоления преграды. В еще меньшей мере может быть определена степень сложности в отношении противоположений по месту образования преграды. Ципф в качестве примера приводит противоположение *m—n* и, исходя из того

<sup>1</sup> G.K. Zipf, *Psycho-biology of language*, стр. 68 и сл. (ср. нашу рецензию в «Slovo a slovesnost», II, 1936, стр. 252 и сл.)

обстоятельства, что во многих языках *n* встречается чаще, чем *m*, считает возможным заключить, что *m* «сложнее» *n*. Однако *m* произносится с сомкнутыми губами и опущенной нёбной занавеской; органы речи, таким образом, находятся в состоянии полного покоя (за исключением напряжения голосовых связок); произношение же *n* (помимо напряжения голосовых связок, которое является общим как для *m*, так и для *n*) связано еще и с поднятием кончика языка к зубам или альвеолам и, как правило, с соответствующим движением нижней челюсти. Следовательно, и эта теория должна быть решительно отвергнута, по крайней мере в приведенной выше редакции.

Обе рассмотренные теории являются уязвимыми главным образом потому, что они пытаются объяснить фонологические факты с помощью биологических, то есть внеязыковых причин. Впрочем, теория Ципфа может быть «изложена фонологически», на что уже указал Марсель Коэн в своей рецензии на книгу Ципфа<sup>1</sup>. В фонологической редакции эта теория могла бы звучать примерно так: «из двух членов привативной оппозиции немаркированный член в связной речи встречается чаще, чем маркированный». В общем и целом такую формулировку можно было бы считать соответствующей действительности. Однако ее ни в коем случае не следует рассматривать как правило, не терпящее исключений. Необходимо различать нейтрализуемые и ненейтрализуемые оппозиции, а также принимать во внимание объем нейтрализуемости. В русском языке, где противоположение твердых и мягких согласных имеет место в отношении двенадцати пар, сформулированное выше правило справедливо лишь для одиннадцати пар: твердые *p, b, f, v, t, d, s, z, m, n, ɣ* фактически встречаются гораздо чаще, нежели соответствующие им мягкие: *p', b', f', v', t', d', s', z', m', n', ɣ'* (в отношении примерно 2:1). Но в отношении пары *l:l'* это правило недействительно; палатализованное *l'* в русском языке встречается чаще, чем непалатализованное (*l:l'* = 42:58). Конечно, не случайно, что оппозиция *l—l'* нейтрализуется только перед *e*, тогда как такие оппозиции, как *p—p', t—t'* и т.д., нейтрализуются и в других положениях (перед апикальными, сибилантами, палатализованными лабиальными). Корреляция звонкости в русском языке нейтрализуема: в исходе слова перед паузой или перед словами, которые начинаются с сонорного, возможны только глухие шумные, благодаря чему они характеризуются как немаркированный член данной корреляции. Однако фонема *v* (а равным образом соответствующая ей мягкая *v'*) занимает особое положение: с одной стороны, она не может находиться в исходе и середине слова перед глухими шумными, заменяясь в этом случае соответствующей глухой фонемой *f*; с другой стороны, однако, перед *v* возможны глухие согласные (ср. *твой, свадьба, закваска* и др.), что совершенно исключено в отношении других звонких шумных. Иными словами, фонема *v* оказывает на другие шумные совсем не то действие, какое оказывает маркированный член корреляции звонкости. С этим связано то обстоятельство, что фонема *v* встречается примерно в четыре раза чаще, чем *f*, тогда как в других чле-

<sup>1</sup> Ср. Marcel Cohen в BSL, XXXVI, 1935, 10.

нах рассматриваемой корреляции звонкая фонема встречается примерно в три раза реже, чем глухая<sup>1</sup>.

Все приведенные Ципфом примеры можно свести к предложенной выше формуле. Ибо в языках, где есть корреляция звонкости, глухие шумные являются такими же немаркированными членами, какими являются непридыхательные шумные в языках, где есть корреляция придыхания. То, что здесь речь идет не о придыхании как таковом, а об оппозитивных отношениях, свидетельствуют такие языки, как лезгинский (кюринский), в которых придыхательные смычные являются немаркированными членами корреляции интенсивности<sup>2</sup>: здесь придыхательные смычные встречаются, как правило, чаще, чем соответствующие им непридыхательные ( $p^h$  1,8 :  $P$  0,8;  $t^h$  5,2 :  $T$  2,2;  $k^h$  8,8 :  $K$  0,7;  $c^h$  9,0 :  $C$  0,1), и только у глубоковелярных отношение оказывается обратным ( $q^h$  1,6 :  $Q$  3,8); следует, однако, заметить, что в противоположность всем прочим оппозициям по интенсивности оппозиция  $q^h$  :  $Q$  в послелударных слогах не нейтрализуется.

Если не подлежит никакому сомнению, что различие между маркированным и немаркированным членами оппозиции, с одной стороны, и между нейтрализуемыми и ненейтрализуемыми оппозициями, с другой стороны, оказывает свое действие на частотность фонем, то столь же очевидно, что одного этого факта недостаточно для объяснения отношений частотности. В разных языках всегда имеются оппозиции, привативный характер которых не может быть доказан объективно. Так, например, корреляция звонкости является во французском языке привативной и нейтрализуемой, однако она подлежит только диссимилятивной нейтрализации, которая обусловлена контекстом (тип *a*), причем выбор представителя архифонемы обусловлен внешними обстоятельствами, в силу чего немаркированный характер какого-либо члена этой оппозиции не может быть установлен объективно<sup>3</sup>. В общем же французские глухие шумные встречаются чаще соответствующих звонких (примерно 60:40), но в каждом отдельном случае отношение оказывается разным:  $\dot{z}$  и  $v$  встречаются гораздо чаще, чем  $\dot{s}$  и  $f$ ;  $d$  и  $t$  примерно одинаково часты, тогда как в других парах ( $p$ — $b$ ,  $k$ — $g$ ,  $s$ — $z$ ) глухой встречается чаще звонкого.

#### 4. Фактическая и теоретическая частотность

Было бы безнадежно пытаться установить строгую закономерность для фонемной частотности, поскольку эта частотность является результатом целого ряда факторов. Абсолютные цифры фактической частотности имеют лишь второстепенное значение. Ре-

<sup>1</sup> Помимо этого, звонкое  $\dot{z}$  встречается в русском языке чаще глухого. Это исключение, однако, не наблюдается у тех русских, которые произносят *щ* как  $\dot{z}\dot{s}$ .

<sup>2</sup> Для статистики лезгинских фонем подсчеты были произведены по сказке № 5, взятой из приложения к работе П.К. Услара, Кюринский язык, «Этнография Кавказа», стр. 291—299.

<sup>3</sup> Ср. A. Martinet в TCLP, VI, стр. 51 и сл.



альную значимость имеет только отношение этих чисел к числам, выражающим теоретически рассчитанную частотность. Следовательно, фактическому подсчету фонем в том или ином тексте должен быть предпослан тщательный расчет теоретических возможностей (с учетом правил нейтрализации и сочетаемости). Представим себе язык, в котором определенная оппозиция согласных фонем нейтрализуется в начале и исходе слова, причем в нейтрализуемом положении выступает только немаркированный член. В таком языке немаркированный член данной оппозиции может быть, следовательно, в начале каждого слога, а, кроме того, еще в исходе слова, тогда как маркированный член может встречаться в начале каждого слога, за исключением первого. Если в таком языке среднее число слогов в слове равно  $\alpha$ , то следует ожидать, что частотность немаркированного члена относится к частотности маркированного члена как  $\alpha+1$  к  $\alpha-1$ . В чеченском языке, где геминированные согласные встречаются только в середине слова (как и в большинстве других языков с корреляциями геминации) и где слово состоит в среднем из 1,9 слога (по крайней мере в сказках), частотность геминированных согласных должна относиться к частотности соответствующих негеминированных как 9:29 (то есть примерно как 1:3). В действительности подсчет дает следующие цифры:

tt : t	12 : 90	(4 : 30)
qq : q	6 : 45	(4 : 30)
čč : č	25 : 59	(13 : 30)
ll : l	16 : 32	(15 : 30) <sup>1</sup>

Получается, таким образом, что геминированные čč и ll встречаются фактически чаще, а геминаты tt и qq реже, чем это ожидалось теоретически. В том же чеченском языке в составе смычных выделяется корреляция рекурсии; такая корреляция возможна лишь в начале слова; в середине и конце слова она нейтрализуется (в качестве представителя архифонемы выступает нерекурсивный смычный). Таким образом, маркированные члены этой оппозиции возможны лишь в начале слова; обозначив общее число слогов в тексте через  $\beta$ , а среднее число слогов в слове через  $\alpha$ , получим

<sup>1</sup> Подсчеты производились по тексту № IV в собрании K. Bouda, Tschetschenische Texte (= Mitteilungen des Seminars für orientalische Sprachen zu Berlin, Jahrg., XXXVIII, Abt II, Westasiatische Studien, Berlin, 1935, стр. 31–35); в отношении tt, t, qq, q, čč, č подсчет производился с учетом всего текста; для ll, l были взяты только первые 300 слов.

число  $\frac{\beta}{\alpha}$ , выражающее частотность маркированного члена. Соответствующий немаркированный член возможен в начале каждого слога и, кроме того, в исходе каждого слова; частотность этого члена может быть выражена числом  $\beta + \frac{\beta}{\alpha}$ . Теоретически вычисленные частотности относятся, таким образом, друг к другу, как  $\frac{\beta}{\alpha}$  к  $\beta + \frac{\beta}{\alpha}$ , то есть как 1 к  $\alpha + 1$ . Поскольку среднее число слогов в чеченском слове выражается цифрой 1,9, то отношение принимает вид 1 : 2,9. Реально же мы имеем следующие цифры:

č' : t	33 : 90	(11 : 30)
k' : k	38 : 47	(24 : 30)
q' : q	21 : 45	(14 : 30)
c' : c	17 : 97	(5 : 30)
č' : č	5 : 59	(2,5 : 30)
p' : p	? : 27	(?) <sup>1</sup>

В общем фактическое отношение частот для рекурсивных и нерекурсивных смычных примерно соответствует теоретически вычисленному отношению ( $114 : 365 = 0,9 : 2,9$ ), однако отдельные пары значительно отклоняются от теоретических расчетов, причем в обоих направлениях; но все же немаркированные члены всегда более часты, нежели маркированные.

Вычисление теоретических возможностей не всегда оказывается столь легким, как в приведенных выше примерах. Не следует, однако, бояться технических трудностей, возникающих при расчетах. Ведь цифры фактической частотности приобретают определенную значимость только путем сравнения их с цифрами в озимой частотности, добытой путем таких вычислений; они показывают, насколько полно используется та или иная фонема в данном языке.

При фонологически-статистическом обследовании текста необходимо принимать во внимание не только то, как часто встречается та или иная фонема вообще, но и то, как часто встречается она в определенных положениях. Если, например, немаркированный член нейтрализуемой оппозиции особенно часто встречается в нейтрализованном положении (где он представляет архифонему), то это говорит о том, что соответствующая оппозиция реализуется редко; но если данный член оппозиции особенно часто выступает

<sup>1</sup> Рекурсивное *p* в обследованных текстах не встретилось ни разу.

в релевантной позиции (чаще, чем это можно было бы ожидать теоретически), то это обстоятельство свидетельствует об особой склонности к употреблению данной оппозиции. Аналогичным образом можно было бы установить статистически и степень использования различных ненейтрализуемых оппозиций. Во многих языках мы встречаем такие положения в слове, при которых возможно лишь ограниченное число фонем и, следовательно, ограниченное число смысловоразличительных оппозиций. Фактическая частотность подобных позиций может либо превысить теоретическую, либо быть ниже ее; в зависимости от того или иного отношения этих частотностей можно заключить, сколь часто или сколь редко используется та или иная оппозиция в данном языке.

Следовательно, грубая глобальная статистика фонем должна быть заменена более тонкой, специфической статистикой, объектом которой являются уже не фонемы, а оппозиции. Ибо здесь, как, впрочем, и всюду в фонологии, нужно постоянно помнить, что подлинным объектом фонологического исследования являются не фонемы, а оппозиции.

### *5. Фонологическая статистика словарного состава*

Из всего сказанного выше со всей очевидностью вытекает, что одного статистического обследования текстов далеко не достаточно, чтобы составить себе представление об относительной употребительности различных фонологических элементов. Такое обследование должно быть восполнено статистическим рассмотрением словарного состава. Последнее со своей стороны должно постоянно учитывать отношения фактического к теоретически возможному. В. Матезиус и Б. Трнка уже предприняли в этом направлении весьма важную попытку. Из работ В. Матезиуса со всей очевидностью вытекает важность подобного рода исследований для фонологической характеристики языков. В этом легко убедиться, если сопоставить слова ряда языков, состоящие из двух фонем. В немецком языке в начале слова допустимы 18 согласных (b, p, m, d, t, n, k, g, c, z, š, f, v, ř, h, r, l, j), а в исходе слова — 14 (p, m, t, n, k, ř, x, c, s, ř, f, š, r, l); все без исключения ударные гласные (10 фонем, если не различать ä и е) возможны как в начале, так и в исходе слова; недопустимы сочетания j + i, au + r, au + ř, eu + r, eu + ř, ai + r, ai + ř, o + ř, ö + ř. Таким образом, с теоретической точки зрения в немецком языке возможны  $(18 \times 10) - 1 = 179$  слов типа «согласный + гласный» и  $(14 \times 10) - 8 = 132$  слова типа «гласный + согласный» (безотносительно к слогоделению). Факти-

чески первый тип («согласный + гласный») представлен в немецком 57 словами (du, Kuh, zu, Schuh, wo, loh, roh, Bau, Tau, kau, Gau, Pfau, Vau, Sau, schau, hau, lau, rauh, die, nie, Vieh, wie, zieh, sie, hie, lieh, mäh, Tee, näh, Weh, Zeh, See, je, geh, bei, weih, zeih, sei, reih, leih, Küh', Höh', neu, scheu, Heu, Leu, Reuh', da, nah, sah, ja, названия букв: Be, Pe, De, Ha, Ka), а второй тип («гласный + согласный») — 37 словами (Uhr, Ohr, ob, Aug', Auch, aus, auf, ihr, im, in, ich, iß, er, El, Em, En, eng, Eck, ätz, es, Esch, Eid, ein, eil, Eich', Eis, Eul', euch, Aar, Aal, am, an, ach, aß, Aff', ab, Asch'). Во французском языке в начале слова могут быть 15 согласных (b, p, d, t, g, k, v, f, s, š, ž, m, n, r, l), а в исходе — 18 (b, p, d, t, g, k, v, f, z, s, š, ž, m, n, ŋ, r, l, j); что касается гласных, то в закрытых слогах их может быть 12 (u, o, ə, a, e, i, ø, y, ô, â, ê, ø), а в открытых — 13 (те же самые + e). Сочетания типа «носовой гласный» (ö, ê, ø, â) + m, n, ŋ, r, l, j невозможны. Теоретически в этом языке возможны  $15 \times 13 = 195$  слов типа «согласный + гласный» и  $(12 \times 18) - (4 \times 6) = 192$  слова типа «гласный + согласный». В действительности же первый тип представлен 142 словами, а второй — 50 словами. Иначе: в немецком языке в типе «согласный + гласный» реализуется лишь 31,8% теоретически возможных слов, а во французском языке — 73%. Что касается типа «гласный + согласный», то здесь процент реализации теоретически возможных слов в обоих языках оказывается примерно одинаковым: 28% в немецком и 26% во французском. Однако, тогда как в немецком слова этого типа составляют 40% всех односложных слов из двух фонем, во французском языке такие слова составляют лишь 26%. Мы видим, таким образом, что даже в столь узких рамках отчетливо проявляется индивидуальность каждого языка. В. Матезиус, сопоставлявший чешский с немецким в отношении использования фонологических средств, установил, между прочим, что среди слов, состоящих из двух или трех фонем, слова с начальным гласным составляют в немецком 25,2%, а в чешском только 8,2%. Далее оказалось, что немецкий предпочитает скопления согласных в исходе слова, а чешский — в его начале.

Все указанные особенности, накладывающие свой особый отпечаток на любой язык, можно выразить в числах. При таком методе исследования словарного состава в числах можно выразить и степень смысловозначительного использования («функциональной нагрузки») отдельных фонологических оппозиций и среднюю нагрузку фонем вообще. При этом оказывается, что все языки в этом отношении делятся на «бережливые» и «расточительные». В «бережливых» языках слова, отличающиеся друг от друга всего лишь

одной фонемой, весьма многочисленны, а процент реализации теоретически возможных сочетаний фонем очень высок. В «расточительных» языках наблюдается склонность отличать слова друг от друга с помощью многих фонологических средств и реализовывать лишь незначительную часть теоретически возможных сочетаний фонем.

На фоне фонологически-статистического обследования словарного запаса фонологически-статистическое исследование связного текста приобретает новый смысл. Числа, выражающие частотность, приобретают как бы двойную соотнесенность. Теперь уже речь идет о том, насколько использованы в связной речи те теоретические возможности, которые реализованы в словаре и даны в правилах сочетаемости фонем. Чем больше фонем в определенном типе слов, тем выше число теоретически возможных слов этого типа. Статистическое исследование словарного состава показывает, какой процент теоретических возможностей реализуется, иными словами, какое число сочетаний фонем данного типа обладает определенной лексической значимостью. Оно, однако, ничего не говорит о фактической частотности слов этого типа в нормальной, связной речи. Сведения об этом может дать лишь статистическое обследование текстов. Причем может оказаться, что типы слов с высоким процентом реализации теоретических возможностей менее часты, чем слова с незначительным процентом реализации. Есть ли здесь какие-либо общезначимые законы или же все языки в данном отношении различны — об этом пока ничего определенного сказать нельзя, поскольку фонологическая статистика еще очень мало этим занималась. Во всяком случае необходимо остерегаться поспешных заключений и теоретизирований в этой области.

В заключение следовало бы еще указать, что лексической статистике приходится преодолевать часто те же трудности, что и статистике текстов. Не все части словарного состава одинаковы и сопоставимы. Существуют технические выражения, известные лишь ограниченному кругу специалистов, хотя такие слова и не являются заимствованиями в обычном смысле. Следует ли привлекать подобные термины для статистики? Далее, есть слова, которые в своей литературной форме употребляются разве только в словарях, а фактически живут лишь в своем диалектном облике, поскольку они по своему значению принадлежат диалектам (различные сельскохозяйственные термины и т.п.). В каком звуковом облике должна учитывать их статистика? Подобного рода проблемы возникают перед лексической статистикой в любом языке. Для некоторых восточных литературных языков подобные вопросы оказываются прямо-таки роковыми. Во всяком случае не следует представлять себе вещи слишком упрощенно.